Medical Image Processing, Reconstruction and Restoration

Concepts and Methods

Jiří Jan

Taylor & Francis
Taylor & Francis Group

Boca Raton London New York Singapore

A CRC title, part of the Taylor & Francis imprint, a member of the
Taylor & Francis Group, the academic division of T&F Informa plc.

Signal Processing and Communications

1. Digital Signal Processing for Multimedia Systems, *edited by Keshab K. Parhi and Takao Nishitani*
2. Multimedia Systems, Standards, and Networks, *edited by Atul Puri and Tsuhan Chen*
3. Embedded Multiprocessors: Scheduling and Synchronization, *Sundararajan Sriram and Shuvra S. Bhattacharyya*
4. Signal Processing for Intelligent Sensor Systems, *David C. Swanson*
5. Compressed Video over Networks, *edited by Ming-Ting Sun and Amy R. Reibman*
6. Modulated Coding for Intersymbol Interference Channels, *Xiang-Gen Xia*
7. Digital Speech Processing, Synthesis, and Recognition: Second Edition, Revised and Expanded, *Sadaoki Furui*
8. Modern Digital Halftoning, *Daniel L. Lau and Gonzalo R. Arce*
9. Blind Equalization and Identification, *Zhi Ding and Ye (Geoffrey) Li*
10. Video Coding for Wireless Communication Systems, *King N. Ngan, Chi W. Yap, and Keng T. Tan*
11. Adaptive Digital Filters: Second Edition, Revised and Expanded, *Maurice G. Bellanger*
12. Design of Digital Video Coding Systems, *Jie Chen, Ut-Va Koc, and K. J. Ray Liu*
13. Programmable Digital Signal Processors: Architecture, Programming, and Applications, *edited by Yu Hen Hu*
14. Pattern Recognition and Image Preprocessing: Second Edition, Revised and Expanded, *Sing-Tze Bow*
15. Signal Processing for Magnetic Resonance Imaging and Spectroscopy, *edited by Hong Yan*

16. Satellite Communication Engineering, *Michael O. Kolawole*
17. Speech Processing: A Dynamic and Optimization-Oriented Approach, *Li Deng*
18. Multidimensional Discrete Unitary Transforms: Representation: Partitioning and Algorithms, *Artyom M. Grigoryan, Sos S. Agaian, S.S. Agaian*
19. High-Resolution and Robust Signal Processing, *Yingbo Hua, Alex B. Gershman and Qi Cheng*
20. Domain-Specific Processors: Systems, Architectures, Modeling, and Simulation, *Shuvra Bhattacharyya; Ed Deprettere; Jurgen Teich*
21. Watermarking Systems Engineering: Enabling Digital Assets Security and Other Applications, *Mauro Barni, Franco Bartolini*
22. Biosignal and Biomedical Image Processing: MATLAB-Based Applications, *John L. Semmlow*
23. Broadband Last Mile Technologies: Access Technologies for Multimedia Communications, *edited by Nikil Jayant*
24. Image Processing Technologies: Algorithms, Sensors, and Applications, *edited by Kiyoharu Aizawa, Katsuhiko Sakaue and Yasuhito Suenaga*
25. Medical Image Processing, Reconstruction and Restoration: Concepts and Methods, *Jiří Jan*
26. Multi-Sensor Image Fusion and Its Applications, *edited by Rick Blum and Zheng Liu*
27. Advanced Image Processing in Magnetic Resonance Imaging, *edited by Luigi Landini, Vincenzo Positano and Maria Santarelli*

Published in 2006 by
CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

Printed in the United States of America on acid-free paper
10 9 8 7 6 5 4 3 2 1

International Standard Book Number-10: 0-8247-5849-8 (Hardcover)
International Standard Book Number-13: 978-0-8247-5849-3 (Hardcover)
Library of Congress Card Number 2004063503

Library of Congress Cataloging-in-Publication Data

Jan, Jirí.
Medical image processing, reconstruction and restoration : concepts and methods / by Jirí Jan.
p . cm. -- (Signal processing and communications ; 24)
Includes bibliographical references and index.
ISBN 0-8247-5849-8 (alk. paper)
1. Diagnostic imaging--Digital techniques. I. Title. II. Series.

RC78.7.D53J36 2005
616.07'54--dc22 2004063503

informa
Taylor & Francis Group
is the Academic Division of Informa plc.

Visit the Taylor & Francis Web site at
http://www.taylorandfrancis.com

and the CRC Press Web site at
http://www.crcpress.com

Preface

Beginning with modest initial attempts in roughly the 1960s, digital image processing has become a recognized field of science, as well as a broadly accepted methodology, to solve practical problems in many different kinds of human activities. The applications encompass an enormous range, starting perhaps with astronomy, geology, and physics, via medical, biological, and ecological imaging and technological exploitation, up to the initially unexpected use in humane sciences, e.g., archaeology or art history. The results obtained in the area of digital image acquisition, synthesis, processing, and analysis are impressive, though it is often not generally known that digital methods have been applied. The basic concepts and theory are, of course, common to the spectrum of applications, but some aspects are more emphasized and some less in each particular application field. This book, besides introducing general principles and methods, concentrates on applications in the field of medical imaging, which is specific for at least two features: biomedical imaging often concerns internal structures of living organisms inaccessible to standard imaging methods, and the resulting images are observed, evaluated, and classified mostly by nontechnically oriented staff.

The first feature means that rather specific imaging methods, namely, tomographic modalities, had to be developed that are entirely dependent on digital processing of measured preimage data and that utilize rather sophisticated theoretical backgrounds stemming from the advanced signal theory. Therefore, development of new or innovated image processing approaches, as well as interpretation of more complicated or unexpected results, requires a deep understanding of the underlying theory and methods.

Excellent theoretical books on general image processing methods are available, some of them mentioned in references. In the area of medical imaging, many books oriented toward individual clinical branches have been published, mostly with medically interpreted case studies. Technical publications on modality-oriented specialized methods are frequent, either as original journal papers and conference proceedings or as edited books, contributed to by numerous specialized authors and summarizing recent contributions to a particular field of medical image processing. However, there may be a niche for books that would respect the particularities of biomedical orientation while still providing a consistent, theoretically reasonably exact, and yet comprehensible explanation of the underlying *theoretical concepts* and *principles of methods* of image processing as applied in the broad medical field and other application fields.

This book is intended as an attempt in this direction. It is the author's persuasion that a good understanding of concepts and principles forms a necessary basis to any valid methodology and solid application. It is relatively easy to continue studying and even designing specialized advanced approaches with such a background; on the other hand, it is extremely difficult to grasp a sophisticated method without well understanding the underlying concepts. Investigating a well-defined theory from the background also makes the study enjoyable; even this aspect was in the foundation of the concept of the book.

This is a book primarily for a technically oriented audience, e.g., staff members from the medical environment, interdisciplinary experts of different (not necessarily only biomedical) orientations, and graduate and postgraduate engineering students. The purpose of the book is to provide *insight*; this determines the way the material is treated: the rigorous mathematical treatment — definition, lemma, proof — has been abandoned in favor of continuous explanation, in which most results and conclusions are consistently derived, though the derivation is contained (and sometimes perhaps even hidden)

in the text. The aim is that the reader becomes familiar with the explained concepts and principles, and acquires the idea of not only believing the conclusions, but also checking and interpreting every result himself, though perhaps with informal reasoning. It is also important that all the results would be interpreted in terms of their "physical" meaning. This does not mean that they be related to a concrete physical parameter, but rather that they are reasonably interpreted with the purpose of the applied processing in mind, e.g., in terms of information or spectral content. The selection of the material in the book was based on the idea of including the established background without becoming mathematically or theoretically superficial, while possibly eliminating unnecessary details or too specialized information that, moreover, may have a rather time-limited validity.

Though the book was primarily conceived with the engineering community of readers in mind, it should not be unreadable to technically inclined biomedical experts. It is, of course, possible to successfully exploit the image processing methods in clinical practice or scientific research without becoming involved in the processing principles. The implementation of imaging modalities must be adapted to this standard situation by providing an environment in which the nontechnical expert would not feel the image processing to be a strange or even hostile element. However, the interpretation of the image results, namely, in more involved cases, as well as the indication of suitable image processing procedures under more complex circumstances, may be supported by the user's understanding of the processing concepts. It is therefore a side ambition of this book to be comprehensible enough to enable appreciation of the principles, perhaps without derivations, even by a differently oriented expert, should he be interested.

It should also be stated what the book is not intended to be. It does not discuss the medical interpretation of the image results; no casuistic analysis is included. Concerning the technical contents, it is also not a theoretical in-depth monograph on a highly specialized theme that would not be understandable to a technically or mathematically educated user of the imaging methods or a similarly oriented graduate student; such specialized publications may be found among the references. Finally, while the book may be helpful even as a daily reference to concepts and methods, it is not a manual on application details and does not refer to any particular program, system, or implementation.

The content of the book has been divided into three parts. The first part, "Images as Multidimensional Signals," provides the

introductory chapters on the basic image processing theory. The second part, "Imaging Systems as Data Sources," is intended as an alternative view on the imaging modalities. While the physical principles are limited to the extent necessary to explain the imaging properties, the emphasis is put on analyzing the internal signals and (pre)image data that are to be consequently processed. With respect to this goal, the technological solutions and details of the imaging systems are also omitted. The third part, "Image Processing and Analysis," starts with tomographic image reconstruction, which is of fundamental importance in medical imaging. Another topical theme of medical imaging is image fusion, including multimodal image registration. Further, methods of image enhancement and restoration are treated in individual chapters. The next chapter is devoted to image analysis, including segmentation, as a preparation for diagnostics. The concluding chapter, on the image processing environment, briefly comments on hardware and software exploited in medical imaging and on processing aspects of image archiving and communication, including principles of image data compression.

With respect to the broad spectrum of potential readers, the book was designed to be as self-contained as possible. Though background in signal theory would be advantageous, it is not necessary, as the basic terms are briefly explained where needed. Each part of the book is provided with a list of references, containing the literature used as sources or recommended for further study. Citation of numerous original works, though their influence and contribution to the medical imaging field are highly appreciated, was mostly avoided as superfluous in this type of book, unless these works served as immediate sources or examples.

The author hopes that (in spite of some ever-present oversights and omissions) the reader will find the book's content to be consistent and interesting, and studying it intellectually rewarding. If the basic knowledge contained within becomes a key to solving practical application problems and to informed interpretation of results, or a starting point to investigating more advanced approaches and methods, the book's intentions will have been fulfilled.

Jiří Jan
Brno, Czech Republic

Acknowledgments

This book is partly based on courses on basic and advanced digital image processing methods, offered for almost 20 years to graduate and Ph.D. students of electronics and informatics at Brno University of Technology. A part of these courses has always been oriented toward biomedical applications. Here I express thanks to all colleagues and students, with whom discussions often led to a better view of individual problems. In this respect, the comments of the book reviewer, Dr. S.M. Krishnan, Nanyang Technological University Singapore, have also been highly appreciated.

Most of medical images presented as illustrations or used as material in the derived figures have been kindly provided by the cooperating hospitals and their staffs: the Faculty Hospital of St. Anne Brno (Assoc. Prof. P. Krupa, M.D., Ph.D.), the Faculty Hospital Brno-Bohunice (Assoc. Prof. J. Prasek, M.D., Ph.D.; Assoc. Prof. V. Chaloupka, M.D., Ph.D., Assist. Prof. R. Gerychova, M.D.), Masaryk Memorial Cancer Institute Brno (Karel Bolcak, M.D.), Institute of Scientific Instruments, Academy of Sciences of the Czech Republic (Assoc. Prof. M. Kasal, Ph.D.), and Brno University of Technology (Assoc. Prof. A. Drastich, Ph.D., D. Janova, M.Sc.). Their courtesy is highly appreciated. Recognition notices are only placed with figures that contain

original medical images; they are not repeated with figures where these images serve as material to be processed or analyzed. Thanks also belong to former doctoral students V. Jan, Ph.D., and R. Jirik, Ph.D., who provided most of the drawn and derived-image figures.

The book utilizes as illustrations of the described methods, among others, some results of the research conducted by the group headed by the author. Support of the related projects by grant no. 102/02/0890 of the Grant Agency of the Czech Republic, by grants no. CEZ MSM 262200011 and CEZ MS 0021630513 of the Ministry of Education of the Czech Republic, and also by the research centre grant 1M6798555601 is acknowledged.

Contents

Part I

Images as Multidimensional Signals

Part I provides the theoretical background for the rest of the book. It introduces the concept of still images interpreted as two-dimensional signals, as well as the generalization to multidimensional interpretation of moving images and three-dimensional (spatial) image information. Once this general notion is introduced, the signal theoretical concepts, after generalization to the two-dimensional or multidimensional case, can be utilized for image processing and analysis. This concept proved very successful in enabling the formalization (and consequently optimization) of many approaches to image acquisition, processing, and analysis that were originally designed as heuristic or even not feasible.

A characteristic example comes from the area of medical tomographic imaging: the intuitively suggested heuristic algorithm of image reconstruction from projections by back-projection turned out to be very unsatisfactory, giving only a crude approximation of the proper image, with very disturbing artifacts. Later, relatively complex theory (see Chapter 9) was developed that led to a formally derived algorithm of *filtered* back-projection, widely used nowadays, that is theoretically correct and provides very good images, even

under practical limitations. Both algorithms are quite similar, with the only difference being the filtering of each individual projection added to the original procedure in the later method — seemingly an elementary step, but probably impossible to discover without the involved theory. The alternative methods of image reconstruction from projections rely heavily on other aspects of the multidimensional signal theory as well.

Part I introduces the basic image processing concepts and terminology needed to understand further sections of the book. Broader and deeper treatment of the theory can be found in the numerous literature that is partly listed in the references to this section, e.g., in [4], [5], [6], [18], [22], [23], [25], [26]. Other sources used but not cited elsewhere are [1], [2], [8], [12], [14]–[17], [19], [21], [24].

In context of the theoretical principles, we shall introduce the concepts of two-dimensional systems and operators, two-dimensional transforms, and two-dimensional stochastic fields. The text is conceived to be self-contained: the necessary concepts of the one-dimensional signal theory will be briefly included, however, without detailed derivations. A prior knowledge of the signal theory elements, though definitely advantageous, is thus not necessary. With respect to the purpose of the book, we shall mostly limit ourselves to the two-dimensional case; the generalization to three- and four-dimensional cases is rather straightforward and will be mentioned where necessary.

This theoretical section is subdivided into two similarly structured chapters. The first chapter deals with the theory of images in continuous space, often briefly denoted as analogue images. Besides being necessary as such, because some later derivations will need this concept, the theory seems also to be more easily comprehensible, thanks to intuitive interpretations — what is perceived by humans is the analogue image. The second chapter deals with *discrete images,* i.e., discrete-space (sampled) images, the values of which are also quantized, as only these can be represented in and processed by computers. The reader should realize that, on one hand, there are many similarities and analogies between the worlds of continuous and discrete images; on the other hand, discretization changes some basic theoretical properties of the signals and operators. Therefore, the notion of a "densely enough" sampled discrete image being equivalent to its analogue counterpart is false in principle and often misleading, though the similarities between both areas may occasionally be utilized with advantage; it includes the fact that all real-world images can be represented digitally without any loss in the information content.

1

Analogue (Continuous-Space) Image Representation

1.1 MULTIDIMENSIONAL SIGNALS AS IMAGE REPRESENTATION

1.1.1 General Notion of Multidimensional Signals

The *two-dimensional continuous-space signal* is simply defined as a function of two continuous variables,

$$f(x, y). \tag{1.1}$$

The physical meaning of the function value and variables is arbitrary. In the present context, such a function is usually interpreted as spatially variable brightness (or gray-scale degree) of a planar still image, dependent on the position determined by the two coordinates *x, y* forming the position vector $\mathbf{r} = (x, y)$. With respect to the interpretation of variables, the function is often denoted as a spatial-domain function. Theoretically, it is possible to consider spatially unlimited images, covering the complete infinite plane (x,y); practically, however, image

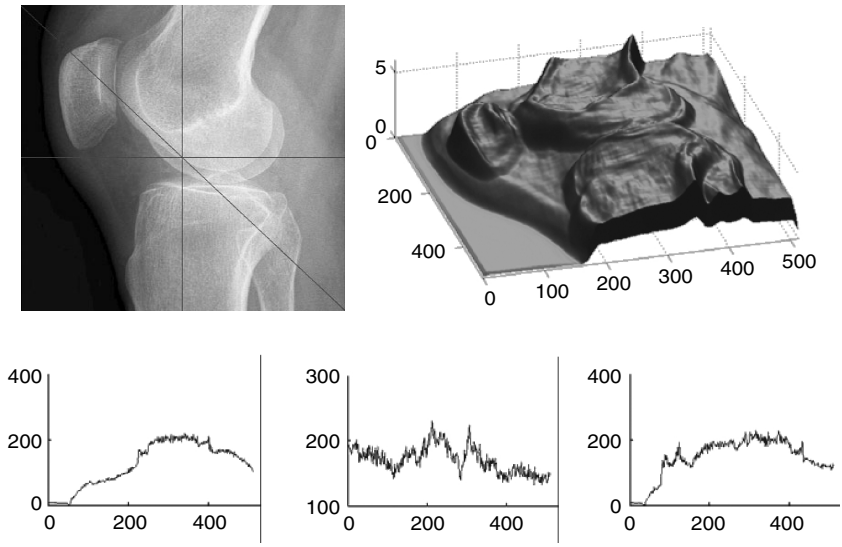


Figure 1.1 Different representations of a two-dimensional signal: gray-scale image, surface representation, and profiles along indicated horizontal, vertical, and diagonal lines.

size (definition extent) is always limited, usually to a rectangle with the origin of coordinates in the center, e.g.,

$$x \in \langle -x_{\max}, x_{\max} \rangle, \qquad y \in \langle -y_{\max,}\, y_{\max} \rangle. \tag{1.2}$$

A function of the type of Equation 1.1 may be represented in different ways, as can be seen in Figure 1.1, either as a gray-scale image or as a surface, the z-component of which is given by the function value, or by means of profiles — one-dimensional functions describing the function values on a line $ax + by + c = 0$, often horizontal or vertical, e.g., for one of the spatial variables constant.

We shall deal mostly with gray-scale images, as they form the output of practically all medical imaging modalities, so that the above definition should suffice. In medicine, color is mostly used only for emphasizing the contrast of originally gray-scale images via false colors or for labeling, both usually using no formally derived algorithms to determine the colors. Nevertheless, it should be mentioned that a color image can be represented by a vector function

$$\mathbf{f}(x,y) = [f_R(x,y), f_G(x,y), f_B(x,y)]^T, \tag{1.3}$$

the components of which describe the brightness of the individual color components, e.g., red, green, and blue. (Differently colored or interpreted triads of components are also used, e.g., brightness, hue, and color saturation.) Each of the components constitutes a gray-scale image and may be treated as such. In real-color image processing, occasionally though still rarely appearing in medical applications, the images may be treated as vector valued, and thus the correlation among the color components utilized.

In the context of *vector-valued signals*, a special case should be mentioned: *complex-valued signals* of the type (in two-dimensional case)

$$\mathbf{f}(x, y) = f_{re}(x, y) + j\, f_{im}(x, y). \tag{1.4}$$

Although natural images are real-valued, the generalization to the concept of complex-valued images is useful theoretically* and may even simplify computations in some cases. Again, each of both parts (real and imaginary) can be interpreted as a gray-scale image.

The two-dimensional concept (Equation 1.1) may be generalized to the case of a *multidimensional signal*:

$$f(\mathbf{x}) \tag{1.5}$$

where $\mathbf{x}$ is a vector with an arbitrary number of components. While $\mathbf{x}$ may even be a scalar (like time in the case of one-dimensional time-dependent signals), in our context it will have mostly two (like above) to four components. The physical meaning of the components depends on the context; a three-dimensional vector may represent three spatial coordinates, $\mathbf{x} = (x, y, z)^{\mathrm{T}}$, as in the case of spatial data provided by some tomographic imaging systems, or two spatial variables and time, $\mathbf{x} = (x, y, t)^{\mathrm{T}}$, in the case of time-variable (moving, animated) two-dimensional planar image. The four-dimensional case means mostly $\mathbf{x} = (x, y, z, t)^{\mathrm{T}}$, related to three-dimensional image data, variable in time (as, e.g., in real-time three-dimensional tomography). The difference between a vector-valued and vector-dependent (multidimensional) signal and the possibility of combining both properties should be well understood.

Obviously, the higher the dimensionality, the more demanding the relevant processing will be regarding both memory requirements and computational complexity aspects. A few years ago,

*Computed images may become complex valued, e.g., in magnetic resonance imaging (MRI).

two-dimensional processing was at the edge of possibilities when more demanding methods were applied, while today, the enormous increase in easily available computational resources enables the solving of complex three-dimensional and even four-dimensional problems. It may be expected that the future development in this direction will enable the processing and analyzing of multidimensional signals by highly sophisticated methods, utilizing complex inner spatial and temporal correlations among the signal components. Presently, such approaches may seem only hypothetical, or even infeasible, but the continuing increase in the available computing power will definitely turn them into practical tools, as may be learned from history.

1.1.2 Some Important Two-Dimensional Signals

Among one-dimensional signals, the *harmonic (cosine) signal* has a basic importance. The two-dimensional analogy of this basic signal, with a similar significance, is the function

$$f(x,y) = A\cos(u_0 x + v_0 y + \varphi), \tag{1.6}$$

which represents a stripy image, an example of which is depicted in Figure 1.2. The interpretation of the function parameters can be arrived at by the following simple analysis. If v_0 is zero, obviously all the image profiles parallel with the x-axis are identical sinusoids with the amplitude A and the (spatial) period $P_x = 2\pi/u_0$, thus forming vertical stripes in the image representation. Correspondingly, the parameter u_0 is denoted (angular) *space frequency*, measured in radians per meter (rad m^{-1}); the corresponding (plain) space frequency $u_0/2\pi$ (m^{-1}) determines how many stripes there are in 1 m in the x-direction. The remaining parameter, phase φ, determines the shift d_x of the nearest maximum (ridge) from the origin of coordinates, as a fraction of the harmonic function period,

$$d_x = P_x \frac{\varphi}{2\pi} = \frac{\varphi}{u_0}. \tag{1.7}$$

The situation for $u_0 = 0$ and $v_0 \neq 0$ is similar, with the image function rotated by 90° with respect to the first case.

If both spatial frequencies are nonzero, the stripes are oblique, with the angle ϑ between the x-axis and the line perpendicular to the stripes being

$$\vartheta = \arctan\left(\frac{v_0}{u_0}\right); \tag{1.8}$$

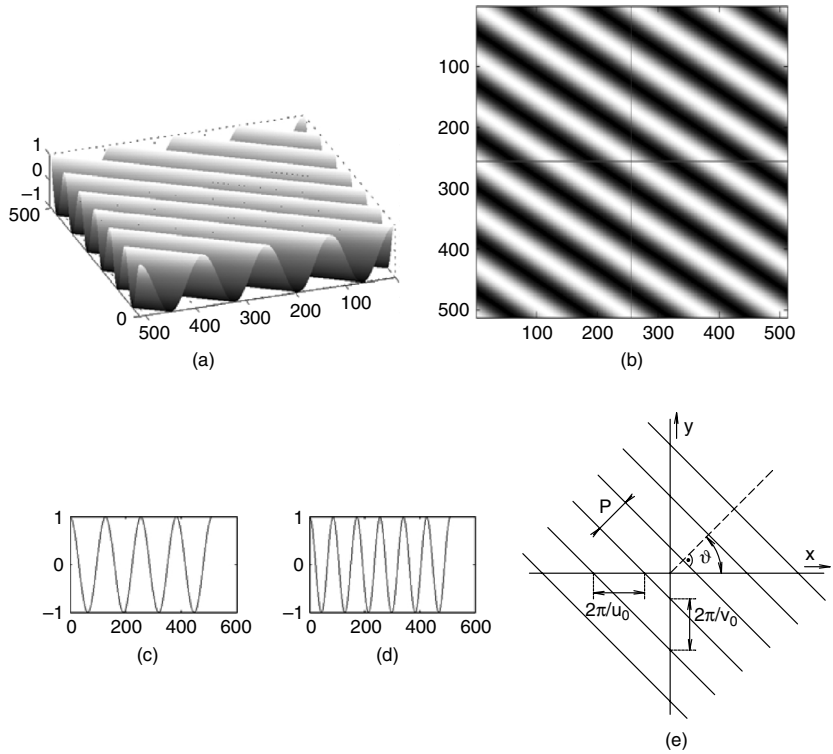


Figure 1.2 Example of a two-dimensional harmonic function in different representations: (a) surface, (b) gray scale, (c) horizontal, and (d) vertical profiles, and (e) schematic representation.

thus, the ratio of both frequencies determines the orientation (direction) of the stripes. This can easily be derived from Figure 1.2 (right, bottom) — the ridges of the stripes are characterized by

$$u_0 x + v_0 y + \varphi = 0 \pm k2\pi, \tag{1.9}$$

which is obviously a family of linear equations describing oblique parallel lines. The distance among the lines — the period P of the harmonic function — is

$$P = 2\pi \Big/ \sqrt{u_0^2 + v_0^2} \tag{1.10}$$

and consequently, the resulting absolute (angular) spatial frequency determining the spatial density of stripes is $w_0 = \sqrt{u_0^2 + v_0^2}$. The phase

φ determines the distance of the nearest ridge to the origin, expressed as a fraction of the period P,

$$d = P\frac{\varphi}{2\pi} = \frac{\varphi}{w}. \tag{1.11}$$

That is, by changing the phase, the stripes are shifted with respect to the origin, perpendicularly to the ridges.

Another particularly important theoretical two-dimensional signal is the δ*-distribution* (the *Dirac impulse*). It is defined as follows:

$$\delta(x, y) = \begin{cases} \infty & \text{for} \quad x = 0 \wedge y = 0 \\ 0 & \text{for} \quad x \neq 0 \vee y = 0 \end{cases} \tag{1.12}$$

while

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \delta(x, y)\mathrm{d}x\,\mathrm{d}y = 1. \tag{1.13}$$

The Dirac impulse may be described as an infinitely narrow impulse at the origin of coordinates, but nevertheless with the weight of 1. The image interpretation is the black x-y plane with a bright infinitesimally small point at the origin, the total luminosity of which is unitary. Such an image may be considered the limit case of a sequence of images formed by a bright square centered on the origin, the total luminosity of which is kept constant while the size is decreasing. This way, the brightness, constant over the square area, is increasing without limits. When the unit-square function is defined as

$$\mathrm{rect}(x, y) = \begin{cases} 1 & \text{for} \quad |x| \leq 0.5, |y| \leq 0.5 \\ 0 & \text{else} \end{cases}, \tag{1.14}$$

then the sequence may be expressed as

$$\delta_m(x, y) = m^2\,\mathrm{rect}(mx, my) \tag{1.15}$$

where each member fulfills Equation 1.13 when $\delta(x, y)$ is substituted by $\delta_m(x, y)$. The limit of the sequence is

$$\lim_{m\to\infty} \delta_m(x, y) = \delta(x, y). \tag{1.16}$$

This notion is very useful even theoretically, as in every phase of the limiting process, the function is piece-wise continuous and finite. Many useful theorems can be derived for piece-wise continuous

finite representation (Equation 1.15) and the results easily extended for the impulse distribution as the limiting case.

The most important properties of the Dirac impulse are:

- *Sifting property*:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\delta(x-\xi,y-\eta)\mathrm{d}x\,\mathrm{d}y = f(\xi,\eta). \tag{1.17}$$

  The proof is straightforward: the integrand is nonzero only for $x = \xi$, $y = \eta$, then $f(\xi, \eta)$ is independent of integration variables and may be set ahead of the integral, which is then equal to 1 according to Equation 1.13.
- Shift to an arbitrary position (ξ, η): $\delta(x - \xi, y - \eta)$, obviously, the impulse function is nonzero only when both input variables are zero.
- Expression via two-dimensional discrete Fourier transform (see below):

$$\delta(x,y) = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-j(ux+vy)}\mathrm{d}u\,\mathrm{d}v. \tag{1.18}$$

This last lemma is left without proof.

1.2 TWO-DIMENSIONAL FOURIER TRANSFORM

1.2.1 Forward Two-Dimensional Fourier Transform

The continuous *two-dimensional Fourier transform* (FT) of a two-dimensional spatial-domain function $f(x,y)$ is a function of two variables $F(u, v)$, defined by the double integral

$$F(u,v) = \mathrm{FT}\{f(x,y)\} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j(ux+vy)}\mathrm{d}x\,\mathrm{d}y. \tag{1.19}$$

The resulting function is often called the *two-dimensional spectrum*. We shall return to the physical interpretation of the spectral values in Section 1.2.3. Let us only say here that the variables u, v are (angular) *spatial frequencies*; thus, the result is a frequency-domain function that may be (and mostly is) complex valued.

The two-dimensional FT may be obtained also by double application of one-dimensional Fourier transform*. The first transform leads to a mixed-domain function

$$F_1(u,y) = \int_{-\infty}^{\infty} f(x,y)e^{-jux}\,\mathrm{d}x, \tag{1.20}$$

while the second transform leads to the frequency-domain spectrum,

$$\begin{aligned} F_2(u,v) &= \int_{-\infty}^{\infty} F_1(u,y)e^{-jvy}\mathrm{d}y = \int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty} f(x,y)e^{-jux}\mathrm{d}x\right]e^{-jvy}\mathrm{d}y \\ &= \mathrm{FT}\{f(x,y)\}. \end{aligned} \tag{1.21}$$

The indicated order of integration has been chosen arbitrarily, and the other one gives the same result.

An example of simple artificially synthetized images and their spectra is shown in Figure 1.3. As the images are formed by a white square or a rectangle on black background, the profiles of the image through the white area are isolated rectangular pulses, and thus the partial one-dimensional FT is of the type $\sin(x)/x$ (see, e.g., [9]). Consequently, as can easily be shown, the respective two-dimensional spectrum of the left upper image is

$$F(u,v) = \mathrm{FT}\{\mathrm{rect}(x,y)\} = \frac{\sin u/2}{u/2}\frac{\sin v/2}{v/2} = \mathrm{sinc}(u/2, v/2), \tag{1.22}$$

which leads to the depicted stripy structure of the amplitude spectrum, when interpreted as a gray-scale image. The phase spectrum is zero due to the image symmetry. The similar image below of a rectangle shows how the width of the original-domain pattern influences the behavior of the spectral function: the narrower the changes in the original (space) domain, the wider and more slowly changing is the spectrum in the respective direction.

Figure 1.4 presents a natural image provided by a standard X-ray examination and its amplitude spectrum and phase spectrum,

*The one-dimensional Fourier transform of a one-dimensional function $f(t)$ is defined as $F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\mathrm{d}t$, where t usually has the physical meaning of time in seconds and ω of (angular) frequency in radians per second; $F(\omega)$ is then called the (complex) spectrum of the (time-dependent) signal $f(t)$. On its properties, see the rich literature on basic signal theory, e.g., [9].

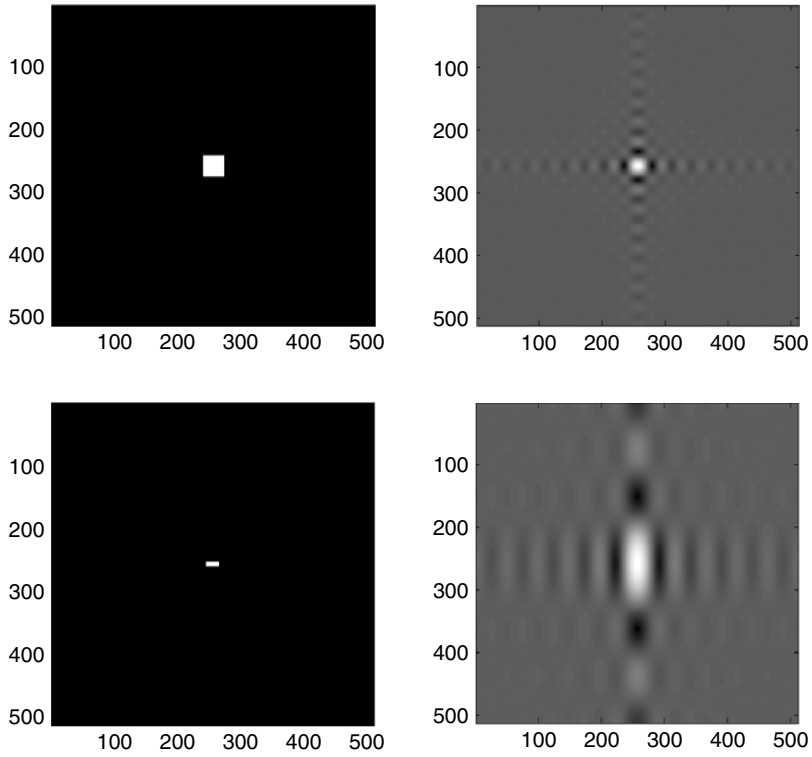


Figure 1.3 Example of two artificial images (left) and their spectra (right). Images and spectra are supposedly unlimited; naturally, only parts can be depicted.

both represented as gray-scale images using the indicated gray scales. Similarly, as in the previous figure, the amplitude spectrum is log-transformed before being converted to gray scale in order to suppress the enormous dynamic range of the amplitudes.

For the transform to exist, $f(x,y)$ must have at most a finite number of extremes and a finite number of discontinuities in any finite area $\Delta x \Delta y$; it must not have infinite discontinuities and must be absolutely integrable over the entire x-y plane. All natural images are represented by functions fulfilling these requirements.

Nevertheless, unnatural images appearing in theoretical considerations need not fulfill the FT existence conditions, but extension of the FT definition is available. Often, such an image function can be shown to be the limit case of a sequence of functions. If each

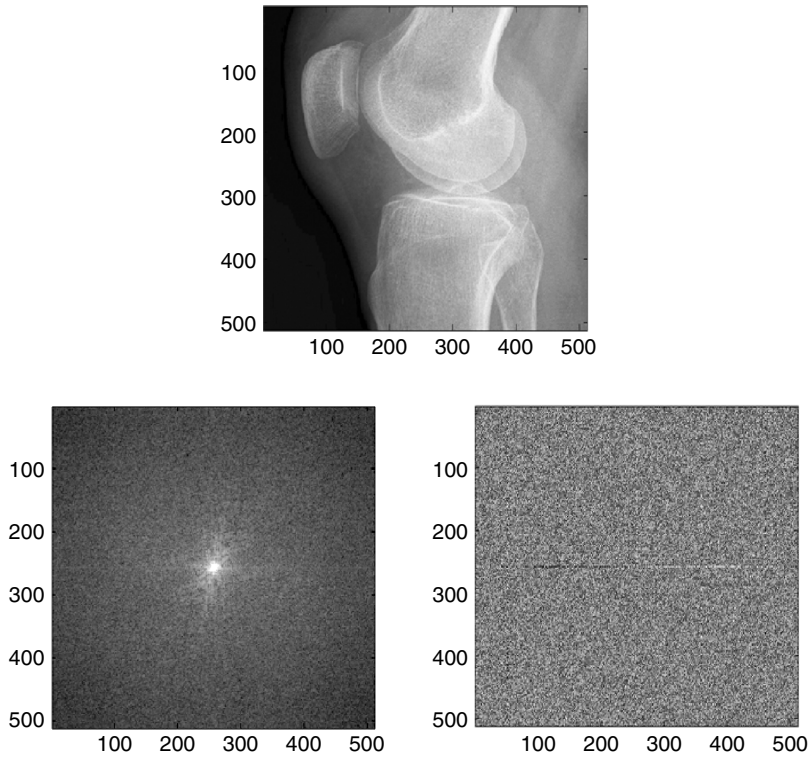


Figure 1.4 Example of a natural image (above) and its amplitude and phase spectra (from left, cropped to a finite size as approximated by two-dimensional discrete Fourier transform (DFT); see Chapter 2).

member function of the sequence fulfills the requirements, then the spectrum of each member can be derived. This way, a sequence of spectra is formulated. If this sequence of spectral functions converges to a limit function, this limit is then regarded the spectrum of the original unnatural image.

The Dirac distribution is an example of an image that does not meet the requirements, as it has an unlimited discontinuity at the origin. However, when we consider the Dirac impulse as the limit case of the sequence (Equation 1.15), the sequence of the respective spectra, derived from Equation 1.22,

$$F_m(u,v) = m^2 \frac{1}{m^2} \operatorname{sinc}(u/2m, v/2m) \tag{1.23}$$

has the limit

$$\lim_{m\to\infty} F_m(u,v) = 1, \tag{1.24}$$

which is the spectrum of the impulse. This can be immediately seen, when realizing that the main lobe of the spectra (Equation 1.23) is widening with m, while its maximum remains equal to 1. This important result says that the spectrum of the two-dimensional impulse function (of a point source in the image) contains all the frequencies equally.

1.2.2 Inverse Two-Dimensional Fourier Transform

The Fourier transform is an inversible transform,

$$f(x,y) = \mathrm{FT}^{-1}\{F(u,v)\} = \mathrm{FT}^{-1}\{\mathrm{FT}\{f(x,y)\}\}. \tag{1.25}$$

It can easily be proved by multiplying both sides of Equation 1.19 by $e^{j(us+vt)}$ and then double integrating along u and v, which gives

$$\begin{aligned}&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j(us+vt)}\mathrm{d}u\,\mathrm{d}v \\ &\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(e^{j(us+vt)}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j(ux+vy)}\mathrm{d}x\,\mathrm{d}y\right)\mathrm{d}u\,\mathrm{d}v.\end{aligned} \tag{1.26}$$

By changing the order of integration and utilizing Equation 1.18, we obtain

$$\begin{aligned}&\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(f(x,y)\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-j(u(x-s)+v(y-t))}\mathrm{d}u\,\mathrm{d}y\right)\mathrm{d}x\,\mathrm{d}y \\ &\quad = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)4\pi^2\delta(x-s,y-t)\mathrm{d}x\,\mathrm{d}y = 4\pi^2 f(s,t).\end{aligned} \tag{1.27}$$

Substituting x, y for s, t provides the formula of the *inverse two-dimensional Fourier transform*:

$$f(x,y) = \mathrm{FT}^{-1}\{F(u,v)\} = \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)\,e^{j(ux+vy)}\mathrm{d}u\,\mathrm{d}v. \tag{1.28}$$

Thus, the inverse transform of the spectrum of a particular original image yields exactly this image.

If the generally complex-valued spectrum was provided by two-dimensional FT of a real-valued image, the result of the inverse transform is naturally also real-valued. On the other hand, every spectrum (providing that it obeys similar requirements as the originals, as mentioned in the previous section) obviously defines an image. If the spectrum is an arbitrarily defined frequency-domain function, the inverse transform may yield a complex-valued spatial-domain function, which may be interpreted as two images — of the real part and of the imaginary part (or alternatively, images of the amplitude and of the phase). The interpretation of such a result then depends on the investigated problem.

1.2.3 Physical Interpretation of the Two-Dimensional Fourier Transform

The two-dimensional Fourier transform may be interpreted in terms of harmonic image components, similarly to its one-dimensional counterpart, which is interpreted in terms of harmonic signal components.

Let us find the spectral representation of the generic harmonic function (Equation 1.6). It may be decomposed, according to Euler's theorem, into two complex conjugate harmonic parts:

$$f(x,y) = A\cos(u_0 x + v_0 y + \varphi) = \frac{1}{2}A[e^{j(u_0 x + v_0 y + \varphi)} + e^{-j(u_0 x + v_0 y + \varphi)}]. \quad (1.29)$$

Because the Fourier transform is linear, each of the parts may be transformed individually and the resulting spectrum of $f(x, y)$ obtained as the sum of the partial spectra. These spectra have to be defined in the sense of a sequence limit, as mentioned above, because the original functions are not absolutely integrable due to periodicity. However, it can be easily shown by the inverse transform that the spectrum $F_1(u,v)$ of the first part $f_1(x,y)$ is an impulse function. Really,

$$\begin{aligned} \mathrm{FT}^{-1}\{F_1(u,v)\} &= \frac{1}{4\pi^2}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[2A\pi^2\delta(u-u_0, v-v_0)e^{j\varphi}\right]e^{j(ux+vy)}\mathrm{d}u\,\mathrm{d}v \\ &= \frac{A}{2}e^{j(u_0 x + v_0 y + \varphi)}, \end{aligned} \quad (1.30)$$

when utilizing the sinking property of δ-distribution. Thus, the term in brackets is the spectrum $F_1(u,v)$ of the first part of the real

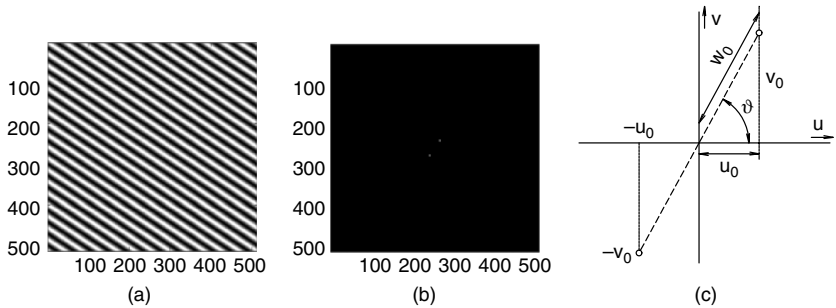


Figure 1.5 Spectrum of a harmonic function: (a) original, (b) amplitude spectrum, and (c) schematic plot (magnified).

harmonic function. Similarly, we also obtain the spectrum of the second part, and the resulting spectrum is

$$\begin{aligned} F(u,v) &= \mathrm{FT}\{A\cos(u_0 x + v_0 y + \varphi)\} = F_1(u,v) + F_2(u,v) \\ &= 2A\pi^2\left[\,\delta(u-u_0, v-v_0)e^{j\varphi} + \delta(u+u_0, v+v_0)e^{-j\varphi}\,\right]. \end{aligned} \tag{1.31}$$

Such a spectrum is schematically depicted in Figure 1.5; it is formed of two impulses with the complex conjugate weights $2\pi^2 A e^{\pm j\varphi}$ located symmetrically with respect to the origin. The absolute value of the weights corresponds to the amplitude of the original harmonic function (i.e., to the contrast of the strips in the image representation). The phase expresses the distance d (Equation 1.11) of the nearest ridge of the image function from the origin of the x-y plane (see Figure 1.2e). The orientation of the stripes is given by the direction of the join of both spectral impulses; the distance of the impulses determines the absolute frequency, i.e., spatial density of stripes. This way, the spectral impulses couple provides all the parameters describing the two-dimensional harmonic function.

Obviously, the continuous spectrum can be understood as an infinite set of such pairs of impulses, each of them describing parameters of a different harmonic function. The limit sum — i.e., the integral of all these (infinitesimally small) harmonic components — is therefore the original image function $f(x, y)$. This is exactly the physical interpretation of Equation 1.28 for the inverse Fourier transform. Here, the real (cosine) harmonic functions are expressed via Equation 1.29 by summed pairs of complex harmonic functions that are the base functions of the transform. Obviously, the values

in centrally symmetrical pairs of the spectrum must be complex conjugate, should the original image be real-valued,

$$F(u,v) = F^*(-u,-v). \tag{1.32}$$

Naturally, this relation does not apply generally in the case of complex-valued images when the spectrum may be asymmetrical.

Another important property of the two-dimensional spectrum can be shown quite visually. As already mentioned and can be seen in Figure 1.2e and Figure 1.5c, the angle of the direction of stripes (perpendicular to ridges) with respect to spatial axes in the original image and the orientation of the join of the impulse pair relative to the frequency axes are identical. This applies to an arbitrary harmonic function and the corresponding pair of spectral impulses. As any image can be regarded an (infinite) sum of harmonic components or, alternatively, may be represented by corresponding spectral pairs, it is obvious that rotation of the original image by an angle corresponds, in the frequency domain, to the rotation of the two-dimensional spectrum by the same angle. Let us note that this implies rotational symmetry of the two-dimensional spectrum in the case of the rotationally symmetric image (e.g., isotropic point-spread function (PSF) — see later)*.

Let us refer preliminarily to Chapter 2, where we shall see that the discrete two-dimensional Fourier transform is interpreted quite similarly, with the only difference being that there is a finite number of the base harmonic functions, and thus the integrals are replaced by a plain double sum of a finite number of components.

1.2.4 Properties of the Two-Dimensional Fourier Transform

The main properties of the two-dimensional FT are as follows:

- Linearity:

$$\mathrm{FT}\left\{\sum_i a_i f_i(x,y)\right\} = \sum_i a_i \mathrm{FT}\{f_i(x,y)\} = \sum_i a_i F_i(u,v) \tag{1.33}$$

*Obviously, both rotationally symmetric functions, the image, and the corresponding spectrum may then be described by a one-dimensional function — the radial density: $f(x,y) = g(r)$, $F(u,v) = G(w)$ with $r = \sqrt{x^2 + y^2}$, $w = \sqrt{u^2 + v^2}$. It can be shown that the relation of the original function and the spectrum is given by the Hankel transform of the first-order, $G(w) = \int_0^\infty g(r) r J_0(wr) dr$, where $J_0(x)$ is the Bessel function of the first kind and index 0.

- Change of scale:

$$\mathrm{FT}\{f(ax,by)\} = \frac{1}{ab} F\left(\frac{u}{a},\frac{v}{b}\right) \tag{1.34}$$

(Compare the spectra of the square and the rectangle in Figure 1.3.)

- Shift:

$$\begin{aligned} &\mathrm{FT}\{f(x-a,y-b)\} = F(u,v)e^{-j(ua+vb)} \\ &\mathrm{FT}\{f(x,y)e^{-j(u_0x+v_0y)}\} = F(u-u_0,v-v_0) \end{aligned} \tag{1.35}$$

- Rotation by 180°:

$$\mathrm{FT}\{\mathrm{FT}\{f(x,y)\}\} = f(-x,-y) \tag{1.36}$$

- *Convolutional property*:

$$\begin{aligned} &\mathrm{FT}\{f(x,y) * g(x,y)\} = F(u,v)G(u,v) \\ &\mathrm{FT}\{f(x,y)\ g(x,y)\} = \frac{1}{4\pi^2} F(u,v) * G(u,v), \end{aligned} \tag{1.37}$$

where the convolution of two two-dimensional functions is defined as

$$f(x,y) * g(x,y) = f * g|(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s,t)g(x-s,y-t)\mathrm{d}s\ \mathrm{d}t. \tag{1.38}$$

(The second symbolic expression for the two-dimensional result of the convolution operation is formally more adequate, though less frequently used than the left-most one.) On the physical interpretation of convolution, see later. The convolutional property says that the product in the spectral domain corresponds to the convolution in the spatial domain and vice versa. It is the fundamental feature of the two-dimensional FT; most linear analysis of image processing is based on it. The proof of this crucially important theorem is discussed in Section 1.3.3, where it is supported by immediate physical interpretation.

- *Parseval's theorem* for generally complex image functions:

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)g^*(x,y)\mathrm{d}x\ \mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)G^*(u,v)\mathrm{d}u\ \mathrm{d}v. \tag{1.39}$$

This simplifies in case of a single image, i.e., $f(x, y) = g(x, y)$, to the *law of total energy preservation*,

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |f(x,y)|^2 \, \mathrm{d}x \, \mathrm{d}y = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} |F(u,v)|^2 \, \mathrm{d}u \, \mathrm{d}v, \tag{1.40}$$

as the square of the signal or spectrum value may be formally regarded as local energy.

- Spectrum of the correlation integral:

$$\mathrm{FT}\{C_{fg}(x,y)\} = F^*(u,v)G(u,v), \tag{1.41}$$

where the correlation integral $C_{fg}(x,y)$ is defined as

$$C_{fg}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^*(s-x, t-y)g(s,t)\mathrm{d}s \, \mathrm{d}t; \tag{1.42}$$

the integrand is simply the product of values of images f, g mutually shifted by (x,y). It finds its use as an estimate of the cross-correlation function in correlation analysis of homogeneous and ergodic fields, as a similarity measure and also in derivation of certain theoretical concepts (see later). When $f(x,y) = g(x,y)$, the autocorrelation integral is computed and Equation 1.41 becomes, obviously,

$$\mathrm{FT}\{C_{ff}(x,y)\} = |F(u,v)|^2. \tag{1.43}$$

Although this might be (and erroneously often is) considered the Wiener–Khintchin theorem — the relation between the autocorrelation function and the power spectrum, this is, strictly taken, not the case*.

*The right-hand side is not the power spectrum of a stochastic field (see later), but only an individual power spectrum of a particular realization $f(x,y)$ of the field. Correctly, the mentioned theorem expresses a relation between the ensemble mean values on both sides taken over all possible realizations. Only in the special case of homogeneous and ergodic fields, when Equation 1.42 really becomes the autocorrelation function, can Equation 1.43 be considered the theorem.

The proofs of the above theorems are not difficult; however, except for the derivation of the convolution theorem in Section 1.3.3, they will be omitted in favor of brevity.

1.3 TWO-DIMENSIONAL CONTINUOUS-SPACE SYSTEMS

1.3.1 The Notion of Multidimensional Systems

In analogy with one-dimensional systems, working in continuous time and accepting one-dimensional (time-dependent) signals as inputs while producing other one-dimensional signals as outputs, *two-dimensional systems* are defined as working with two-dimensional signals. In the frame of this book, the signals concerned will mostly be two-dimensional still images, of which both independent variables are spatial; such systems then are called *image processing systems*. Naturally, the concept can easily be generalized to multidimensional cases, with the third dimension again a spatial variable (in case of three-dimensional spatial image data processing) or time, when systems processing time-variable planar images are considered. Four-dimensional systems are usually interpreted as processors of time-variable three-dimensional image data.

In the following paragraphs, we shall limit ourselves to two-dimensional systems with a single input and a single output, as schematically depicted in Figure 1.6. The system action can be described, without any detailed information on the internal structure of the system, in terms of the external signals by means of the *input–output description*:

$$g(x,y) = \mathrm{P}\{f(x,y)\}; \tag{1.44}$$

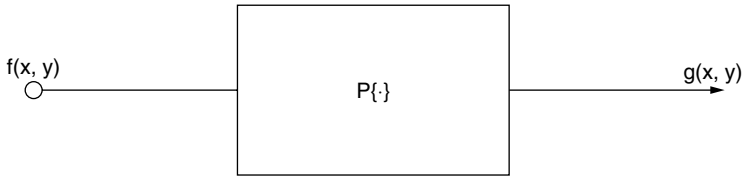


Figure 1.6 Two-dimensional image processing system — input–output representation.

the input image $f(x,y)$ is processed by the operator P, providing the output image $g(x,y)$. The character of the operator may be arbitrary; below, an attempt to show several classification views can be found. Let us note that in analogy with the dynamic one-dimensional systems, where both input and output signals are functions of time, even in the case of two-dimensional systems there is usually no necessity to distinguish between input-plane and output-plane coordinates.

Every operator is characterized by the *definition domain*, formed by the set of acceptable input images $\{f_1(x,y), f_2(x,y),\ldots, f_N(x,y)\}$, by the *output domain* (the set of possible output images), $\{g_1(x,y), g_2(x,y),\ldots, g_M(x,y)\}$, and by the *correspondence rule*, assigning an output to each acceptable input. The extent N of the input set as well as of the output set (M) may be finite, or one or both may be infinite. If both sets are finite, it would be possible, at least theoretically, to construct a correspondence table; however, to realize the operator this way is usually a clumsy method, and naturally not applicable to the case of infinite input or output domain. Usually, the operator is described by means of a formula or an algorithm that describes how to derive the output image based on values of the input image. Properties of the formula or algorithm determine the classification of the operator.

There are several classification aspects; the two-dimensional systems may be classified according to,

- Linearity: The system is *linear* if the operator fulfills the *principle of superposition*,

$$\mathrm{P}\left\{\sum_f a_i f_i(x,y)\right\} = \sum_i a_i \mathrm{P}\{f_i(x,y)\}, \tag{1.45}$$

  where a_i are arbitrary coefficients. All the systems not fulfilling this equality are nonlinear. Though most of the natural systems are, strictly taken, nonlinear, there is a common tendency to analyze them approximately as linear, since the methodology of linear system analysis and synthesis is highly evolved, and the needed formalism belongs to basic engineering tools. The nonlinear methods are more difficult to design and analyze, and there is no generic unifying theory of nonlinear processing (this probably cannot even be due to the extreme scope of, and diversity in, the nonlinear class). However, the nonlinear

operators yield in many cases better results and some special classes are therefore topical subjects of research.

- Extent of input data per output point*:
  - *Point-wise operators*: Each output point is influenced only by the correspondingly placed input point value; such operators are mostly nonlinear and serve to contrast (or color) transforms.
  - *Local operators*: An output point is influenced by values of a certain (local) area surrounding the corresponding input point; these operators are most frequently used in basic image processing for image enhancement, like sharpening and noise reduction, and for basic analysis like edge detection or morphological operations.
  - *Global operators*: Each output point is determined by all values of the input image; this is the most generic concept, often used in theoretical considerations and in advanced methods of image reconstruction or restoration. Also, integral transforms producing two-dimensional spectra belong to this class.
- Variability of the operator in space:
  - *Space-invariant systems*: The operator is invariable on the entire space (i.e., *x-y* plane in two-dimensional systems); e.g., convolutional systems are of this kind (see later).
  - *Space-variable systems*: The operator changes its properties when working in different parts of the space (*x-y* plane); e.g., adaptive filters are of this kind.
- Isotropy:
  - *Isotropic systems*: The realized operator is isotropic with respect to processed images.
  - *Anisotropic systems*: The operator is direction dependent.
- Dimensionality: The common image processing systems are obviously two-dimensional; the *system dimensionality* is determined by the number of dimensions of the processed signals, as mentioned above.

---

*This aspect corresponds to the notion of inertiality in dynamic systems working in the time domain. However, as there is nothing like natural flow of time in spatial variables, the concept of causality, necessary in the dynamic systems, is meaningless in image processing. If introduced artificially, as in recursive image processing, it enforces an unnatural anisotropy.

1.3.2 Linear Two-Dimensional Systems: Original-Domain Characterization

Linear methods form a substantial part of the image processing methodology, and the corresponding theory is well elaborated. The linear methods and corresponding operators realized by linear continuous-space systems constitute a formalized basis of many simple and advanced approaches, or are used at least as standards for comparison. In the following paragraphs, we shall deal with them in a certain degree of detail.

In order to express the properties of a generic linear system by means of clearly defined characteristics, we shall use the equality based on the sinking property of the Dirac impulse,

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s,t)\delta(x-s,y-t)\mathrm{d}s\,\mathrm{d}t. \tag{1.46}$$

According to this expression, the input image can be obviously interpreted as a limit sum (integral) of infinitesimally small impulse components.

When this expression is substituted for $f(x, y)$ into Equation 1.44, we obtain

$$g(x,y) = \mathrm{P}\left\{\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s,t)\delta(x-s,y-t)\mathrm{d}s\,\mathrm{d}t\right\}. \tag{1.47}$$

As the operator P is supposed to be linear, as is the operator of integration, they may be interchanged according to the principle of superposition, yielding

$$\begin{aligned} g(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \mathrm{P}\{f(s,t)\delta(x-s,y-t)\}\,\mathrm{d}s\,\mathrm{d}t \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s,t)\mathrm{P}\{\delta(x-s,y-t)\}\mathrm{d}s\,\mathrm{d}t, \end{aligned} \tag{1.48}$$

because $f(s, t)$ is, for particular values of integration variables, a constant factor independent of x and y.

The term

$$h(x,y,s,t) = \mathrm{P}\{\delta(x-s,y-t)\} \tag{1.49}$$

obviously represents the response of the analyzed system to the input image formed by the Dirac impulse situated at (s, t); therefrom the name of the h-function — the *impulse response* or the *point-spread function* (PSF) of the system. It is naturally a two-dimensional function of x, y, as follows from Equation 1.44; besides that, its shape obviously depends on the impulse position in the input image, too. In the general case, a two-dimensional linear system is thus described by an infinite number of differing impulse responses.

The output (Equation 1.48) can be rewritten as

$$g(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x,y,s,t) f(s,t)\,\mathrm{d}s\,\mathrm{d}t. \tag{1.50}$$

This important result is usually denoted as a *superposition integral* with an apparent physical interpretation: the value $g(x,y)$ at each point (x,y) of the resulting image is additively contributed to by every point (s,t) of the input image, the particular (infinitesimal) contribution being proportional to the input value $f(s,t)$ weighted by the factor $h(x, y, s, t)$ influenced by both positions, of the contribution source and of its target. Taken this way, the function $h(x, y, s, t)$ may be considered a four-dimensional weight function — a complete, but rather complicated description of a generic linear two-dimensional system.

An important special class of systems is formed by *isoplanar* (*space-invariant*) *systems*, whose impulse response has the special form

$$\mathrm{P}\{\delta(x-s, y-t)\} = h(x-s, y-t). \tag{1.51}$$

The shape of such a point-spread function is evidently invariable with respect to the position of the input impulse; only the position of the whole function is shifted by the same vector (s, t) in the output plane as is the source impulse in the input plane from the coordinate origin. The generally four-dimensional impulse response function degenerates in case of isoplanar systems to a two-dimensional function — a substantial and practically very welcome simplification.

The superposition integral simplifies for isoplanar systems to

$$\begin{aligned} g(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-s, y-t) f(s,t)\,\mathrm{d}s\,\mathrm{d}t \\ &= h(x,y) * f(x,y) = h * f|(x,y), \end{aligned} \tag{1.52}$$

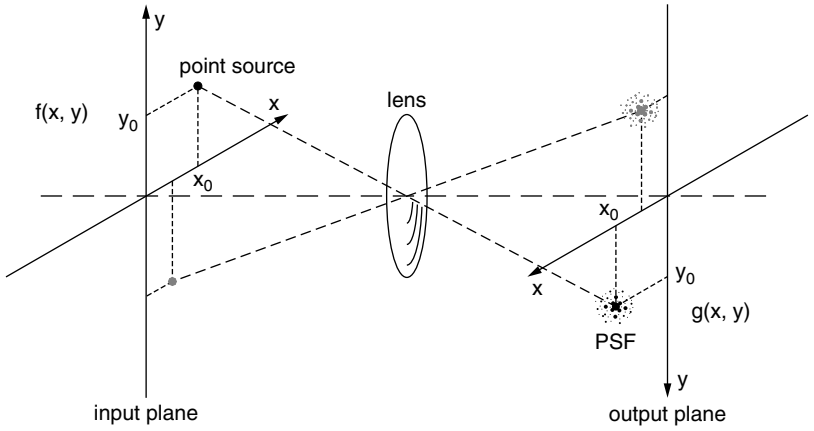


Figure 1.7 A scheme of a simple (optical) two-dimensional system — forming of the output image as a sum of weighted and shifted impulse responses.

which is the *convolution integral* describing the already mentioned operation of *convolution* between two images, as symbolized by the star operator. We can thus conclude this derivation with the statement that the output of a linear space-invariant two-dimensional system is the two-dimensional convolution of its PSF and the input image.

This last result can be physically interpreted as forming the output image additively from the (infinitely many infinitesimally weak) point-spread functions of identical shape, but differently weighted and spatially shifted, as schematically depicted for two input points in Figure 1.7. Each point of the input image may be considered an impulse source with the weight $f(x, y)$, to which the system responds with the PSF, weighted by the value of $f(x, y)$ and correspondingly shifted — its origin located at the point x, y in the output image. These individual responses are then summed into the result so that the output image is assembled of these weighted PSFs. This clearly shows that the system influences (and mostly deteriorates) the spatial resolution in the image — a point is imaged as a PSF wisp; a sharp edge (gray value step) becomes a correspondingly smeared band with a continuous transition from darker to brighter area.

Theoretically, the PSF may cover the complete x-y plane, justifying the above statement of the influence of each input point to every output point. However, in practical imaging or image processing systems, the PSF has important values usually only for small arguments; i.e., the size of PSF is limited. Clearly, it may be then stated

that a point (x, y) in the output image is influenced only by a small area, with the size of the same order as the PSF size, around the point (x, y) in the input image. This enables the realization or description of most of the convolution tasks by the simpler local operators, instead of global ones.

1.3.3 Linear Two-Dimensional Systems: Frequency-Domain Characterization

Both system signals—the input and output images—may be equivalently described in the frequency domain by their spectra. It will be shown now that a simple description in the frequency domain is possible also for the system itself, as far as it is linear and space invariant.

When both sides of the original-domain description of such a system (Equation 1.52) are two-dimensionally Fourier transformed, we obtain

$$\begin{aligned} G(u,v) &= \mathrm{FT}\{h * f\,|(x,y)\} \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-s,y-t)f(s,t)\,\mathrm{d}s\,\mathrm{d}t\right] e^{-j(ux+vy)}\,\mathrm{d}x\,\mathrm{d}y \end{aligned} \tag{1.53}$$

and, by interchanging the order of integration,

$$G(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s,t)\left[\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-s,y-t)e^{-j(ux+vy)}\,\mathrm{d}x\,\mathrm{d}y\right]\mathrm{d}s\,\mathrm{d}t. \tag{1.54}$$

The bracketed term is evidently the two-dimensional Fourier transform of the function $h(\ldots)$ shifted by the vector (s, t), which can be expressed according to Equation 1.35 by the spectrum $H(u, v)$ of $h(x, y)$, multiplied by a phase factor corresponding to the shift,

$$G(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s,t)H(u,v)e^{-j(us+vt)}\,\mathrm{d}s\,\mathrm{d}t. \tag{1.55}$$

After setting $H(u, v)$ out of the integral, as independent of integration variables, we see that the remaining integral is the two-dimensional FT of the original image $f(\ldots)$, so that finally,

$$G(u,v) = H(u,v)F(u,v). \tag{1.56}$$

Let us note that this derivation may serve as a proof of the general convolutional property (Equation 1.37) of the two-dimensional Fourier transform, when only substituting symbols.

The last result has a simple but important physical interpretation. The spectrum $G(u, v)$ of the output image is given by the product of the input spectrum $F(u, v)$ and the spectrum $H(u, v)$ of the system impulse response $h(x, y)$. The last spectrum is the required frequency-domain characteristic of the system, called the *frequency transfer function*, or the *frequency response*. Its physical meaning is obvious: for each spatial frequency (u, v) (a combination of directional spatial frequencies), the transfer function value at this frequency determines how a particular harmonic component with these frequencies will be transferred via the system. In greater detail, the amplitude $|H(u, v)|$ specifies the degree of amplification (or attenuation) of the particular harmonic component by the system, i.e., the ratio of the output image content at this frequency with the content of input. The phase $\mathrm{Arg}(H(u, v))$ determines how this component is phase shifted, i.e., how the real-valued stripy structure corresponding to a symmetrical pair (u, v) and $(-u, -v)$ is shifted with respect to the input, both relative to origin of the x-y plane (see Section 1.1.2).

Examples of the output images derived from an original image by three linear isoplanar systems (filters) can be seen in Figure 1.8. The frequency responses of the filters are on the last row, while the corresponding derived images in which some frequencies were suppressed are on the previous line. The first system is a low-pass filter suppressing high frequencies so that the details are lost. On the contrary, the second system, a high-pass filter, suppresses the low frequencies carrying the information on brightness of greater areas; only the edges represented by high frequencies remain. The last example is a narrowband filter allowing only frequencies close to a certain mean frequency — the smaller the passband areas, the closer is the result to a single harmonic function.

1.3.4 Nonlinear Two-Dimensional Continuous-Space Systems

As mentioned in Section 1.3.1, the multidimensional systems described generally by Equation 1.44 may be nonlinear. Several types of the nonlinear systems are routinely applied in image processing; of them, two deserve explicit mention in the continuous-space context: *point operators* providing contrast or color transforms,

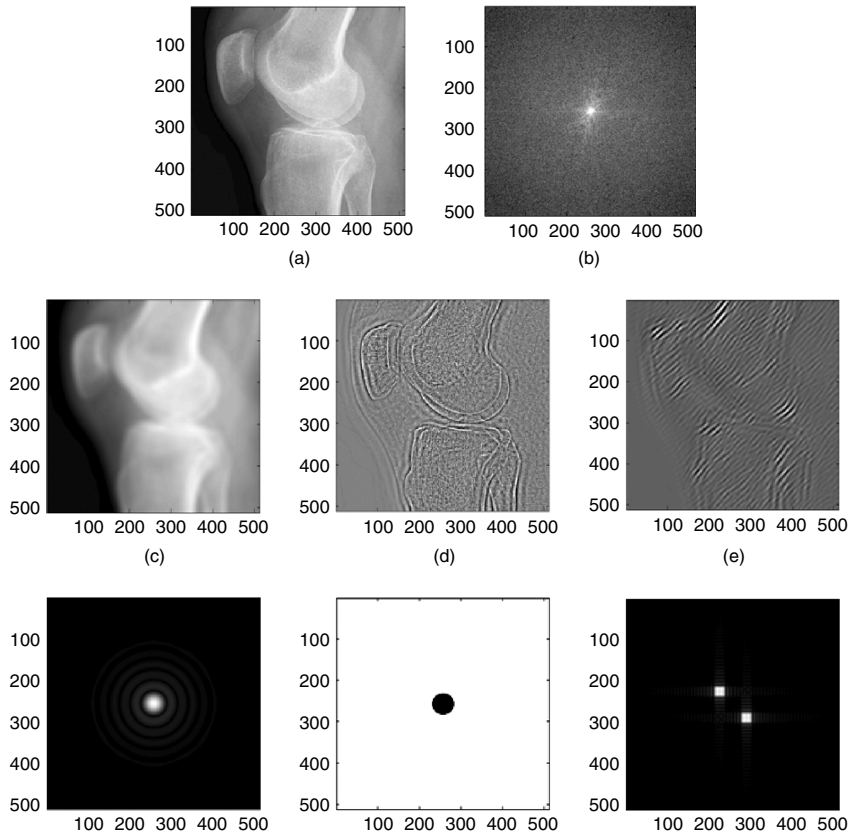


Figure 1.8 Influence of the system frequency response to image transfer: (a) original image; (b) original amplitude spectrum; (c) low-pass-filtered, (d) high-pass-filtered, and (e) narrowband-filtered versions with the corresponding frequency responses below.

and *homomorphic systems*. The two generic concepts are briefly dealt with below. Some of the commonly used nonlinear operators, such as order statistics filters, are more easily defined in discrete-space representation; these will be treated in Section 2.2.2.

1.3.4.1 Point Operators

Point (point-wise) operators, as the name suggests, are simple nonlinear operators of the type

$$g(x,y) = \mathcal{N}_{\mathbf{r}}(f(x,y)), \qquad \text{or} \qquad \mathbf{g}(x,y) = \overline{\mathcal{N}}_{\mathbf{r}}(f(x,y)) \tag{1.57}$$

so that the output value is solely dependent on the value of the corresponding point in the input image. Here $\mathcal{N}_{\mathbf{r}}(x, y)$ is an unambiguous (but not necessarily inversible) function of a single variable, the definition extent of which is given by the range of the input image values, while the range of the function values corresponds to the image values of output. The left equation describes a *gray-scale transform* (also often called *contrast transform*), while the right equation, where the vector function value represents a color $\mathbf{g} = [g_R, g_G, g_B]$, leads to a *false color presentation* of the original gray-scale image. The index $\mathbf{r} = (x, y)$ at the function symbol represents the possibility of the $\mathcal{N}_{\mathbf{r}}$-function form being space variable, thus providing an *anisoplanar* contrast (or color) transform. The function forms may be given *a priori* by concrete requirements on the spatial variability, or they may depend on the local properties of a particular image. In this case, the operators are called *adaptive contrast (or color) transforms*. Obviously, if the argument in the right relation of Equation 1.57 was a vector $\mathbf{f}(x, y)$ representing a color, the operator would represent a color transform of color images, i.e., a mapping from the three-dimensional input color space to a similar output space.

Though point operators are generally conceptually simple and easy to execute, they constitute a powerful and very important class of operators used in image processing, and in medical applications particularly. The effect of contrast or color transforms on the image readability as far as important feature detection concerns may be dramatic, thus substantially influencing the diagnostic value. An illustrative example of isoplanar contrast enhancement can be seen in Figure 1.9.

(a) (b)

Figure 1.9 Example of contrast enhancement: (a) original and (b) transformed image.

Different methods of contrast transform aimed at specific tasks in medical imaging will be described in Section 11.1 on image contrast enhancement.

1.3.4.2 Homomorphic Systems

The homomorphic systems are, in a sense, a generalization of linear systems in that they fulfill the *generalized superposition principle*.

Let M{…} be a generally nonlinear two-dimensional operator, realizing the transform $g(x,y) = \mathrm{M}\{f(x,y\}$. A system implementing such an operator (or the operator itself) is called *homomorphic* if it fulfills the generalized principle of superposition

$$\begin{aligned} \mathrm{M}\{f(x,y) @ g(x,y)\} &= \mathrm{M}\{f(x,y)\} \,\$\, \mathrm{M}\{g(x,y)\} \\ \mathrm{M}\{c \;\&\; f(x,y)\} &= c \,\#\, \mathrm{M}\{f(x,y)\}, \end{aligned} \tag{1.58}$$

where @ and $ are some, for a concrete system given, *characteristic operations* among two-dimensional image functions; @ is called the *input operation*, while $ is the *output operation* of the system. The & and # symbols are some operations between a scalar and a two-dimensional function that are fixed for a particular system. A homomorphic system can thus be schematically represented as in Figure 1.10, where the input and output operations are indicated.

It is easy to show that linear systems in fact form a special class of the homomorphic systems: replacing both characteristic operations by addition, and the & and # operations by multiplication, leads to the principle of superposition (Equation 1.45). Another example of a homomorphic system is the nonlinear point operator $g(x,y) = \exp(f(x,y))$, for which addition is the input characteristic

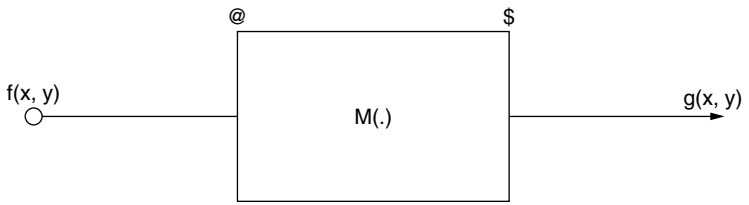


Figure 1.10 Schematic representation of a homomorphic system with indicated characteristic operations. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

operation and multiplication is the output operation; then & means multiplication and # raising to a power.

The homomorphic systems constitute a rather generic class of nonlinear systems, the main advantage of which is that they can be decomposed into a cascade of a linear part and some, usually not too complicated, nonlinear parts. This is enabled via *canonical realization*. This notion is based on the supposition that for each characteristic operation, e.g., @, it is possible to find a homomorphic subsystem with the inversible operator φ (with the same definition range), the output operation of which is addition:

$$\begin{aligned} \varphi\{f(x,y)@\, g(x,y)\} &= \varphi\{f(x,y)\} + \varphi\{g(x,y)\}, \\ \varphi\{c\ \&\ f(x,y)\} &= c\ \varphi\{f(x,y)\}. \end{aligned} \tag{1.59}$$

Though it is sometimes not easy to find such an operator for a given @, it turns out that it should always be possible. Obviously, the system, realizing the inverse operator, would fulfill

$$\begin{aligned} \varphi^{-1}\{f(x,y)+g(x,y)\} &= \varphi^{-1}\{f(x,y)\}@\,\varphi^{-1}\{g(x,y)\}, \\ \varphi^{-1}\{c\ f(x,y)\} &= c\ \&\ \varphi^{-1}\ \{f(x,y)\}. \end{aligned} \tag{1.60}$$

Similarly, it is supposed that an analogous subsystem with the inversible operator ψ is available to each output operation \$. A cascade of two systems with φ and φ^{-1} naturally has a unit transfer and may be included anywhere in the signal path without changing the overall system properties. Thus, the scheme of Figure 1.10 may be expanded to an equivalent scheme, as on the upper part of Figure 1.11. The three inner blocks here may be considered another subsystem, as indicated by the dashed rectangle; this subsystem is obviously linear, as both the input and output operations are additions. As a result, we arrive at the representation on the bottom of the figure, consisting of a linear core subsystem surrounded by two generally nonlinear subsystems that provide for conversion of the input characteristic operations to addition, and vice versa for the output operation. Such a representation is advantageous in separating all the nonlinearity of the system to outer blocks, so that the inner block may be designed using the large body of linear system theory. A class of homomorphic systems may then be defined that is marked out by concrete outer blocks, i.e., concrete input and output operations, while the linear core may be modified, diversifying the systems belonging to the same class.

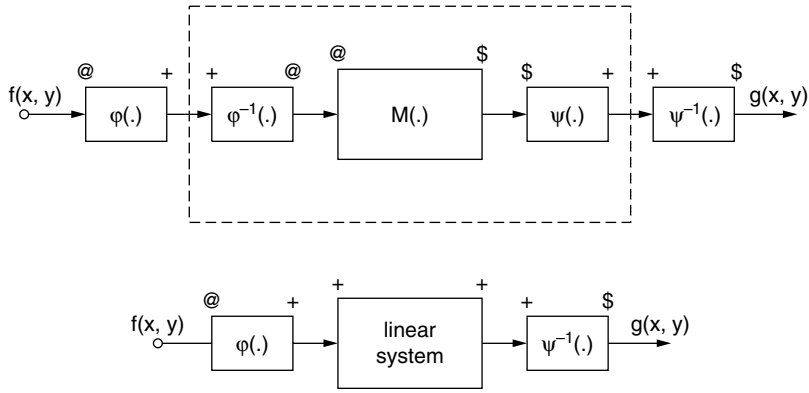


Figure 1.11 Canonical realization of a homomorphic system. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

When the input and output operations are point-wise (in analogy to memoryless one-dimensional systems), the outer subsystems are point-wise as well, so that all the system memory is then concentrated in the linear part. This may further simplify the design and analysis of the system.

In the area of image processing, the homomorphic systems are used primarily for filtering in cases when the input image is a mixture of two (or more) components to be separated, which are mixed by an operation different from addition. The generic schematic representation on the lower part of Figure 1.11 still applies, with the only difference of the outer blocks being inversion of each other, as shown in Figure 1.12. The input block thus converts the nonadditive mixture to an additive one, suitable for linear filtering that yields (when well designed) just the component that is needed, while the undesirable

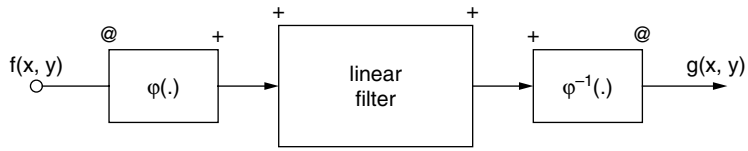


Figure 1.12 Generic two-dimensional homomorphic filter. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

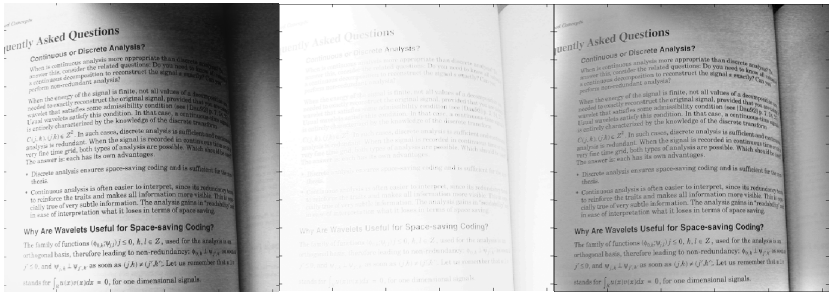


Figure 1.13 Example of an unevenly illuminated image, its log-transformed version with illumination component removed by low-frequency component suppression, and the resulting image.

components are suppressed. The linear system output, however, is distorted by the input conversion, and the rectifying inverse conversion that is provided by the output block is necessary.

A concrete simple application example is the image filtering aiming to suppress influence of nonuniform illumination in an image where the useful information is in two-dimensional reflectivity of the objects (Figure 1.13). Obviously, the measured image is the product of illumination and reflectivity. It may be supposed that the illumination component varies relatively slowly on the x-y plane, the lack of fast changes causing its spectrum to be limited only to rather low spatial frequencies. However, applying a linear low-frequency attenuating filter would fail, as the components are not linearly separable. In contrast, the homomorphic approach converts, by the input logarithmic transform φ, the composite image first into an additive mixture, which can be handled by the linear filter. The output of the linear filter contains primarily only the reflection component, but logarithmically distorted. This must be corrected for by the inverse exponential output conversion φ^{-1}. The cautious reader will notice that, exactly taken, the linear filter must be designed with respect to the properties of the spectra of log-converted components, not the original ones. However, in the mentioned example, like in many other cases suitable for homomorphic filtering, the formulation of the spectral properties is vague anyway, and the differences between the spectra of converted and nonconverted components are not critical. A smooth and monotonous character of the nonlinear transform supports this. The linear filter frequency response must be designed to be smoothly variable,

without any dramatic cutoffs, as there is no distinct frequency border between the useful and disturbing components.

Another example concerns the suppression of disturbance of the useful image by convolution with an almost unknown smearing two-dimensional function. Then the convolutive mixture $g(x, y) = f(x, y) * h(x, y)$ is to be converted into the additive one. Theoretically, this can be done by applying the cascade of the two-dimensional Fourier transform providing the multiplicative mixture $G(u, v) = F(u, v)H(u, v)$, followed by the previously mentioned logarithmic transform, and finally by the inverse two-dimensional FT, returning the image to the original domain (where it is subsequently filtered). This way, the convolutive mixture is converted into the two-dimensional *complex cepstrum*

$$\begin{aligned} g'(x,y) &= \mathrm{FT}^{-1}\{\log(G(u,v))\} \\ &= \mathrm{FT}^{-1}\{\log(F(u,v))\} + \mathrm{FT}^{-1}\{\log(H(u,v))\}, \end{aligned} \tag{1.61}$$

consisting of two linearly separable image-domain components, each derived from one of the input components. Obviously, the cepstral converter is not memoryless, and thus contains a part of the system memory; also, the proper definition of the logarithm of the complex-valued argument must be observed. Nevertheless, the possibility of converting the convolutive signal into the additive one is interesting, as it enables, at least in principle, suppression of the smearing component not by inverse (or modified inverse, e.g., Wiener) filtering, but, in the cepstral domain, by suitable linear filtering forming the inner part of a homomorphic filter. Such a filter can be designed based on much weaker problem identification than the inverse filter.

Naturally, the homomorphic filters are realized mostly digitally, but the principle is not limited to such a realization.

1.4 CONCEPT OF STOCHASTIC IMAGES

So far, we dealt with deterministic images defined by concrete functions of the type $f(x, y)$, so that every image was defined and treated separately. However, it is a frequent situation in practice that images similar in some aspects form classes that may be characterized by certain common properties. It may be advantageous then to design image processing and analyzing procedures not as image specific, but rather as suitable for the class as a whole.

A stochastic field, generating a class of images, is a probabilistic concept that finds frequent use in image filtering, restoration, and

analysis. A stochastic field generates images based on a usually large image set, of which one image is chosen accidentally so that the result—sc., *realization of the field*—can be considered *a priori* a *stochastic image* (before it is known). Naturally, once an image function is generated this way, it represents a deterministic image; nevertheless, only probabilities of appearance of individual images are (at most) known ahead of time, so that the processing or analysis must be designed with respect to the common features of the whole image class.

1.4.1 Stochastic Fields as Generators of Stochastic Images

The definition of a stochastic field* is based on the notion of *family of images* (more precisely, of image functions). Such a family $\mathbf{f}_w(x,y) = \{f_{w_i}(x,y)\}$ is a countable (possibly even infinite) set of mutually different image functions $f_{w_i}(x,y) = f_{w_i}(\mathbf{r})$. The family—the *basis of the stochastic field*—contains all possible image functions that can be expected in a given application. Which of the functions is chosen as the concrete realization depends on the result of a discrete random experiment—sc., *associated experiment*—with a countable set of possible results $W = \{w_1, w_2, \ldots, w_n, \ldots\}$ that appear with a given probability distribution. Obviously, in order that each experiment yields an image, the mapping $W \to \mathbf{f}$ has to be unambiguous, but not necessarily inversible—more results of the experiment may lead to the same image realization.

An example of an image function family is schematically depicted in Figure 1.14, formed of an indexed set of K image functions, e.g., a pile of slides. The k-th function, which will become the actual image realization, is selected from the set according to the result w_k of an associated experiment, for example, tossing one of the lots indexed $1 \ldots K$. Though seemingly primitive, such visualization is useful in developing a good understanding of stochastic field concept applications.

In practice, the random experiment is usually hidden in complicated mechanisms of generating, transferring, or distorting the image. As an example, a concrete image, provided in the frame of screening by an X-ray system under a stable adjustment, is influenced by the instantaneous imaging properties of the system (anode

*A similar development of the one-dimensional concept of stochastic *processes* as sources of random signals may be found in literature on signal processing; e.g., [7] provides a parallel explanation.

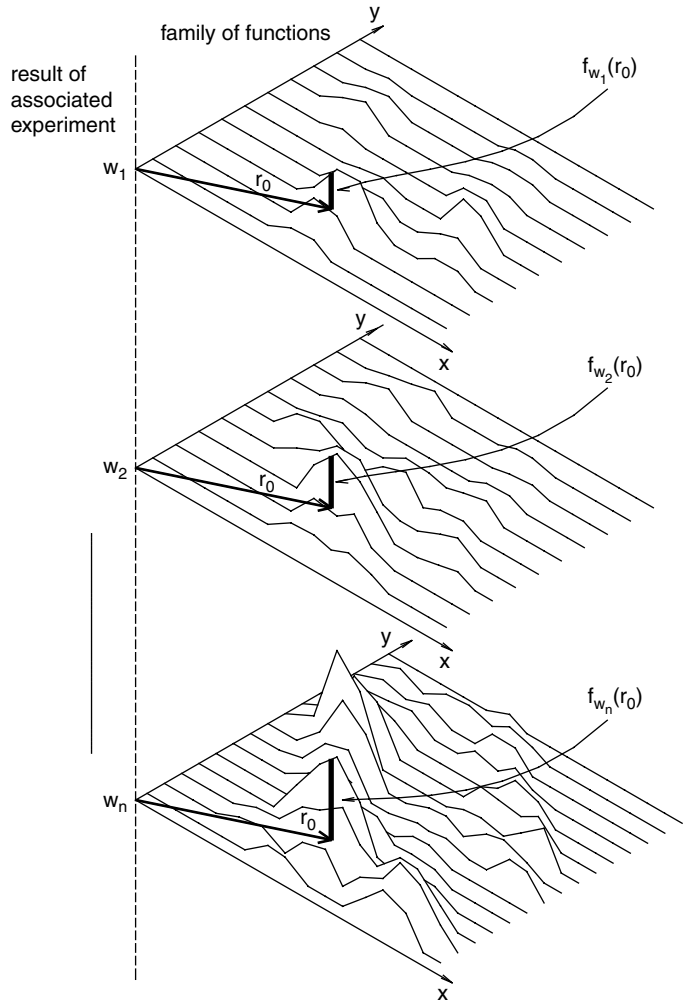


Figure 1.14 Schematic example of a family of two-dimensional functions.

voltage, size of focus spot, imaging geometry) that may be subject to random fluctuations due to unstable electricity supply, different patient properties, and momentary quantum noise. Nevertheless, all the images would have a similar character, which may be characterized by certain statistical parameters and functions. Obviously, a set of cardiological ultrasonic scans will be very different, though the individual ultrasonographic images again will be mutually similar

and thus form a family of images, naturally with different properties from the previous family. It may be predicted that efficient processing methods will be different for the two image classes, though fixed procedures may be used for all images of a particular class.

When choosing a particular position vector $\mathbf{r}_0 = (x_0, y_0)$, the corresponding function values $f_{w_i}(\mathbf{r}_0)$ in all images may be regarded as possible realizations of a random variable $f_w(\mathbf{r}_0)$ dependent on the result w of the associated experiment. Such a *local variable* may be characterized by its probabilistic or statistical properties; as these characteristics apply to a particular position $\mathbf{r}_0$ in the image set, they are called *local characteristics*. Obviously, infinitely many such random variables constitute the continuous stochastic field.

The most important local characteristic, completely describing the individual variable, is the one-dimensional *local distribution function* $P_f(z, \mathbf{r})$, describing the probability $\mathcal{P}$ of the individual realization sample $f_{w_i}(\mathbf{r})$ taking on at most the value of the real parameter z,

$$P_f(z,\mathbf{r}) = \mathcal{P}\{f_{w_i}(\mathbf{r}) \le z\}. \tag{1.62}$$

The derivative of the distribution function is the *local probability density* $p_f(z, \mathbf{r})$,

$$p_f(z,\mathbf{r}) = \frac{\partial P_f(z,\mathbf{r})}{\partial z}. \tag{1.63}$$

The most important parameters related to these two functions are the moments of probability density function, primarily the *local mean value* of $f_w(\mathbf{r})$,

$$\mu_f(\mathbf{r}) = \mathcal{E}_w\{f_{w_i}(\mathbf{r})\} = \int_{-\infty}^{\infty} z\, p_f(z,\mathbf{r})\mathrm{d}z \tag{1.64}$$

and the *local variance*,

$$\sigma_f^2(\mathbf{r}) = \mathcal{E}\left\{\left|f_w(\mathbf{r}) - \mu_f(\mathbf{r})\right|^2\right\} = \int_{-\infty}^{\infty} [z - \mu_f(\mathbf{r})]^2\, p_j(z,\mathbf{r})\mathrm{d}z. \tag{1.65}$$

To estimate the characteristics statistically, a (high) number M of realizations of the stochastic field (sc., an *ensemble* of realizations) must be provided and the values of $f_w(\mathbf{r})$ measured. The probability density can then be approximated by the ensemble histogram

values (see Section 2.4 on discrete stochastic fields). The local mean may be estimated by the *ensemble average*,

$$\mu_f(\mathbf{r}) \approx \frac{1}{M}\sum_{i=1}^{M} f_{w_i}(\mathbf{r}) \tag{1.66}$$

and the local variance by the ensemble average of another quantity,

$$\sigma_f^2(\mathbf{r}) \approx \frac{1}{M}\sum_{i=1}^{M} \left| f_{w_i}(\mathbf{r}) - \mu_f(\mathbf{r}) \right|^2. \tag{1.67}$$

In the limit for $M \to \infty$, these averages become the exact *ensemble mean* values.

The above characteristics describe the local random variable as an isolated item; obviously, infinitely many such characteristics would be needed to describe the stochastic field from this aspect. We shall see later that discretization of images leads to a finite number of random variables, and thus enables the characterization of the stochastic fields by a finite number of characteristics.

However, the local characteristics do not say anything about relations among the image values in different locations. The relations between two locations $\mathbf{r}_1$ and $\mathbf{r}_2$ are described by the two-dimensional *joint distribution function*,

$$P_f(z_1, z_2, \mathbf{r}_1, \mathbf{r}_2) = \mathcal{P}\{f_{w_i}(\mathbf{r}_1) \le z_1 \wedge f_{w_i}(\mathbf{r}_2) \le z_2\} \tag{1.68}$$

and the two-dimensional *joint probability distribution*,

$$p_f(z_1, z_2, \mathbf{r}_1, \mathbf{r}_2) = \frac{\partial^2 P_f(z_1, z_2, \mathbf{r}_1, \mathbf{r}_2)}{\partial z_1 \, \partial z_2} \tag{1.69}$$

As visible in Equation 1.68, the probability concerns the case of both image values taken from the same image indexed w_i. The relations among different realizations of the same field are usually not of interest. Among the moments of the two-dimensional probability distribution, the mixed second-order moment, called *correlation*,

$$R_{ff}(\mathbf{r}_1, \mathbf{r}_2) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} z_1 z_2 \, p(z_1, z_2, \mathbf{r}_1, \mathbf{r}_2) \mathrm{d}z_1 \mathrm{d}z_2 \tag{1.70}$$

is the most important. We shall return to it in the next section.

Theoretically, it is possible to investigate the relations among more (say m) locations by simple generalization of the above expressions, obtaining m-dimensional probability functions. The multidimensional probability functions provide the most complete description of the stochastic fields. On the other hand, not only are these functions more difficult to represent, but also the statistical analysis needed to obtain their estimates becomes rather cumbersome and hardly feasible with increasing m. They are therefore rarely used; we will not continue in this direction until we switch to discrete images (Section 2.4), where the concept may be developed a step further with some useful applications.

So far, we investigated internal relations of a single stochastic field. It is, however, possible to generalize the above considerations to the case of two (joint) *concurrent stochastic fields*, $f_w(\mathbf{r})$ and $g_w(\mathbf{r})$, that are both controlled by the same associated random experiment. In this case, analogous development as above leads to the notion of *mutual relations* among fields, again in the sense that the interest is limited to relations between realization images in pairs selected by the same result w_i of the associated experiment. Particularly, the joint distribution function becomes

$$P_{fg}(z_1, z_2, \mathbf{r}_1, \mathbf{r}_2) = \mathcal{P}\{f_{w_i}(\mathbf{r}_1) \le z_1 \wedge g_{w_i}(\mathbf{r}_2) \le z_2\} \tag{1.71}$$

and the probability density is given by Equation 1.69 when only substituting the single index f by double indices fg.

1.4.2 Correlation and Covariance Functions

As already mentioned, the relations between values at pairs of positions, whether on individual images of a single field or on corresponding images in couples obtained from two concurrent fields, are completely described by two-dimensional joint distribution functions or corresponding two-dimensional probability densities. However, the total amount of information in even a minimal set of them for different position couples is enormous, and also, it would be extremely laborious to provide statistically based estimates of these probability functions. This is the reason for using just the mixed second-order moments of the joint probability density — correlations — instead, like mere local mean values and variances are used as simplified substitutes for the complete local distribution functions.

The *correlation* R_{ff} between two random variables $f_w(\mathbf{r}_1)$, $f_w(\mathbf{r}_2)$ is defined as the ensemble mean of products of values $f_{w_i}(\mathbf{r}_1)$, $f_{w_i}(\mathbf{r}_2)$

always taken from the same realization at the given places $\mathbf{r}_1$, $\mathbf{r}_2$, over all possible realizations:

$$R_{ff}(\mathbf{r}_1, \mathbf{r}_2) = \mathcal{E}_w\{f_{w_i}(\mathbf{r}_1) f_{w_i}(\mathbf{r}_2)\} \approx \frac{1}{M} \sum_{i=1}^{M} f_{w_i}(\mathbf{r}_1) f_{w_i}(\mathbf{r}_2). \tag{1.72}$$

It can be shown to be related to the joint probability density of the pair of variables by Equation 1.70. While this relation is of theoretical interest, it does not provide a useful way for estimating the correlation, as the joint probability is usually unknown. On the other hand, the mean value in Equation 1.72 can easily be approximated by an ensemble average, as indicated by the last expression.

Similar is the *covariance* operator C_{ff},

$$\begin{aligned} C_{ff}(\mathbf{r}_1, \mathbf{r}_2) &= \mathcal{E}_w\{(f_{w_i}(\mathbf{r}_1) - \mu_f(\mathbf{r}_1))\,(f_{w_i}(\mathbf{r}_2) - \mu_f(\mathbf{r}_2))\} \\ &= R_{ff}(\mathbf{r}_1, \mathbf{r}_2) - \mu_f(\mathbf{r}_1)\mu_f(\mathbf{r}_2) \\ &\approx \frac{1}{M} \sum_{i=1}^{m} (f_{w_i}(\mathbf{r}_1) - \mu_f(\mathbf{r}_1))(f_{w_i}(\mathbf{r}_2) - \mu_f(\mathbf{r}_2)). \end{aligned} \tag{1.73}$$

The covariance differs from correlation only by the product of the mean values of both variables; thus, in the case of at least one of the variables being centered (having zero mean), the covariance and correlation are identical.

The covariance may be rather transparently interpreted. If both variables are dependent, though stochastically, we may expect that the differences from the respective mean values will be mostly of the same sign (positively related variables) or mostly of opposite signs (inversely related variables). In the first case, most products in the mean (Equation 1.73) will then be positive, leading to a positive value of the average, i.e., of the covariance estimate. With inverse relation, the products will be mostly negative, thus giving a negative covariance estimate value. Strong relations with most products of the same sign provide high absolute covariance values, while products of randomly changing sign, yielding the low absolute value of the covariance estimate, indicate weak dependencies. Independent variables may be expected to have equally frequent differences with identical signs as differences with opposite signs; the mean then approaches zero. Variables, the covariance of which is zero, are called *uncorrelated*; while it can be shown that *independent* random variables are always uncorrelated, the opposite is not

generally true. Still, it is often necessary to make a hypothesis that variables whose absolute covariance has been found to be small are independent, as the covariance may be the only available information. The hypothesis should be confirmed independently, e.g., by the consistency of some consequential results.

Though the covariance indicates qualitatively the (non)existence of dependence, more quantitative measure is sometimes needed. This is the *correlation coefficient*

$$\rho(\mathbf{r}_1, \mathbf{r}_2) = \frac{C_{ff}(\mathbf{r}_1, \mathbf{r}_2)}{\sigma_f^2(\mathbf{r}_1)\sigma_f^2(\mathbf{r}_2)}, \tag{1.74}$$

which assumes values between +1 and −1, the extremes indicating the deterministic (positive or inverse) mutual dependence between the variables.

The values of correlation or covariance naturally depend on the positions $\mathbf{r}_1$, $\mathbf{r}_2$ as the local variables have generally space-variant properties. This way, Equations 1.72 and 1.73 may be interpreted as definitions of four-dimensional functions of these position vectors, i.e., of four spatial variables x_1, y_1, x_2, y_2. Both functions concern the internal relations inside the same stochastic field; this is often indicated by the prefix *auto-*, as in *autocorrelation function* and *autocovariance function*. These functions are of basic importance in many branches of image processing, modeling, and restoration.

When analyzing mutual relations in a couple of concurrent fields, the correlation and covariance functions are defined accordingly to the explanation preceding Equation 1.71 and specified by the prefix *cross-*; thus, the *cross-correlation function* can be evidently obtained by modification of Equation 1.72 as

$$R_{fg}(\mathbf{r}_1, \mathbf{r}_2) = \mathcal{E}_w\{f_{w_i}(\mathbf{r}_1)g_{w_i}(\mathbf{r}_2)\} \approx \frac{1}{M}\sum_{i=1}^{M} f_{w_i}(\mathbf{r}_1)g_{w_i}(\mathbf{r}_2), \tag{1.75}$$

while the *cross-covariance function*, in analogy to Equation 1.73, can be given as

$$\begin{aligned} C_{fg}(\mathbf{r}_1, \mathbf{r}_2) &= \mathcal{E}_w\{(f_{w_i}(\mathbf{r}_1) - \mu_f(\mathbf{r}_1))(g_{w_i}(\mathbf{r}_2) - \mu_g(\mathbf{r}_2))\} \\ &= R_{fg}(\mathbf{r}_1, \mathbf{r}_2) - \mu_f(\mathbf{r}_1)\mu_g(\mathbf{r}_2) \\ &\approx \frac{1}{M}\sum_{i=1}^{M}(f_{w_i}(\mathbf{r}_1) - \mu_f(\mathbf{r}_1))(g_{w_i}(\mathbf{r}_2) - \mu_g(\mathbf{r}_2)). \end{aligned} \tag{1.76}$$

The properties and applications of auto- and cross-functions are rather analogous but not identical. We shall mention the differences in the special case of homogeneous fields in the next section.

1.4.3 Homogeneous and Ergodic Fields

So far, we discussed properties of generic fields characterized by having all their (even joint and therefore multidimensional) probabilistic characteristics generally spatially variable. This is characteristic for a generic *inhomogeneous field*. In particular, the correlation and covariance functions are of local but spatially variant character as well, as they depend on the concrete positions forming the couple.

However, it is often felt intuitively that this concept is too generic for a given problem because, in a substantial part of stochastic fields met in practice, the local characteristics are obviously not dependent on the position in the image area, or the dependence is weak. This observation has led to formulation of the concept of homogeneous fields.

Most generally, the *strict-sense homogeneous stochastic field* is defined by having all its multidimensional probabilistic characteristics (distribution functions including joint ones, probability densities, and their moments) of any order invariant with respect to spatial shift. This means that all the distributions will not change if all the analyzed positions are shifted in the image plane by the same difference vector, or in other words, all the characteristics are independent of the choice of the position of the origin of coordinates,

$$\begin{aligned} &P_f(z_1, z_2, \ldots, z_m, \mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_m) \\ &\quad = P_f(z_1, z_2, \ldots, z_m, \mathbf{r}_1 + \mathbf{d}, \mathbf{r}_2 + \mathbf{d}, \ldots, \mathbf{r}_m + \mathbf{d}) \end{aligned} \tag{1.77}$$

for any m and an arbitrary position difference vector $\mathbf{d}$.

It is practically very difficult, if not impossible, to verify this field property generally. Therefore, weaker definitions of homogeneity are used; namely, the definition of the *wide-sense homogeneous stochastic field* requires only that the local mean and the autocorrelation function are space invariant in the above sense, i.e.,

$$\begin{aligned} &\mu_f(\mathbf{r}) = \mu_f(\mathbf{r} + \mathbf{d}), \\ &R_{ff}(\mathbf{r}_1, \mathbf{r}_2) = R_{ff}(\mathbf{r}_1 + \mathbf{d}, \mathbf{r}_2 + \mathbf{d}) = R_{ff}(\mathbf{r}_1 - \mathbf{r}_2) = R_{ff}(\Delta\mathbf{r}), \end{aligned} \tag{1.78}$$

which is often possible to verify based on a known mechanism of image generation or on statistical analysis. The dimensionality of the autocorrelation function has decreased to two — it depends on the components of the difference vector only, not on the absolute positions $\mathbf{r}_1$, $\mathbf{r}_2$; this is a very welcome simplification enabling both easier identification of the function for a particular stochastic field and easier visualization and interpretation.

The autocorrelation function can be shown to be symmetric with respect to the origin of coordinates (the components of $\Delta\mathbf{r}$). Really,

$$\begin{aligned} R_{ff}(\Delta\mathbf{r}) = R_{ff}(\mathbf{r}_1 - \mathbf{r}_2) &= \mathcal{E}_w\{f_{w_i}(\mathbf{r}_2 + \Delta\mathbf{r}) f_{w_i}(\mathbf{r}_2)\} \\ &= \mathcal{E}_w\{f_{w_i}(\mathbf{r}_2) f_{w_i}(\mathbf{r}_2 + \Delta\mathbf{r})\} = R_{ff}(-\Delta\mathbf{r}). \end{aligned} \tag{1.79}$$

An example of an autocorrelation function is presented in Figure 1.15. The speckled structure of ultrasonographic images that form the basis of the analyzed stochastic field causes the periodic-like character of the function.

Sometimes, namely in the area of formalized image restoration, other definitions of wide-sense homogeneity are used as well, usually more strict, e.g., the requirement of not only local mean, but also the local variance or even the complete local probability distribution being spatially independent. Occasionally, even

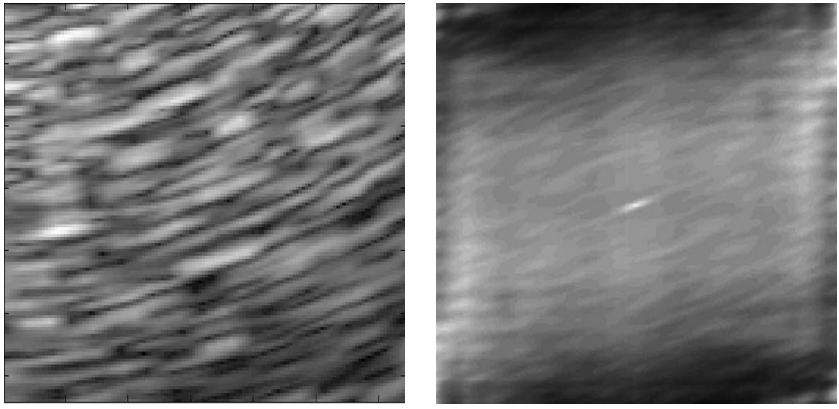


Figure 1.15 Autocorrelation function (right) of a stochastic field formed by a set of ultrasonographic images (example on left).

weaker definitions are used, e.g., requiring only the spatial invariance of the local mean, or conversely, allowing the local mean to be spatially variant while constraining the correlation function to the two-dimensional case. Such definitions are nonetheless rarely used.

When the relationship between two concurrent stochastic fields is investigated, it is also advantageous when the relations can be expressed in a simpler form than by the generic four-dimensional cross-correlation function. Similarly as above, the *wide-sense mutual homogeneity* of two fields f and g may be defined, meaning

$$R_{fg}(\mathbf{r}_1, \mathbf{r}_2) = R_{fg}(\mathbf{r}_1 + \mathbf{d}, \mathbf{r}_2 + \mathbf{d}) = R_{fg}(\Delta\mathbf{r}). \tag{1.80}$$

The interchange of variables that was possible in Equation 1.79 cannot be done in the case of the cross-correlation function; the order of the variables $\mathbf{r}_1$, $\mathbf{r}_2$ is therefore important and the function lacks symmetry,

$$R_{fg}(\Delta\mathbf{r}) \neq R_{fg}(-\Delta\mathbf{r}); \qquad \text{however,} \qquad R_{fg}(\Delta\mathbf{r}) = R_{gf}(-\Delta\mathbf{r}). \tag{1.81}$$

The requirement of homogeneity, when introduced in an image processing method, narrows the class of acceptable random images but obviously simplifies the correlation analysis (and consequently enables the spectral analysis), which may justify the limitation. However, the mean values needed in probabilistic characteristics must still be estimated by *ensemble* averages; this requires acquisition of many image realizations. It is at least cumbersome, and may be difficult or even impossible in practical tasks, in which only a few or even a single image is available for analysis. In such cases, the *spatial averages*, provided from a single image, may be the only alternative, although this approach requires a certain caution.

Primarily, it is clear that a necessary condition required for using spatial averages is the homogeneity of the stochastic field with respect to the investigated parameter (and all involved parameters) — a spatial average obviously cannot provide any valid estimate of spatially variable parameters. However, homogeneity is not a sufficient condition, as the following example demonstrates: Let us have two families of images consisting of randomly taken photographs (so that bright and dark areas are situated randomly on the image area, thus providing for homogeneity of the local mean brightness). Exposition of images of the first family was automatic, so that all the images

are equally dark on average calculated over the image area, while the images of the other family were exposed randomly, some being very light, others very dark. The local mean, as the local ensemble average across the whole family, is homogeneous in both groups (perhaps even equal in both families). When providing spatial averages over individual images of the first group, all will be equal and obviously equal also to the ensemble average. On the other hand, the average brightnesses of individual images in the second group are apparently varying with the image and generally not equal to the (spatially independent) ensemble average. Consequently, a spatial average over a single image is not a good estimator of the local mean value in the latter case, although the field is homogeneous with respect to the mean.

It is obvious that estimating the probabilistic characteristics of homogeneous stochastic fields by means of spatial averages is acceptable only in a narrower class of random images. This subclass of homogeneous fields is formed by *ergodic fields*. In the strict sense, an ergodic field is a homogeneous field for which any probabilistic characteristic can be provided either by using ensemble mean values or by means of spatial mean values over any of its individual realizations, with the same result. It is usually impossible to check completely this demanding requirement; thus, the *wide-sense ergodicity* is commonly understood to concern only the mean values $\mu_f(\mathbf{r}) = \mu_f$ and the autocorrelation function $R_{ff}(\mathbf{r}_1, \mathbf{r}_2) = R_{ff}(\Delta\mathbf{r})$. Let us specify the terms more precisely.

The individual spatial mean for a realization image is defined as

$$E_{w_i} = \lim_{S\to\infty} \frac{1}{S} \int_{\mathrm{S}} f_{w_i}(x,y)\mathrm{d}x\,\mathrm{d}y, \tag{1.82}$$

where S is the surface extent of the area S, in the limit covering the whole x-y plane. If

$$E_{w_i} = E_{w_j} = \mu_f, \quad \forall i,j \tag{1.83}$$

i.e., the individual spatial mean values are all equal and also equal to the ensemble mean, the stochastic field is called *ergodic with respect to mean*.

The individual correlation function of an image is defined by means of the autocorrelation integral derived from Equation 1.42 as

$$C_{w_i}(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f^*_{w_i}(s-x,t-y) f_{w_i}(s,t)\mathrm{d}s\mathrm{d}t. \tag{1.84}$$

If all the individual functions are identical and equal to the ensemble-derived autocorrelation function,

$$C_{w_i}(x,y) = C_{w_j}(x,y) = R_{ff}(x,y), \quad \forall i,j, \tag{1.85}$$

the field is *ergodic with respect to auto-correlation function*.

Mutually ergodic fields are defined correspondingly; they must fulfill analogous conditions concerning the cross-correlation function $R_{fg}(x, y)$ and cross-correlation integrals (Equation 1.42).

The concept of ergodic fields provides the practically important possibility to obtain the probabilistic characteristics of a stochastic field by statistical analysis of a single image. If the image is of unlimited size, the estimates are improved with increasing S and converge to proper values when $S \to \infty$; with limited size images, the estimates suffer with a variance, characterizing the extent of possible errors. There are, nonetheless, usually no means to check the ergodicity of a concrete field, of which only a single realization is available. If no relevant conclusion can be made based on the character of the problem (e.g., on a known mechanism of image generation), the only test is the consistence of the consequential results. In other words, initially we accept the hypothesis that the field is ergodic (or mutually ergodic in case of a couple of fields). The results based on the estimates relying on ergodicity should then be taken with reserve unless the hypothesis can be confirmed in a convincing way, e.g., by the consistency of results arrived at by different ways. Such checks should always be done; otherwise, a risk exists that the results based on the ergodicity hypothesis are misleading.

Should the check be unavailable, or the hypothesis seem dubious, it is advisable to provide more data, e.g., another image, and to compare the estimates of individual characteristics.

1.4.4 Two-Dimensional Spectra of Stochastic Images

1.4.4.1 Power Spectra

Each of the realizations of a stochastic field naturally has, as every image, its two-dimensional spectrum

$$F_{w_i}(u,v) = \mathrm{FT}\{f_{w_i}(x,y)\}. \tag{1.86}$$

The spectra of individual realizations form another family of two-dimensional functions, parallel with the family of image realizations. They are mutually different, $F_{w_i}(u,v) \neq F_{w_j}(u,v)$, if $i \neq j$ (naturally

providing $f_{w_i}(x,y) \neq f_{wj}(x,y)$), but as is frequently the case, they may have some features in common that are typical for individual spectra of most of the field realizations. These similarities are characterized by certain spatial-frequency areas emphasized in most images, while other spectral areas are mostly suppressed. One may be tempted to look for the common spectral features in the mean spectrum, practically approximated by averaging the spectra of all the available realizations. Unfortunately, this approach turns out to be useless, as the spectra are complex and the individual values for each particular frequency combination (u,v) have random phases; in such a case, the mean is zero and, consequently, the averages tend to zero as well.

This problem can be eliminated by considering, instead of complex spectra, the real-valued *individual power spectra*,

$$S_{w_i}(u,v) = |\mathrm{FT}_{2\mathrm{D}}\{f_{w_i}(x,y)\}|^2 = |F_{w_i}(u,v)|^2. \tag{1.87}$$

Such functions may obviously be averaged in the sense that, for each frequency combination (u, v), the mean of the respective values of individual spectra is found,

$$\begin{aligned} S_{ff}(u,v) &= \mathcal{E}_w\left\{|F_{w_i}(u,v)|^2\right\} = \mathcal{E}_w\{F^*_{w_i}(u,v)F_{w_i}(u,v)\} \\ &\approx \frac{1}{M}\sum_{w=w_1}^{w_M}|F_{w_i}(u,v)|^2; \end{aligned} \tag{1.88}$$

the last expression providing the estimate of the mean by averaging the individual two-dimensional power spectra of a finite number of available image realizations. Such a *power spectrum* is real and nonnegative and represents the average power distribution of the stochastic field in the frequency domain. An example may be seen in Figure 1.16.

In the signal theory, a fundamental lemma is proved, called the *Wiener–Khintchin theorem*, here presented in its two-dimensional version:

$$S_{ff}(u,v) = \mathrm{FT}_{2\mathrm{D}}\{R_{ff}(\Delta\mathbf{r})\} \qquad \text{or} \qquad R_{ff}(\Delta\mathbf{r}) = \mathrm{FT}^{-1}_{2\mathrm{D}}\{S_{ff}(u,v)\}, \tag{1.89}$$

i.e., the autocorrelation function and the power spectrum of a stochastic field form a two-dimensional Fourier transform pair. It opens an alternative way to estimate the power spectrum via estimation of autocorrelation function, followed by two-dimensional FT, besides the above average in the spectral domain.

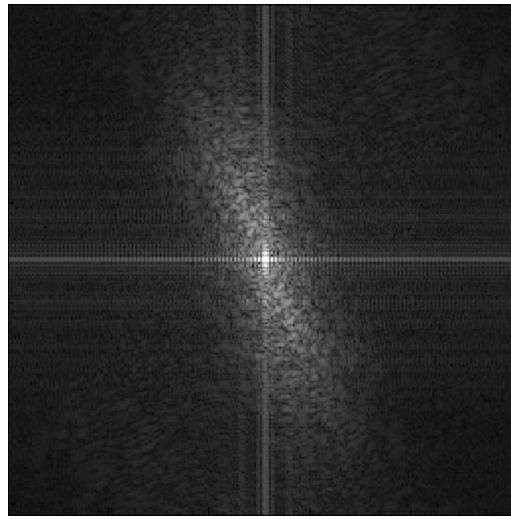


Figure 1.16 Power spectrum of the stochastic field from Figure 1.15.

In consequence of the symmetry of the autocorrelation function, the power spectrum has zero phase (i.e., is real-valued), in accordance with Equation 1.88.

1.4.4.2 Cross-Spectra

The above concept can easily be generalized for the analysis of mutual relations of two concurrent stochastic fields,

$$S_{fg}(u,v) = \mathrm{FT}_{2\mathrm{D}}\{R_{fg}(\Delta\mathbf{r})\} \qquad \text{or} \qquad R_{fg}(\Delta\mathbf{r}) = \mathrm{FT}_{2\mathrm{D}}^{-1}\{S_{fg}(u,v)\}. \quad (1.90)$$

This relation defines a new frequency-domain function called *cross-spectrum* as the spectral counterpart of the cross-correlation function. In analogy with Equation 1.88, it can also be expressed as

$$\begin{aligned} S_{fg}(u,v) &= \mathcal{E}_w\{F^*_{w_i}(u,v)G_{w_i}(u,v)\} \\ &\approx \frac{1}{M}\sum_{w=w_1}^{w_M} F^*_{w_i}(u,v)G_{w_i}(u,v). \end{aligned} \quad (1.91)$$

The interpretation of the cross-spectrum, though not difficult, deserves a few sentences to enlighten its physical meaning (often misunderstood in literature). Primarily, it should be clear that the

individual spectra are random (as their corresponding original images are), so that the variables $F_w(u,v)$, $G_w(u,v)$ for a particular frequency (u,v) are also random. Then, the cross-spectral value at this two-dimensional frequency is, as the ensemble mean of the product of their realizations, obviously the correlation of these variables, according to Equation 1.91. The cross-spectrum may therefore be interpreted as a two-dimensional function of two-dimensional frequency, describing the cross-correlation between corresponding spectral values of both fields.

The correlation values are formed as the main values (approximated by averages in the estimate) of products, with each individual complex-valued product obviously having the argument given by the difference of phases of factors *F, G.* In case of purely random phase differences, the average of the complex products tends to zero, as the individual contributing terms cancel mutually. When there is a prevailing phase relation, the contributions add together, forming a complex number with the argument equal to the average of the phase difference.

The correlation at a particular two-dimensional frequency (u, v) becomes or approaches zero in two cases—primarily when most of the products in the mean are zero, i.e., when at least one of the factors is almost always zero. This means that the spectral component at the particular frequency is mostly missing at least in one of the corresponding images. It is then clear that no important mutual relationship can exist at this frequency. The correlation also becomes zero if the phase relation of the frequency components is random, as mentioned above, even when the components are mostly present in both images. The zero correlation at a particular frequency may be interpreted as an indication of a lack of relationship between both stochastic fields at this frequency, should it be due to either of the two reasons.

On the other hand, the correlation at a particular two-dimensional frequency obviously reaches its maximum if there is a fixed phase relationship among the spectral components at this frequency in both images; in this case, the magnitude of the correlation reveals the average product of intensities of these components in both fields.

The correlation interpretation of the cross-spectrum values at individual frequencies thus leads to the following conclusion: the individual spectral values indicate, by their magnitudes, the average strength of the relationship between the spectral components and, by their arguments, the prevailing phase relation between these components in corresponding pairs of images.

1.4.5 Transfer of Stochastic Images via Two-Dimensional Linear Systems

This topic is another source of frequent misunderstanding: a stochastic image is only a notion, interpreted in terms of stochastic fields, which cannot be submitted to a physical (though perhaps computational) two-dimensional system as an input. The concept of randomness is contained in the associated stochastic experiment selecting the concrete realization, the result of which is not known ahead of time. This background is, however, irrelevant with respect to the two-dimensional deterministic system that, by definition, is only capable of processing deterministic images. Thus, the section title does not make sense, unless interpreted symbolically as follows.

Though a stochastic image $f_w(x, y)$ as such is not applicable to the input of a two-dimensional system, any realization $f_{w_i}(x, y)$ of the stochastic field is a standard deterministic image, which can be submitted to the system for processing. The system then yields, as its response, an output image,

$$g_{w_i}(x, y) = \mathrm{P}\{f_{w_i}(x, y)\}, \tag{1.92}$$

by application of the operator P characterizing the system. According to the stochastic field definition, $f_{w_i}(x, y)$ is a member of a family of functions, $f_w(x,y)$. Similarly, the output image $g_{w_i}(x, y)$ may be considered a member of another family of functions $g_w(x, y)$ that forms a basis of another stochastic field, controlled by the same associated experiment as the field generating the input images. The input field might, at least hypothetically, generate successively a sequence of input images that would generate a corresponding sequence of output images gradually building the output image family. This way, we obtain two concurrent stochastic fields, defining two random images: the input image and the output image. Whether the images were really processed in sequence or in parallel by a large set of identical systems, or even not at all, is in the end irrelevant. Important is the fact that, to each image of the input family, an output image might be produced and determined. In this sense, we shall understand the relation among the input and output families,

$$g_w(x, y) = \mathrm{P}\{f_w(x, y)\}, \tag{1.93}$$

as a symbolic formalism (a family of equations), expressing Equation 1.92 for all i and enabling the simplification of the notation.

It is now possible to formulate the task of this section as finding the characteristics of the output stochastic field when the

characteristics of the input field and two-dimensional system are known — in other words, closer to the section title, to find the parameters of the output random image based on the parameters of the input random image and the operator P.

Processing of random images by two-dimensional space-invariant linear systems is of particular interest. In this case, Equation 1.92 becomes the convolution

$$g_{w_i}(x,y) = h * f_{w_i}\,|(x,y), \tag{1.94}$$

where $h(x, y)$ is the impulse response (PSF) of the system. We shall derive some of the probabilistic characteristics of the output field (output random image) with respect to the input characteristics.

The mean value of the input family is an image

$$\mathcal{E}_w\{f_{w_i}(x,y)\}, \tag{1.95}$$

the values of which are local ensemble mean values. According to Equation 1.94, the output mean is

$$\mathcal{E}_w\{g_{w_i}(x,y)\} = \mathcal{E}_w\{f_{w_i}(x,y) * h(x,y)\} = \mathcal{E}_w\{f_{w_i}(x,y)\} * h(x,y), \tag{1.96}$$

where we utilized the deterministic nature of the PSF. The *output mean* image is thus derived from the input mean (which may be regarded as a deterministic image) by the same operation of convolution as applied to any other deterministic image. Similarly, we could derive the image of the output local variance.

Further, Fourier transforming Equation 1.94 yields the spectrum of the output image, $G_{w_i}(u,v) = H(u,v)F_{w_i}(u,v)$, where $H(u,v)$ is the frequency response of the system. Also, because $G^*_{w_i}(u,v) = H^*(u,v)\ F^*_{w_i}(u,v)$, the power spectrum of an individual output image can be expressed as

$$G_{w_i}(u,v)G^*_{w_i}(u,v) = |G_{w_i}(u,v)|^2 = |H(u,v)|^2\ |F_{w_i}(u,v)|^2, \tag{1.97}$$

from where we obtain the power spectrum of the output field by applying the mean value operator to both sides,

$$S_{gg}(u,v) = \mathcal{E}\left\{|G_{w_i}(u,v)|^2\right\} = |H(u,v)|^2\ \mathcal{E}\left\{|F_{w_i}(u,v)|^2\right\} = |H(u,v)|^2\ S_{ff}(u,v). \tag{1.98}$$

The power spectrum of the output random image is thus given by the input power spectrum modified by multiplication with the square of the amplitude frequency transfer function of the system.

This theorem describes the *transfer of the stochastic image* via a linear system in the frequency domain.

The cross-spectrum describing frequency-domain relations between the input and output random images is

$$\begin{aligned} S_{fg}(u,v) &= \mathcal{E}_w\{F_w^*(u,v)G_w(u,v)\} \\ &= \mathcal{E}_w\{F_w^*(u,v)H(u,v)F_w(u,v)\} = H(u,v)S_{ff}(u,v). \end{aligned} \tag{1.99}$$

The cross-spectrum is thus given by the input power spectrum modified by the system as any other image in the original domain. By inverse Fourier transforming the last equation, we obtain its original-domain counterpart, the two-dimensional version of the *Wiener–Lee theorem*,

$$R_{fg}(x,y) = h * R_{ff}|(x,y). \tag{1.100}$$

Equations 1.98 and 1.99 enable the identification of the properties of a linear two-dimensional system by random images, i.e., by allowing the system to process a sufficient number of realizations (e.g., of random noise images) so that the spectral estimates are reliable enough. Equation 1.98 enables estimation of only the amplitude frequency response of the two-dimensional system, while Equation 1.99 provides an estimate of the complete frequency transfer, including phase characteristic, though at the cost of a more complicated measurement. In the first case, the input and output power spectra may be measured independently (should the process generating the images be stationary), while the second case requires a concurrent measurement at the input and output. Equation 1.100 shows that it is possible to identify the PSF of the system by means of the correlation technique, which, as can be proved, is insensitive to any signals that are uncorrelated with the measuring input images (usually noise realizations), either external signals or noise components generated internally in the system.

1.4.6 Linear Estimation of Stochastic Variables

We shall conclude this section on stochastic images with a simple mathematical result of fundamental importance that finds frequent applications in the image processing area, namely, in image restoration. Using this principle substantially simplifies many later derivations, and its intuitive use may lead to novel ideas.

The problem concerns estimating a realization value of a stochastic variable g on the basis of some other stochastic variables f_i, $i = 1, 2, \ldots, N$, the values of which are assumed to be measurable and known, in contrast to g that is not available for measurement. Naturally, we can expect a reasonable estimate only if there exist some dependencies among the variables. Though it may not be optimal with respect to particular properties of the problem, for which a nonlinear estimate might be preferable, usually only linear estimates are used thanks to their easier mathematical tractability.

The linear estimate $\hat{g}$ can be expressed as the weighted sum

$$\hat{g} = \sum_{i=1}^{N} a_i f_i = \mathbf{a}^{\mathrm{T}}\mathbf{f}, \tag{1.101}$$

where $\mathbf{f} = [f_1, f_2, \ldots, f_N]^{\mathrm{T}}$ is the vector of the known variable values and $\mathbf{a} = [a_1, a_2, \ldots, a_N]^{\mathrm{T}}$ is the vector of fixed coefficients (weights) to be determined so that a suitable error criterion is met. The commonly used standard criterion is the mean quadratic error of the estimate with respect to the actual value g,

$$\varepsilon(\mathbf{a}) = \mathcal{E}\{(g - \hat{g}(\mathbf{a}))^2\} = \mathcal{E}\left\{\left(g - \sum_{i=1}^{N} a_i f_i\right)^2\right\} \rightarrow \mathrm{Min}, \tag{1.102}$$

taken over all possible realizations of the variables. The minimum is reached for $\mathbf{a}_{\mathrm{opt}}$, fulfilling the condition of zero gradient,

$$\nabla\varepsilon(\mathbf{a}_{\mathrm{opt}}) = \mathbf{0}, \tag{1.103}$$

which is a necessary (and with respect to circumstances, also sufficient) condition. It represents a set of linear equations

$$\frac{\partial\varepsilon}{\partial a_i} = \mathcal{E}\left\{\frac{\partial}{\partial a_i}(g - \hat{g}(\mathbf{a}))^2\right\} = -2\mathcal{E}\{(g - \hat{g}(\mathbf{a}))f_i\} = 0, \quad \forall i; \tag{1.104}$$

thus

$$\mathcal{E}\{(g - \hat{\mathrm{g}}(\mathbf{a}))f_i\} = 0, \quad \forall i. \tag{1.105}$$

This relation is called the *principle of orthogonality*, and its interpretation is as follows: should the linear estimate of a stochastic variable g be optimum in the mean square error sense, the error

variable ($g - \hat{g}(\mathbf{a})$) must be statistically orthogonal to (in other words, uncorrelated with) each of the random variables f_i, on which the estimate is based.

Converting the requirement in Equation 1.102 into Equation 1.105 is a step often met in derivations of image processing methods. The number N of the variables entering the estimate is, of course, not limited. The principle therefore also applies in cases where the estimate is formulated by an integral instead of a finite sum, as the integral can be regarded as just the limit case of a sum. This will be used with advantage later.

2

Digital Image Representation

2.1 DIGITAL IMAGE REPRESENTATION

2.1.1 Sampling and Digitizing Images

2.1.1.1 Sampling

In order to be handled by computers, the images must be digitized. For the analogue continuous-space continuous-value two-dimensional image (Equation 1.1), it means *discretization* of two kinds: spatial sampling and amplitude quantizing. This way, the analogue function is replaced by its discrete representation: a matrix of numbers. In the present chapter, we shall analyze theoretically the concept of discretization, its consequences and related limitations, and ways in which to convert the discrete representation back into the analogue image form. As before, we shall mostly cover two-dimensional still images; the results can be easily generalized to multidimensional cases.

The *ideal spatial sampling* can be represented by multiplying the analogue image by a two-dimensional sampling signal $s(x,y)$,

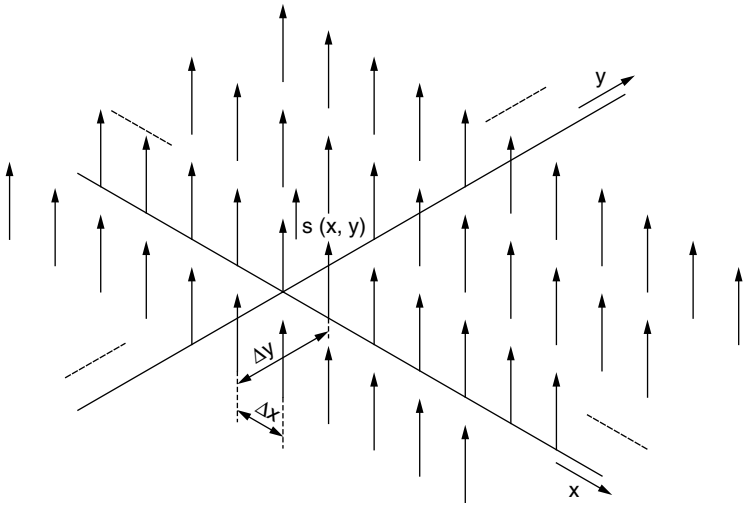


Figure 2.1 Two-dimensional sampling function. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

which is formed of an infinite number of periodically repeated Dirac impulses (Figure 2.1),

$$s(x,y) = \sum_{i=\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(x - k\Delta x, y - i\Delta y). \tag{2.1}$$

The sampled image may then be expressed as the product of the analogue image ("to be sampled" image) and the sampling function,

$$f_s(x,y) = f(x,y)s(x,y) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(k\Delta x, i\Delta y)\delta(x - k\Delta x, y - i\Delta y). \tag{2.2}$$

Obviously, only those discrete values of the analogue image are needed that correspond to the nodes of the sampling grid. The information carried by the intermediate image values therefore seems to be lost; we will show that this does not need to be the case if certain

conditions are observed, so that the image matrix (so far of unlimited size),

$$\bar{\bar{\mathbf{f}}} = [{}_{s}f_{i,k}], \qquad {}_{s}f_{i,k} = f(k\Delta x, i\Delta y) \tag{2.3}$$

carries the complete image content*.

The spectral representation of the sampling signal is

$$S(u,v) = \mathrm{FT}\{s(x,y)\} = \frac{4\pi^2}{\Delta x \Delta y} \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \delta(u - kU, v - iV)$$
$$U = \frac{2\pi}{\Delta x}, \qquad V = \frac{2\pi}{\Delta y}. \tag{2.4}$$

Here, U, V are called (angular) *sampling frequencies*. The formal proof will be omitted, but it is easy when transforming the spectral formula back into the original domain by the inverse two-dimensional FT, utilizing the sifting property of the Dirac impulse. A qualitative explanation is simple: the discrete spectrum corresponds to the original function that is two-dimensionally periodical, with the basic frequencies and their multiples — harmonic frequencies — given by the original-domain periods Δx, Δy. The extent of the spectrum is infinite, as the δ-impulses in the original domain contain even infinite frequencies. The spectrum of the sampled image will then be

$$\begin{aligned} F_s(u,v) &= \mathrm{FT}\{f(x,y)s(x,y)\} = \frac{1}{4\pi^2} F * S \,|(u,v) \\ &= \frac{1}{\Delta x \Delta y} \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} \int\int F(u-s, v-t)\, \delta(s - kU, t - iV)\, \mathrm{d}s\, \mathrm{d}t \\ &= \frac{1}{\Delta x \Delta y} \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} F(u - kU, v - iV), \end{aligned} \tag{2.5}$$

*From now on, the matrices and vectors will often be represented in equations like $\bar{\bar{\mathbf{f}}}$ or $\bar{\mathbf{f}}$, respectively, in order to distinguish between both image representations. In the text, where the meaning is clear from the context, usually plain bold symbols (like $\mathbf{f}$) will be used. The form $[f_{i,k}]$ means the matrix that is formed of the elements $f_{i,k}$.

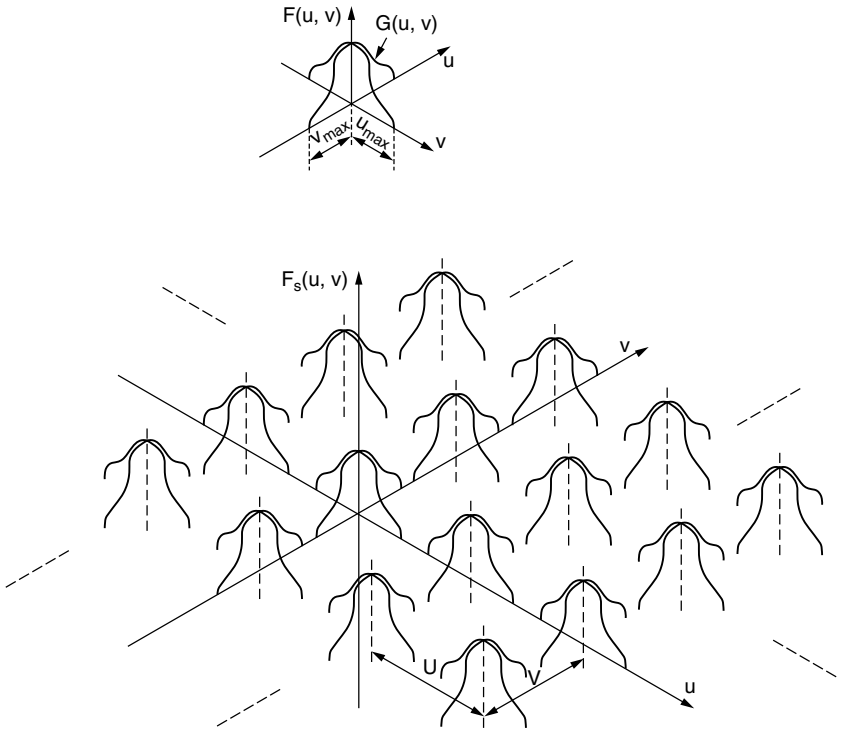


Figure 2.2 Schematic band-limited two-dimensional spectrum of an analogue image (above) and the unlimited periodic spectrum of the sampled image. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

where $F(u, v)$ is the spectrum of the analogue image and the integral has been simplified using the sifting property of the Dirac impulse. The spectrum of the sampled image is thus formed by periodically repeated replicas of $F(u, v)$ centered at the positions of spectral lines of the sampling function. The spectral representations of both analogue and sampled images are schematically depicted in Figure 2.2.

Let the maximum frequencies of the analogue image in the x and y directions be u_{max}, v_{max}. It can be seen that the neighboring replicas of spectra in $F_s(u, v)$ would interfere should U or V be insufficient to separate them; in such a case, it would not be possible to recover the original spectrum and, consequently, even the original

continuous image from its sampled version. The conditions preventing the interference are

$$U > 2u_{\max}, \qquad V > 2v_{\max}, \tag{2.6}$$

which means for the sampling intervals in the original domain,

$$\Delta x < \frac{\pi}{u_{\max}}, \qquad \Delta y < \frac{\pi}{v_{\max}}. \tag{2.7}$$

These conditions are called the *two-dimensional sampling theorem*. When we denote $u_N = U/2$, $v_N = V/2$ as *Nyquist frequencies*, the sampling theorem may be interpreted as follows: the two-dimensional rectangular sampling with the spatial (angular) sampling frequencies U, V allows working with images having frequency limits under the respective Nyquist frequencies. This fundamental theorem must be observed; otherwise, an unrecoverable distortion called *aliasing* appears. The name comes from the fact that due to frequency overlap and interference of the spectral replicas, high frequencies manifest themselves as low ones, thus causing a disturbing *moiré* pattern in the reproduced image. Consequently, not only high-frequency details above the Nyquist limits are lost, but even the image content under the limits may be substantially distorted.

From the spectral representation, it is clear that no image information is lost by sampling as far as the sampling theorem is fulfilled. As any natural image has a limited spectrum, $u_{\max} < \infty$, $v_{\max} < \infty$, it is always possible to find a sufficient density of sampling to preserve the complete information content. Nevertheless, limits of technical equipment sometimes may not allow sufficient sampling frequencies with respect to image properties. Then it is necessary to limit the frequency content of the image by a low-pass *antialiasing filtering* that stops the unwanted frequencies above the Nyquist limit (though it naturally leads to a certain loss of detail). This way, the unrecoverable aliasing distortion is prevented. It should be stressed that the antialiasing filtering has to be applied to the analogue image before sampling; thus, it is not a simple task. It may be approximated, e.g., by certain defocusing of the image, which may paradoxically improve the appearance of the resulting discrete image. Also, imperfect (nonpoint) sampling may positively contribute (see below).

An example of the marked aliasing in a sampled image is presented in Figure 2.3. The original image has an increasingly

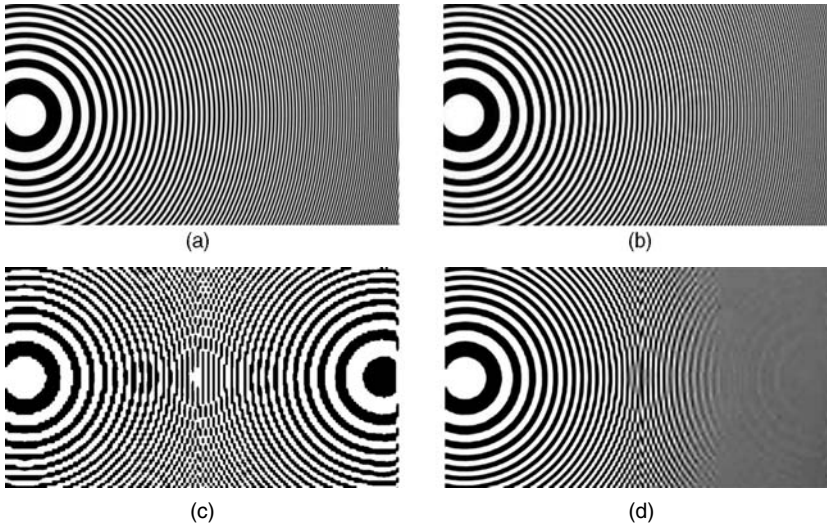


Figure 2.3 Aliasing as a consequence of insufficient sampling density: (a) original (well sampled image, (b) half of density of sampling, (c) quarter density of sampling, and (d) quarter density of sampling after image smoothing (i.e., after simple low-pass filtering).

dense structure, which is well represented in the version with the highest sampling density (certain residual moiré in this image is due to typographical limitations). The half and quarter densities are obviously insufficient, thus misreproducing a smaller or greater part of the image from the dense side. When the image has been smoothed before sampling, which is equivalent to simple low-pass filtering, the higher frequencies of the original image spectrum were suppressed, and even the quarter sampling became reasonably adequate — the moiré artifacts are substantially suppressed.

The above sampling theory is based on the notion of point sampling by infinitesimally narrow Dirac impulses, over the infinite x-y plane. The situation is idealized primarily in two aspects: the size of any real picture is finite, and practical sampling systems do not provide point values, but always take a weighted two-dimensional integral of brightness from a finite area around the sampling point. Let the spatial sensitivity of the sampling sensor (the sensing spot description) $p(x, y)$ be isoplanar. Then, the value measured by the

sensor at the position $(k\Delta x, i\Delta y)$ is obviously

$$f_r(k\Delta x, i\Delta y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)p(x-k\Delta x, y-i\Delta y)\mathrm{d}x\,\mathrm{d}y = f(x,y) * p(-x,-y), \tag{2.8}$$

where the second factor of the integrand is the shifted spatial characteristic of the sensor. Comparing this with Equation 1.52, we see that the result is the same as if the image, preliminarily processed by a two-dimensional linear filter with the point-spread function (PSF) equal to $p(-x, -y)$, had been ideally sampled. With respect to usual symmetry of the PSF, the rotation by 180° may be irrelevant. Thus, the spectrum of the hypothetical image before sampling is

$$F_r(u,v) = F(u,v)P(-u,-v) \tag{2.9}$$

where $P(u, v)$ is the spectrum of the PSF of the sensor. Because this spectrum, i.e., the frequency response of such sampling, is usually of a low-pass character, the imperfect sampling may even be advantageous, as it may be interpreted as a certain antialiasing filtering followed by the ideal sampling. It is an error, occasionally met in literature, to interpret the finite sampling spot effect as influencing the amplitudes of the higher-order spectral replicas (sometimes explained by the s.c. first or second type sampling, which actually has nothing to do with discrete signals). In fact, the result is a pure discrete image that can be expressed by Equation 2.2, where only $f(x,y)$ is replaced by its filtered version $f(x,y)*p(-x, -y)$.

The influence of the finite size of the sampled image may be taken into account by multiplying the infinite sampled image by the unitary window $w(x,y)$ of the size of the real image, thus finally obtaining the imperfectly sampled finite size image as

$$f_{sw}(x,y) = (f(x,y) * p(-x,-y))s(x,y)w(x,y). \tag{2.10}$$

The resulting spectrum $F_{sw}(u, v)$ is then

$$F_{sw}(u,v) = (F(u,v)P(-u,-v)) * S(u,v) * W(u,v), \tag{2.11}$$

where $W(u, v)$ is the spectrum of the window, of the same type as Equation 1.22 (modified for nonunit size of the image). The influence

of the finite image size is thus only a certain smearing of the complete spectrum (a "leakage" to improper frequencies) as a consequence of the last convolution.

2.1.1.2 Digitization

To obtain numbers tractable by computers, the sample values obtained by sampling are consequently digitized by an analogue-to-digital (A/D) converter. As any finite code can represent only limited-precision numbers, the measured values are rounded (or cut off) to the nearest of the available levels (nowadays, mostly 256, 4096, or 65,536 levels corresponding to 8-, 12-, or 16-bit fixed-point representation are used). This means an unrecoverable nonlinear distortion, causing *quantization noise* — an additive image formed by the rounding errors. As the original sample values are distributed supposedly continuously, the error distribution is even, with the maximum error given by plus or minus half the quantizing step when rounding is used (twice as much for cutoff quantizing). The quantization noise may be rather high when the number codes are short; e.g., with today's most common 8-bit coding, the maximum error is about 2×10^{-3} (or 4×10^{-3}) of the gray-scale amplitude range. This is under human resolution of brightness differences, thus negligible for display, but it may be prohibitive for some more advanced processing methods. That is why the contemporary professional systems use 12-, 16-, or even 24-bit gray-scale coding (three times that for color images); this way the quantization noise may be substantially suppressed providing that quality A/D converters are used.

As any natural image has limited dimensions, the result of digitization is a finite set of numbers representing the sample values — picture elements, *pixels*[*], that may be arranged into a matrix. Contemporary systems use matrices of sizes approximately in the range between ~0.3 and (rather rarely) ~16 Mpixels (Mpx), corresponding to square matrices of 512 × 512 to 4096 × 4096; however, many systems use the rectangular (nonsquare) format. This image size, together with the bit depth, determines the memory requirements (unless data compression is used): e.g., a 5-Mpx 16-bit

[*]The term *pixel* may also mean a displayed picture element, usually a square of a constant brightness corresponding to a sample value (see also the chapter on image presentation).

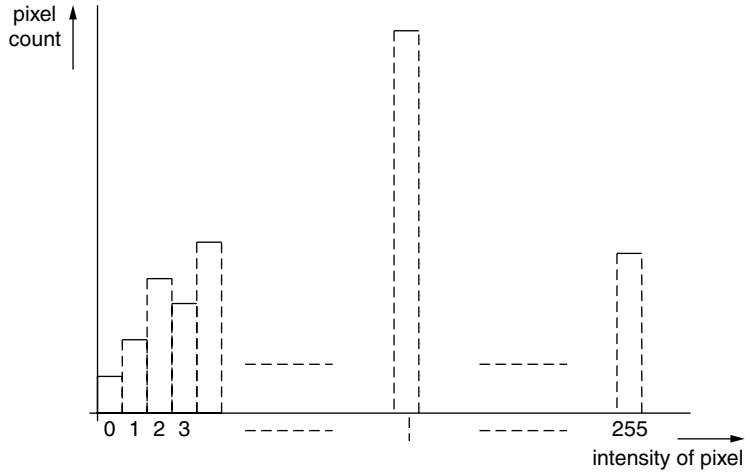


Figure 2.4 Schematic plot of a histogram of an image with a single-byte pixel representation.

gray-scale image requires 10 Mbytes of memory space, and a color image three times as much.

The important concept of the *gray-scale histogram* should be mentioned in connection with quantization. In a finite sampled and quantized (i.e., discrete-level) image, it is possible to count pixels with a concrete intensity, and to provide a plot of such counts belonging to different intensity levels (Figure 2.4).

More formally, the one-dimensional histogram of a discrete-intensity image A with q degrees of gray (or brightness) is the vector

$$\mathbf{h}^A:\ h_l^A = \text{count}(A_l) = \sum_{A_l} 1,\ l = 0,1,\ldots,q-1,$$
$$A_l = \{a_{i,k} : ((a_{i,k,} \in A) = l\}. \tag{2.12}$$

That is, the l-th component of the histogram carries the information on how many pixels at the l-th intensity level are present in the image; A_l is thus the set of pixels a_{ik} acquiring the intensity l. Should the image be a realization of a homogeneous and ergodic stochastic field, a component of the normalized histogram approximates

the probability P_l of a concrete intensity l in a generic pixel of image A,

$$P_l = P\{(a_{i,k} \in A) = l\} \approx \frac{h_l^A}{\sum_l h_l^A}, \ \forall i,k. \tag{2.13}$$

The histogram provides information on the distribution of intensity levels in the image that may serve primarily to assess the adjustment of the imaging system with respect to utilizing the available dynamic range of pixel values. A narrow histogram envelope with wide zero-count ranges at either end of the gray scale means a low-contrast and unnecessarily rough quantization, while a histogram with excessively high values at either end of the scale indicates over- or underexposure. Secondarily, the histogram may be used in proposing suitable contrast transforms as image-enhancing procedures (see Section 11.1).

As mentioned above, the most natural representation of a discrete finite image of $N \times M$ samples is a matrix of the sample values:

$$\bar{\bar{\mathbf{f}}} = \begin{bmatrix} f_{00} & \cdots & f_{0,M-1} \\ \vdots & f_{i,k} & \\ & & \vdots \\ f_{N-1,0} & \cdots & f_{N-1,M-1} \end{bmatrix}, \quad f_{i,k} = f(i,k)^*. \tag{2.14}$$

Strictly taken, according to Equation 2.3, indexing depends on the choice of the coordinate origin position in the *x-y* plane; practically, the image matrices are mostly indexed as if the origin were at the upper left corner with (0, 0) index, as used in the last expression. Though this scheme leads mostly to simpler indexing algorithms, indexing starting from (1, 1) is also often used, as, e.g., in the MATLAB® environment. In both cases, the coordinate directions are aiming down and to the right.

For certain purposes, it is advantageous to arrange the same numbers into a column vector, obtained by scanning the matrix usually by columns,

$$\bar{\mathbf{f}} = [f_{00},...,f_{N-1,0},f_{01},...,f_{N-1,1},...,f_{0,M-1},...,f_{N-1,M-1}]^{\mathrm{T}}. \tag{2.15}$$

*The indices i, k may be presented in both indicated styles, depending on the context. The upper index T means transposed matrix.

The algorithm of converting the matrix form into a vector, and vice versa, is obvious; it is nevertheless sometimes useful, in formal analytical derivations, to express the matrix–vector conversion by a closed formula. As shown in [18], the formulae can be based, particularly for $M = N$, on the auxiliary vector $\mathbf{v}_n$ of the N elements, having 1 at the n-th position while all other elements are zero, and on the auxiliary matrix $\mathbf{N}_n$, sized $N^2 \times N$, formed of N submatrices, of which only the n-th submatrix is a unit matrix while others are zero. Then it is possible to express the form conversion as

$$\bar{\mathbf{f}} = \sum_{n=0}^{M-1} \mathbf{N}_n \bar{\bar{\mathbf{f}}} \mathbf{v}_n \quad \text{or inversely} \quad \bar{\bar{\mathbf{f}}} = \sum_{n=0}^{M-1} \mathbf{N}_n^T \bar{\mathbf{f}} \mathbf{v}_n^T. \tag{2.16}$$

2.1.2 Image Interpolation from Samples

As pointed out in the previous section, theoretically it is possible to reconstruct the original continuous image *exactly* from its samples. The interpretation in the frequency domain is straightforward: as all the spectral replicas in the sampled image spectrum are identical and equal to the original spectrum of the analogue image (up to a multiplicative constant), it suffices to remove all the replicas except that in the original position (base-band). Obviously, this may be done by a suitable two-dimensional low-pass *reconstruction filter* with the frequency response

$$R(u,v) = \begin{cases} \Delta x \Delta y & \text{in the baseband area} \\ 0 & \text{outside of the area} \end{cases}, \tag{2.17}$$

as schematically illustrated in Figure 2.5. The form of the filter base (usually a rectangle or a circle) is arbitrary as far as it covers completely (and evenly) the frequency extent of the original image and rejects any other spectral components. Because no ideal low-pass (LP) filter exists, sufficient spacing of the spectral replicas is necessary, allowing a certain transition band of the filter between its passband and stop-band; no spectral components should exist in the transition band. Sufficiently high sampling frequencies U and V with a necessary reserve must be provided in the sampling phase by choosing appropriately high spatial sampling densities.

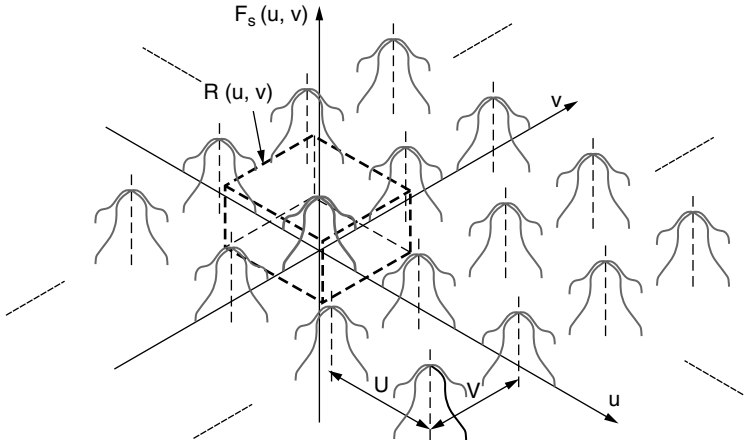


Figure 2.5 Spectral representation of image reconstruction from samples. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

The action of the reconstruction filter is interpreted in the original domain as the convolution, the reconstructed analogue image being

$$f_r(x,y) = f_s * r|(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_s(x,y)r(x-s,y-t)\mathrm{d}s\mathrm{d}t, \qquad (2.18)$$

where $r(x,y) = \mathrm{FT}_{2\mathrm{D}}^{-1}\{R(u,v)\}$ is the PSF of the reconstruction filter. Substituting Equation 2.2 for $f_s(x,y)$, we obtain

$$\begin{aligned} f_r(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\sum_i\sum_k f(k\Delta x, i\Delta y)\delta(s-k\Delta x, t-i\Delta y)r(x-s,y-t)\mathrm{d}s\mathrm{d}t \\ &= \sum_i\sum_k f(k\Delta x, i\Delta y)r(x-k\Delta x, y-i\Delta y), \end{aligned} \qquad (2.19)$$

where the last expression was derived by interchanging the integration and summing, and utilizing the sifting property of the Dirac function. The final sum may obviously be interpreted as interpolation among the sample values, with interpolation functions being

shifted versions of the reconstruction filter PSF. It is interesting to notice that the interpolation is always exact when Equation 2.17 applies, independently on the filter base form, influencing the PSF shape. As examples, the rectangular filter sized $2u_m \times 2v_m$ and the circular-base filter with a diameter w_m have the impulse responses (i.e., the interpolation functions)

$$r(x,y) = \frac{u_m v_m}{\pi^2} \frac{\sin(u_m x)}{u_m x} \frac{\sin(v_m x)}{v_m x} \quad \text{and}$$
$$r(x,y) = 2\pi w_m \frac{J_1(w_m \sqrt{x^2 + y^2})}{\sqrt{x^2 + y^2}}, \quad \text{respectively.} \tag{2.20}$$

Naturally, because the displayed image is finally analogue, the low-pass filtering should be, exactly taken, also realized in continuous space. The display chain thus should contain a digital-to-analogue (D/A) converter, providing discrete analogue brightness values, a display unit presenting the discrete image as a grid of controlled intensity point sources (thus already an analogue, though spatially discrete image), and finally an analogue (e.g., optical) filter that interpolates among the sample values, thus yielding a continuous-space image. The last step would require an expensive optical computer; although such systems do exist, usually the LP analogue filtering is substituted for by using a display (monitor) of points dense enough not to be resolvable by the observer's sight, and the interpolation, when needed, is performed on the digital side (see also Section 10.1.2).

2.2 DISCRETE TWO-DIMENSIONAL OPERATORS

The discrete two-dimensional operators, usually realized in the digital environment, may be described in terms of the input–output relation

$$g_{i,k} = \mathrm{P}_{i,k}\{\bar{\bar{\mathbf{f}}}\}. \tag{2.21}$$

In its most general form, it means that each element $g_{i,k}$ of the output matrix $\mathbf{g}$ depends on all elements of the input matrix.

Such operators may be denoted as *global*; the indices at the operator symbol indicate that, in the general case, the operator may be space variant. Most two-dimensional discrete transforms are of the global kind, as are some complicated filtering approaches as well.

Many common operators derive an output element based only on certain surroundings of the corresponding element in the input matrix. These are called *local operators.* Examples may be linear or nonlinear two-dimensional local filters, sometimes called mask operators, that form the basic means of most standard image processing packages. Obviously, the global operators may be considered local with the largest possible input area.

The *point (point-wise) operators*, described by

$$g_{i,k} = \mathrm{P}_{i,k}\{f_{i,k}\} \tag{2.22}$$

(in the common case of identical-size input and output matrices), form the other extreme. These operators are mostly nonlinear and obviously represent contrast (or color) transforms not touching other aspects.

In comparison with one-dimensional operators applied to time-dependent signals, the two-dimensional operators may be considered a generalization, but with some important differences in features and formulations. Primarily, the notion of causality does not apply to image data processing, as there is no preferential direction in the x-y plane (while time flow is unidirectional). Thus, while only past samples may be used in real-time filtering, most two-dimensional operators utilize the inputs positioned multidirectionally around the input position corresponding to the calculated element of the output. Exceptions may be, e.g., recursive two-dimensional filters or Markovian image-generating models that introduce artificially preferred directions in the image plane.

The operators mentioned in the previous paragraphs all correspond to finite impulse response (FIR) type processing of time-dependent signals. The one-dimensional filtering of the infinite impulse response (IIR) type may find its two-dimensional counterpart in the mentioned area of recursive two-dimensional processing or modeling, which, however, does not form the main stream in present applications. Variability of a one-dimensional operator in time has its counterpart in space-variant operators, and conversely, isoplanar image operators correspond to time-invariant one-dimensional operators.

2.2.1 Discrete Linear Two-Dimensional Operators

2.2.1.1 Generic Operators

In Part I, we deal mostly with linear theory, as it is more easily mathematically tractable; thus, we shall also devote attention primarily to linear two-dimensional operators, on which a substantial part of the image processing is still based. In order to simplify notation, we shall suppose that the processed images are square, e.g., input (output) images of the size $N \times N$ ($M \times M$) pixels. This does not mean any limitation with respect to principles of imaging; generalization to the case of nonsquare rectangular images is straightforward.

In analogy with continuous-space linear operators, every output pixel value $g_{i,k}$ is generally supposed to be given by a linear combination of all pixel values of the input image matrix $\bar{\bar{\mathbf{f}}}$,

$$g_{i,k} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} f_{m,n} h_{m,n,i,k}, \qquad n,m = 0,1,\ldots M-1, \tag{2.23}$$

where $h_{m,n,i,k}$ are weights determining the influence of a particular input pixel on the calculated output pixel value. A particular operator P is defined by the side sizes of the input and output images N, M, and by the four-dimensional data structure carrying the $N^2 \times M^2$ weights $h_{m,n,i,k}$ An equivalent vector equation between the vector representation of the input and the output vector is

$$\bar{\mathbf{g}} = \bar{\bar{\mathbf{P}}} \bar{\mathbf{f}}, \tag{2.24}$$

where $\bar{\bar{\mathbf{P}}}$ is the operator matrix, sized $M^2 \times N^2$ (rows $\times$ columns, hence very large). This is the most general form of a discrete linear two-dimensional operator that encompasses, among others, two-dimensional global filtering, or mask operators, including space-variant versions, and also two-dimensional unitary transforms. Equation 2.24 is formally very simple, and thus suitable for formal derivations*. On the other hand, the vector form of image data is less lucid than the matrix form, and from the computational complexity point of view, the direct use of Equation 2.24, though correct, would be

*As an example, when a regular matrix $\mathbf{P}$ characterizes a distortion of the image $\mathbf{f}$, then it is obviously (though rather theoretically) possible to obtain the restored original image $\mathbf{f}$, knowing the distorted image $\mathbf{g}$, simply by inverting the equation, e.g., $\bar{\mathbf{f}} = \bar{\bar{\mathbf{P}}}^{-1} \bar{\mathbf{g}}$.

obviously rather cumbersome ($M^2 \times N^2$ multiplications and additions) and memory-intensive (only $\mathbf{P}$ itself requires $M^2 \times N^2$ memory locations). Similar difficulties apply also to the equivalent matrix equation that may be derived from Equation 2.24 utilizing Equation 2.16 to express $\mathbf{g}$ and $\mathbf{f}$ (see [18]),

$$\begin{aligned}\bar{\bar{\mathbf{g}}} &= \sum_m \bar{\bar{\mathbf{M}}}_m^T (\bar{\bar{\mathbf{P}}}\bar{\bar{\mathbf{f}}})\bar{\mathbf{u}}_m^T = \sum_m \bar{\bar{\mathbf{M}}}_m^T \left(\bar{\bar{\mathbf{P}}} \sum_n \mathbf{N}_n \bar{\bar{\mathbf{f}}} \mathbf{v}_n \right) \bar{\mathbf{u}}_m^T = \\ &= \sum_m \sum_n \left(\bar{\bar{\mathbf{M}}}_m^T \bar{\bar{\mathbf{P}}} \bar{\bar{\mathbf{N}}}_n \right) \bar{\bar{\mathbf{f}}} \left(\bar{\mathbf{v}} \bar{\mathbf{u}}_m^T \right),\end{aligned} \tag{2.25}$$

where the matrix $\mathbf{M}$ and the vector $\mathbf{u}$ are counterparts of $\mathbf{N}$ and $\mathbf{v}$, but for the image matrix sized $M \times M$. It is a formal relation between input and output matrices useful for analytical derivations, but rather impractical for direct computations.

The enormous computational complexity of both matrix forms is why special cases encompassed in Equation 2.24 are studied individually, which offers certain simplifications and leads to better computational schemes.

2.2.1.2 Separable Operators

Separability is a special property of a class of linear operators. Most important among them are two-dimensional unitary transforms; other examples are separable filters. A *separable operator* can be decomposed into two cascaded operators: a *column operator* deriving each column of the output matrix solely from the corresponding column of the input matrix, and a *row operator* operating similarly on rows.

In order to determine the number of operations in this case, let us consider the characteristics of the matrices of both types of partial operators, $\mathbf{P}_c$ and $\mathbf{P}_r$, respectively. It is obvious that the input and output matrices of a (this way defined) separable operator must be equally sized (i.e., $M = N$ in our case). As the vectors $\mathbf{f}$ and $\mathbf{g}$ are built of N-long columns sequentially, it can easily be seen that the matrix $\mathbf{P}$ is formed of submatrices $\mathbf{P}_{m,n}$, each of the size $N \times N$ and containing the weights for the computation of the m-th output column based on the n-th input column. There are obviously $N \times N$ such submatrices in $\mathbf{P}$. It is easy to see that the matrix $\mathbf{P}_c$, sized $N^2 \times N^2$, consists of nonzero submatrices $\mathbf{P}_{n,n}$ (only along the main diagonal) and therefore has only $N \times N^2$ nonzero elements in these N

submatrices; other submatrices are obviously zero. Similarly, it can be found by inspection that nonzero elements of $\mathbf{P}_r$ are situated solely on the main diagonals of all the submatrices $\mathbf{P}_{m,n}$ so that $\mathbf{P}_r$ also has only N^3 nonzero elements; these elements, though scattered in $\mathbf{P}_r$, may be understood as forming $N \times N$-sized smaller matrices containing the weights of row-to-row computations. Therefore, realizing the separable operator as the serial (cascade) application of both partial operators defined by sparse matrices,

$$\bar{\mathbf{g}} = \bar{\bar{\mathbf{P}}}_C(\bar{\bar{\mathbf{P}}}_R\bar{\mathbf{f}}), \tag{2.26}$$

requires $2N^3$ memory locations and operations (reduction by $N/2$ in comparison with the generic case, typically of the order 10^2 to 10^3).

A special class of separable operators with *invariable submatrices* $\mathbf{P}_C$ and $\mathbf{P}_R$, ($\mathbf{P}_C = \mathbf{P}_{n,n}$, $\forall n$ and similarly adequately for $\mathbf{P}_R$) means a further simplification. Such an operator obviously needs only $2N^2$ memory locations for all the operator weights; however, the total number of operations remains. The action of either the column or row operators may be described by a matrix equation,

$$\bar{\bar{\mathbf{g}}}_c = \bar{\bar{\mathbf{P}}}_C\bar{\bar{\mathbf{f}}}, \quad \bar{\bar{\mathbf{g}}}_r = \bar{\bar{\mathbf{f}}}\bar{\bar{\mathbf{P}}}_R^T, \tag{2.27}$$

respectively. The cascade of both operators, generally expressed by the vector equation (Equation 2.26), simplifies then to the comfortable matrix equation form:

$$\bar{\bar{\mathbf{g}}} = \bar{\bar{\mathbf{P}}}_C\bar{\bar{\mathbf{f}}}\bar{\bar{\mathbf{P}}}_R^T. \tag{2.28}$$

Which of the two partial operators is performed first is thus irrelevant. Let us note, finally, that the operator matrix $\mathbf{P}$ may be expressed in terms of $\mathbf{P}_C$ and $\mathbf{P}_R$ as the direct (Kronecker) product,

$$\bar{\bar{\mathbf{P}}} = \bar{\bar{\mathbf{P}}}_C \otimes \bar{\bar{\mathbf{P}}}_R, \tag{2.29}$$

formed of submatrices $\bar{\bar{\mathbf{P}}}_{m,n} = P_C(m,n)\bar{\bar{\mathbf{P}}}_R$.

2.2.1.3 Local Operators

The *local operators* form another important special subclass of linear operators. They are characterized by being limited in the extent of the input values used for the calculation of each output pixel. On the difference from Equation 2.23, the relevant inputs form only

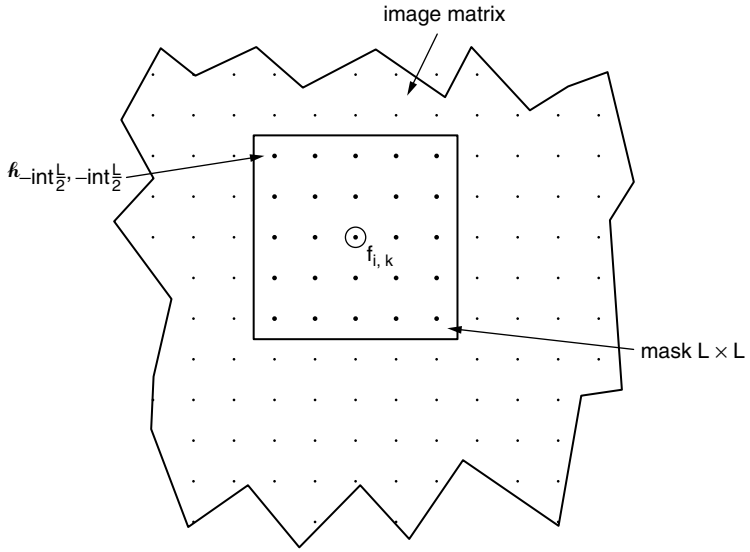


Figure 2.6 Action of a mask operator.

a $L \times L$-sized submatrix of the input matrix ($L < N$, usually $L << N$). An individual output pixel is then given by

$$g(i,k) = \sum_{m=i-L/2}^{i+L/2} \sum_{n=k-L/2}^{k+L/2} f(m,n)\, h_{i,k}(m-i, n-k). \tag{2.30}$$

Because L is usually chosen odd, in order that the $L \times L$-sized submatrix **h** would have a central element, $L/2$ should be understood as its cutoff value, int($L/2$). As in Equation 2.23, the weights are generally dependent on both the position of the calculated output pixel and the source (input) pixel; this is expressed here differently, as will be immediately clarified when introducing the notion of mask operators.

The local operators are also denoted as *mask operators*. It follows from the following visual description of a local operator action (Figure 2.6). The $L \times L$ matrix **h** of the weights may be considered a mask that is to be placed on the input matrix **f** so that the central element of **h** coincides with $f_{i,k}$. The corresponding output element $g_{i,k}$* is then the weighted sum (Equation 2.30) of products

*In case of identically sized input and output matrices; in other cases, the indexing may be different (see below).

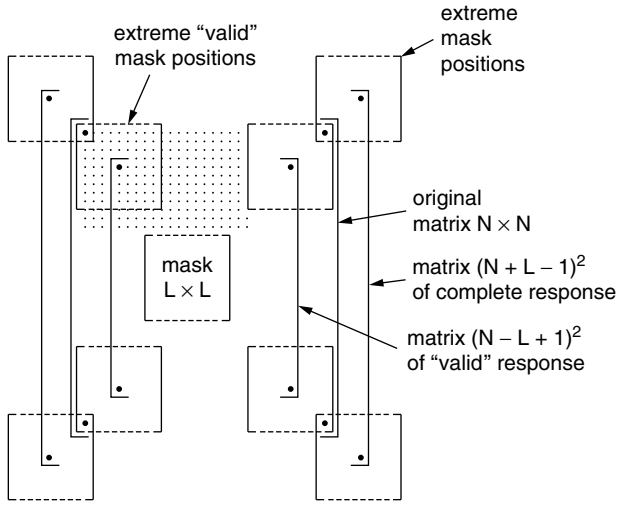


Figure 2.7 Sizes of output matrices of a mask operator.

of the values of **f** covered by the mask, with the matching mask weights. The mask is step by step shifted on the image area, thus gradually providing all the output image values.

The meaning of indices of h in the last expression can be understood as follows. Here, the matrix **h** is indexed from its center, denoted (0,0); thus, the range of both indices is $\pm\text{int}(L/2)$ and the meaning of the mask indices in parentheses is obvious. In order to encompass anisoplanar (spatially variant) operators, the mask is considered spatially variant, generally different for each position (i, k), as indicated by the lower indices at h. Note that the generic local operator is then described by N^2 matrices $\mathbf{h}_{i,k}$*; although this means a large amount of data — N^2L^2 items, this is still $(N/L)^2$ times lower than N^4 needed for a generic linear operator. As N is of the order 10^2 to 10^3, while usually $L < 10$, the saving in both computational complexity and memory requirements may be of the order 10^2 to 10^4.

The range of indices in the output image depends on the chosen definition (Figure 2.7). When only output values completely supported by the input values are taken as relevant, the output

*The number may be up to $(N + L - 1)^2$; see below.

matrix (sometimes denoted as *valid*) is obviously smaller by int(L/2) rows or columns on each side, so that the resulting side size is $N - L + 1$. When, on the other hand, all output results (Equation 2.30) are accepted if at least one term in the sum contains a valid value of $f(\ldots)$, while external input values are supposed zero (or extrapolated), the matrix is on each side greater by int(L/2); thus, the matrix (called sometimes *full*) has a side of $N + L - 1$ elements. The most frequently used definition considers just the output values contained in the matrix of the original size (the *same* matrix). The practical advantage is that even an arbitrarily long chain of mask operations does not change the image size, though the marginal bands along matrix sides are partly distorted due to lack of input information.

2.2.1.4 Convolutional Operators

An important special class of local operators is constituted by *space-invariant (isoplanar) operators*, for which $\mathbf{h}_{i,k} = \mathbf{h}$, $\forall i,k$. These are the most frequently met operators in practice. The computational requirements are basically the same as for a space-variant operator of the same mask size (though saving some overhead operations); however, the operator is fully determined by only a single matrix **h**, thus lowering memory requirements. It is easy to see that the result (Equation 2.30) can be rewritten as

$$g(i,k) = \sum_{m=-\text{int}\,L/2}^{L/2} \sum_{n=L/2}^{L/2} f(i-m,\, k-n)h(-m,-n), \tag{2.31}$$

which can be recognized as the discrete convolution of finite matrices **f** and **h**.

Really, the generic definition of the operation of *discrete convolution* of two matrices (of generally unlimited size) is

$$g(i,k) = f * fi|(i,k) = fi * f|(i,k) = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f(i-m,k-n)\, fi(m,n). \tag{2.32}$$

Obviously, the extent of summation is limited by the size of the smaller matrix; when taking $fi(m, n) = h(-m, -n)$, i.e., rotating the weight matrix by 180°, the equivalence of Equations 2.31 and 2.32

is apparent. To find another physical interpretation of $fi(m, n)$, find the response of the operator to a *discrete unit impulse* defined as

$$u(i,k) = \begin{cases} 1 & \text{for} \quad i = 0, j = 0 \\ 0 & \text{otherwise} \end{cases}, \tag{2.33}$$

situated at i_0, k_0. According to Equation 2.33, the output to $u(i - i_0, k - k_0)$ is

$$\begin{aligned} g(i,k) &= \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} u(i - i_0 - m, k - k_0 - n)\, fi(m,n) \\ &= fi(i - i_0, k - k_0), \end{aligned} \tag{2.34}$$

based on the sinking property ($u(\ldots)$ is nonzero only when both variables are zero). Thus, the discrete two-dimensional function $fi(i, k)$ is the operator's impulse response (PSF), here shifted to the position of the source impulse. The weight matrix (the mask) **h** is therefore just the 180° rotated matrix of the PSF.

It may be concluded that a space-invariant discrete linear operator realizes the finite discrete convolution of the input image with the 180° rotated weight matrix that completely characterizes the concrete operator. These operators are therefore also called *convolutional operators*. An example of such an operator matrix together with the operator's response to an image formed by isolated points of different intensities is shown in Figure 2.8. Application of the same operator to a fragment of a natural image can be found in Figure 2.9. Note that the resulting images contain negative pixel values due to the concrete type of the (difference) operator, so that the gray scale used had to be adapted correspondingly.

A convolutional operator may be realized directly in the original domain according to Equation 2.31; alternatively, it may be calculated via the frequency domain using discrete Fourier transform (Section 2.3.2) and its circular-convolution property.

Note that sometimes anisoplanar (space-variant) linear mask operators are also denoted as convolutional though space variant. Obviously, these operators cannot be realized via frequency domain.

Naturally, the local operators, as a special class of linear operators, can also be expressed in the vector form (Equation 2.24) with the operator matrix **P** having certain special properties, the most prominent of them being sparsity. Particularly in the case of (isoplanar) convolutional operators, the matrix becomes block circulant,

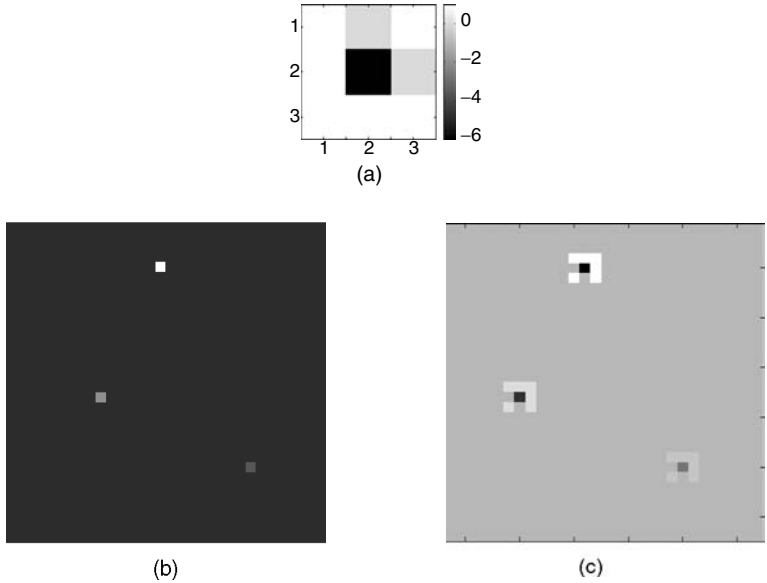


Figure 2.8 Space-invariant impulse response of a convolutional operator: (a) 3×3 mask (magnified), (b) an image sized 32×32 pixels with isolated point sources, and (c) the output image: individual weighted impulse responses can be recognized.

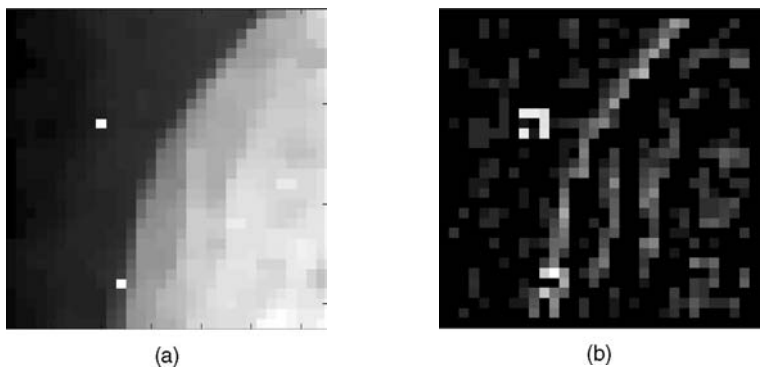


Figure 2.9 Application of the convolutional operator defined by the mask in Figure 2.8: (a) input image and (b) output image.

which is of interest in advanced theoretical considerations, e.g., in the frame of restoration methods.

The convolutional operators are frequently used in image processing, and effective realization is thus of crucial importance. Though the computational complexity of the conventionally realized local operators, $\sim N^2L^2$, is advantageous in comparison with the complexity of the generic operators, proportional to N^4, still they are rather demanding. When the mask size L is not very small, say $L > 10$, it is advantageous to realize the convolution operations *via frequency domain*, as will be described in Section 2.3.2.

2.2.2 Nonlinear Two-Dimensional Discrete Operators

2.2.2.1 Point Operators

Discrete point operators (as a counterpart of their continuous-space version, Section 1.3.4.1) represent, in the context of gray-scale images, just the *contrast transforms* (or, in other words, brightness transforms) like

$$g_{i,k} = N(f_{i,k}), \tag{2.35}$$

where the function N determines the character of the transform. Different types of the practically used functions together with their effect will be shown in Section 11.1. Obviously, the spatial relations in the image remain untouched; the point operators only change the gray scale, in which the image is presented.

Point operators for color images — *color transforms* — may be generally formulated as given by a vector valued function $\mathcal{N}$ of a vector argument,

$$\mathbf{g}_{i,k} = [\mathcal{N}_R(\mathbf{f}_{i,k}), \mathcal{N}_G(\mathbf{f}_{i,k}), \mathcal{N}_B(\mathbf{f}_{i,k})] = \mathcal{N}(\mathbf{f}_{i,k}), \tag{2.36}$$

where $\mathbf{g}_{i,k}$ and $\mathbf{f}_{i,k}$ are three-component color pixel values. In the frame of this book, a simplified form should be particularly mentioned,

$$\mathbf{g}_{i,k} = [\mathcal{N}_R(f_{i,k}), \mathcal{N}_G(f_{i,k}), \mathcal{N}_B(f_{i,k})] = \mathcal{N}(f_{i,k}), \tag{2.37}$$

that transforms a single gray-scale value into a color shade and is often used also in medical imaging to enhance the contrast or to differentiate among different objects. However, the obtained colors

have nothing to do with the real coloring of the scene, and therefore Equation 2.37 is often called a *false color transform*.

The discrete gray-scale point operators are most naturally expressed by means of two-column tables with one column containing all the possible input pixel values, while the other column represents the corresponding output values. Thus, e.g., for 8-bit gray-scale representation, the table has the form

$$\begin{array}{cc} 0 & N(0) \\ 1 & N(1) \\ 2 & N(2) \\ \vdots & \vdots \\ 255 & N(255), \end{array} \tag{2.38}$$

the second column items being naturally replaced by concrete function values, usually in the same value extent as the input. The table for the false color transform naturally has three output columns:

$$\begin{array}{cccc} 0 & N_R(0) & N_G(0) & N_B(0) \\ 1 & N_R(1) & N_G(1) & N_B(1) \\ 2 & N_R(2) & N_G(2) & N_B(2) \\ \vdots & \vdots & & \\ 255 & N_R(255) & N_G(255) & N_B(255). \end{array} \tag{2.39}$$

Such tables are used as *lookup tables*, enabling interactively in real-time controlled transforms. The input column then corresponds to a memory address, and the memory content represents the output values that may be flexibly changed according to the operator's requirements, mediated by an interface, e.g., a joystick.

We shall return to applications and possibilities of the lookup table design in Section 11.1.

2.2.2.2 Homomorphic Operators

The discrete-space homomorphic operators are direct counterparts of the operators realized by continuous-space homomorphic systems (Section 1.3.4.2). The principle of operation remains the same except that

all the operations concern isolated pixels instead of the *x-y* plane continuum. It should be noted that, in canonical representation, the inner linear part of the system may be formed by a linear discrete operator, as described in Section 2.2.1. The outer nonlinear parts are often constituted by discrete point operators (see previous paragraph), e.g., logarithmic or exponential converters; sometimes, however, they may be quite complex, like in the case of homomorphic blind deconvolution. When such partial analogue operators contain integral transforms, like Equation 1.61, these are to be replaced by corresponding discrete transforms, in the mentioned case by two-dimensional discrete Fourier transform (2D DFT). It is then necessary to observe the differences in the behavior of the continuous-space and discrete transforms and to modify the procedures (or the result interpretation), respectively. In the mentioned case, it is important to realize that the DFT has the circular (periodic) convolution property instead of the linear convolution property belonging to the integral FT.

2.2.2.3 Order Statistics Operators

Order statistics operators constitute a relatively wide class of nonlinear local operators. They are based on the notion of a defined neighborhood and on the statistics of the image values belonging to this neighborhood. These statistics can be expressed by the *local histogram* of the neighborhood (on gray-value histograms; see Section 2.1.1.2), which may also provide a fast way for computation of order statistics operators.

Topologically, an order statistics operator may be considered a mask operator (similarly as in Section 2.2.1.3), with the mask given by a neighborhood of a chosen shape (not necessarily square), in which the position of the reference pixel, corresponding to the position of the output pixel in the output image, must be specified. One of the input image values belonging to the neighborhood becomes the output, to be assigned to the respective output image position. The rule determining which of the pixels covered by the mask would become the output defines the type of filter.

We shall limit ourselves to the *generalized median filters*, which are all based on sorting of the set of input values from the neighborhood, providing a sequence

$$f_{i_1,k_1}, f_{i_2,k_2}, \ldots, f_{i_L,k_L}, \tag{2.40}$$

in which $f_{i_j,k_j} \leq f_{i_{(j+1)},k_{(j+1)}}$, $j = 1 \ldots L-1$. Obviously, the order in this series has nothing to do with spatial arrangement of the values in

the neighborhood. Nevertheless, experience shows that the essential spatial relations among pixels remain preserved in the output image as far as the chosen neighborhood is reasonably sized with respect to important details of the image.

The *plain median filter* (or simply *median filter*) selects the middle term of the sequence — the median — as the output (note that L should be an odd number). This is obviously a nonlinear operation, not easy to analyze with respect to the filter properties that are rather variable and dependent on the concrete input values in each step. However, the median filter used for noise suppression shows some very good properties, as is demonstrated in Figure 2.10. Primarily, on the difference to linear smoothing, it preserves sharp edges, thus preventing image smearing. It completely removes little objects smaller than the filter mask, while

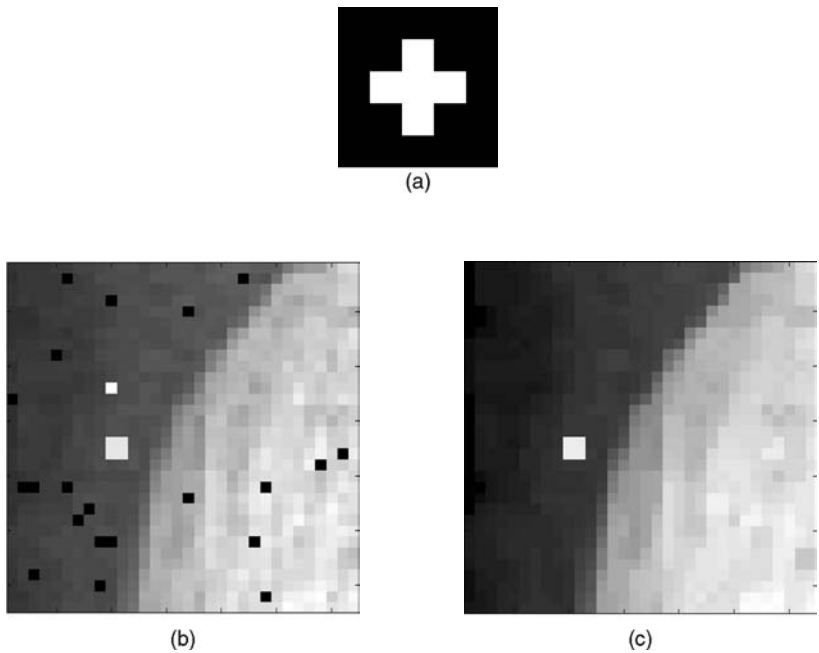


Figure 2.10 Effect of median filtering to image details: (a) chosen neighborhood (magnified), (b) image to be filtered, and (c) output of filtering.

greater objects remain untouched. It has been observed that objects comparable in size with the mask are preserved better when having a shape similar to that of the mask.

Although conceptually simple, the median filters are computationally rather intensive, the most demanding step being sorting, which has to be repeated for all mask positions.

The generalization of the median filter consists of choosing another, say j-th, term of the ordered sequence as the output. The result naturally differs from the plain median output ($j = \text{int}(L/2 + 1)$), the extremes being the filters for $j = 1$ and $j = L$, sc., *minimum* and *maximum filters*. As the ordering is the most demanding part of the filtering, it is possible, without much extra effort, to provide output images for more than a single value of j.

2.2.2.4 Neuronal Operators

Operators, based on *neural networks*, are a relatively new concept in image processing and still are a subject of lively research. They find their use in image processing (neuronal two-dimensional learning filters using mostly feedforward networks), in restoration (optimization based on minimization of energy of a feedback network), or in image analysis, namely, in classification and recognition tasks (self-organizing maps, as well as learning multilayer feedforward networks). A detailed description of this area is beyond the scope of this book (see, e.g., [3], [10], [11], [13], [27], [28]); however, a brief introduction, as follows, is necessary (for a more thorough introductory treatment focusing on signal processing applications, see [7]).

Artificial neural networks may be, from our viewpoint, considered nonlinear systems with multiple inputs and usually also multiple outputs, thus realizing a mapping from the input vector space to the output vector space, $\{\mathbf{x}\} \rightarrow \{\mathbf{y}\}$. Each network has a set of nodes used as inputs, to which the elements of the input vector (i.e., real or binary numbers) are applied, while the output nodes carry the output vector elements, acquiring real or binary values as well, depending on application. Some types of networks have spatially separated inputs and outputs; others may yield the output vector on the same nodes used before as inputs (temporal separation). The mapping is defined by means of certain network internal parameters, namely *weights* that are usually time variable — constantly, intermittently, or in a certain period of work after which they are fixed.

The desirable behavior (i.e., the mapping) of a neural network is typically obtained during a *learning period* of work. The *learning* is based on a training set of input vectors $\{\mathbf{x}_k\}$ that are successively applied to the input of the network. In principle, two different modes of learning are possible: supervised learning (learning with a teacher) or unsupervised learning (self-learning, self-organization).

The *supervised learning* requires that for each input vector $\mathbf{x}_k$ of the learning set, the corresponding (desirable) output vector $\mathbf{y}_k$ is known. The actual output of the network $\mathbf{y}_k'$ naturally differs (at least initially) from $\mathbf{y}_k$, and the difference vector $\mathbf{e}_k = \mathbf{y}_k - \mathbf{y}_k'$ controls in a way the modification of the weights in each step (processing of a particular couple $\mathbf{x}_k$, $\mathbf{y}_k$). During processing of a sequence of different couples, the mapping gradually (usually very slowly and not monotonically) converges to the desired one, or rather to its approximation. In comparison with the deterministic operators dealt with so far that are designed to have the desirable properties directly, it is a heavy-going procedure; however, this approach enables the obtaining of the needed operators even without identifying and analyzing their properties, simply by showing (usually repeatedly) enough examples of the required mapping to the network.

The *unsupervised learning* has a different area of application: it is used when the desirable mapping is not known and should be derived automatically based on automatic feature recognition (classification) of the input vectors that are again successively shown to the network, without any information on the desirable output; the training set then naturally does not contain the output part. The features on which the organization should be based are mostly also unknown ahead of time, although their automatic choice may be partly influenced by the designed mechanism of the network parameter modification. This mechanism should be obviously designed so that the mapping converges to a more or less fixed state; the mechanism (and the philosophy behind it) may partly determine the character of the resulting mapping that is nevertheless derived autonomously. This way, as a typical example, a new classification scheme may be obtained, independent on any *a priori* taxonomy.

The artificial neuronal networks consist of simple computing elements called *neurons* (Figure 2.11), which execute a simple operation

$$y = f\left(\sum_{i=0}^{N} w_i x_i\right) = f(\mathbf{w}^T\mathbf{x}) = f(\alpha), \tag{2.41}$$

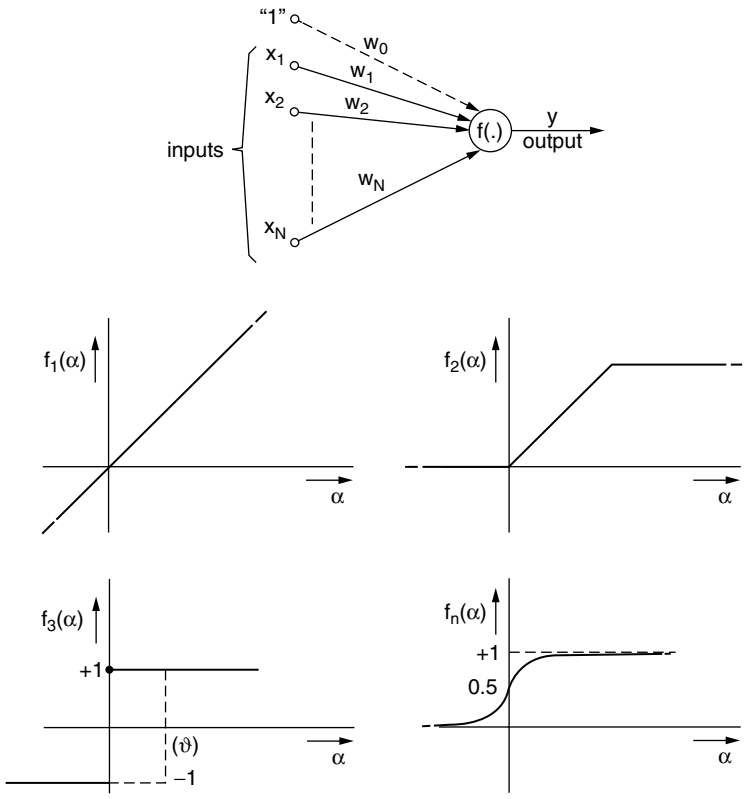


Figure 2.11 Isolated artificial neuron and its typical output functions. (From Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

where **x** is the vector of the neuron inputs, **w** contains its weights, and the scalar output y is determined by a generally nonlinear (usually monotonously increasing or step) function $f(\ldots)$; two examples are depicted. The argument α of the function is called *activation* of the neuron. The form of the function is fixed and typically the same in all neurons of a network. The neuron thus maps the N-dimensional input vector space onto a range of real numbers, or — in case of the step function — on a binary set, e.g., {0,1}. The external function of N variables realized by a neuron is dependent on the weights that are to be adjusted, mostly in the process of supervised learning, which consists of applying a series of input vectors one after another to the neuron input. If the desirable

mapping is known, the difference between the desired output y_d and the actual output y of the neuron for a particular input vector $\mathbf{x}$ may be used to iteratively correct the weight vector,

$$_{n+1}\mathbf{w} = {}_{n}\mathbf{w} + \mu(y_d - y)\mathbf{x}. \tag{2.42}$$

Here, μ is a chosen constant influencing the speed (and reliability) of convergence. This algorithm, called the *δ-rule*, may be shown to converge close to the desired mapping, though rather slowly. Other learning algorithms are used as well, but nevertheless less frequently.

Although the computational and classification possibilities of an individual neuron are rather limited, the potential increases enormously when numerous elements form a mutually interconnected structure, the neural network. Many different architectures of neural networks have been published, of which three classes seem to be most important in the signal and image processing area: multilayer feedforward networks (often called back-propagation nets), feedback networks, and self-organizing maps.

The multilayer *feedforward neural networks* (Figure 2.12) are characterized by unidirectional flow of information from the input

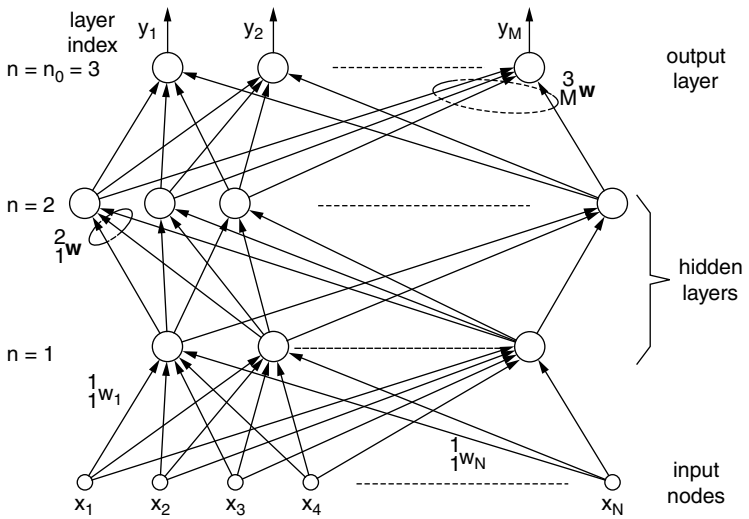


Figure 2.12 Architecture of a three-layer feedforward network. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

layer of neurons to the second layer, and from there to the next one, etc.; there is theoretically no response delay — the output vector $\mathbf{y}_k$ is the instant reaction of the network to an input vector $\mathbf{x}_k$. The weights are adjusted to provide an approximation of the desired mapping, typically by supervised learning using the method of *error back-propagation*: the elements of the error vector $\mathbf{e}_k =_d \mathbf{y}_k - \mathbf{y}_k$ propagate after every learning step back, from the output layer to all lower layers, being distributed proportionally to the weights they pass. This procedure allows determination of the individual error for each neuron in every step. Once the local error is known, each individual neuron weight vector $\mathbf{w}$ can be modified by the δ-rule, as its local input vector is known as well. This way, it is possible to modify all the weights in the network after every learning step. It has been shown by Widrow, using certain rough approximations, that the back-propagation learning is basically the steepest-descent optimization toward the minimal mean square error of the whole network. Many modifications of the algorithm exist aiming at faster and possibly monotonous learning. After many (thousands to even millions) such learning steps, the mapping realized by the network approaches usually the desired one; to achieve such a high number of learning steps with a reasonably sized training set, its application is repeated after every exhaustion, in the same or in a random order of individual training vectors. A series of steps exhausting all learning vectors is called a *learning epoch*; many epochs are usually needed to reach an acceptable response of the network.

Obviously, each element of the output vector is a nonlinear function of N variables. Considering the *Kolmogoroff theorem* that any given function of N variables can be expressed by a set of linear combinations and a repeatedly used nonlinear monotonous function of one variable (which is exactly what a multilayer neural network provides), it may be deduced that such a network can in principle realize any desired mapping and has therefore very generic use. However, the theorem does not say anything about the needed arrangement of the computational units and their count, i.e., the architecture or structure of the network. Thus, it is still a matter of experiments to design an optimal or at least acceptable structure for a desired mapping.

Multilayer networks, as the most straightforward learning systems, find numerous applications in image processing as local operators (filters) and classifiers. In both cases, a discrete image, its section, or a modification (transform) of image data form the input vector. In case of a neural filter, the output is usually a scalar, like

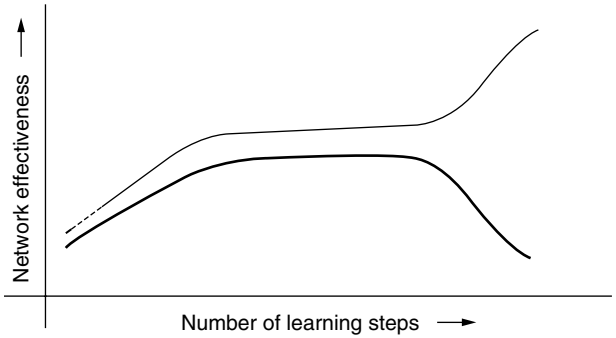


Figure 2.13 Network effectiveness as a function of duration of learning. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

in other local filters, describing (successively) individual pixels of the output filtered image. The classifiers may have a vector output of which the maximal element indicates the proper classification among the other possibilities corresponding to other output vector elements.

While it applies to any type of learning, the multilayer networks are particularly inclined to the phenomenon of *overtraining*, schematically depicted in Figure 2.13. Here, a typical development of network effectiveness, defined possibly as the inverse of the mean square error or as the percentage of proper classifications, is shown as a function of the duration of learning. The two curves correspond to effectiveness with respect to the training set (thin line) and to a testing set, similar but not including the training vectors. After a period of more or less steady improvement, a plateau is reached in both cases, when there is either slow or almost no progress; it can be interpreted as exhaustion of relevant features that all have been already recognized and implemented via weight adjustment. The effectiveness corresponding to the testing set must naturally be expected to be lower, but the character of both curves is similar so far. The further development is dramatically different: while the effectiveness regarding the training set may still increase substantially, the quality of response to the testing set vectors declines. It is explained as a consequence of the network acquiring detailed knowledge on the training set that goes beyond the relevant typical features, thus losing the needed ability of *knowledge generalization*. Prevention may consist of stopping the learning once the plateau is reached, which requires a continuous check of effectiveness during

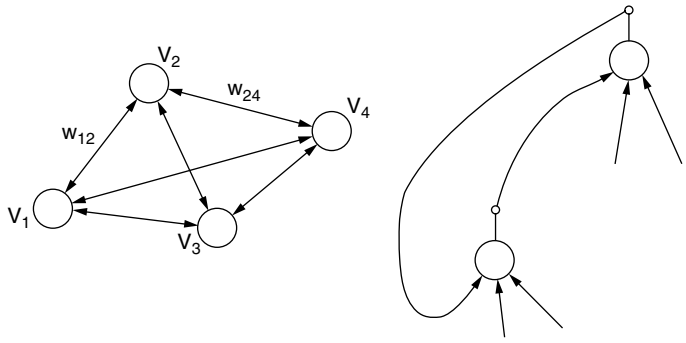


Figure 2.14 Fully interconnected feedback network. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

learning. Another approach is in diversifying the training vectors by adding a certain amount of noise so that the irrelevant details are constantly changing. The most elegant possibility is in designing (experimentally) a network with a bottleneck in the form of a layer with a low number of neurons, which blocks transfer of undesirable details while not hindering the relevant information.

Another architecture used in image processing is the *feedback network* of mutually interconnected neurons (Hopfield nets, Boltzmann machines, and the like). An example of such a simple structure is shown in Figure 2.14; only binary neurons are usually considered (i.e., with bistate outputs). The bidirectional connections here mean that the output of each neuron forms inputs to all other neurons (as on the right). The network of say N neurons works sequentially in discrete time. In each step, the state of the network, defined as the vector of all neuronal outputs, may change in only a single neuron. Obviously, the network can have 2^N different states. The neuron that has a potential of changing its state in a particular step is chosen randomly and its state is or is not changed, depending on its inputs and weights.

The initial state of the network is given by a (binary) input vector; in the initial step, the individual neurons are forced to proper states V_i. Then the network works autonomously, potentially changing its state in each step. The fundamental property of such a network is that a quantity called *network energy* E,

$$E = -\sum_i V_i \alpha_i, \tag{2.43}$$

determined by the network state **V** together with the interconnection weights, influencing the activations α_i of the individual neurons, may only decrease or maximally remain unchanged in the individual steps of working. Consequently, there is a clear tendency of the network toward a state (or states) with minimum energy. Once such a minimum (*attractor*) is reached — either the absolute minimum or a local minimum without any possibility to reach a state with a lower energy in one step — the network remains in this state. The state vector thus reached may be considered the output corresponding to the given input.

Experimenting with such a network may be interpreted in different ways. Primarily, the simple input–output mapping as described may be the goal; however, as the behavior of the network is random, and the number of attractors may be more than one, the response is generally also random. This leads to an idea of using the network as *associative memory* — when there is a number of different attractors, it may be expected that the attractor closest (i.e., most similar, in a sense) to the input vector will be reached, thus deriving the response from incomplete or distorted information. Naturally, the network must be properly trained in advance to possess desired attractors.

In other applications, the tendency toward the minimal attractor may be utilized for optimization problems; successful applications were also published in the area of image restoration. The principle is in formulating the restoration criterion so that it corresponds to the network's energy, while the image information is in a way represented by the network state. It may then be expected that, with certain probability, the absolute (or at least a reasonably low) minimum will be reached, representing a restored image. There is a danger that the network will be trapped in a local minimum (false attractor) that does not represent a reasonable restoration; it is practically very difficult to prevent generating such attractors in the process of learning or designing the network.

It would be very desirable to equip the network with the ability to escape from such a false attractor; this can be done via modifying the neurons from deterministic to stochastic ones. The obtained networks, called *Boltzmann machines*, were shown to be able to proceed, with a certain relatively low probability, even against the energy descent, thus hopefully getting to a state from which a better attractor may be reached. The degree of neuron randomness may be controlled and is gradually reduced during the experiment, which corresponds to the techniques denoted as simulated annealing; it

can be shown that the probability of reaching the optimal attractor increases this way.

The third mentioned type of networks — the *self-organizing maps* — finds its application primarily in classification or recognition tasks. As this field is rather marginal in the frame of this book, we shall only state that a typical representative of this class — the *Kohonen map* — solves two problems. It finds an optimum vector quantization of the input (e.g., image) space and simultaneously generates a low-dimensional space of representatives of input vectors, where it may be easier to detect newly discovered classes of input vectors (e.g., images). The first problem appears in image data compression (see Section 14.2); the other is met in image analysis, e.g., in texture classification used a.o. in tissue characterization (Section 13.1.3).

2.3 DISCRETE TWO-DIMENSIONAL LINEAR TRANSFORMS

Any concrete image (sc., *original image*) can be considered a linear combination of many component images of the same size as the original. When the component images are chosen according to certain strict mathematical rules (orthonormality and completeness), these *basis images* form a complete set, the *transform base*. The original image may be then expanded (or decomposed) in the series of differently weighted basis images, and such a decomposition of a particular image is unique. The coefficients, weighting the basis images so that the total sum of the series gives the original image, are called *spectral coefficients*. The complete coefficient set is denoted the *discrete spectrum* of the original image. Because the spectrum usually has the same size as the original image, it is sometimes also called the *transformed image*.

The transforms that will be described below are inversible; i.e., the discrete spectrum can be inversely transformed, which yields the original image. Thus, the information contents of the original image and of its spectrum are identical (when neglecting the influence of rounding errors in the computations).

Discrete two-dimensional transforms belong to the class of global linear two-dimensional discrete operators, described in Section 2.2.1. From another viewpoint, they constitute the class of two-dimensional unitary transforms, as will be defined below. After formulating the generic properties of the unitary transforms, we shall briefly describe the definitions and properties of the most frequently used two-dimensional discrete transforms.

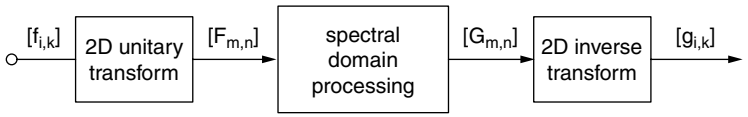


Figure 2.15 Frequency-domain processing.

The unitary transforms are extremely important in image processing, both as self-standing procedures, providing spectra that may consequently be used as feature sets in analysis, classification, or recognition tasks, and as constituents of more complex procedures. The references to such use of discrete two-dimensional transforms will be frequent in Part III of the book; let us only state here the basic notion of the frequency-domain analysis and processing (Figure 2.15). The original two-dimensional discrete signal (image matrix) is submitted to a forward unitary transform that yields the corresponding two-dimensional spectrum of the image (the term *spectrum* is used here in a generic sense, meaning the two-dimensional transform of the input image, thus not necessarily the Fourier frequency-domain spectrum). The spectrum may be then assessed as such or analyzed by a higher-level procedure, e.g., a classification algorithm in the frame of some structure analysis where the spectral coefficients are used directly or via some grouping or processing as classification features.

The spectral-domain processing, on the other hand, may be characterized as modifying the spectrum in a way corresponding to a concrete purpose and then transforming the modified spectrum by the inverse transform back into the original (image) domain. The modification of the spectrum means any processing from a very wide class, the most common being multiplying the spectrum term by term by a matrix that may be considered the spectral transfer function. Obviously, when all spectral coefficients remain untouched, as corresponds to unit transfer, the resulting output image will be identical to the original. On the other hand, some spectral coefficients (or rather sets of them — spectral areas) may be suppressed or vice versa enhanced, which is denoted as spectral-domain filtering. A special case is the filtering in the Fourier frequency domain, where the transfer function represents frequency transfer of the processing system (see above), the transfer function being the two-dimensional DFT of the isoplanar two-dimensional impulse response of the system. In other words, the output image will be the convolution of the input image and another image that

may be considered the impulse response; hence, such processing is also called convolution via frequency domain (or fast convolution). Other examples of spectral-domain processing may be found in the area of image data compression, where the spectral components (e.g., of discrete cosine or wavelet transform) are weighted according to their subjective importance in the image; the least important (invisible) components may even be completely omitted, thus achieving the desired data reduction. The compressed spectral data must be, of course, later inverse transformed, thus completing the mentioned chain, though the decompressing inverse transform may be spatially and temporally displaced.

All the two-dimensional unitary transforms are separable and, in this sense, may be considered direct generalizations of the corresponding one-dimensional transforms: a two-dimensional unitary transform is simply a cascade of two one-dimensional transforms of the same kind applied consecutively on rows and then columns of the image data matrix (or in the reversed order). It is helpful to realize this when considering the properties of a transform and interpreting its results. The definitions and most important properties of one-dimensional transforms will therefore be included below. However, it is beyond the frame of this book to describe the one-dimensional transforms in greater detail; refer to books on digital signal processing, e.g., [7], [20], etc. A brief explanation of the concept will be devoted to both forms of discrete one-dimensional wavelet transform (WT), which is still rather new and often misinterpreted.

2.3.1 Two-Dimensional Unitary Transforms Generally

The two-dimensional unitary transforms, as global discrete linear operators, are all described by Equations 2.23 and 2.24, though the notation is usually slightly different: the output matrix elements are often denoted by the capital letter corresponding to the lowercase letter denoting the input; the four-dimensional weight set of the forward transform, called the *transform core*, may be denoted **a** and the corresponding transform matrix **A**; the indices in the original domain that correspond to the spatial coordinates will be denoted i, k, while the indices in the transformed (spectral) domain may be m, n. It should be stressed that the indices only correspond (via sampling intervals) to spatial or spectral coordinates; thus, m, n are not directly frequencies even in the case of two-dimensional

DFT, when these indices are often named u, v. The output matrix of a unitary transform has the same format as the input matrix; for the sake of notational simplicity, we shall suppose that both matrices are square, of the size $N \times N$. An element of the output matrix of such a unitary transform (called a *spectral coefficient*) can thus be expressed as

$$F_{m,n} = \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} f_{i,k} a_{i,k,m,n}, \qquad i,k,m,n = 0,1,...,N-1. \tag{2.44}$$

It can be rewritten into the vector form, corresponding to Equation 2.24, when scanning both input and output matrices by columns, as

$$\bar{\bar{\mathbf{F}}} = \bar{\bar{\mathbf{A}}} \bar{\mathbf{f}}. \tag{2.45}$$

All used two-dimensional unitary transforms are separable operators with invariable submatrices, i.e., according to Equation 2.29,

$$\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{A}}}_R \otimes \bar{\bar{\mathbf{A}}}_C, \tag{2.46}$$

so that the matrix relation as Equation 2.28 applies,

$$\bar{\bar{\mathbf{F}}} = \bar{\bar{\mathbf{A}}}_R \bar{\bar{\mathbf{f}}} \bar{\bar{\mathbf{A}}}_C^T, \tag{2.47}$$

where all the matrices are merely of the $N \times N$ size. Both partial transforms are identical, $\mathbf{A}_R = \mathbf{A}_C$, and represent the corresponding one-dimensional transforms applied successively to rows and columns or vice versa. The property of separability, besides simplifying interpretation of the obtained spectra, also substantially cuts down the computational requirements (by $N/2$, i.e., typically by a factor of the order 10^3). Each of the partial one-dimensional transforms has in principle the computational complexity $\sim N^2$. Most of the one-dimensional transforms (including DFT) can be calculated using fast algorithms based on factorization of the transform kernel into sparse matrices, which lowers the computational complexity to almost only linear dependence, $\sim N \log_2 N$. The explanation of the principles of these algorithms can be found in digital signal processing literature, e.g., in Chapter 3.3.2 of [7].

The individual transforms are characterized by different kernel matrices $\mathbf{A}$. Nevertheless, all the unitary transforms share some common properties, the most important of them being *reversibility*.

As the core matrix **A** of all unitary transforms is nonsingular, it is possible to write

$$\bar{\mathbf{f}} = \bar{\bar{\mathbf{A}}}^{-1}\bar{\mathbf{F}}. \tag{2.48}$$

The kernel of the inverse unitary transforms has the property

$$\bar{\bar{\mathbf{A}}}^{-1} = \bar{\bar{\mathbf{A}}}^{*T}, \tag{2.49}$$

so that the matrix of the inverse transform can be simply obtained by conjugation and transposition of the forward kernel. The inverse transform may then be expressed by

$$f_{i,k} = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} F_{m,n}\,{}_{I}a_{m,n,i,k}, \qquad m,n,i,k = 0,1,...,N-1 \tag{2.50}$$

where ${}_{I}a$ are elements of $\mathbf{A}^{-1}$. All unitary transforms also share the property of *energy conservation* that follows from Equation 2.49,

$$\sum_{i=0}^{N-1}\sum_{k=0}^{N-1}\left|f_{i,k}\right|^2 = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1}\left|F_{u,v}\right|^2, \qquad m,n,i,k = 0,1,...,N-1. \tag{2.51}$$

With respect to stochastic images (or stochastic fields generating the images), the unitary transforms have, in a greater or lesser degree, the important properties of *energy compaction* and *element decorrelation*. The mean energy of a pixel in a discrete image (or of a spectral coefficient in a discrete spectrum) is given by the ensemble mean of its squared absolute value; the correlations among the pixels or among the spectral coefficients form the respective correlation matrices. It can be shown, via deriving the mean energy of the spectral coefficients in terms of the mean energy of the original pixels using the transform relations, that the distribution among the spectral coefficients is considerably uneven even if the original field is homogeneous with respect to the local energy. It means that the transforms have the tendency to pack a majority of the mean energy to only a few spectral coefficients, which is advantageous from the aspect of image data compression (see Section 14.2). Similarly, it can be shown that the transforms tend to produce the spectral correlation matrices with small off-diagonal terms (cross-correlations among the spectral coefficients), compared to the elements on the main diagonal

(i.e., variances of individual coefficients), even if the cross-correlations among original pixels are important. Such decorrelation in the spectral domain is also often desirable, e.g., in adaptive processing and restoration tasks. Individual transforms show these properties in different degrees; it becomes one of the criteria of practical utility.

2.3.2 Two-Dimensional Discrete Fourier and Related Transforms

2.3.2.1 Two-Dimensional DFT Definition

Two-dimensional DFT is the discrete counterpart of the two-dimensional integral Fourier transform, as explained in Section 1.2. The *forward two-dimensional DFT* is defined according to Equation 2.44 as a complete set of the spectral coefficients

$$\{F_{m,n}\} = \mathrm{DFT}_{2\mathrm{D}}\{f_{i,k}\}, \quad \text{where}$$
$$F_{m,n} = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} f_{i,k}\, e^{-j\frac{2\pi}{N}(mi+nk)}, \qquad i,k,m,n = 0,1,\ldots,N-1. \tag{2.52}$$

Obviously, the complex weights forming the basis functions of the transform are

$$\begin{aligned} a_{i,k,m,n} &= \exp\left(-j\frac{2\pi}{N}(mi+nk)\right) \\ &= \cos\left(\frac{2\pi}{N}(mi+nk)\right) - j\sin\left(\frac{2\pi}{N}(mi+nk)\right). \end{aligned} \tag{2.53}$$

In the matrix notation, the two-dimensional DFT, as a separable transform with invariable submatrices, is described by Equation 2.47, $\bar{\bar{\mathbf{F}}} = \bar{\bar{\mathbf{A}}}_R \bar{\bar{\mathbf{f}}} \bar{\bar{\mathbf{A}}}_C = \bar{\bar{\mathbf{A}}}_C \bar{\bar{\mathbf{f}}} \bar{\bar{\mathbf{A}}}_R$, where

$$\bar{\mathbf{A}}_R = \bar{\mathbf{A}}_C = \frac{1}{\sqrt{N}}\begin{bmatrix} W^0 & W^0 & W^0 & \cdots & W^0 \\ W^0 & W^1 & W^2 & \cdots & W^{N-1} \\ W^0 & W^2 & W^4 & \cdots & W^{2(N-1)} \\ & & & \cdots & \\ W^0 & W^{N-1} & W^{2(N-1)} & \cdots & W^{(N-1)^2} \end{bmatrix}, \quad W = e^{-j\frac{2\pi}{N}}. \tag{2.54}$$

The partial transforms, represented by the matrices $\mathbf{A}_R, \mathbf{A}_C$, are both the one-dimensional DFTs, as massively used in signal processing. The matrices are symmetric, $\mathbf{A}_R = \mathbf{A}_R^T$, and unitary, $\mathbf{A}_R^{-1} = \mathbf{A}_R^*$. Very efficient algorithms for DFT have been developed based either on decimation in the original or spectral domain (in fact, on expansion of the transform matrix into a cascade of sparse matrices) or on the number theoretical approach; these algorithms are in common denoted as fast Fourier transform (FFT). It should be clear that the FFT is not a special transform, but only a generic label for efficient algorithms to compute the DFT. While the two-dimensional definition algorithm (Equation 2.52) has the complexity proportional to N^4, the complexity of the one-dimensional–FFT-based algorithm of two-dimensional DFT is only of the order $N^2 \log_2 N$; for a typical $N = 10^3$, it means a speed-up by about five decimal orders. Even in comparison with the classical separable algorithm, the complexity of which is ~$2N^3$, the improvement is still by about two orders. The interested reader may find more details in the rich relevant literature (e.g., [6], Chapters 3.3 and 9.2 of [7], [18], [20]).

It can easily be shown that the *inverse two-dimensional DFT* yields the set of individual pixel values according to

$$\{f_{i,k}\} = \mathrm{DFT}_{2D}^{-1}\{F_{m,n}\}, \quad \text{where}$$

$$f_{i,k} = \frac{1}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} F_{m,n} e^{j\frac{2\pi}{N}(mi+nk)}, \qquad i,k,m,n = 0,1,...,N-1. \tag{2.55}$$

The weight coefficients ${}_I a\ldots$ are complex conjugates of $a\ldots$,

$${}_I a_{m,n,i,k} = \exp\left(j\frac{2\pi}{N}(mi+nk)\right). \tag{2.56}$$

2.3.2.2 Physical Interpretation of Two-Dimensional DFT

So far, the transform has been introduced as a mapping among a pair of matrices, i.e., a purely numerical calculation. In many applications, this view is sufficient and even preferable. It is, however, useful to interpret the two-dimensional DFT transform also in terms of physical reality and, in this way, to relate it to the integral two-dimensional FT, especially in considerations associated with frequency analysis.

In accordance with Section 2.1.1, let the sampling intervals in the original domain be Δx, Δy so that the corresponding angular sampling frequencies are $U = 2\pi/\Delta x$, $V = 2\pi/\Delta y$. The allowable extent of frequencies in the processed image is, according to the sampling theorem (Equation 2.6),

$$u \in (-U/2, U/2), \qquad v \in (-V/2, V/2); \tag{2.57}$$

to obtain a discrete spectrum, these intervals in the frequency domain have to be sampled as well. As the spectrum matrix has the same size as the original image, the number of samples in this range should be $N \times N$; thus, the sampling intervals in the frequency domain become

$$\Delta u = U/N = 2\pi/(N\Delta x), \qquad \Delta v = V/N = 2\pi/(N\Delta y) \tag{2.58}$$

and consequently

$$m\Delta u\, i\Delta x = mi\frac{2\pi}{N}, \qquad n\Delta v\, k\Delta y = nk\frac{2\pi}{N}. \tag{2.59}$$

It is then possible to rewrite Equation 2.52 so that the meaning of the samples involved would be better visible,

$$F(m\Delta u, n\Delta v) = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} f(i\Delta x, k\Delta y)e^{-j(i\Delta x\, m\Delta u + k\Delta y\, n\Delta v)}, \quad \text{i.e.,}$$

$$F(u_m, v_n) = \frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} f(x_i, y_k)e^{-j(x_i u_m + y_k v_n)}, \quad i,k,m,n = 0,1,\ldots,N-1. \tag{2.60}$$

Here, the spectral coefficients are related to concrete spatial frequencies with respect to the sampling densities and the image size in the original domain; the input image samples are related to concrete original-domain positions as well.

The inverse counterpart to Equation 2.60 obviously becomes

$$\begin{aligned} f(x_i, y_k) &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} F(u_m, v_n)e^{j(x_i u_m + y_k v_n)} \\ &= \frac{1}{N}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} |F(u_m, v_n)|\, e^{j(x_i u_m + y_k v_n + \mathrm{Arg}(F(u_m, v_n)))}, \end{aligned} \tag{2.61}$$

and by comparison with Equation 1.29, we see that the basis functions, of which the original image is now assembled, are sampled versions of the spatial complex harmonic functions with the frequencies (u_m, v_n), $\forall m,n$. As far as the spectrum is symmetrical (as is the spectrum of every real-valued image),

$$F_{m,n} = F^*_{-m,-n}, \tag{2.62}$$

the partial basis functions form complex conjugate pairs, which in turn constitute the phase-shifted two-dimensional cosine functions, the properties of which have been analyzed in Section 1.1.2. The inverse two-dimensional DFT thus provides a linear combination of the harmonic basis functions where the weights are the absolute values of the spectral coefficients, while the position of the nearest ridge (light stripe) of each basis function with respect to the origin of spatial coordinates is determined by the phase of the respective spectral coefficient. Complementarily, the forward two-dimensional DFT decomposes the original sampled image into the (sampled) basis images and provides the respective weights, i.e., spectral coefficients $F_{m,n}$.

It can be easily seen that when m, n go beyond $N/2$ and approach the limit of $N - 1$, the frequencies of the respective basis functions exceed the limits given by the sampling theorem, and consequently, aliasing appears. Thus, the functions for, e.g., $n = N/2 \pm k$ have effectively the same frequency (though generally different phases); the maximally dense harmonic functions correspond to the frequencies $U/2$, $V/2$, thus having $f_x = \frac{1}{2\Delta x}$, $f_y = \frac{1}{2\Delta y}$ cycles per meter. It implies that the highest frequency $(U/2, V/2)$ is represented in the center of the spectral matrix, while the lowest frequencies are situated in the corners, the (0, 0)-frequency (image constant-value component) being in the upper left corner. As an example, the real-valued basis functions for the small 8×8 two-dimensional DFT are depicted in Figure 2.16, together with several examples of 128×128 bases, as obtained by pairing the complex conjugate exponential basis functions (below). An example of the two-dimensional DFT spectrum can be seen in Figure 1.4 that is, in fact, an approximation of the two-dimensional integral Fourier transform by two-dimensional DFT (the only visible difference being, in this case, the finite size of the discrete spectra).

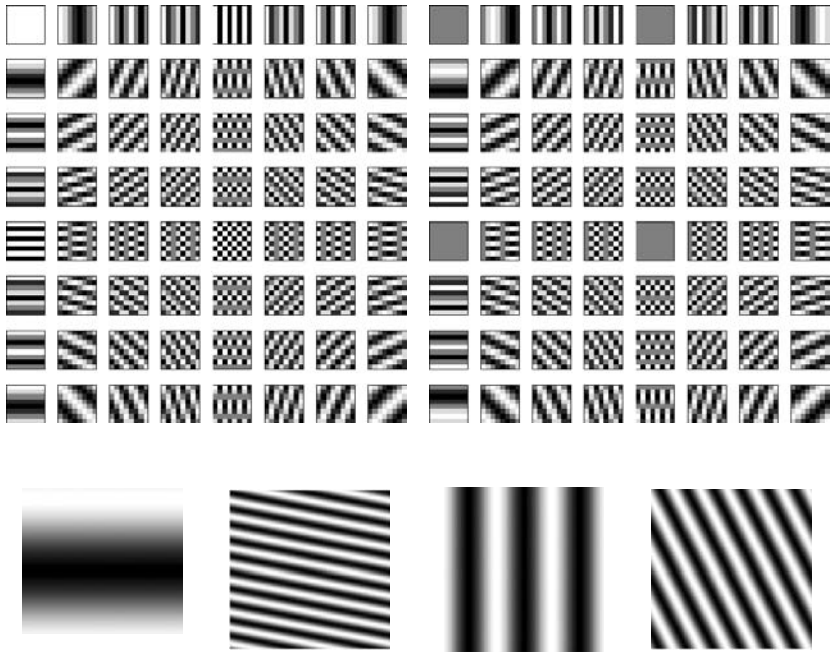


Figure 2.16 Above: Real and imaginary parts of basis functions of the 8×8 two-dimensional DFT. Below: Selected real-valued bases (obtained by pairing the complementary complex two-dimensional exponentials) of 128×128 two-dimensional DFT (for harmonics 0-1, 2-10, 3-0, 6-3).

While this arrangement of the spectrum is perfect for frequency-domain image processing, in cases when the spectrum is to be evaluated visually, as in spectral analysis tasks, it is more convenient if the (0, 0)-frequency is situated in the center of the spectral matrix. The highest frequencies are then in the corners, and the representation approximates the integral two-dimensional FT representation in the (u,v)-coordinates. This can easily be achieved by reshuffling the quadrant submatrices of the output two-dimensional transform; this can be understood as reindexing the matrix obtained in the index range $m,n \in \langle -N/2 + 1, N/2 \rangle$, to $m,n \in \langle 0, N - 1 \rangle$ thanks to the periodicity. The same result will be obtained when a checkerboard matrix consisting of ±1s multiplies the input image matrix (term by term) before the transformation is performed (the proof is left out).

2.3.2.3 Relation of Two-Dimensional DFT to Two-Dimensional Integral FT and Its Applications in Spectral Analysis

However, the exact relation of the two-dimensional DFT to the two-dimensional integral FT is still to be found. This can be mediated by the two-dimensional *discrete-space Fourier transform* (DSFT) that is defined as the function of the continuous spatial frequencies u, v,

$$F(u,v) = \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(i\Delta x, k\Delta y) e^{-j(i\Delta x\, u + k\Delta y\, v)}. \tag{2.63}$$

This is the exact two-dimensional integral spectrum of the sampled, spatially unlimited image: really, when considering Equation 2.2 for the sampled image, and realizing that two-dimensional FT is a linear operator, the spectrum may be obtained by transforming the image, formed by weighted and shifted impulses, term by term,

$$\begin{aligned} F(u,v) &= \mathrm{FT}_{2\mathrm{D}} \left\{ \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(i\Delta x, k\Delta y) \delta(x - i\Delta x, y - k\Delta y) \right\} \\ &= \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(i\Delta x, k\Delta y) \mathrm{FT}_{2\mathrm{D}} \{\delta(x - i\Delta x, y - k\Delta y)\} \\ &= \sum_{i=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} f(i\Delta x, k\Delta y) e^{-j(i\Delta x\, u + k\Delta y\, v)}. \end{aligned} \tag{2.64}$$

We have utilized here that $\mathrm{FT}_{2\mathrm{D}}\{\delta(0, 0)\} = 1$, and the shift property (Equation 1.35) of the FT. Comparison of the last expression with the discrete spectrum (Equation 2.60) shows that (disregarding the normalization factor $1/N$) the two-dimensional DFT is formed by *exact* samples of the two-dimensional integral Fourier transform of the *spatially limited* and *sampled* image. This is a fundamental conclusion, often neglected in image spectral analysis. The two-dimensional DFT is thus no vaguely approximate integral spectrum; it is the exactly defined discrete spectrum that may (but often need not) also be interpreted as the sampled version of a precise integral spectrum. However, the obtained discrete spectrum is distorted with respect to what is usually required in

spectral analysis — the continuous-frequency spectrum of the unlimited continuous-space image.

It can easily be explained and quantified how the spatial limiting (cutoff, cropping) of the image and its sampling change the result. The cropping of the image to a finite size is equivalent to multiplying the original unlimited image by a rectangular unit window covering the trimmed image. The spectrum of such a window has been shown in Figure 1.3; when the window is relatively large, the main lobe of the spectrum and the extent of its side lobes are rather narrow, covering only a small part of the spectral domain. Anyway, the spectrum of the windowed (trimmed) image becomes the convolution of the ideal unlimited-image spectrum with the window spectrum, thus smearing each single-frequency trace. This way, the existing spectral components not only influence values at proper frequencies, but also contribute to improper frequency coefficients. This phenomenon is called *leakage* and degrades the frequency resolution in the spectrum. It can in principle only be suppressed by increasing the image size. Alternatively, it is possible to limit the leakage to distant frequencies by sacrificing partly the frequency resolution: when applying a soft-edge window (like separable Gaussian, Hamming, or Hann windows), the side lobes of the window spectrum may be substantially suppressed at the cost of partly widening the main lobe.

Sampling the original image leads, as explained in Section 2.1.1 and formulated by Equation 2.5, to periodization of the spectrum and, consequently, to possible interference near to Nyquist frequencies $U/2, V/2$. It appears primarily when the sampling is not dense enough to fulfill the sampling theorem, which leads to *primary aliasing* (this error should not happen), but also due to leakage and consequential aliasing. Both phenomena usually manifest themselves as moiré patterns. This can be prevented primarily by an (expensive) increase in sampling density; a secondary help is to use antialiasing filters prior to sampling and to suppress the leakage by smooth windowing, as mentioned above.

The following discretization of the so far continuous spectrum $F(u,v) \rightarrow F(u_m,v_n)$ may hide the leakage under special circumstances, namely, when a particular narrowband (single-frequency) image component has the frequency (u,v) that falls on a node of the frequency-domain sampling grid, $u = u_m$, $v = v_n$. Then all the other samples of the smeared response would become zero, and the leakage seemingly disappears. Of course, it does not apply to other spectral components whose frequencies do not match the sampling nodes.

The described phenomena are illustrated in Figure 2.17, where this development is followed. Panel (a) presents an (theoretically unlimited) image containing a single real-valued harmonic component and its amplitude spectrum, formed by two nonzero bright spots symmetrical with respect to the origin of frequency coordinates. In panel (b), a rectangular window, intended to limit the size of the image, and its spectrum are shown. The spatially limited image as the product of images in panels (a) and (b), and its spectrum as the convolution of the spectra in panels (a) and (b), are presented in panel (c), where the leakage from the proper frequencies can be seen. Influence of sampling in the image domain is demonstrated in panel (d), where both the sampled image and the corresponding periodic spectrum are shown; the slight aliasing is due to insufficient density of sampling. Finally, panels (e1) and (e2) depict the sampled versions of the spectra together with the consequential effect in the original domain — periodization of the trimmed image. In the case (e1), the image frequency matches one of the nodes of the frequency-domain sampling grid, while the case (e2) shows the more common case, when the frequency falls in among the nodes. It should be understood that a natural image consists of many such components, and each of them is a subject of a similarly distorted spectral presentation.

2.3.2.4 Properties of the Two-Dimensional DFT

It should be understood that the one-dimensional DFT is nothing other than the *exact* finite Fourier series of the band-limited function described by the samples f_n, $n = 0 \ldots N - 1$, and periodically extended, $f_{n+N} = f_n$.* Consequently, the two-dimensional DFT is a two-dimensional Fourier series with analogous properties:

- *Periodical extension* of both the image and the spectrum,

$$f_{i,k} = f_{i+N,k+N}, \quad F_{m,n} = F_{m+N,n+N}, \quad \forall i,k,m,n \tag{2.65}$$

*See [7], Chapter 9.2.1. The requirement of band limitation applies to the function extended over the period borders; i.e., if there is a jump between the left and right values, the transform considers an interpolated function among valid samples that is band-limited, but different from the original signal.

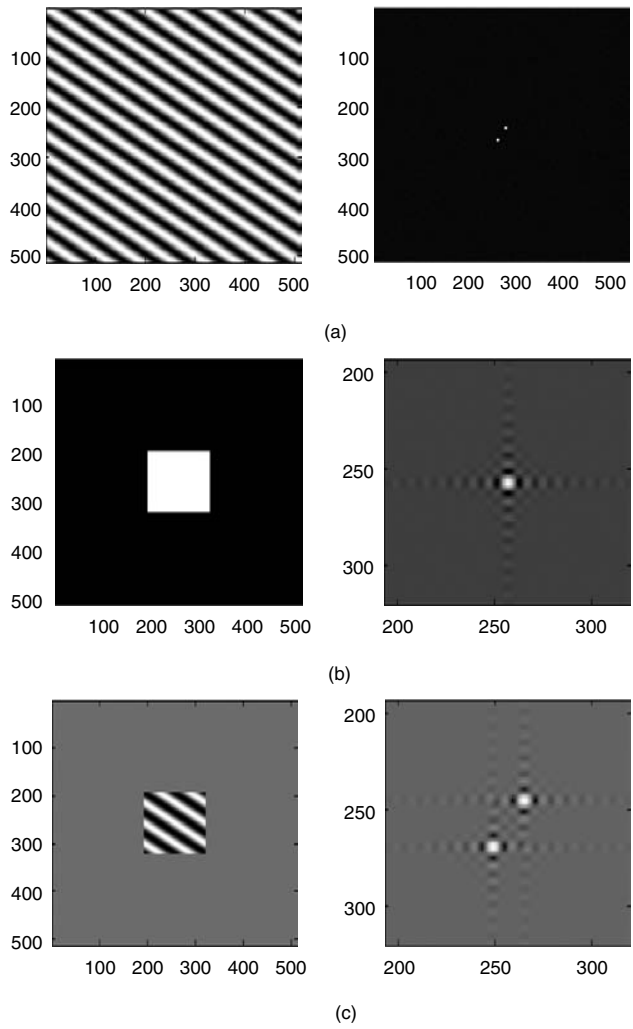


Figure 2.17 Development of the two-dimensional DFT spectrum of a simple (monofrequency) image from the integral two-dimensional FT spectrum (note the different scales in the individual figures for better display). (a) Theoretically unlimited monofrequency image and its amplitude spectrum. (b) Rectangular window and its spectrum. (c) Spatially limited image and its spectrum. (d) Sampled image and its spectrum. (e1) Sampled version of the spectrum (right) and the corresponding periodic image (left — more periods; center — a period magnified) in case the image frequencies coincide with nodes of the spectral sampling grid. (e2) Similar to e1, but when the image frequencies fall in among the sampling nodes.

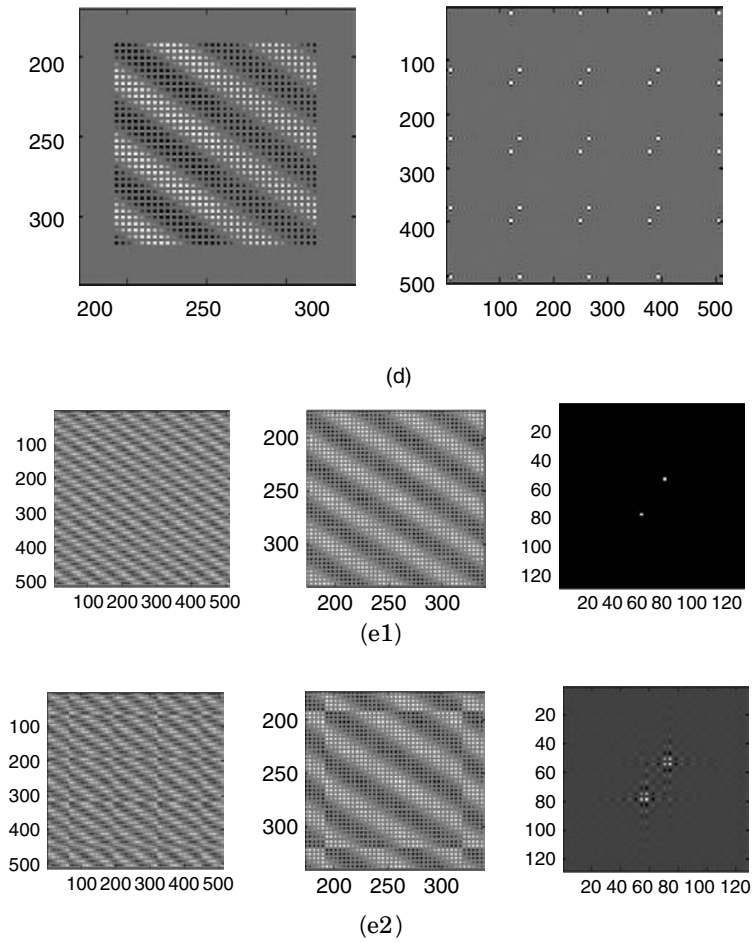


Figure 2.17 (Continued).

so that even the negative indices are meaningful; the effects of possible left–right and top–bottom discontinuities in the original domain manifest themselves by bright axes in the amplitude spectra, representing the numerous artificial spectral components necessary to describe the sharp edges at the image boundaries.

- The output transform matrix may either be presented as provided by the definition, with low frequencies in the corners (the (0,0)-frequency in the upper left corner), or, thanks to

the periodicity, be rearranged, with (0,0)-frequency in the center and the highest frequencies in the corners (see above).

- The two-dimensional DFT of real-valued images exhibits *conjugate symmetry*,

$$F_{-m+kN,-n+kN} = F^{*}_{m,n}, \qquad \forall m,n,k. \tag{2.66}$$

Thanks to this property, two quadrants of the spectrum carry the complete spectral information (demonstrating the preservation of information: input of N^2 real numbers is represented by the output of $N^2/2$ complex numbers (N^2 real components)), which offers computational and memory savings.

- The two-dimensional DFT transform is *separable*, *symmetric*, and *unitary*,

$$\begin{aligned} &\bar{\bar{\mathbf{A}}}_{2DDFT} = \bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{A}}}_R \otimes \bar{\bar{\mathbf{A}}}_C = \bar{\bar{\mathbf{A}}}_{1DDFT} \otimes \bar{\bar{\mathbf{A}}}_{1DDFT}, \\ &\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{A}}}^{T},\ \bar{\bar{\mathbf{A}}}^{-1} = \bar{\bar{\mathbf{A}}}^{*}. \end{aligned} \tag{2.67}$$

- The two-dimensional DFT is *exactly* equal to the sampled version of the two-dimensional integral FT of the finite-size and sampled original image (see above).
- The *basis images* of the two-dimensional DFT are the two-dimensional harmonic functions.
- Two-dimensional DFT possesses the two-dimensional *circular convolution property*,

$$\begin{aligned} \mathrm{DFT}_{2\mathrm{D}}\{f \otimes g|_{i,k}\} &= \mathrm{DFT}_{2\mathrm{D}}\{f_{i,k}\}\ \mathrm{DFT}_{2\mathrm{D}}\{g_{i.k}\} \\ &= \{F_{m,n}G_{m,n}\},\ \forall i,k,m,n. \end{aligned} \tag{2.68}$$

As indicated by the last equation, the product of spectra is the term-by-(corresponding)-term product, not the matrix product. The symbol $\otimes$ means the operation of *circular convolution* of two $(N \times N)$-sized matrices,

$$f \otimes g\,|_{i,k} = \sum_{p=0}^{N-1}\sum_{q=0}^{N-1} f_{[i-p],[k-q]}\, g_{p,q}, \qquad i,k,p,q \in \langle 0, N-1\rangle, \tag{2.69}$$

where $[p] = p \bmod N \in \langle 0, N-1\rangle$ is the principal value of the index. The *modulo* operation reflects the already

mentioned input image periodicity imposed by the two-dimensional DFT; the values for negative (or else excessive) indices are defined this way. The proof of the theorem is straightforward but tedious: $\{F_{m,n}\}$ and $\{G_{m,n}\}$ are expressed as the two-dimensional DFTs of **f** and **g** according to Equation 2.52, their product inverse transformed according to Equation 2.55, and the result in the form of a multiple sum then rearranged so that partial sums of exponential terms are obtained that, being discrete impulse functions, filter out certain products in the sum, thus finally leading to Equation 2.69. The circular convolution theorem has important implications for convolutional operations realized via the frequency domain (see below).

- The *compaction* and *decorrelation properties* of the two-dimensional DFT are average (some other transforms perform better in these respects).

2.3.2.5 Frequency Domain Convolution

Although the circular convolution (Equation 2.69) is not identical to the linear convolution (Equation 2.32) realized by two-dimensional space-invariant linear systems, it is possible to obtain the same numerical results via either way if certain rules are observed. In comparison with the linear discrete convolution, the difference is in the periodicity of data imposed by the circular convolution: not only the input matrices are taken as being two-dimensionally periodically extended with the period, determined by the size $N \times N$ of the used two-dimensional DFT, but the same applies naturally to the output matrix as well. Let $M \times M$ and $L \times L$ denote the sizes of the original input matrices, the latter being smaller (the mask). The linear convolution *full* output matrix (see above) has the size $(M + L - 1)^2$, while the output size of the circular convolution is always given by the transform size $N \times N$. It is obviously possible to prevent the disturbing effects of periodicity by choosing sufficiently large $N \times N$ matrices so that the desired output matrix fits in. A single period would thus encompass all the useful data, and no interperiodic interference appears. With respect to the full output matrix size, the transform size must be chosen so that $N \geq (M + L - 1)$; usually the nearest higher integer power of 2 is used to enable realization of the two-dimensional DFT by the most efficient fast algorithms.

Naturally, the size of all the involved matrices must be equal in order that the spectra may be multiplied term by term. Increasing

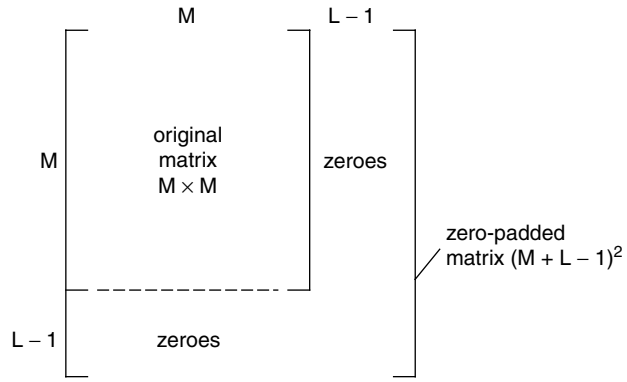


Figure 2.18 Padding of input matrices by zeros.

both input matrices to the $N \times N$ size is achieved by padding with zeros (Figure 2.18). These zeros then ensure that the terms in the circular convolution (Equation 2.69) that would otherwise cause the interperiod interference are eliminated, and the resulting matrix contains the results equal to those provided by the linear convolution (Equation 2.32).

Two-dimensional convolution via the frequency domain consists of two-dimensional DFT — transforming both (zero-padded) input matrices, then providing the term-by-term product, and finally inverse transforming the obtained spectrum,

$$\begin{aligned} \{F_{m,n}\} &= \mathrm{DFT}_{2D}\{\{f_{i,k}\}'\}, \qquad \{H_{m,n}\} = \mathrm{DFT}_{2D}\{\{h_{i,k}\}'\} \\ \{g_{i,k}\} &= \mathrm{DFT}_{2D}^{-1}\{F_{m,n}H_{m,n}\}, \end{aligned} \tag{2.70}$$

where $\{h_{i,k}\}'$ means the zero-padded matrix $\{h_{i,k}\}$. Because the DFTs may be calculated with the computational complexity $\sim N^2 \log_2 N$, the total computational load, including the spectra multiplication, is $\sim N^2 (3 \log_2 N + 1)$, which compares favorably with the classical algorithm complexity M^2L^2 if the smaller (mask) matrix is not very small. (Usually $N \cong M$, for which the fast convolution is then theoretically preferable if about $L^2 > 3 \log_2 N + 1$; for $N = 10^3$, taking into account the overhead computations, this *fast convolution* turns out to be beneficial for roughly $L > 10$.) Though conceptually more complex, the fast convolution of greater matrices via frequency domain may be much simpler computationally, thus enabling realization of sophisticated procedures hardly feasible otherwise.

2.3.2.6 Two-Dimensional Cosine, Sine, and Hartley Transforms

The two-dimensional DFT has many desirable analytic properties, but the necessity of complex-valued computation may be considered a disadvantage. The transforms presented in this section are related to two-dimensional DFT in that they are also based on harmonic basis functions; however, they use only real-valued calculations. Though the cosine, sine, and Hartley transforms are self-standing autonomously defined transforms, they can also be interpreted in terms of two-dimensional DFT of modified images, which is advantageous with regard to understanding some of their properties.

The *cosine transform* of an image matrix $\{f_{i,k}\}$, may be regarded the two-dimensional DFT of an evenly symmetrical image (Figure 2.19) obtained by adding three mirror-symmetrical versions of the original image forming the quadrant D of the symmetrical image matrix $\{{}_{\text{sym}}f_{i,k}\}$, $i, k \in \langle -N, N\rangle$. Thus, the values in the individual quadrants are ${}_{\text{sym}}f_{i,k} = f_{-i-1,-k-1}$, ${}_{\text{sym}}f_{i,k} = f_{-i-1,k}$, ${}_{\text{sym}}f_{i,k} = f_{i,k-1}$, and ${}_{\text{sym}}f_{i,k} = f_{i,k}$ in quadrants A, B, C, and D, respectively. This way, each value of the original image is present four times, including the element $f_{0,0}$. (The odd symmetry, in which all the quadrants share a single element $f_{0,0}$, leads to certain computational complications and is less frequently used.) The two-dimensional DFT of such a symmetrical image is naturally real-valued. Coming out of Equation 2.52, we obtain by obvious modification

$$ {}_{sym}F_{m,n} = \frac{1}{2N}\sum_{i=-N}^{N}\sum_{k=-N}^{N} {}_{sym}f_{i,k}\, e^{-j\frac{2\pi}{2N}\left(m\left(i+\frac{1}{2}\right)+n\left(k+\frac{1}{2}\right)\right)}, \qquad (2.71) $$

where the frequency increment is halved due to doubled image size, and the 1/2 values reflect the shift of the spatial coordinate origin from the position of $f_{0,0}$. Nevertheless, the frequency resolution remains the same as that for the two-dimensional DFT of the original image, as there is no new information in the modified image; the intermediate frequency-domain coefficients are merely interpolated. In comparison with two-dimensional DFT of the original image, the directional information on the spectral components is partially lost, as the symmetry leads to four different orientations of each symmetrisized harmonic function, as can be derived from the figure.

The spectral matrix $\{{}_{\text{sym}}F_{m,n}\}$ is naturally of the same size as the symmetrisized image, but due to the symmetry (Equation 2.66) of the (real) spectral values, the contents of all four spectral

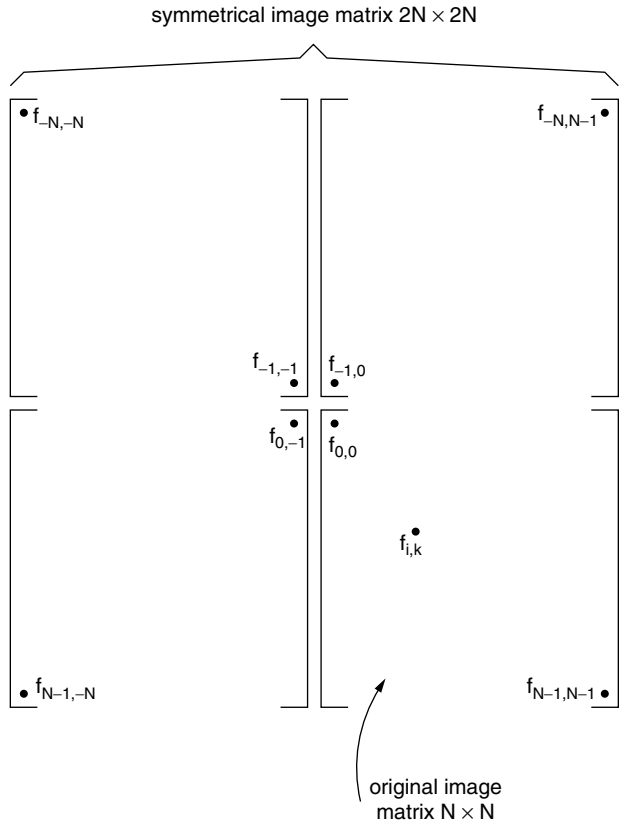


Figure 2.19 Forming of (evenly) symmetrical image.

quadrants are identical; thus, just a single quadrant sized $N \times N$ carries the complete information. The computation may be simplified by pairing the complex conjugate exponential terms, thus obtaining the real-valued cosine basis functions. This way, a new transform is defined, called the *discrete cosine transform* (DCT), as

$$\{F_{m,n}\} = \mathrm{DCT}_{2\mathrm{D}}\{f_{i,k}\}, \quad \text{where}$$

$$F_{m,n} = \frac{2}{N} K_m K_n \sum_{i=0}^{N-1} \sum_{k=0}^{N-1} f_{i,k} \cos\left(\frac{\pi}{N} m\left(i+\frac{1}{2}\right)\right) \cos\left(\frac{\pi}{N} n\left(k+\frac{1}{2}\right)\right),$$

$$i,k,m,n = 0,1,...,N-1, \quad K_0 = 1/\sqrt{2}, \quad K_m = 1 \quad \text{for} \quad m = 1,2,\ldots,N-1. \tag{2.72}$$

Obviously, the two-dimensional cosine transform is separable with invariable submatrices, $\bar{\bar{\mathbf{F}}} = \bar{\bar{\mathbf{A}}}_R \bar{\bar{\mathbf{f}}} \bar{\bar{\mathbf{A}}}_C$. The identical column and row transform $N \times N$ matrices are

$$\bar{\bar{\mathbf{A}}}_R = \bar{\bar{\mathbf{A}}}_C = \frac{1}{\sqrt{N}}[a_{k,n}] = \frac{1}{\sqrt{N}}\left[c_k \cos\left(\frac{\pi}{N} k\left(n+\frac{1}{2}\right)\right)\right], \qquad (2.73)$$

$$c_0 = 1, \quad c_k = \sqrt{2}, k \neq 0.$$

The corresponding inverse transform is

$$\{f_{i,k}\} = \mathrm{DCT}_{2D}^{-1}\{F_{m,n}\}, \quad \text{where}$$

$$f_{i,k} = \frac{2}{N}\sum_{m=0}^{N-1}\sum_{n=0}^{N-1} K_m K_n F_{m,n} \cos\left(\frac{\pi}{N} m\left(i+\frac{1}{2}\right)\right)\cos\left(\frac{\pi}{N} n\left(k+\frac{1}{2}\right)\right), \qquad (2.74)$$

$$i,k,m,n = 0,1,\ldots,N-1, \quad K_m \text{ as above.}$$

As an example, the basis functions for the small 8×8 two-dimensional DCT are depicted in Figure 2.20, together with an example of a DCT spectrum.

It should be emphasized that the two-dimensional DCT, though it may be interpreted in terms of the two-dimensional DFT, is an autonomous transform defining a particular mapping between the

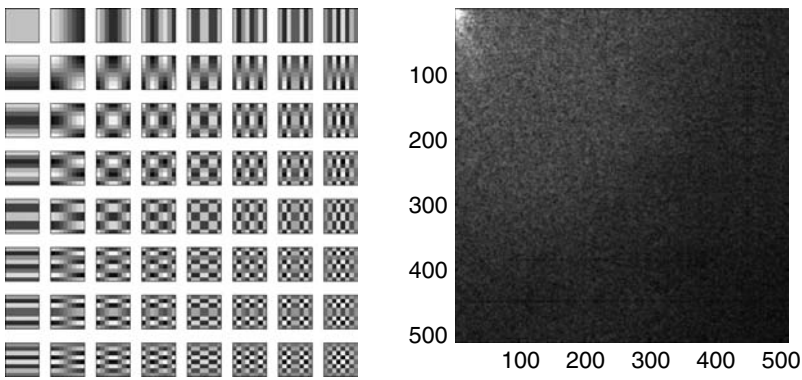


Figure 2.20 Basis functions of the 8×8 two-dimensional DCT (left) and the DCT spectrum (right) of the image in Figure 1.4.

original and spectral-domain matrices sized $N \times N$. Its main properties may be summarized as follows:

- The spectrum is real-valued; all the computations concern only real numbers.
- The spectral extent covers only positive frequencies in the range between zero and the Nyquist frequency in each direction (i.e., $U/2$, $V/2$).
- The doubled frequency resolution is only apparent; every second value may be interpolated from neighbors.
- The directional information on the harmonic components of the image is partly lost.
- Fast algorithms of two-dimensional DCT exist, similar in properties to those of the two-dimensional DFT.
- The compaction and decorrelation properties of the two-dimensional DCT are excellent. It may be explained by the fact that, for natural images that can be modeled by first-order highly correlated Markov chains, the DCT approximates well the optimal Karhunen–Loeve transform. Two-dimensional DCT is therefore massively used in the nowadays common image data compression algorithms and in the related standards, e.g., JPEG and MPEG.

The two-dimensional *discrete sine transform* is in a sense a counterpart of the two-dimensional DCT, as it can be interpreted in terms of the two-dimensionally DFT of the antisymmetrically extended (two-dimensionally doubled) image. Because it is also a separable transform with invariable submatrices, it suffices to state the general form of an element of the $N \times N$ submatrix,

$$\bar{\bar{\mathbf{A}}}_R = \bar{\bar{\mathbf{A}}}_C = \sqrt{\frac{2}{N+1}}[a_{k,n}] = \sqrt{\frac{2}{N+1}}\left[\sin\left(\frac{\pi(k+1)(n+1)}{N+1}\right)\right]. \quad (2.75)$$

The properties of the two-dimensional sine transform are similar to those of two-dimensional DCT. It also has very good compaction property for images that can be modeled as Markov sequences with a lower intrinsic correlation than in the case of two-dimensional DCT.

Still another real-valued transform related to two-dimensional DFT is the two-dimensional *Hartley transform*. Being separable, it can be defined by

$$\bar{\bar{\mathbf{A}}}_R = \bar{\bar{\mathbf{A}}}_C = \frac{1}{\sqrt{N}}[a_{k,n}] = \frac{1}{\sqrt{N}}\left[\sin\left(\frac{2\pi kn}{N}\right) + \cos\left(\frac{2\pi kn}{N}\right)\right]. \quad (2.76)$$

In a sense, the Hartley transform is a compromise between the sine and the cosine transforms, its basis functions again being real harmonic functions, but phase shifted by $\pm\pi/4$ with respect to the pure sine or cosine basis.

2.3.3 Two-Dimensional Hadamard–Walsh and Haar Transforms

2.3.3.1 Two-Dimensional Hadamard–Walsh Transform

Two-dimensional *Hadamard transform* (HT) is a separable unitary transform with invariable submatrices that are defined recursively for $N = 2^p$ (integer p) as follows:

$$\begin{aligned} \bar{\bar{\mathbf{A}}}_R &= \bar{\bar{\mathbf{A}}}_C = \mathbf{H}_{(N)}, \\ \mathbf{H}_{(2)} &= \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \\ \mathbf{H}_{(2N)} &= \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{H}_{(N)} & \mathbf{H}_{(N)} \\ \mathbf{H}_{(N)} & -\mathbf{H}_{(N)} \end{bmatrix}. \end{aligned} \tag{2.77}$$

Obviously, the transform values are real as well as all calculations that do not include multiplication, except for scaling the final result. Fast algorithms to compute the two-dimensional HT (or two-dimensional HWT; see below) are available. It can easily be proved that the Hadamard transform is symmetric and orthogonal, and thus also unitary, $\mathbf{H} = \mathbf{H}^T = \mathbf{H}^{-1}$.

The one-dimensional basis functions corresponding to the Hadamard matrices may be considered sampled and amplitude-scaled versions of rectangular (binary) functions valued ±1. The number of sign changes per N samples of these row functions, called the row's "sequency," characterizes an average frequency of the function and may be used as the characteristic spectral-domain variable. However, the sequences are not ordered monotonously in the Hadamard matrices. By reordering the rows so that the sequences increase monotonously, a modified transform is obtained named *Hadamard–Walsh transform* (HWT). The two-dimensional basis functions of the size 8×8 are shown in Figure 2.21; note that the first group of the functions provides a view of a representation of individual one-dimensional Hadamard functions.

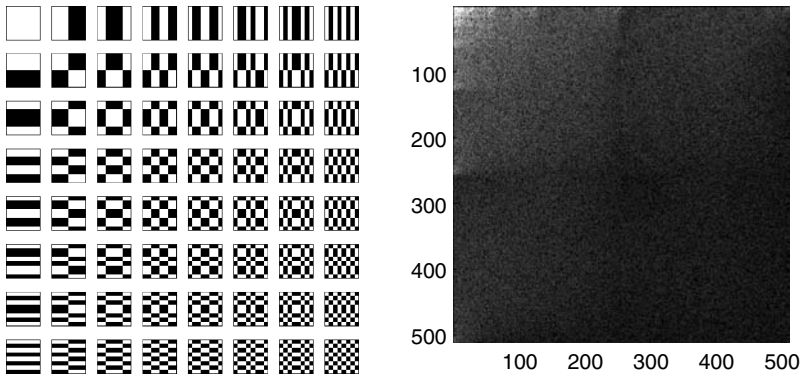


Figure 2.21 Basis functions of the 8×8 two-dimensional Hadamard–Walsh transform (left) and the H-W transform (right) of the image in Figure 1.4.

The Hadamard (Hadamard–Walsh) transform is extremely fast and simple to implement, so that it is used namely in coding applications. Its compaction properties are rather good for images with high in-frame correlations. On the other hand, the efforts to utilize the HWT as an approximation of the DFT in amplitude spectral analysis practically failed, as the HWT spectrum is only a very rough approximation of the Fourier amplitude spectrum.

A brief note should be devoted to the *slant transform* that is related in a sense to the Hadamard transform in having the same initial matrix $\mathbf{S}_2 = \mathbf{H}_2$ and being defined recursively, $\mathbf{S}_{n+1} = \mathbf{M}\{\mathbf{S}_n\}$, where $\mathbf{M}$ is a relatively complicated matrix operator (for details, see, e.g., [6]). On the other hand, its basis functions are not samples of rectangular shapes, but rather are formed of samples of piece-wise linear slant sections of repeated or alternating slope. The transform has been designed with the requirement of being realizable by a fast algorithm, and with the first basis function being a linear (monotonously increasing) sequence. The other basis functions are naturally crossing the zero level more than once; again, all the sequences up to $N - 1$ are present. The transform shows a very good to excellent compaction property for common images.

2.3.3.2 Two-Dimensional Haar Transform

The two-dimensional Haar transform is similar to the two-dimensional Hadamard transform in that the basis functions are also sampled rectangular waves; however, the character of the binary

waves is different. As the Haar transform is also a separable operator, the difference can be demonstrated on the simpler one-dimensional case. While the one-dimensional Hadamard–Walsh functions are distributed over the whole range of N samples, this does not apply to most of the Haar functions that are supported (i.e., of nonzero value) on only a part of that range. This changes the situation substantially: the concrete spectral coefficient, based on such a basis function, depends on only a part of the one-dimensional signal course (e.g., on a spatially limited part of an image row), thus carrying certain spatial localization information in the spectral value. Naturally, the available spatial resolution depends on the extent of the (nonzero) support that is different in individual Haar basis functions. This is a new feature, conceptually different from all other transforms discussed so far, which derived each spectral coefficient from all the N image samples. Hence, no spatial localization of the spectral information has been possible so far.

The one-dimensional discrete Haar functions of the length N are defined for $k = 0, 1,\ldots,N-1$, where $N = 2^n$ (integer n), based on decomposition of $k = 2^p + q - 1$ (integer p, q), which is unique. The k-th function is defined as

$$ {}_{(N)}h_{k,l} = \frac{1}{\sqrt{N}} \begin{cases} 2^{\frac{p}{2}} & \text{when} \quad \frac{l}{N} \in \left(\frac{q-1}{2^p}, \frac{q-0.5}{2^p} \right) \\ -2^{\frac{p}{2}} & \text{when} \quad \frac{l}{N} \in \left(\frac{q-0.5}{2^p}, \frac{q}{2^p} \right), \\ 0 & \text{otherwise} \end{cases} \quad k, l = 0,1,\ldots,N-1. \tag{2.78} $$

This defines the one-dimensional Haar transform matrices $\mathbf{H}_{(N)} = [{}_{(N)}h_{k,l}]$, which in turn form the separable two-dimensional Haar transform $\bar{\bar{\mathbf{A}}} = \bar{\bar{\mathbf{A}}}_R \otimes \bar{\bar{\mathbf{A}}}_C$ using the row and column submatrices,

$$ \bar{\bar{\mathbf{A}}}_R = \bar{\bar{\mathbf{A}}}_C = \mathbf{H}_{(N)}. \tag{2.79} $$

As it is visible from Figure 2.22, the Haar coefficients are given by differences of averages of neighboring intervals (of neighboring two-dimensional spatial areas; see Figure 2.23). As every spectral coefficient carries information on its spatial position, the spectrum is formed of

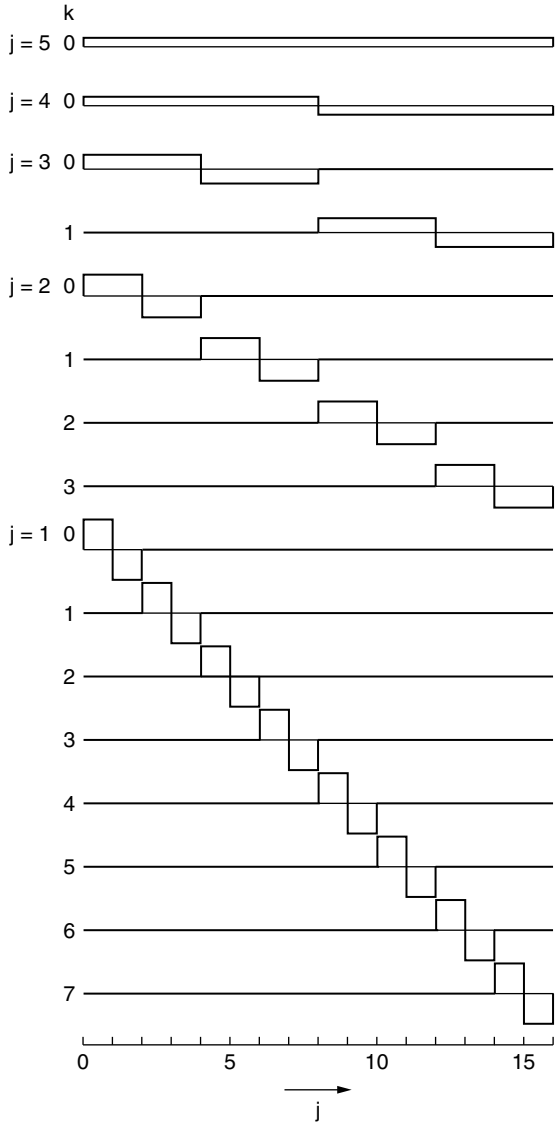


Figure 2.22 Family of one-dimensional Haar basis functions for $N = 16$. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

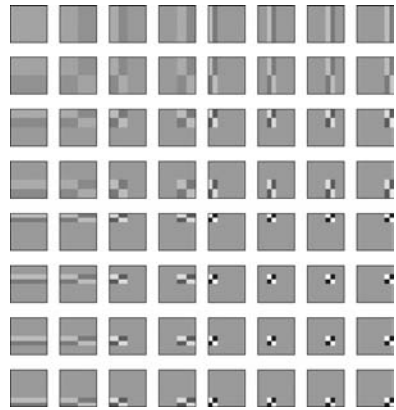


Figure 2.23 Basis functions of two-dimensional Haar transform for $N = 8$.

variously scaled differentiated images (Figure 2.24). The transform is inversible; i.e., it contains the complete information on the image.

The properties of the two-dimensional Haar transform follow from its definition:

- It is a real-valued and orthogonal transform, $\mathbf{H} = \mathbf{H}^T = \mathbf{H}^{-1}$.
- As multiplication is only needed for rare amplitude normalization and the basis functions are rather sparse, it is very fast.
- It enables spatial localization of the spectral features (with differing resolution in different coefficients).

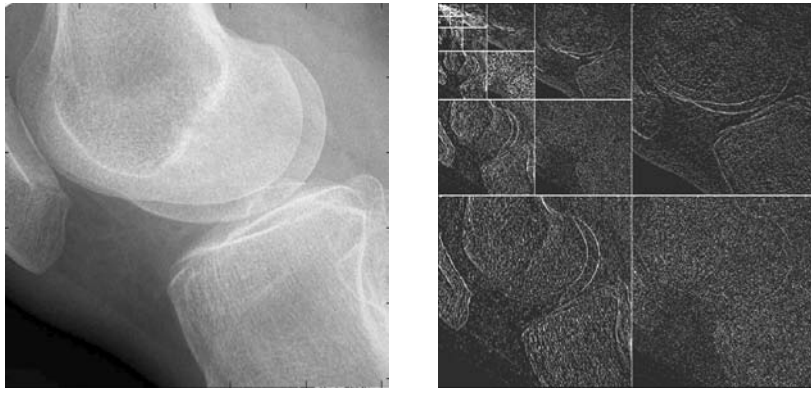


Figure 2.24 Haar transform (right) of the x-ray image from Figure 1.4 (left).

- The compaction properties of the Haar transform are rather poor.
- Though designed about a century ago, the Haar functions turn out nowadays to be the simplest of the standard Daubechies wavelets, used commonly in topical wavelet transform analysis.

2.3.4 Two-Dimensional Discrete Wavelet Transforms

2.3.4.1 Two-Dimensional Continuous Wavelet Transforms

The family of *wavelet transforms* (WT) is characterized by certain common properties of their basis functions, primarily the finite support in both the frequency and original domains, and the scalability. An example of the full-featured one-dimensional wavelet family (Haar functions) has already been given in Section 2.3.3. The two-dimensional wavelet transforms are separable, and therefore it suffices to describe the properties of one-dimensional transforms that apply either to rows or columns of the processed matrices. The two-dimensional basis functions are obtained as combinations of the one-dimensional versions, as visible, e.g., in Figure 2.23, where the one-dimensional functions form the profiles of the functions in the left column and upper row.

Common wavelets are designed as usually fast oscillating functions of short extent of nonzero values (short support in the original domain). It means that only the short interval of the original signal (e.g., of a row profile of the input image) influences the respective spectral coefficient. This way, certain positional information belongs to the coefficient; however, several mutually shifted versions of the same wavelet function are needed in order to describe the complete signal (e.g., a complete row). The corresponding spectral coefficients thus describe the same frequency content, but for different parts of the signal. On the other hand, it may be shown that a wavelet (exactly its reversed version) should act as the impulse response of a filter, thus extracting the frequency-domain information for a certain frequency band. It is desirable that the frequency-domain support (the interval of nonzero frequency response of the filter) is also finite and short*. This is, precisely taken, in contradiction to

*This is not well fulfilled by the Haar functions.

the requirement of finite original-domain support; thus, a compromise must be admitted: either only one of the supports is finite and the other one is then only approximately limited, with more or less negligible values outside of a certain interval, or both supports are compromised. There is a limit on available resolution in the original-domain and frequency-domain coordinates that is linked to the Heisenberg principle of uncertainty: the more compact the wavelet is in one domain, the more dilated it is in the other, as follows from the scaling theorem of the Fourier transform.

It turns out that most of the natural images can be well described by a small number of two-dimensional WT coefficients, as the sharp edges, lines, and details are more easily represented by a low number of wavelet functions than, e.g., by FT coefficients related to harmonic bases. In contrast, the Fourier or other transforms, based on infinite-support basis functions, need a much higher number of components to achieve the same degree of accuracy in approximating details. In other words, the two-dimensional WTs have generally very good compaction properties for images; that is why they are finding an increasingly important role in image data compression.

A basic property of a wavelet is its *scalability*. This means that a family of wavelets that forms the basis of a one-dimensional WT is completely derived from a single one-dimensional *mother wavelet* $\psi(x)$. A particular wavelet function,

$$\psi_{a,s}(x) = \frac{1}{\sqrt{a}}\psi\left(\frac{x}{a} - s\right), \tag{2.80}$$

is controlled by two parameters, the scale a and the shift s. As far as these parameters are continuous, an infinite set of derived wavelets is available. An example of a mother wavelet and two derived modifications can be seen in Figure 2.25.

According to scaling properties of the integral Fourier transform,

$$\mathrm{FT}\left\{\psi\left(\frac{x}{a}\right)\right\} = a\Psi(aw), \qquad \text{where} \qquad \Psi(w) = \mathrm{FT}\{\psi(x)\}, \tag{2.81}$$

so that the spectrum of the wavelet shrinks by the scale factor a when the wavelet is dilated by a. If the mother wavelet has only a finite (or approximately finite) frequency support, the border frequencies of the band covered by the derived wavelet are shifted correspondingly.

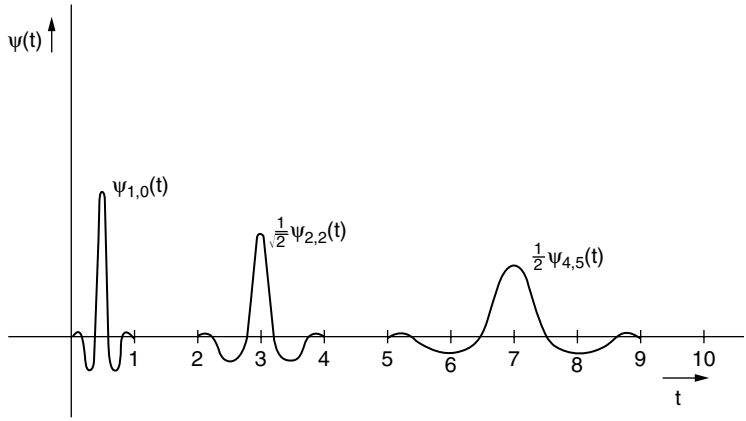


Figure 2.25 A mother wavelet and its scaled and shifted versions. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

In this way, the frequency range covered by a particular wavelet can be controlled by the scale a. The influence of the shift parameter s is obvious: simply shifting the wavelet along the spatial coordinate. Thus, a concrete WT coefficient describes a concrete frequency content determined by a in a particular space interval, given by s. The resulting WT spectrum thus belongs to the category of time-frequency (space-frequency in case of images, and more precisely, space-scale) analytic methods. This way, the spectrum is two-dimensional for a one-dimensional input signal, as shown in Figure 2.26.

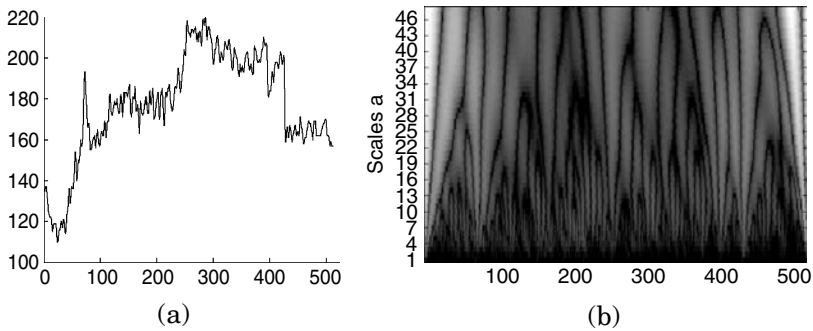


Figure 2.26 Example of one-dimensional discrete CWT spectrum (b) of an image profile. (a) Profile of a row in the image in Figure 2.24.

The *continuous WT* is defined as the correlation integral between the analyzed function and the generic modification of the mother wavelet function,

$$\mathrm{WT}\{f(x)\} = S_{WT}(a,s) = \frac{1}{\sqrt{a}} \int_{x=-\infty}^{\infty} f(x)\psi\left(\frac{x}{a} - s\right)\mathrm{d}x. \qquad (2.82)$$

Naturally, the numerical computation of the transform, based on the sampled input signal, is based on a sum rather than on integration, though the commonly used abbreviation CWT (continuous wavelet transform) may suggest the latter. Therefore,

$$\mathrm{CWT}\{f_i\} = F_{CWT}(a,s) = \frac{1}{\sqrt{a}} \sum_{i=0}^{N-1} f_i\ \psi\left(\frac{x_i}{a} - s\right) \qquad (2.83)$$

is a WT counterpart of DSFT, being continuous (though two-dimensional) in the spectral domain while coming out of the discrete input. However, the spectrum in continuous variables a, s cannot be computed numerically; thus, the (a,s)-space must be sampled in a way. If a two-dimensional equidistant grid is used, we obtain (in the above sense) a counterpart of DFT — discrete spectral coefficients derived from a discrete input. The commonly used name "continuous wavelet transform" is obviously improper and misleading; in order to maintain certain terminological consistency, let us denote the sampled version of Equation 2.83 as *discrete CWT* (though *DCWT* may sound strange). It is obviously

$$\mathrm{DCWT}\{f_i\} = \{F_{CWT}(m\Delta a, n\Delta s)\} = \left\{\frac{1}{m\Delta a} \sum_{i=0}^{N-1} f_i\ \psi\left(\frac{x_i}{m\Delta a} - n\Delta s\right)\right\}. \qquad (2.84)$$

This is what produces the two-dimensional gray-scale plots of WT spectra of one-dimensional signals (including Figure 2.26b); if the sampling grid is fine enough, the result resembles the continuous spectrum.

In the two-dimensional case of WT, the corresponding spectrum of the two-dimensional image is four-dimensional (the spectral space has two scale factors and two shifts), and thus not easy to represent visually. Therefore, the regularly sampled DCWT is rarely used in image processing.

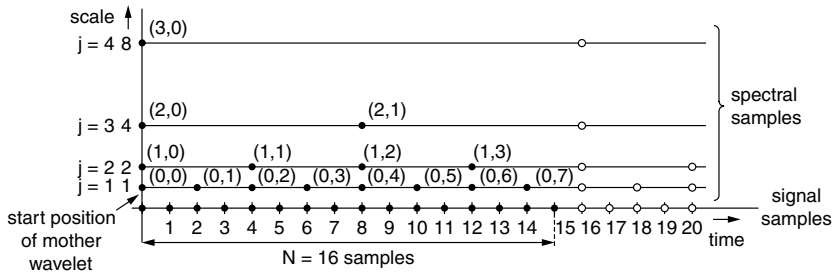


Figure 2.27 Dyadic sampling in the scale-space domain. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE Press, London, 2000. Courtesy IEE Press, London.)

2.3.4.2 Two-Dimensional Dyadic Wavelet Transforms

Another possibility for sampling the CWT defined by Equation 2.83 is the *dyadic sampling*,

$$a = 2^m, \quad s = na = n2^m, \quad m = 1, 2, \ldots, \log_2 N, \quad k = 0, 1, \ldots, (N/2^m - 1) \tag{2.85}$$

as schematically depicted in Figure 2.27. The scale variable is here sampled in a dyadic (octave) sequence, while the spatial axis is divided linearly, with the density that corresponds to the scale. If the support of the mother wavelet (for $m = n = 0$) is $x \in \langle 0, 1 \rangle$, then the shift step on each scale is just equal to the respective wavelet support, thus covering the signal completely on every scale level. The sampling density in space decreases with increasing scale; this is physically sound as the transform coefficients on higher levels need a longer spatial correlation interval to determine the resemblance index $F_{\mathrm{CWT}}(a, s)$. The example of a wavelet family chosen on the dyadic scale is the Haar basis function set in Figure 2.23.

Sampling the CWT at the nodes of the dyadic grid and normalizing the magnitude gives the *dyadic wavelet transform* (DWT),

$$\mathrm{DWT}\{f_i\} = \left\{F_{CWT}(2^m, n2^m)\right\} = \left\{\frac{1}{\sqrt{N}\,2^m} \sum_{i=0}^{N-1} f_i \;\; \psi\left(\frac{x_i}{2^m} - n2^m\right)\right\}. \tag{2.86}$$

When denoting the samples of the (m,n)-th wavelet as

$$ {}_{m,n}w_i = \frac{1}{\sqrt{N}\,2^m}\,\psi\left(\frac{x_i}{2^m} - n2^m\right), \tag{2.87} $$

the DWT may be rewritten as a purely discrete transform between the number sequences,

$$ \mathrm{DWT}\{f_i\} = \{F_{DWT}(m,n)\} = \left\{\sum_{i=0}^{N-1} f_i\ {}_{m,n}w_i\right\}. \tag{2.88} $$

This formula is independent of the physical meaning of the coordinates. Obviously, for a given sequence length N, there are $\log_2 N + 1$ different scales, including the last one, formally denoted $F_{\mathrm{DVT}}(\log_2 N + 1, 0)$, which provides a constant basis function.

The one-dimensional *inverse dyadic wavelet transform* (IDWT) synthesizes the original signal as a (two-dimensional) linear combination of the basis wavelets,

$$ \mathrm{DWT}^{-1}\{F_{\mathrm{DWT}}(m,n)\} = \{f_i\} = \left\{\sum_{m=0}^{\log_2 N+1}\ \sum_{n=0}^{N/2^m-i} F_{\mathrm{DWT}}(m,n)\ {}_{m,n}w_i\right\}. \tag{2.89} $$

Providing that the chosen wavelets fulfill certain strict requirements (as do all the commonly used published families like Haar, Daubechies, biorthogonal, coiflets, symlets, etc.), the one-dimensional DWT is a unitary, real, and thus orthogonal transform, the core matrix of which contains the wavelet samples. Any two-dimensional DWT can be expressed as the separable transform based on the one-dimensional transform submatrices in the standard way.

However, the DWT may also be interpreted with advantage in terms of operation of quadrature-mirror filter banks. Besides providing a different view—decomposing the signal gradually into pairs of signals (details—approximation), this interpretation also enables the execution of only a *partial DWT* (see example in Figure 2.28), decomposing the signal only to a certain level, which may be sufficient and even advantageous. This interpretation also provides a way to fast algorithms for the transform. The details are beyond the scope of this book and can be found, e.g., in [27]; a simple link exists between the complete DWT decomposition and unitary transforms [7]. The DWT has excellent compaction properties for most of

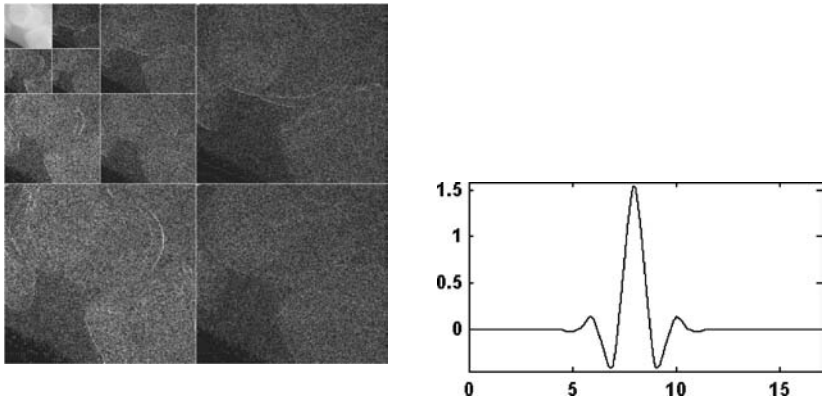


Figure 2.28 Partial DWT (left) of the image in Figure 2.24, using the biorthogonal wavelet (right).

the natural images; this applies even to complete images, which eliminates the need to code individual sub-blocks of the image independently. The field of wavelet-based image data compression is presently very active, and the blockless DWT-based approach has already become a part of standardized compression methods. A lot of DWT-based spectral-domain processing is also reported on. However, many of the published methods are rather intuitive as to the processing in the spectral-domain concerns (e.g., filtering with intuitively designed modification of the DWT coefficients); also, the choice of a concrete type of wavelets is often only heuristic or random, without deeper reasoning.

2.3.5 Two-Dimensional Discrete Karhunen–Loeve Transform

The two-dimensional *discrete Karhunen–Loeve transform* (DKLT), sometimes also called *Hotelling transform*, is a transform that, although it is not easy to compute, serves as the golden standard in the sense that its compaction and decorrelation properties are optimal. A particular two-dimensional DKLT applies to a stochastic image (i.e., a family of images forming a stochastic field), and the transform matrix is dependent on the properties of the field, namely, on the autocorrelation function that is four-dimensional in the generic case. Thus, the transform matrix is case dependent, in the sense that for different stochastic fields, the two-dimensional DKLTs are different.

Generally, two-dimensional DKLT is not a separable transform, but if the autocorrelation matrix (sized $N^2 \times N^2$) of the image represented by $N^2 \times 1$ vector is separable, the transform becomes separable as well. It is the usual approach to approximate the correlation matrix in this way, as otherwise the computations become extremely demanding. We shall restrict ourselves to this case, so that briefly introducing only the one-dimensional DKLT will be sufficient.

Let a column (or a row) $\mathbf{u}$ of an image be an $N \times 1$ stochastic vector (e.g., a realization of a one-dimensional stochastic field) with the autocorrelation matrix

$$\mathbf{R}_{uu} = \mathcal{E}\{\mathbf{u}\mathbf{u}^{*\mathrm{T}}\} \tag{2.90}$$

derived as the ensemble mean. Should the transform be identical for all columns, all the columns have to belong to the same stochastic field. A unitary transform $\mathbf{\Phi}$ is sought, yielding the discrete spectrum

$$\mathbf{v} = \mathbf{\Phi}\mathbf{u}, \tag{2.91}$$

formed of the spectral coefficients that are mutually uncorrelated. The autocorrelation matrix of the spectrum is then a diagonal matrix

$$\begin{aligned} \mathbf{R}_{vv} &= \mathcal{E}\{\mathbf{v}\mathbf{v}^{*\mathrm{T}}\} = \mathcal{E}\{\mathbf{\Phi}\,\mathbf{u}\mathbf{u}^{*\mathrm{T}}\mathbf{\Phi}^{*\mathrm{T}}\} \\ &= \mathbf{\Phi}\,\mathcal{E}\{\mathbf{u}\mathbf{u}^{*\mathrm{T}}\}\mathbf{\Phi}^{*T} = \mathbf{\Phi}\,\mathbf{R}_{\mathrm{uu}}\mathbf{\Phi}^{*T} = \Lambda, \end{aligned} \tag{2.92}$$

so that

$$\mathbf{R}_{\mathrm{uu}}\mathbf{\Phi}^{*T} = \mathbf{\Phi}^{*T}\Lambda, \tag{2.93}$$

as $\mathbf{\Phi}^{-1} = \mathbf{\Phi}^{*\mathrm{T}}$ for a unitary transform. It is derived in matrix theory that the last equation will be satisfied (not exclusively) when the columns of the matrix $\mathbf{\Phi}^{*T}$ are eigenvectors of the matrix $\mathbf{R}_{\mathrm{uu}}$. Thus, the design of the KLT consists of identification of the autocorrelation matrix $\mathbf{R}_{\mathrm{uu}}$ and determination of its eigenvectors — a relatively demanding problem.

The desired transform (DKLT) is therefore

$$\mathbf{v} = \mathrm{DKLT}\{\mathbf{u}\} = \mathbf{\Phi}\mathbf{u}, \qquad \mathbf{u} = \mathrm{DKLT}^{-1}\{\mathbf{v}\} = \mathbf{\Phi}^{*\mathrm{T}}\mathbf{v}. \tag{2.94}$$

Note that DKLT is not the only transform having the complete decorrelation property, as its matrix is not a unique solution of Equation 2.93.

The main properties of the discrete Karhunen–Loeve transform are as follows:

- The spectral coefficients have all zero ensemble mean and are mutually uncorrelated, as follows from the above derivation.
- The DKLT has an excellent compaction property: among all the unitary transforms **A** sized $N \times N$, it packs the maximum average energy into any chosen subset of first spectral coefficients v_k, $k = 0, 1,\ldots,m < N - 1$. Thus, when $E_m(\mathbf{Au})$ means the energy contained in the first m spectral coefficients of an arbitrary unitary transform **A** of an original **u,** then

$$E_m(\mathbf{\Phi u}) \geq E_m(\mathbf{Au}) \tag{2.95}$$

- The error caused by zeroing higher-indexed spectral coefficients (sc., basis restriction error) is minimal among all unitary transforms **A**. More precisely, when $\mathbf{v} = \mathbf{Au}$ and the shortened spectrum **w** is

$$\mathbf{w} = \left\{ w_k = \begin{cases} v_k, & k = 0,1,\ldots,m-1 \\ 0, & k \geq m \end{cases} \right\}, \tag{2.96}$$

the modified original-domain vector becomes $\mathbf{z} = \mathbf{A}^{-1}\mathbf{w}$. The mean of square length of the error vector,

$$\mathcal{E}\{|\mathbf{u} - \mathbf{z}|^2\}, \tag{2.97}$$

characterizes the error due to spectrum shortening. It can be proved that this error is minimum when DKLT is used, $\mathbf{A} = \mathbf{\Phi}$. It has important implications in assessing the efficiency of image data compression methods.

The generalization of DKLT to two dimensions generally requires provision of a large four-dimensional data set carrying the autocorrelation function values of the type

$$r(i,k;i',k') = \mathcal{E}\{u_{i,k}u_{i',k'}\}. \tag{2.98}$$

This is itself extraordinarily difficult; also, the $N^2 \times N^2$ eigenvalue problem to be solved in the frame of a generic two-dimensional DKLT derivation is computationally extremely demanding. As already mentioned, the two-dimensional transform becomes more practical when the autocorrelation function is separable into row

and column correlation functions, the former being common for all rows, while the latter is common to all columns,

$$r(i,k;i',k') = r_C(i,i')r_R(k,k'). \tag{2.99}$$

In this case, the two-dimensional transform matrix $\mathbf{\Phi}$ is separable as well, consisting of the one-dimensional DKLTs derived separately for rows and for columns,

$$\mathbf{\Phi}, = \mathbf{\Phi}_C \otimes \mathbf{\Phi}_R, \tag{2.100}$$

so that the DKL spectra from the original image (and vice versa) are

$$\mathbf{V} = \mathbf{\Phi}_C^{*\mathrm{T}}\mathbf{U}\mathbf{\Phi}_R^*, \qquad \mathbf{U} = \mathbf{\Phi}_C\mathbf{V}\mathbf{\Phi}_R^T, \tag{2.101}$$

where all the matrices are of the size $N \times N$, making both derivation of the transform matrices and the transform computations feasible. Naturally, the requirements of separability, and of all columns (and all rows), each belonging to a respective single stochastic process, mean severe limitations.

Further details on DKLT and two-dimensional DKLT and their relations to other transforms may be found, e.g., in [6].

2.4 DISCRETE STOCHASTIC IMAGES

The analysis of discrete stochastic images, generated by discrete stochastic fields, is very similar to the concept of continuous-space continuous-value stochastic images, introduced in Section 1.4, so that most of the concepts and results presented there are adequately applicable to the discrete case as well. Is is recommended that the reader refer to this section and try to derive the discrete-case individual properties and rules based on the differences stated below.

The main differences stem rather obviously from the following two fundamental facts:

- The space sampling, together with the finite size of images, means a finite number of pixels in the discrete image. Thus, a stochastic field is constituted by only a finite number N^2 of stochastic variables (pixel values), in case of two-dimensional $N \times N$ images.
- The pixel magnitudes, i.e., the realization values of the stochastic variables, are quantized to a certain discrete scale of a finite size, say to J levels. Tables of probabilities therefore replace the continuous probability density functions.

Thus, the individual probability distributions are expressed by one-dimensional probability vectors, while multidimensional tables (two-dimensional matrices, etc.) express the joint probabilities.

2.4.1 Discrete Stochastic Fields as Generators of Stochastic Images

A discrete stochastic field consists obviously of a *finite family of discrete images* (image matrices). Such a family $\mathbf{f}_w = \{\mathbf{f}_{w_l}\}$ —the *basis of the discrete stochastic field*—is a finite set of M mutually different image matrices $\boldsymbol{f}_{w_l} = [f_{w_l}(i,k)] = [f_{w_l}(\mathbf{r})]$ that can be expected in a given application. Which of the functions is chosen as the concrete realization depends again on the result of the *associated experiment*, with a finite set of possible results $W = \{w_1, w_2, \ldots, w_l, \ldots, w_{L-1}\}$ that appear with a given probability distribution. Again, the mapping $W \rightarrow \mathbf{f}$ has to be unambiguous (but not necessarily inversible) in order that each experiment yields an image.

The pixel values $f_{w_l}(\mathbf{r}_0)$ (the $\mathbf{r}_0$ being defined by concrete indices, $\mathbf{r}_0 = [i_0, k_0]^{\mathrm{T}}$) in all images may be regarded as realizations of a random variable $f_w(\mathbf{r}_0)$ dependent on the result w of the associated experiment. The *local characteristics* at the position $\mathbf{r}$ are now:

- The *local probabilities*:

$$p_f(z_j, \mathbf{r}), \quad \text{with} \quad \sum_{\forall j} p_f(z_j, \mathbf{r}) = 1 \tag{2.102}$$

- The *local mean value* of $f_w(\mathbf{r})$:

$$\mu_f(\mathbf{r}) = \mathcal{E}_w\{f_w(\mathbf{r})\} = \sum_{\forall j} z_j p_f(z_j, \mathbf{r}) \tag{2.103}$$

- The *local variance*:

$$\sigma_f^2(\mathbf{r}) = \mathcal{E}_w\left\{|f_{w_l}(\mathbf{r}) - \mu_f(\mathbf{r})|^2\right\} = \sum_{\forall j} (z_j - \mu_f(\mathbf{r}))^2 p_f(z_j, \mathbf{r}). \tag{2.104}$$

The local probability distribution can be approximated, based on an ensemble of realizations, by the *ensemble histogram*, which is the vector $_{\mathbf{r}}\mathbf{h} = [_{\mathbf{r}}h_j]$ of counts (statistical frequencies) of each

discrete value z_j in the ensemble of pixel values at $\mathbf{r}$. Obviously, the individual probabilities are approximated as

$$p_f(z_j, \mathbf{r}) \approx \frac{{}^{\mathbf{r}}h_j}{L}, \qquad \text{with} \qquad p_f(z_j, \mathbf{r}) = \lim_{L\to\infty} \frac{{}^{\mathbf{r}}h_j}{L} \tag{2.105}$$

for a statistically stable environment. The local mean may be estimated by ensemble averages analogously as in the continuous-space case.

The relations between two locations $\mathbf{r}_1, \mathbf{r}_2$ are described by the two-dimensional joint probability distribution

$$p_f(z_j,\ z_k, \mathbf{r}_1,\ \mathbf{r}_2) = \mathcal{P}\{f_w(\mathbf{r}_1) = z_j \wedge f_w(\mathbf{r}_2) = z_k\}, \tag{2.106}$$

where both values are to be taken from the same realization. As already mentioned in Section 1.4, the relations between more than two locations are rarely investigated due to practical obstacles. However, thanks to the final number of images that can be represented in a given matrix with a given discrete scale of gray, it is possible, at least theoretically, to construct the finite table of probabilities of concurrent appearances of prescribed values at all pixels of the image, i.e., the joint probabilities of the order N^2. It is interesting that such a table would be obviously the complete vector of probabilities of all individual images representable in the given matrix.

When two *concurrent stochastic fields* $f_w(\mathbf{r})$ and $g_w(\mathbf{r})$, both controlled by the same associated random experiment, are under analysis, the *mutual relations* among them are also expressed by joint probabilities. Particularly, the joint probabilities for pairs of points are

$$p_{fg}(z_j, z_k, \mathbf{r}_1, \mathbf{r}_2) = \mathcal{P}\{f_w(\mathbf{r}_1) = z_j \wedge g_w(\mathbf{r}_2) = z_k\}. \tag{2.107}$$

2.4.2 Discrete Correlation and Covariance Functions

The autocorrelation function $R_{ff}(\mathbf{r}_1, \mathbf{r}_2)$ is defined and approximated in a similar way as in the continuous-space case,

$$R_{ff}(\mathbf{r}_1, \mathbf{r}_2) = \mathcal{E}_w\{f_{w_i}(\mathbf{r}_1) f_{w_i}(\mathbf{r}_2)\} \approx \frac{1}{M} \sum_{i=1}^{M} f_{w_i}(\mathbf{r}_1) f_{w_i}(\mathbf{r}_2); \tag{2.108}$$

only the positional vectors $\mathbf{r} = (i, k)$ are interpreted in terms of indices instead of positional coordinates. Quite analogously, the expressions for the *autocovariance function* and the *autocorrelation coefficient*, as presented in Section 1.4.2, may be adapted to the discrete case. Obviously, the same also applies to *cross-correlation* and *cross-covariance functions*.

However, the functions are now discrete as far as the individual variables are concerned. Thus, in the generic case, each of the mentioned functions is represented by a four-dimensional data structure indexed by i_1, k_1, i_2, and k_2, e.g., $\mathbf{R}_{ff}(i_1, k_1, i_2, k_2)$; the function corresponding to an $N \times N$ image would have N^4 entries, thus being rather clumsy and difficult to visualize. A special case, corresponding to a certain class of stochastic images, is the *separable autocorrelation function* with the elements

$$R_{ff}(i_1, k_1, i_2, k_2) = {}_C R_{ff}(i_1, i_2)\ {}_R R_{ff}(k_1, k_2), \tag{2.109}$$

where only two two-dimensional correlation functions sized $N \times N$, characterizing the relations along a column and along a row, are to be identified and kept in memory. Obviously, this does not allow for any variance of the row (or column) correlation function with the row (column) index. Similarly, the covariance function may also be separable.

It should be realized that when an image is represented by an $N^2 \times 1$ vector, the same correlation and covariance functions would be expressed as matrices sized $N^2 \times N^2$, which are called *autocorrelation matrix*, *autocovariance matrix*, *cross-correlation matrix*, etc.

2.4.3 Discrete Homogeneous and Ergodic Fields

The *strict-sense (or wide-sense) homogeneous discrete stochastic fields* are defined by spatial invariance of all (or some selected) probabilistic characteristics, like in the continuous-space case. Naturally, the spatial shifts are limited to those allowed by the spatial sampling grid, and represented by index differences. Thus, for *wide-sense homogeneity*, the mean and variance are to be spatially constant while the discrete auto- and cross-correlation functions are represented by two-dimensional correlation matrices indexed with the elements of $\Delta\mathbf{r} = (\Delta i, \Delta k)$:

$$\mu_f(\mathbf{r}) = \mu_f, \quad \sigma_f^2(\mathbf{r}) = \sigma_f^2, \quad R_{ff}(\mathbf{r}_1, \mathbf{r}_2) = R_{ff}(\Delta\mathbf{r}), \quad R_{fg}(\mathbf{r}_1, \mathbf{r}_2) = R_{fg}(\Delta\mathbf{r}). \tag{2.110}$$

The *(wide-sense) ergodic fields* enable estimation of the mean, variance, and auto- and cross-correlation functions as spatial averages on any image (any concurrent pair of images) of the ensemble,

$$\mu_f \approx \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1} f_{w_l}(i,k), \quad \sigma_f^2 \approx \frac{1}{N^2}\sum_{i=0}^{N-1}\sum_{k=0}^{N-1}(f_{w_l}(i,k)-\mu_f)^2,$$

$$R_{ff}(\Delta i,\Delta k) \approx \frac{1}{(N-\Delta i)(N-\Delta k)}\sum_{i=0}^{N-\Delta i-1}\sum_{k=0}^{N-\Delta k-1} f_{w_l}(i,k) f_{w_l}(i+\Delta i,k+\Delta k),$$

$$R_{fg}(\Delta i,\Delta k) \approx \frac{1}{(N-\Delta i)(N-\Delta k)}\sum_{i=0}^{N-\Delta i-1}\sum_{k=0}^{N-\Delta k-1} f_{w_l}(i,k) g_{w_l}(i+\Delta i,k+\Delta k). \tag{2.111}$$

The upper limits of the correlation sums as well as the denominator of the fraction correspond to the decreasing number of available products in the sums with increasing Δi, Δk. It is obvious that the correlation functions cannot be estimated beyond Δi, $\Delta k = \pm(N-1)$; this determines the maximal extent of the estimate. It should also be noted that the variance of the individual elements of the estimate increases with Δi, Δk as the number of averaged terms decreases.

However, another definition is sometimes used, with the denominator N^2. This would obviously make the estimate biased, albeit the variance of the marginal elements would be improved. However, the resulting matrices may also be interpreted as the unbiased estimates of the *weighted correlation functions*, e.g.,

$$w(\Delta i,\Delta k)R_{ff}(\Delta i,\Delta k) \approx \frac{1}{N^2}\sum_{i=0}^{N-\Delta i-1}\sum_{k=0}^{N-\Delta k-1} f_{w_l}(i,k) f_{w_l}(i+\Delta i,k+\Delta k), \quad \text{where}$$

$$w(\Delta i,\Delta k) = ((N-\Delta i)(N-\Delta k))/N^2; \tag{2.112}$$

the pyramidal weighting function (window) thus just covers the available extent of the estimate. It will be shown in the next section that such a weighted correlation function is needed when estimating the power spectrum based on the autocorrelation function. The same window should also be applied to the cross-correlation function when estimating the cross-spectrum.

2.4.4 Two-Dimensional Spectra of Stochastic Images

2.4.4.1 Power Spectra

The difference in spectral analysis of discrete stochastic images compared to the continuous-space case is in using the two-dimensional DFT instead of the integral transform, with the properties described in Section 2.3.2. Thus, the (complex) discrete spectrum of an individual realization of a stochastic field is the matrix

$$\mathbf{F}_{w_l} = [{}_{w_l}F_{m,n}] = \mathrm{DFT}_{2\mathrm{D}}\{[f_{w_l}(i,k)]\}. \tag{2.113}$$

The real-valued *individual discrete power spectrum* is then defined as

$${}_{ff}\mathbf{S}_{w_l} = [{}_{ff}S_{w_l}(m,n)] = \left[\frac{1}{N^2}\,{}_{w_l}F_{m,n}\,{}_{w_l}F^{*}_{m,n}\right] = \left[\frac{1}{N^2}\,|{}_{w_l}F_{m,n}|^2\right]. \tag{2.114}$$

These individual discrete power spectra form a family of matrices, the mean value of which is the power spectrum of the stochastic field **f**,

$$\mathbf{S}_{ff} = [S_{ff}(m,n)] = \mathcal{E}_w\{{}_{ff}\mathbf{S}_w\} \approx \frac{1}{M}\sum_{w_l=w_1}^{w_M} {}_{ff}\mathbf{S}_{w_l}, \tag{2.115}$$

where the last formula expresses the estimate from M realizations.

The *discrete* version of the *Wiener–Khintchin theorem*, relating the discrete two-dimensional power spectrum to the *weighted* two-dimensional autocorrelation function, is

$$\mathbf{S}_{ff} = \mathrm{DFT}_{2\mathrm{D}}\{[w(\Delta i,\Delta k)R_{ff}(\Delta i,\Delta k)]\}. \tag{2.116}$$

That is, the power spectrum of a discrete stochastic field and its weighted autocorrelation function form a transform pair in two-dimensional DFT. It opens an alternative way to estimating the discrete power spectrum via estimation of the discrete autocorrelation function, besides the estimation via the above average of individual power spectra. The proof of the theorem is straightforward: expressing elements of $\mathbf{S}_{ff}$ via Equations 2.113 to 2.115 and the definition formula of two-dimensional DFT (Equation 2.52) leads to the ensemble mean of a multiple sum, the reshuffling of which yields the right-hand side of Equation 2.116. The role of weighting the correlation function should be observed; the discrete Wiener– Khintchin theorem is often cited incorrectly without the explicit weighting function, which may lead to errors and misinterpretation of results.

2.4.4.2 Discrete Cross-Spectra

Similarly as in the continuous-space case, the above concept can be generalized to analysis of mutual relations of two concurrent (parallel) stochastic fields. The *discrete cross-spectrum* may thus be defined as

$$\mathbf{S}_{fg} = [{}_{fg}S_{m,n}] = \mathrm{DFT}_{2\mathrm{D}}\{[w(\Delta i, \Delta k)R_{fg}(\Delta i, \Delta k)]\}, \tag{2.117}$$

the spectral counterpart of the weighted discrete cross-correlation function. In analogy with Equation 2.115, it can also be expressed as the mean of individual cross-spectra,

$$\mathbf{S}_{fg} = [S_{fg}(m,n)] = \mathcal{E}_w\{{}_{fg}\mathbf{S}_{w_l}\} \approx \frac{1}{M}\sum_{w_l=w_1}^{w_M} {}_{fg}\mathbf{S}_{w_l}; \tag{2.118}$$

the last expression again meaning the estimate from M available realizations.

2.4.5 Transfer of Stochastic Images via Discrete Two-Dimensional Systems

The interpretation of the transfer of a discrete stochastic image by a two-dimensional discrete system realizing an operator P is quite parallel to that presented for the continuous-space case in Section 1.4.5. Thus, the family of equations, representing the relation between the input and output stochastic images, is in the discrete case

$$\mathbf{g}_w = [g_w(i,k)] = \mathrm{P}\{[f_w(i,k)]\} = \mathrm{P}\{\mathbf{f}_w\}, \tag{2.119}$$

where $\mathbf{f}_w$ represents all possible realizations of the input field, and $\mathbf{g}_w$, similarly, all corresponding realizations of the output field.

Processing of discrete random images by a discrete two-dimensional space-invariant linear system performs the convolution on individual realizations that may be summarized in the family of equations

$$\mathbf{g}_w = [g_w(i,k)] = [h * f_w|(i,k)] = \mathbf{h} * \mathbf{f}_w, \tag{2.120}$$

the last expression introducing the notation of convolution between matrices in the sense of Equation 2.32. The matrix $\mathbf{h}$ represents the deterministic two-dimensional PSF of the system in the usual way.

It can easily be shown that, in analogy with the continuous case, the mean value and the power spectrum of the output field

can be expressed in terms of the corresponding characteristics of the input field,

$$\begin{aligned} \mathcal{E}_w\{\mathbf{g}_w\} &= \mathbf{h} * \mathcal{E}_w\{\mathbf{f}_w\}, \\ \mathbf{S}_{gg} &= [S_{gg}(m,n)] = [\,|H(m,n)|^2\, S_{ff}(i,k)]. \end{aligned} \tag{2.121}$$

Similarly, we could derive the matrix containing the output local variance.

The cross-spectrum describing the frequency-domain relations between the discrete random input and output images is

$$\mathbf{S}_{fg} = [S_{fg}(m,n)] = [H(m,n)\, S_{ff}(m,n)]. \tag{2.122}$$

Transforming the last equation by the inverse two-dimensional DFT, we obtain its discrete original-domain counterpart,

$$\mathbf{R}'_{fg} = [wR_{fg}(i,k)] = [h * wR_{ff}\,|\,(i,k)] = \mathbf{h} * (\mathbf{R}'_{ff}), \tag{2.123}$$

where $\mathbf{R}'_{ff}$ is the weighted autocorrelation matrix of the input field, and $\mathbf{R}'_{fg}$ is the weighted cross-correlation matrix.

REFERENCES for Part I

[1] Dougherty, E.R. (Ed.). *Digital Image Processing Methods*. Marcel Dekker, New York, 1994.

[2] Gonzalez, R.C. and Woods, R.E. *Digital Image Processing*. Addison-Wesley, Reading, MA, 1992.

[3] Haykin, S. *Neural Networks*. Prentice Hall, Englewood Cliffs, NJ, 1994.

[4] Jahne, B. *Image Processing for Scientific Applications*. CRC Press, Boca Raton, FL, 1997.

[5] Jahne, B., Haussecker, H., and Geissler, P. (Eds.). *Handbook of Computer Vision and Applications* , Vol. 2. Academic Press, New York, 1999.

[6] Jain, A.K. *Fundamentals of Digital Image Processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[7] Jan, J. *Digital Signal Filtering, Analysis and Restoration*. IEE Press, London, 2000.

[8] Kak, A.C. and Slaney, M. *Principles of Computerized Tomographic Imaging*. SIAM Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[9] Kamen, E.W. *Introduction to Signals and Systems*, 2nd ed. Macmillan Publishing Company, New York, 1990.

[10] Kosko, B. (Ed.). *Neural Networks for Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[11] Kosko, B. *Neural Networks and Fuzzy Systems*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[12] Kreyszig, E. *Advanced Engineering Mathematics*, 4th ed. John Wiley & Sons, New York, 1979.

[13] Lau, C. (Ed.). *Neural Networks, Theoretical Foundations and Analysis*. IEEE Press, New York, 1992.

[14] Madisetti, V.K. and Williams, D.B. *The Digital Signal Processing Handbook*. CRC Press/IEEE Press, Boca Raton, FL, 1998.

[15] MATLAB Image Processing Toolbox, version 2. The Math-Works, Natick, MA, 1997.

[16] MATLAB, version 5.1. The Math-Works, Natick, MA, 1997.

[17] MATLAB Wavelet Toolbox. The Math-Works, Natick, MA, 1996.

[18] Pratt, W.K. *Digital Image Processing*, 3rd ed. John Wiley & Sons, New York, 2001.

[19] Proakis, J.G., Rader, C.M., Ling, F., and Nikias, C.L. *Advanced Digital Signal Processing*. Maxwell Macmillan International, 1992.

[20] Rabiner, L.R. and Gold, B. *Theory and Application of Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1975.

[21] Rektorys, K. *Applied Mathematics*, 6th ed. Prometheus, Prague, 1995.

[22] Rosenfeld, A. and Kak, A.C. *Digital Picture Processing*, 2nd ed. Academic Press, New York, 1982.

[23] Russ, J.C. *The Image Processing Handbook*, 4th ed. CRC Press, Boca Raton, FL, 2002.

[24] Skrzypek, J. and Karplus, W. (Eds.). *Neural Networks in Vision and Pattern Recognition*. World Scientific, Hackensack, NJ, 1992.

[25] Sonka, M. and Fitzpatrick, J.M. *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*. SPIE International Society for Optical Engineering, Bellingham, WA, 2000.

[26] Sonka, M., Hlavac, V., and Boyle, R.D. *Image Processing, Analysis and Machine Vision*, 2nd ed. PWS, Boston, 1998.

[27] Strang, G. and Nguyen, T. *Wavelets and Filter Banks*. Cambridge Press, Wellesley, MA, 1996.

[28] Vaidyanathan, P.P. *Multirate Systems and Filter Banks*. Prentice Hall PTR, Englewood Cliffs, NJ, 1993.

Part II

Imaging Systems as Data Sources

This part of the book provides the necessary link between the image data generation and the world of image reconstruction, processing, and analysis. In the following chapters, we shall comment on different medical imaging modalities, to the extent that is needed to understand the imaging properties of the individual modalities and their requirements with respect to processing of measured data. No attempt will be made to go deeper into the physical background of the individual modalities, nor shall we describe the technical construction of the respective imaging systems. The purpose is solely to explain those features of each modality that determine its imaging properties and limitations, and to comment on intrinsic signal and image data processing, as well as on typical parameters of the provided image data. This should lead the reader to an understanding of the reasons behind the application of this or another data processing approach in the frame of every modality. Nevertheless, because identical or similar data processing approaches are used for different modalities, the chapters concerning the imaging systems will mostly only refer to chapters of Part III when mentioning external image processing methods.

The medical imaging systems can be classified into the following main categories according to imaging medium or basic imaging principle:

- X-ray projection radiography, including digital mammography and digital subtractive angiography (DSA)
- X-ray computed tomography (CT)
- Magnetic resonance imaging (MRI) tomography
- Nuclear imaging (planar gamma-imaging, SPECT, PET)
- Ultrasonic imaging in two and three dimensions (ultrasonography (USG), three-dimensional ultrasound (3D US)), including color flow imaging (CFI)
- Infrared (IR) and optical imaging
- Electron microscopy (transmission electron microscopy (TEM) and scanning electron microscopy (SEM))
- Electrical impedance tomography (IT)

Other modalities appear occasionally, namely in research.

We shall deal with each of the groups individually in a separate chapter, where specific relevant references are cited. Other sources used but not cited are [5], [13], [18], [22], [28], [35], [40], [41], [45], [47], [48], [50]–[57], and [59]. It should be understood that the mentioned modalities are not used exclusively in medicine; many of them find applications in other areas too — in technology and industry (material engineering, micro- and nanotechnology, nondestructive testing), as well as in research and science, such as in biology, ecology, archaeology, etc.

3

Planar X-Ray Imaging

Planar x-ray imaging is considered the historically first medical imaging modality. Its principle, capable of providing directly the images without any reconstruction, is transparent; the digitization appeared in this field rather recently. Of numerous sources — also for more detailed study — let us mention [6], [7], [9], [10], [30], [36], [49].

3.1 X-RAY PROJECTION RADIOGRAPHY

3.1.1 Basic Imaging Geometry

The x-ray projection radiography is the historically oldest medical imaging modality — its history goes back to Roentgen's discovery of x-rays in the last decade of the 19th century. It is also the simplest modality as far as the imaging principle concerns, as can be seen in Figure 3.1. Ideally, the source of radiation is of negligible dimension and therefore can be considered a point radiator in the first approximation. The image is formed of intensity values of the x-rays modified by passing through the imaged object, e.g., a part of the patient's body.

Obviously, the resulting image contains information on the complete three-dimensional volume projected on the two-dimensional

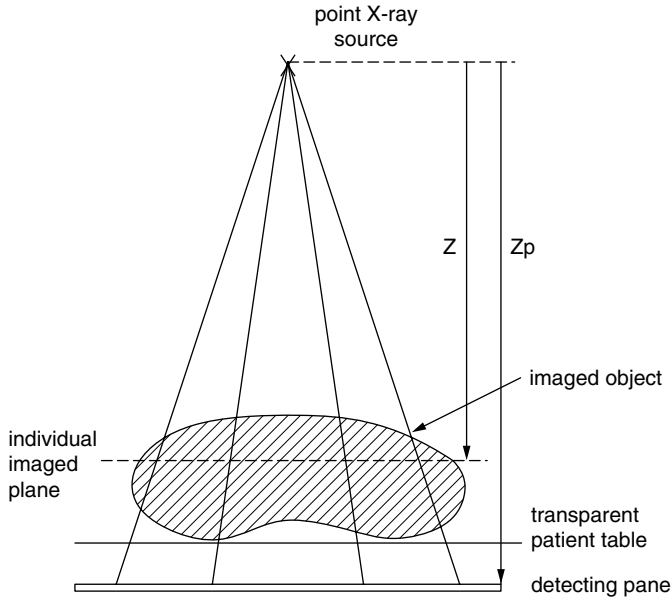


Figure 3.1 Principle of x-ray projection imaging.

plane of the detecting panel (originally a fluorescent screen, film, or image amplifier, currently also a type of digital flat-panel detector). Every pixel in the resulting image ideally represents the intensity of the incident x-ray that carries the information on the total attenuation along the respective ray. The resulting image can be interpreted as a mixture of images of planes parallel with the projection plane. As can be seen from the figure, the individual planes are imaged in different scales, the corresponding linear magnification being

$$m = \frac{z_p}{z}, \tag{3.1}$$

where z and z_p are the distances from the radiation source to a particular imaged plane and to the projection plane, respectively.

It is impossible to separate the image information on the individual planes algorithmically, as the information on the z-coordinate of individual contributions to ray attenuation is missing. Usually, this is a (principally difficult) task for the radiologist analyzing the image to use his imagination in combination with *a priori* anatomical knowledge, in order to evaluate and classify the image information properly,

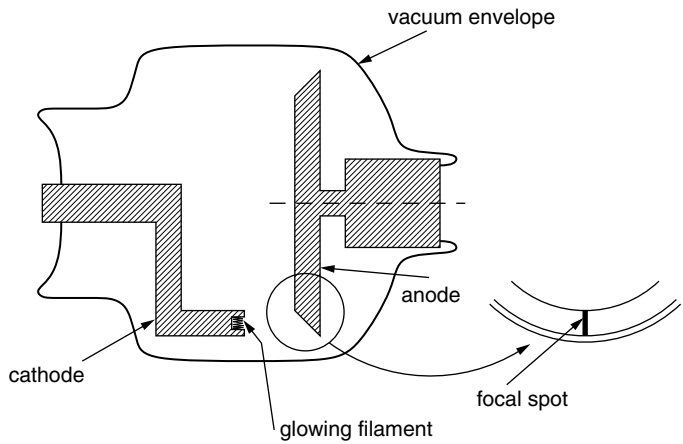


Figure 3.2 Schematic view of an x-ray tube.

taking into account the imaging properties, including the more subtle ones mentioned later.

3.1.2 Source of Radiation

In reality, x-ray imaging is more complicated. The x-rays are produced in x-ray tubes (Figure 3.2) as a result of partial conversion of the energy of highly accelerated electrons into x-rays. Without going into the physics of interaction of the fast electrons with the matter of the anode, let us only state that two components of the x-radiation result. The first component is wideband breaking radiation (*bremsstrahlung*), the intensity of which is linearly decreasing with frequency (i.e., also with photon energy), and the maximum frequency corresponds to the individual energy of incident electrons — it is thus given by the accelerating voltage applied to the x-ray tube (in the range of about 25 to 150 kV). On the lower side of x-ray frequencies, the theoretical linear increase is modified by filtering effects of the tube parts that the radiation must pass through. The other radiation component is so-called *characteristic x-radiation*, a result of the interaction of incident electrons with electron shells of the anode atoms. This radiation component has a line spectrum with frequencies characteristic for a given material of the anode. The total spectrum of the radiated rays may look like that in Figure 3.3.

The properties of the x-radiation used for imaging may be influenced by many factors, primarily the used accelerating voltage:

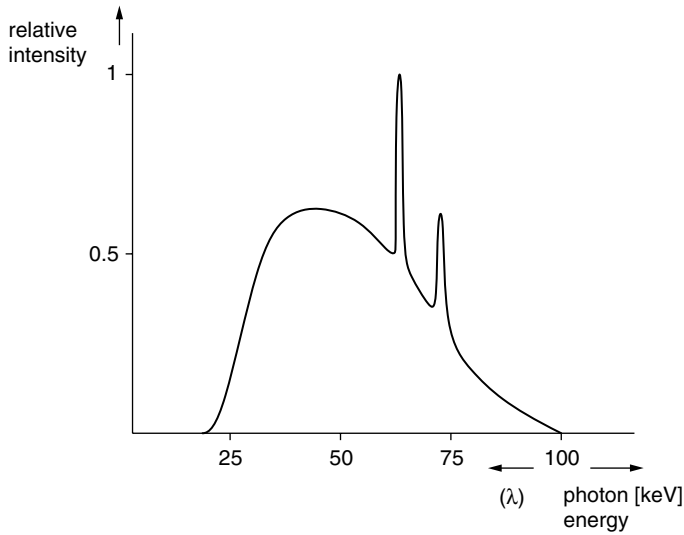


Figure 3.3 A typical x-ray spectrum.

the higher the voltage, the harder the radiation, i.e., with a higher share of short-wave-length high-energy photons that better penetrate materials, including the tissues. The radiation intensity may be influenced, without changing the shape of its spectrum, by adjusting the current of the tube, while the softer part of the rays — often undesirable — may be suppressed by filtering the rays with metallic sheets inserted on the way of radiation, before it reaches the patient's body. On the other hand, in some instances (e.g., in mammography), high-energy components must be removed — this can be done by K-edge filtering. The x-radiation parameters are generally adjusted according to a particular medical task by the radiologist, and therefore can be taken as given, from the image processing point of view.

As the wavelength of even the softest used x-rays is several orders smaller than the resolution of the images, the diffraction phenomena need not be taken into account. From our viewpoint, the quality of x-rays thus does not influence the basic properties of the imaging method. Of course, this does not mean that it has no influence on the quality of concrete images (including noise due to scattered rays), but this is a different aspect. Anyway, other issues should be taken into account when considering postprocessing of the x-ray images.

Primarily, the size of the radiation source is not negligible. As we can derive from Figure 3.2, the radiating spot (physical *focus*) has

basically the shape of a vertically prolonged rectangle. This is due to the necessity to distribute an enormous heat load over a greater area; if the physical focus were made too small, the anode material would melt. Although the lateral size of the focus (perpendicular to the figure) is kept as small as possible by focusing the electrons to a narrow stripe, its length is substantially greater to ease the heat dissipation. The dissipation is further enhanced by fast rotation of the anode, but as the anode is rotationally symmetric, the rotation does not influence the imaging properties and we need not consider it. The effective almost-point focus is provided by the sharp tilt of the anode surface relative to the image-forming rays: as we can see in Figure 3.4, the focus is visible from the image side as a small (roughly square) spot, because the visible length of the longer side d of the rectangle is reduced to

$$d' = d \sin \delta , \tag{3.2}$$

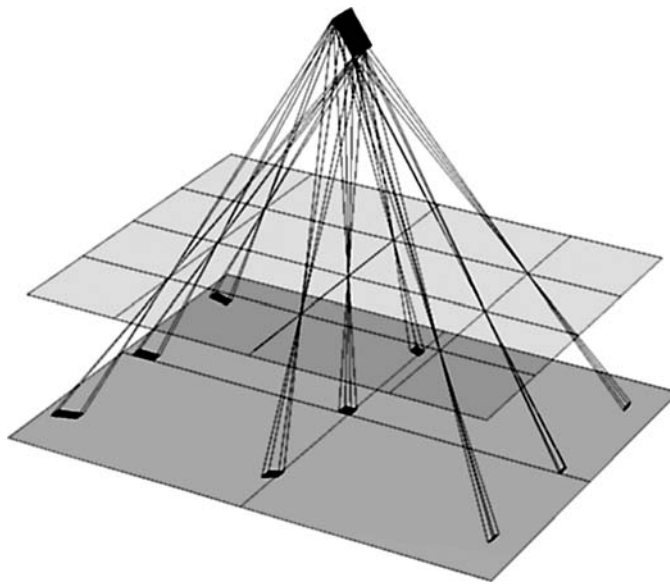


Figure 3.4 Influence of finite size of x-ray focus on the imaging properties — space-variant PSF of x-ray projection imaging. The upper plane is the subject of imaging; the lower plane represents the flat detector of projection (the size of the x-ray focus as well as the differences in the PSF width are intentionally exaggerated).

where δ is the angle between the ray and the anode surface. The nonpoint size of the focus complicates the imaging: a sharp edge is not imaged as such, but is blurred due to the *penumbra effect*, as seen from the figure. Naturally, similar blurring appears also in the perpendicular direction as even the best-focused electron beam always has a certain width. Some x-ray tubes allow the changing of the size of the focus—a smaller one for sharper imaging with lower power while allowing an increase of the power of radiation at the cost of enlarging the focus.

The reader may suggest that this phenomenon causes a nonpoint PSF of the imaging. Really, when we insert an opaque sheet with negligibly small pinhole openings in the position of the imaged plane (Figure 3.4), the response on the projection plane will be spots, corresponding in shape to the scaled and 180° inverted x-ray focus. The distribution of detected intensity inside the spot on the projection plane will correspond to the x-ray brightness distribution in the focal spot, as seen from the position on the image; this is generally not uniform due to geometrical and construction reasons.

Obviously, the shape of the projected focus depends on δ and the similar perpendicular angle β, which are both variable with the position on the image (projection) plane. The steeper the angle δ on the anode side (here, right side) of the image, the more pronounced the variability with δ is. This imposes limits to the choice of δ, depending on the size of the resulting image. Besides that, the brightness of the focus also depends on the angle of view, mainly due to the heel effect; the focus thus behaves as a nonlambertian radiator. Unfortunately, as both the projected size of the focal spot and the visible brightness distribution of the x-ray focus depend on the position in the image, the PSF is space variant. The size of the PSF also obviously depends on the distance of the imaged plane from the x-ray source: though all the PSFs corresponding to impulse objects (pinholes) on a particular ray in different imaged planes are identical in shape and intensity distribution, they are naturally scaled according to Equation 3.1. This means that every imaged plane is not only scaled, but also blurred differently in the projection image—the further the imaged plane is from the projection (detector) plane, the greater is the PSF size and, consequently, the blur.

Due to difficulties with implementation of impulse-type-object (pinhole) imaging, practical evaluations of x-ray imaging properties are based on analysis of step responses—intensity profiles in the image that are perpendicular to differently oriented edges or stripe structures.

The most prominent consequence of the space-variant PSF is a nonuniform detected intensity in the projection plane, even if the imaged object provides uniform attenuation — the anode side of the negative image is generally lighter due to lower x-ray intensity.

3.1.3 Interaction of X-Rays with Imaged Objects

The incident x-ray beam interacts with the material of imaged objects (tissues) in basically three ways: photoelectric effect, coherent Rayleigh scattering, and incoherent Compton scattering. (The fourth way of pair and triplet production appears only with energies higher than those used in medical imaging.) Though physically different, all these stochastic mechanisms are characterized by two phenomena important from the imaging point of view: attenuation of the beam during its passage through the body and scattering of the radiation. Scattering means a generation of x-ray photons — either of the same energy and frequency (Rayleigh) or of lower energy, i.e., softer radiation. The direction of scattered photons is generally different from that of the original x-ray; while Rayleigh scattered photons are preferably oriented more closely to the original direction, namely, at high energies, Compton scattering originates photons of almost all directions, with a high probability of even backscattering at low energies.

Every accomplished interaction means a loss of the interacting photon from the original beam that is therefore attenuated. Let the original monochromatic beam be formed of N photons per second; of it, the number dN of interacting photons when passing a thin layer of a material is proportional to N, to a probability of interaction μ, and to the thickness of the layer dx, thus $dN = -\mu N\, dx$. Integrating this differential equation provides

$$N = N_0 \exp(-\mu x). \tag{3.3}$$

The intensity of the single-frequency beam decreases — in a homogeneous material — exponentially; the material is characterized by μ, which thus can be interpreted as the *linear attenuation coefficient* (m^{-1}), the relative probability of interaction per length unit (thanks to the exclusivity of the mentioned interaction modes, this is the sum of probabilities belonging to individual modes). The attenuation strongly depends on the energy of photons — general tendency: the higher the energy, the lower is μ. This explains the

phenomenon of *hardening* of x-rays by passing a material: softer components are more attenuated so that the central frequency of the spectrum is shifted toward higher values.

The linear attenuation coefficient also depends linearly on the density of the material. This enables us to define the *mass attenuation coefficient* (m^2kg^{-1}) as

$$\kappa = \frac{\mu}{\rho}, \tag{3.4}$$

where ρ is the density of the material ($kg \cdot m^{-3}$). It can be shown that κ increases generally with the atomic number of the material. The mass attenuation coefficient can be computed for a compound material consisting of M chemical elements as

$$\kappa = \sum_{i=1}^{M} w_i \kappa_i, \tag{3.5}$$

where w_i are weight fractions of the elements in the compound.

The quantity measured in projection radiography is the position-dependent intensity $I_p(x, y)$ of a ray passing through a particular point (x, y) on the projection plane. In consequence of Equation 3.3, for monochromatic radiation, it can obviously be expressed in terms of the initial intensity I_0 and the locally dependent attenuation coefficient $\mu(r)$ as

$$I_p(x, y) = I_0 \exp\left(\int_{r=0}^{r_p} \mu(r) \mathrm{d}r \right), \tag{3.6}$$

where r is the radial distance along the ray from the source at $r = 0$, limited by the radial distance r_p of the projection plane. This intensity is determined by the *total attenuation* $A(x, y)$ between the radiation source and the projection plane, which is the argument of the exponential function,

$$A(x, y) = \int_{r=0}^{r_p} \mu(r) \mathrm{d}r. \tag{3.7}$$

When taking into account the usual wideband character of the radiation and the frequency-dependent attenuation, i.e., the hardening of the beam, further integration along frequency would

result, making the expression less comprehensible. This is why Equation 3.6 is often taken as a reasonable approximation (at least qualitative) even for nonmonochromatic radiation cases. Accepting Equation 3.6 as a valid approximation leads to the wide use of logarithmic transform of the measured values, this way converting the intensity data into the *attenuation data* that may be considered a result of (approximately) *linear* imaging. Such data are needed for all postprocessing methods, which rely on linearity, primarily (digital) *subtraction angiography* (DSA; Section 3.2) and *computed tomography* (CT; Chapter 4).

While attenuation is thus the measured parameter, scattering is an undesirable but unavoidable phenomenon. The scattered radiation may reach (in some cases via several interactions) the projection plane and be measured by the detection device as a part of the image data, which increases the background noise of the image (Figure 3.5). The scattered radiation should be prevented from approaching the detector array (or screen); this can be partly influenced via decreasing the generated amount of scatter by imaging as thin an object as possible (aided perhaps by some compression of tissues, e.g., in mammography). The secondary diffused radiation

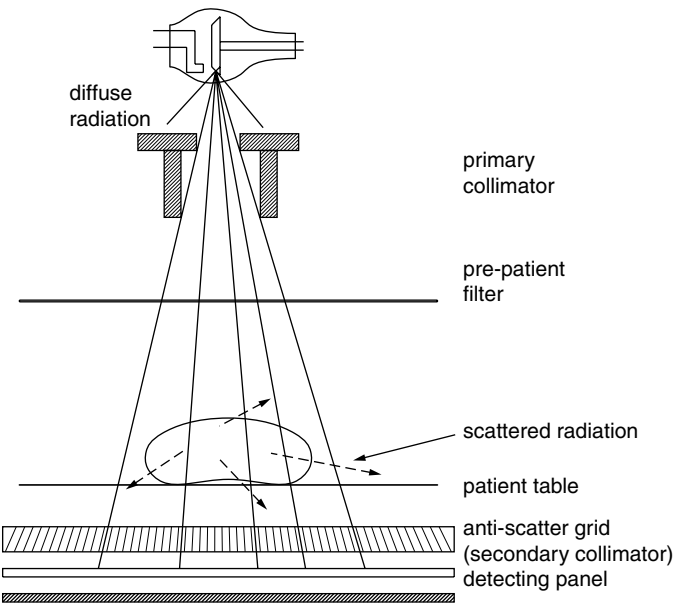


Figure 3.5 Complete x-ray projection arrangement schematically.

is mostly softer than that used for imaging, so that a thin metallic sheet situated between the object and the detector plane would filter it out partially. An effective means for diffuse radiation suppression is the *secondary collimator* formed of narrow stripes of lead arranged to a grid and oriented so that the useful radiation would pass while the differently oriented scattered radiation is mostly absorbed as seen in the lower part of Figure 3.5. To remove the grid structure from the image, the collimator may be slightly moved horizontally during exposition.

The diffuse radiation originating from out-of-focus parts of the x-ray tube that would have a similar noise effect is to be stopped by the *primary collimator* placed near the tube, made of x-ray-opaque material with free opening only for focus-generated rays. The following filter removes the low-energy part of x-rays that is useless from the imaging point of view and would only contribute to patient dose.

3.1.4 Image Detection

Historically, the x-radiation passing the object was visualized by a luminescent screen converting the x-rays into visible light. The same principle is still used to increase efficiency of film recording by means of intensifying foils, or in the first layer of image intensifiers where the emitted light generates free electrons in the following photoemitter layer. Another, more modern possibility is to use a screen covered by storage phosphors that, after being exposed to the x-ray image, carry a latent image, which can be read consequently in a separate reader using laser scanning. All these approaches enable, in principle, digitization of the generated gray-scale image — e.g., by scanning the film or using a TV camera for converting the output of the image amplifier into video signal, to be consequently submitted to analogue-to-digital (A/D) conversion. Nevertheless, although widely used, such complicated arrangements may suffer from many drawbacks, namely, low dynamic range, nonlinear response, possible long time delay or low spatial resolution, often high noise level, and, in the case of using the image amplifier, also geometrical distortion. Though many of these imperfections can be at least partly corrected subsequently by digital processing, better detection methods, providing less impaired image data, have been searched for.

Many contemporary digital x-ray systems use *flat-panel detector* fields (commercially available since the late 1990s) that cover the

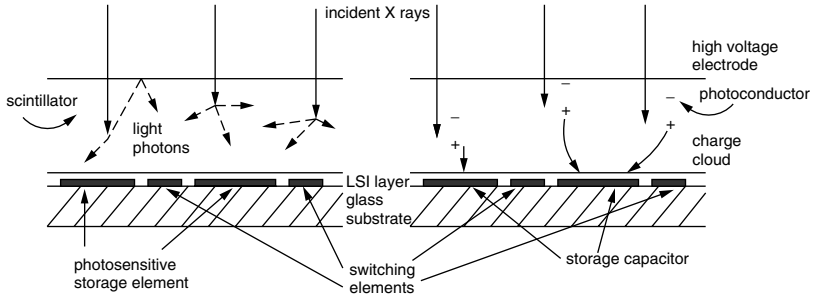


Figure 3.6 Cross-section details of two main types of flat-panel detectors using indirect conversion (left) and direct conversion (right).

whole image area up to or slightly over 40 × 40 cm, and are placed similarly like film in the imaging arrangement. They provide in-place conversion of x-rays into an electric signal, and consequently into a digital sequence of image data. They use either of two physical approaches (Figure 3.6). Two-phase detection consists of *indirect conversion* of the x-radiation, first into light in a luminescent phosphor layer, followed by conversion of the light into local electric charges in the immediately attached grid of storage photodetectors. Alternatively, *direct conversion* of radiation into local electric charges may be used, utilizing the photoelectric phenomenon in a semiconductor layer and directing the charge by an auxiliary electric field to the storage capacitances. In both cases, the locally generated charges are accumulated in the capacitances, formed in the semiconductor substrate of the panel (usually deposited on glass), and arranged in a two-dimensional grid structure in which every capacitor corresponds to a pixel. The local charges are transported sequentially out of the field by means of a switching structure (usually thin-film transistor or diode based), and finally A/D converted into a sequence of numbers forming the image data (Figure 3.7). Obviously, the construction of such devices relies on, besides deep understanding of the physical conversion mechanisms, sophisticated large-scale integration (LSI) technology, which became feasible only recently, thanks to the widely used active-matrix liquid-crystal display (LCD) technology.

The thin flat panels have some obvious advantages over the more conventional detection methods, as they directly replace the film technology and are handier than other arrangements, e.g., image-amplifier-based systems. Sometimes an overlooked virtue of the panels is their insensitivity to parasitic fields (electrical and

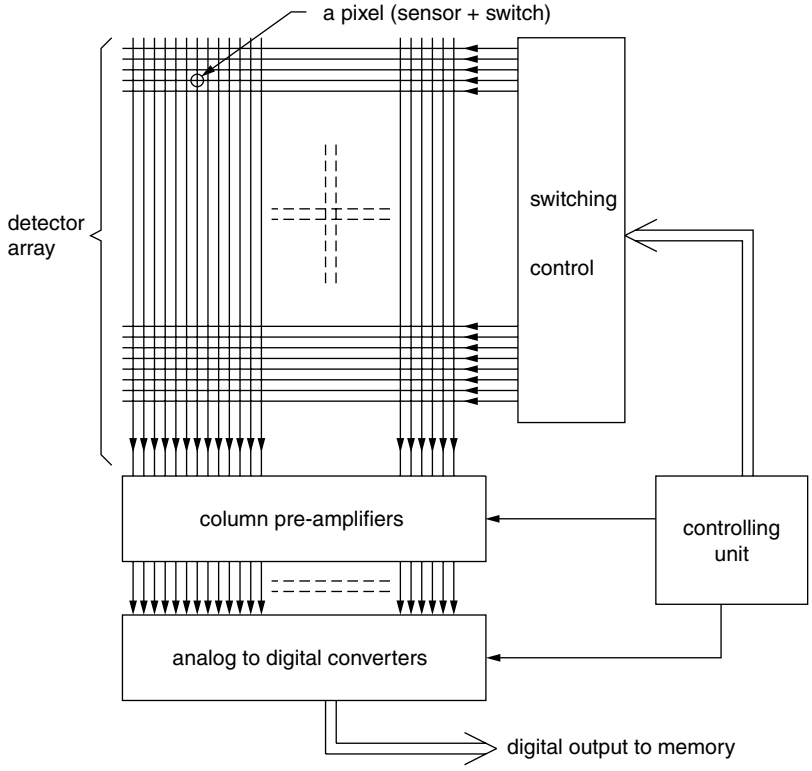


Figure 3.7 Schematic sketch of the flat-panel detector switching and signal processing arrangement.

magnetic fields, light), so that they are well compatible with other neighboring imaging systems, like magnetic resonance imaging (MRI) or CT. Nevertheless, their main advantages lay in excellent imaging properties: primarily, they produce zero geometrical distortion, as every pixel position is physically fixed on the panel and can individually be selected. The theoretical resolution, given by the pixel size (about 50 to 150 μm), already reaches the needed values of about 3 to 10 line pairs per mm (the highest value needed in mammography) and can be well compared with the film. Nevertheless, the real resolution approaches the theoretical limit only in direct-conversion panels, where there exists almost no scattering of charges moving in the layer of the converting semiconductor (mostly

selenium, about 0.5 mm thick), thanks to the directional effect of the imposed electrical field. In indirect-conversion systems, the resolution may be considerably affected by the light scatter in the relatively thick first converting layer (mostly CsI phosphor), though this is to an appreciable extent prevented by a special crystalline structure of the phosphor, imitating vertically oriented parallel optical fibers. Thus, the indirect-conversion systems reach only about half (or less) of the theoretical resolution. The size of the image matrix formed by a flat panel corresponds to its physical size and the pixel dimension; today's common sizes are from about 1500×1500 to 4000×4000.

The high resolution, namely of the direct-conversion type of panels, need not be only an advantage. If the image projected to the panel contains components with the frequency above the Nyquist limit (e.g., noise), aliasing appears that may deteriorate the final image. For this reason, a presampling blur is sometimes artificially introduced in the panels, by means of layers below the converter layer, enabling a lateral diffusion of the generated light or charge to neighboring pixels. It effectively acts as an analogue two-dimensional antialiasing filter that suppresses the components above the Nyquist limit. Unfortunately — due to the imperfection of such filtering — it also deteriorates the useful resolution at the same time; finding a reasonable compromise is not an elementary task.

The dynamic range of both types of panels is large (over 60 dB), with a very good linearity of conversion in the whole range, which is important for consequential processing. Also, the detection quantum efficiency (DQE) of the panels, quantifying the probability of detecting an x-ray photon, is very good (in the range of about 0.6 for indirect-conversion systems or 0.3 for direct-conversion systems at low energy x-ray to under 0.2 for high energies) — generally better than for film detection. This, together with a high geometrical fill factor (approaching 1 for direct- and 0.7 for indirect-conversion panels) and low noise generation in the contemporary LSI circuits, also determines a high signal-to-noise ratio, enabling an efficient use of 12- to 14-bit digitization. The achieved speed of reading not only allows static image detection (radiography), but also can be used in fluoroscopy in dynamic real-time imaging (7 to 60 frames per second). In the dynamic regime, as well as in static radiology with a great extent of sequentially measured radiation, it is important that the previously measured image (or the previous frame in a dynamic sequence) has no influence on the next measurement; it should be totally deleted before acquisition of the next frame.

The storage elements must therefore be completely cleared by respective reading pulses. If there is an undesirable inertia caused either by a lag in photodiodes or by delayed luminiscency in the converting layer, it manifests itself as *carryover* or *ghosting*. It was observed only in indirect-conversion panels, and in such cases, it should be corrected by suitable postprocessing.

3.1.5 Postmeasurement Data Processing in Projection Radiography

The plain projection radiology itself provides, in any of its forms, direct images that do not necessarily need any reconstruction. Typical examples of planar x-ray images are shown in Figure 3.8. Nevertheless, much can be done with the raw images in order to improve the diagnostic readability, i.e., to improve the image quality or to enhance the diagnostically important features. The nonideal PSF (Figure 3.4) is usually only taken into account with respect to uneven illumination of the field (on field equalization, see Section 11.1). The penumbra effects due to finite PSF size are usually neglected as small enough. The systems using image amplifiers and TV cameras always suffer by a certain degree of geometrical distortion, which can be well corrected by restitution (Section 12.2). The x-ray images are formed by relatively low photon flows governed by the Poisson law, so that a certain (sometimes high) level of quantum noise is present, accompanied by further noise components due to thermal effects, variability in circuitry properties, etc. Noise-suppressing procedures may then be needed — either local smoothing or artificial temporal averaging of more images in a sequence (Section 11.3). This is the opposite procedure compared with ghost suppression, which consists of subtracting weighted older versions of the image in a sequence.

When using the flat-panel detectors, a digital postprocessing is always necessary to compensate for the deficiencies of a particular panel — uneven sensitivity and contrast of individual pixels (Section 11.1), dead pixels or lines (nonlinear interpolation or filtering; Section 10.2), etc. When imaging and evaluating structures, the small details of which are decisive (mammography, bone structures), a certain degree of high-pass filtering or local sharpening procedures, perhaps also locally adaptive ones (Sections 11.2 and 12.3), may be used.

As with all modalities, the digital form of the image offers an enormous range of possibilities not only to process the image, but also to analyze it (semi)automatically, as needed for particular

Figure 3.8 Typical x-ray planar projection images. (Courtesy of the Faculty Hospital of St. Anne Brno, Clinic of Radiology, Assoc. Prof. P. Krupa, M.D., Ph.D.)

medical purposes. The derived parameters, including parametric images (that need not closely resemble the original image; see Section 13.1.1) or edge representations (Section 13.1.2), may then serve as a base for computer-aided diagnosis.

3.2 SUBTRACTIVE ANGIOGRAPHY

Subtractive angiography—often called *digital subtractive angiography* (DSA) — is in fact not an individual imaging modality, but rather a specialized type of projection radiography, which uses a preimaging modification of the imaged object by adding a contrast agent to some of its structures and consequently comparing the artificially contrasted images with the noncontrasted ones. Though it it used primarily for visualizing vessel structures (thence its name) and, based on sequences of images, also for evaluation of blood flow and tissue perfusion, its principle is quite general and can be used for other purposes as well. It is a conceptually simple approach, enhancing visualization of structures normally hardly visible, which can be filled with an x-ray-attenuating contrast agent. Basically two images are provided—one before the application of the contrast agent (preinjection image, often called the *mask*), and the other when the agent is properly distributed in the structure to be visualized (e.g., in vascular structures). The mask is then subtracted from the contrast image; this way, the structures that are common to both images (fixed anatomy) are suppressed, leaving only the enhanced image of the contrast-filled structures.

This process can be approximately described in a more quantitative way as follows. The measured intensity in the precontrast image is, according to Equation 3.6,

$$I_p(x,y) = I_0 \exp\left(\int_{r=r_1}^{r_2} \mu(r)\,\mathrm{d}r \right), \tag{3.8}$$

where $r_2 - r_1$ is the total thickness of the imaged tissue on the investigated x-ray. Then, when filling a vessel of a thickness D with the contrast agent of the attenuation coefficient μ_c, the intensity will be obviously given by the product

$$I_{pc}(x,y) = I_0 \exp\left(\int_{r=r_1}^{r_A} \mu(r)\,\mathrm{d}r \right) \exp\left(\int_{r=r_A}^{r_A+D} \mu_c(r)\,\mathrm{d}r \right) \exp\left(\int_{r=r_A+D}^{r_2} \mu(r)\,\mathrm{d}r \right), \tag{3.9}$$

where r_A is the coordinate of the ray entrance to the vessel. When simply subtracting $I_{pc} - I_p$, we would obviously get a complex expression influenced by the attenuation along the whole ray. On the other hand, when applying a logarithmic transform to both previous equations before subtracting, we obtain

$$\log I_{pc} - \log I_p = \int_{r=r_1}^{r_A} \mu(r)\mathrm{d}r + \int_{r=r_A}^{r_A+D} \mu_c(r)\mathrm{d}r + \int_{r=r_A+D}^{r_2} \mu(r)\,\mathrm{d}r - \int_{r=r_1}^{r_2} \mu(r)\,\mathrm{d}r$$

$$= \int_{r=r_A}^{r_A+D} (\mu_c(r) - \mu(r))\,\mathrm{d}r, \tag{3.10}$$

which is exactly what is needed — the net information on the contrast agent distribution (usually, $\mu \ll \mu_c$). Naturally, the precision of this analysis is limited by neglecting the influence of attenuation dependence on the x-ray frequency (photon energy); in other words, the analysis is precise only for monochromatic x-rays. Nevertheless, it can reasonably be used in practice, as the nonlinearity caused by the wideband character of radiation, related also to the hardening effect on the radiation, does not reach a level that would disqualify the linearity hypothesis.

Though the principle is simple, the complete procedure of DSA is more complicated. It consists of providing the images (the precontrast mask and one or more contrast images), followed by the conversion of the obtained intensity images into attenuation images by the point-wise logarithmic contrast transform. The next, very important step is the image registration, which must ensure that the corresponding structures are precisely located at the same positions in both images to be subtracted. If this is not fulfilled perfectly, perhaps due to movement of the patient or his organs during the imaging, the subtraction would enhance not only the contrast-filled structures, but also (and perhaps predominantly) just the differences caused by the imperfect match. This may lead to very disturbing plasticity of the difference images, if not to improper conclusions as to blood flow concerns. The misregistration must therefore be prevented by special presubtraction registration procedures that themselves represent a complex computation and still a challenging problem (Section 10.3). Due to the complicated character of the local shifts, often a rigid registration (consisting of mere shift and rotation) is insufficient, and more efficient but more complicated flexible

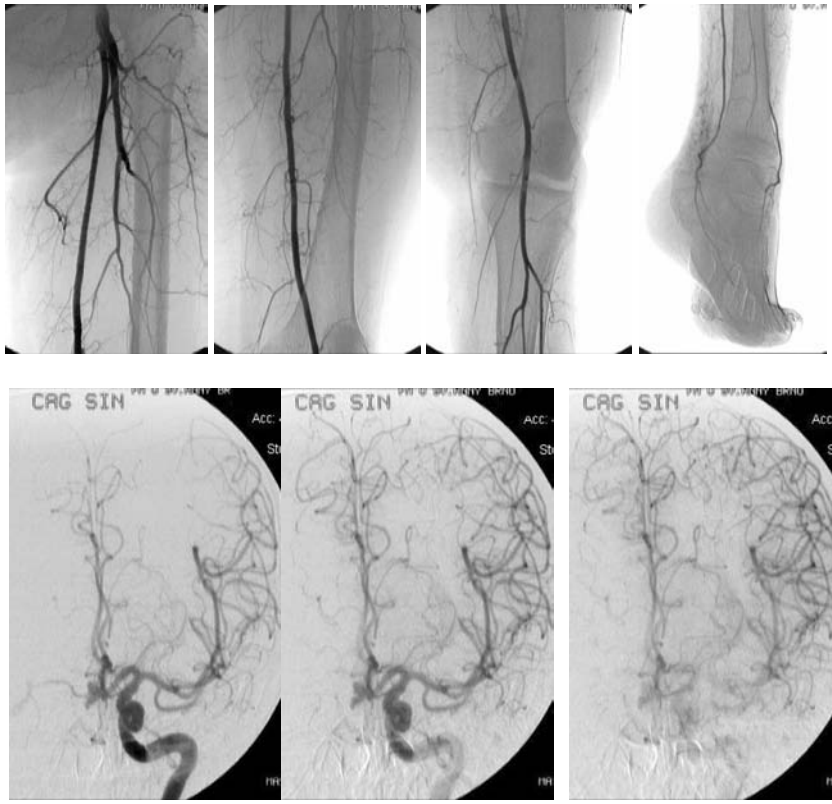


Figure 3.9 Examples of DSA images. Above: Vascularization of an extremity. Below: Image sequence showing gradual filling of the vascular structure by the contrast agent. (Courtesy of the Faculty Hospital of St. Anne Brno, Clinic of Radiology, Assoc. Prof. P. Krupa, M.D., Ph.D.).

registration approaches must be applied. The perfectly matched images are then subtracted, and the contrast scale of the difference image should usually be normalized to the full display extent, as the differences have usually relatively low dynamics (Section 11.1). Typical DSA images are presented in Figure 3.9.

Of course, the same mask can be subtracted sequentially from a series of consecutive contrast images describing, e.g., progressive filling of a vessel structure. In this case, the registration step usually has to be repeated, as the movements in individual images of an image sequence usually differ mutually.

4

X-Ray Computed Tomography

X-ray computed tomography (CT) became the historically first tomographic modality entirely based on digital reconstruction of images. It has qualitatively changed the field of medical imaging. Contemporarily, it is probably the most common computerized tomographic modality, with excellent spatial resolution, fast image acquisition enabling even real-time imaging, and rather generic application. For a deeper study see, besides other rich literature, [4], [6], [7], [9], [10], [21], [26], [27], [30], the main sources used in this chapter.

4.1 IMAGING PRINCIPLE AND GEOMETRY

4.1.1 Principle of a Slice Projection Measurement

Classical x-ray radiography provides attenuation images that are combinations of projections of all layers forming the thickness of the object (see Section 3.1.1). There is no way to separate the information from different layers, except by subjective evaluation using *a priori* anatomical knowledge. It is possible to enhance the image of an individual layer by blurring images of other layers on special x-ray equipment, where the radiation source and detection plane

are moved simultaneously with respect to the patient during exposition in a proper way (classical x-ray tomography). Nevertheless, besides other drawbacks, this does not enable the determination of the spatial attenuation distribution, which would allow the reconstructing of images for any desired plane exactly. *Computed tomography* (CT) solves the problem of determining this inner distribution by measuring individual object-influenced intensity values for many differently oriented rays, and subsequently reconstructing the data on internal attenuation distribution computationally.

The basic principle of CT measurement can be seen in Figure 4.1. The x-ray source, together with primary collimators, provides a fine beam of radiation (ideally an infinitesimally narrow ray) that passes

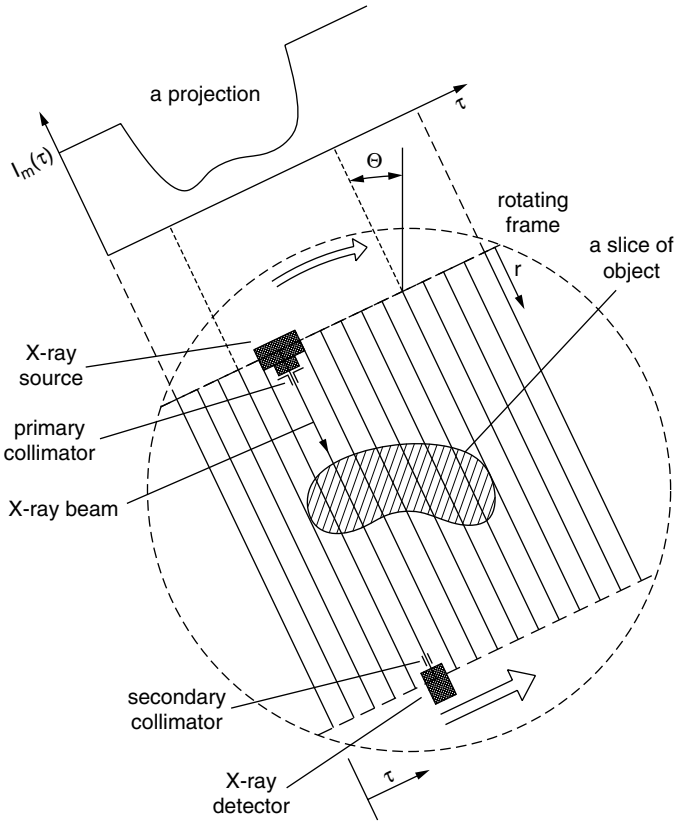


Figure 4.1 Principle of measurement of projections — basic rectangular arrangement.

the object, the intensity of the beam is then measured by a detector. Even with a good primary collimation, the beam is always slightly divergent, thus increasing the diameter of the measured volume of the object when approaching the detector side. This is compensated for by the secondary (detector) collimator, which excludes all the rays outside of the desired beam, in this way also suppressing the radiation scattered along the way through the object. The effective diameter of the resulting beam may be in the range of about 0.5 to 10 mm. This linear arrangement (source–detector) can be moved in the imaged plane with respect to the object perpendicularly to the beam so that the intensity is measured on different parallel rays (for different values of τ). The integral attenuation for each ray position τ is given, corresponding to Equation 3.7, as

$$A(\tau) = \int_{r=0}^{r_p} \mu(\tau, r)\mathrm{d}r = \log \frac{I_m(\tau)}{I_0} \tag{4.1}$$

if the intensity $I_m(\tau)$ measured by the detector is dependent on the initial ray intensity I_0 according to the exponential character of Equation 3.6. The above integral of the linear attenuation coefficient is called a *ray integral* (or line integral). Theoretically, we can obtain a value of the ray integral for a continuous range of τ, so that Equation 4.1 then represents a one-dimensional function of τ, called a *projection*, as depicted in the upper part of Figure 4.1.

The whole measuring arrangement, including the frame enabling the mentioned linear movement, can be rotated as seen in the figure. This way, we may obtain a projection for any angle θ of the measurement coordinates (τ, r) with respect to the object coordinates (x, y). In theory, it is again possible to obtain the projections for a continuum of θ, so that Equation 4.1 can be rewritten as

$$A(\tau, \theta) = \int_{r=0}^{r_p} \mu_\theta(\tau, r)\mathrm{d}r = \int_{x_{\min}}^{x_{\max}} \int_{y_{\min}}^{y_{\max}} \mu(x, y)\, \delta(x\cos\theta + y\sin\theta - \tau)\mathrm{d}x\,\mathrm{d}y, \tag{4.2}$$

where the δ-function selects the ray point set; the limits of x and y are given by the object size. This two-dimensional function is thus an integral transform of the original two-dimensional function (i.e., the attenuation distribution). Equation 4.2 is called *Radon transform*; in the continuous-space formulation, obviously, the task of reconstruction of the original image $\mu(x, y)$ from its projection

representation $A(\tau, \theta)$ is the problem of finding the inverse Radon transform.

As it is practically possible to scan only a finite number of projections, and in turn, each of these can be digitally represented only by a vector of a finite number of ray integrals, the practical task can be formulated as an approximate discrete inverse Radon transform. The respective algorithms will be treated in Section 9.1. Anyway, it should be said that sampling in both measuring coordinates (shift and rotation) must be rather dense in order to suppress as much as possible the artifacts in the final image, which might be caused by undersampling in different phases of reconstruction. The practical values in modern systems are several hundreds to a thousand samples along each coordinate.

The described approach enables, in principle, the reconstruction of two-dimensional slice images of the object. To provide three-dimensional data, consequential slices of the object, usually parallel to the first slide, are to be measured and reconstructed. This requires the object to always be moved after the acquisition of slice data perpendicularly to the imaged plane by an amount that corresponds to the thickness of the slice, as given by the effective diameter of the x-ray beam. (We shall see in the next section that an approximately equivalent, but more effective helical approach is possible.)

4.1.2 Variants of Measurement Arrangement

The arrangement depicted in Figure 4.1 represents the *first generation* of CT equipment that made use of a single pair of the x-ray source and detector. This "pencil ray" probe scanned linearly across the object to provide a projection, and was rotated by about 1° before scanning for another projection, in the angle range of 180° to 240°. The main disadvantage of the first generation was a long scanning time (several minutes), besides the large dose for the patient. The *second-generation* systems improved the situation partially by using a narrow fan beam and a short linear array of detectors, but the linear scanning was still necessary.

Most of the presently used CT systems belong to the *third generation*, of which the most prominent feature is the use of a thin (about 1 to 10 mm) but wide fan beam (30° to 60°) covering the complete slice of the object, as shown in Figure 4.2. The detector array, curved into a circular segment with the center in the radiation focus, is encompassing several hundred to a thousand detectors. Spacing of the detectors is equiangular with respect to the radiation focus, which is advantageous from the image reconstruction viewpoint

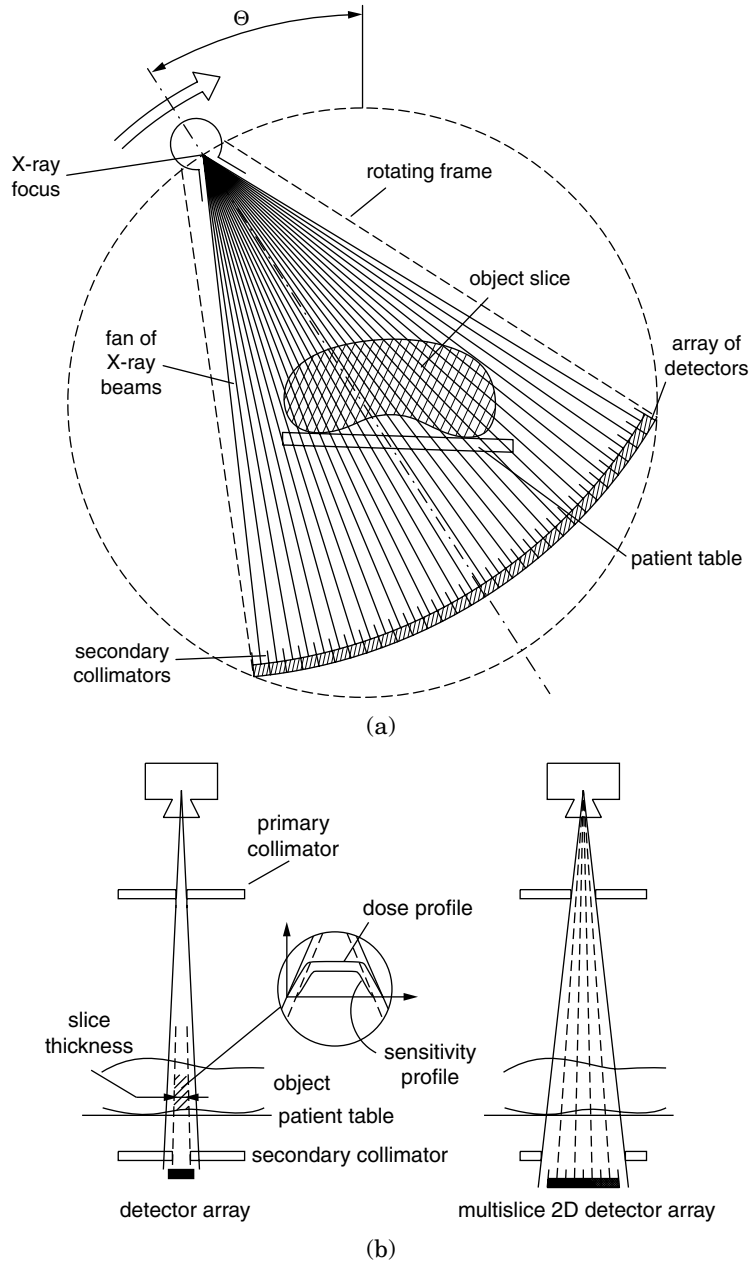


Figure 4.2 Schematic view of the third-generation CT scanner. (a) View of the slice plane. (b) Perpendicular view.

(see Section 9.1.5). The lines connecting the focus with the detectors thus define the individual equiangular rays of the fan. No lateral translation is needed, as the entire slice is scanned at once; thus, only rotational movement is needed. The complete rigid fan arrangement — the x-ray source and the detector array — rotates around a fixed axis, which crosses the object space. Naturally, the inner part of the fan is empty to allow placing of the object — a patient; the bearings are sufficiently far away so that the axial object shift is enabled as needed for scanning of different slices of the patient's body. This concept is complicated mechanically as, apart from others, the energy supply and all measuring signals have to be transferred via a slide ring arrangement to allow continuous rotation. Also, the centrifugal forces are enormous, thus requiring special construction solutions.

This arrangement is the best so far from the viewpoint of image forming. The secondary collimators can be arranged in the optimum way, such as in Figure 3.5, i.e., oriented to the radiation focus, thus refusing most of the scattered rays. The simplicity of only rotational movement allows very fast operation, in the range of 1 second or less per slice. The reconstruction algorithms are somewhat more complicated than those for the first-generation rectangular scanning, but this is more than outweighed by the speed of data acquisition. It should be noticed that while the number of projections per turn can in principle be arbitrary, limited only by the speed of data transfer, the sampling of individual projections is fixed — given by the number of detectors on the ring section.

The fourth- and fifth-generation fan beam systems are similar to those of the third generation as far as the image forming concerns. The concept of the *fourth generation*, as described by Figure 4.3, differs from the previous generation primarily in having a stationary circular array of equidistantly spaced detectors, which eliminates the need of complicated transfer of measuring signals. Only the radiation source is rotating, either inside the detector ring as depicted or possibly also outside, provided that the rays can somehow pass around the detector ring when entering the object space. The fans, forming individual projections, are defined in the opposite way as in the third generation: the vertex of a fan is at a detector and the projection data are acquired gradually, during the x-ray source rotation; obviously, many different fans can be served simultaneously. Nevertheless, the imaging situation is less favorable than that of the third generation because the secondary collimators cannot be efficiently applied since the detectors must accept rays from different angles; thus, the influence of scattered radiation cannot be suppressed that well. Another drawback is the

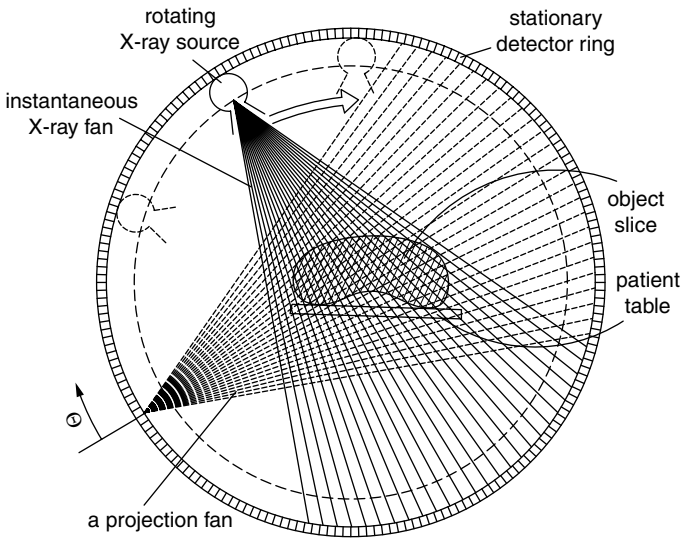


Figure 4.3 Arrangement of the fourth- and fifth-generation CT scanners.

nonequiangular spacing of the rays in a fan, so that some additional interpolation providing the uniform spacing is necessary.

Note that every individual projection has the vertex of its respective fan situated at a detector. In comparison with the third generation, the choice possibilities of sampling are thus reversed: the total number of projections is determined by the number of detectors on the ring, while the sampling density in a fan, given by the sampling rate of continuously changing position of the x-ray focus, can be arbitrarily fine. It also means that the detectors need not be as densely spaced on the ring as those on the sector of the third generation.

The *fifth generation* uses the same imaging arrangement, with the only difference being that the x-ray source is not rotating mechanically. Instead, a large circular metallic anode is situated in place of the radiation source track and the electron beam is electronically swept to reach the anode in the desired instantaneous position of the focus. This naturally requires a huge vacuum envelope, but there are no mechanically moving parts so that the scanning can be very fast. The reconstruction algorithm principle for both the fourth and fifth generations remains practically equivalent to that used in the third generation, with some minor modifications (more interpolation to provide for equiangular sampling of fan projections, a need for more efficient compensation of scatter influence).

For all rotating-fan systems, it is common that every ray in the complete projection set is measured twice from opposite directions during one 360° revolution. It means a redundancy because, as far as the linearity assumption is fulfilled (anyway necessary for image reconstruction), both measurements give the same attenuation value. This may be utilized for faster imaging, by scanning less than a complete turn — it can easily be shown that a scanning angle equal to (180° + fan angle) would provide a complete set of ray integrals. Some systems remove the redundancy by shifting the detector array laterally by half a detector width for the opposite measurement, thus obtaining doubled spatial density of rays. Alternatively, the redundancy may also be used for dividing the complete 360° data into "direct" data and "complementary" data, the latter corresponding to the opposite-direction measurements. The complementary data are obviously provided later during a turn, which can be utilized in helical scanning (see below).

Two substantial improvements, concerning three-dimensional data acquisition in the form of sets of parallel slices, can be implemented in CT systems of the third to fifth generations: helical and multislice scanning.

Helical scanning (Figure 4.4) means that the object is axially moved during the scanning procedure while the measuring fan is rotating. This can easily be arranged for by automatically moving the patient support with respect to the gantry of the scanner. The axial shift during a 360° turn is usually equal to the effective slice thickness w. Every revolution thus basically provides information from a single slice, although not exactly: the individual fans are not scanned all in the same plane of a slice, but rather they are gradually shifted along the z-axis. This can be seen in panel (b) of the figure, where the horizontal axis of the diagram corresponds to the z-shift, while the vertical axis corresponds to the turning angle. The effective slice thickness of the measuring fan is continuously moving along z during each 360° turn, thus covering a slanting stripe in the diagram per revolution. Nevertheless, for a good reconstruction, the projections must all correspond to a certain slice plane perpendicular to z, as represented by the thick line. These data can be provided by one-dimensional interpolation between neighboring measured data, which can be interpreted as longitudinal filtering by a finite impulse response (FIR) filter encompassing the distance corresponding to basically two revolutions (360° interpolation in the figure). Alternatively, the above-mentioned complementary data can be utilized, as they may be closer to the required slice plane than the direct data.

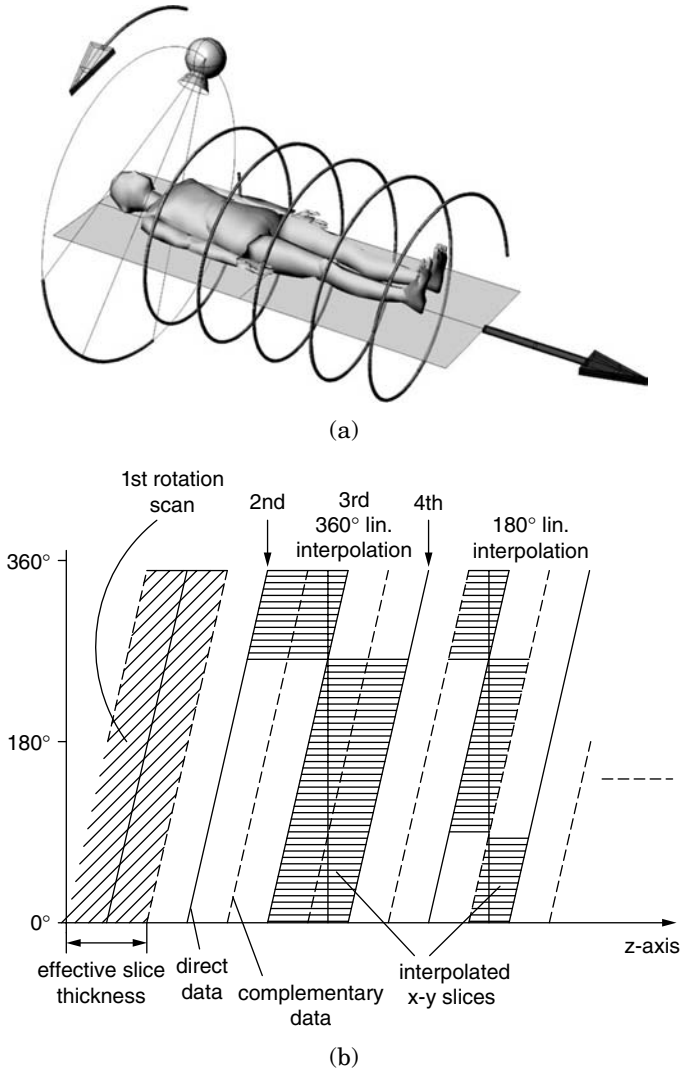


Figure 4.4 Principle of helical scanning: (a) imaging principle and (b) coverage of imaged space along the z-axis.

This way, the density of data along the z-axis is doubled, so that the length of the filter can be shortened to correspond to only a single revolution that effectively provides the needed two sets of data (180° interpolation). Commonly, linear interpolation (triangular filter, as

considered above) is used, but there are also possibilities for a kind of sharpening along the z-axis (i.e., making the slices effectively thinner) by using longer and more complicated filters.

Multislice systems provide more than one slice during a single rotation, usually four slices. It is enabled by using a two-dimensional field of detectors: instead of a single row, several parallel rows of detectors are used, separated in the z-direction by about 0.5 to 2 mm (Figure 4.2b). The number of rows can be even higher than four; in this case, grouping of neighboring detectors is possible to adjust the thickness of the slice to the concrete needs. Naturally, the x-ray fan must cover all the parallel slices at once, and therefore be correspondingly thicker in the range of the object. Because a single x-ray point source is used, the measured slices are not exactly parallel, but rather diverging; this must again be compensated for by longitudinal interpolation. The multislice principle provides immediately three-dimensional data in a limited volume of four slices; a greater extent along the z-axis can be achieved by helical scanning. Obviously, the axial shift during one rotation should now be multiplied—with respect to standard single-slice helical scanning—by the number of simultaneous slices N, thus speeding up the measurement. Nevertheless, to enable use of the complementary data as explained above, and thus to increase the resolution along the z-axis, the shift should only be $(N - 1/2)w$, otherwise, the complementary data become identical to the direct data.

4.2 MEASURING CONSIDERATIONS

4.2.1 Technical Equipment

From the imaging point of view, the x-ray's path, as used for measurements, does not differ in principle from that described in Section 3.1. It should be mentioned that the detector arrays used in CT systems are of two types (Figure 4.5): scintillation type detectors (scintillation crystals followed by integrated photosensitive elements), similar in principle to that described in Section 3.1.4, and rare gas ionization chamber detectors. While the former have greater quantum efficiency of detection, the latter are simpler, have a faster response needed for high-speed scanning, and are partially self-collimating. Both types have an excellent linearity in a large dynamic range, which is needed for good image reconstruction.

Collimation is of primary importance in CT scanning (Figure 4.2). Primary (source) collimators must define the radiation fan of the

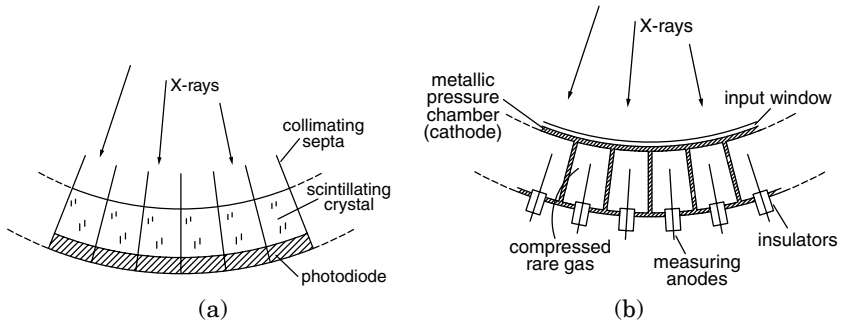


Figure 4.5 Scheme of part of a detector field. (a) Scintillation detectors. (b) Gas ionization detectors.

proper thickness and of a transverse (dose) rectangular (uniform) profile as much as possible, and possibly nondiverging in the z-plane. The secondary (detection) collimation improves the definition of the fan thickness along the z-axis (sensitivity profile) and formulates the width of a ray beam in the slice plane, thus substantially influencing both the (x, y)-resolution in the slice plane and the z-resolution in the perpendicular direction.

Nevertheless, the user cannot subsequently influence these details, and when there is a need to take into account particular properties of an individual system, only input–output identification based on the imaging of some kind of phantoms would provide the needed information.

4.2.2 Attenuation Scale

A real CT scanner uses wideband (polychromatic) radiation so that the notion of the linear attenuation coefficient — being energy (wavelength) dependent — loses its uniqueness and is not directly suitable for evaluating pixel values. It can be replaced by the concept of the *weighted attenuation coefficient*, taking into account the energy spectrum $S(E)$ of the radiation,

$$\mu_m \approx \frac{\int_{R_E} \mu(E) S(E) \, \mathrm{d}E}{\int_{R_E} S(E) \, \mathrm{d}E} \tag{4.3}$$

(integrating over the whole range of involved energies R_E). It has been shown experimentally that this quantity describes well the

values provided by CT scanners as long as the output spectrum of the rays does not differ from the incident spectrum (i.e., if the hardening phenomenon may be neglected). As the differences in linear (or even weighted) attenuation coefficients of different tissues are rather minor, the results of reconstruction are usually expressed in another (Hounsfield) scale by means of *CT numbers* (Hounsfield units). This relative attenuation scale corresponds to the scale of linear attenuation coefficient μ according to

$$\mu_r = CT\ number = 1000\frac{\mu - \mu_{water}}{\mu_{water}}, \tag{4.4}$$

ranging between –1000 for vacuum (or air) and about 3000 for bone tissue. Most soft tissues have CT numbers in the approximate range of only –200 (lung tissue excluded) to about 60; water CT number is naturally 0. When working with weighted attenuation, the attenuation of water is taken at the radiation energy of 72 keV by definition.

4.3 IMAGING PROPERTIES

4.3.1 Spatial Two-Dimensional and Three-Dimensional Resolution and Contrast Resolution

The spatial resolution is not isotropic in CT imaging due to anisotropic methods of image data acquisition and processing. Obviously, the (x-y)-resolution in the slice plane depends primarily on the width of the x-ray beam in this plane and on the density of sampling, i.e., on the number of samples in a projection. This is given by the number of rays (lines) in the fan — therefore by the density of detector field in the third generation and by the sampling rate with respect to the rotational speed in the fourth and fifth generations. The resulting resolution is also influenced by the level of artifacts caused by discrete image reconstruction, and therefore by the total number of projections; this can be arbitrarily chosen in the third generation by selecting the angle increment between consequential scans, while in the fourth and fifth generations it is limited by the number of detectors on the circle.

The geometrical parameters of the beam, which should be ideally infinitesimally thin from the resolution point of view, are determined by the focal spot characteristics, including the smearing caused in the fourth and fifth generations by x-ray source movement, and namely, by collimation — thus the importance of these construction

details. The beam geometry actually used, the design of which may also take into account other aspects (noise level, contrast resolution, patient dose), can be influenced prior to the data acquisition by the radiologist to the extent given by the construction concept. Once the data are acquired, all these parameters are fixed. Nevertheless, the user still has an opportunity to partly influence the resolution during the reconstruction process (see Section 9.1).

The z-axis (axial) resolution is primarily given by the effective thickness of the beam (fan, slice), which is described in detail by the transversal profile, which should ideally be rectangular to define the slice exactly. Unfortunately, the thickness of a slice is not constant along a ray due to limitations in collimation (see above). Even more complicated is the situation in the case of helical scanning, when the slice data are provided by interpolation; this leads to profiles that are more Gaussian shaped than rectangular. On the other hand, the user can partly influence the resulting z-axis resolution subsequently by choosing a suitable longitudinal interpolation filtering method.

Equally important as the space resolution is the *contrast resolution*, which is defined as the ability to detect an area of a CT number a as distinct from the background of a different CT number b. The relative contrast is defined as $(a - b)/b$. The recognizable contrast is strongly dependent on noise level in the image and on the size of the area to be detected. Because the CT systems can measure the attenuation more precisely than classical projection x-ray systems, the contrast resolution is substantially better, allowing the resolution of different types of soft tissue. Imaging properties of a concrete CT system from this point of view, under specified conditions (x-ray spot size, slice thickness, patient dose, corresponding noise level, details of reconstruction algorithm, etc.), can be summarized in the *contrast–detail diagram*, such as that in Figure 4.6. Here, the limiting geometrical resolution can be seen on the left (for high-contrast details), while the minimum size of a detail recognizable under a given contrast (difference in CT numbers) can be read from the curve.

4.3.2 Imaging Artifacts

Like in projection radiography, the x-ray intensity measurement is subject to errors due to *quantum noise*, governed by Poisson distribution of detected photons. The standard deviation of the noise is therefore proportional to $1/\sqrt{N}$, where N is the mean number of photons in a measured beam. The instantaneous flow of photons is

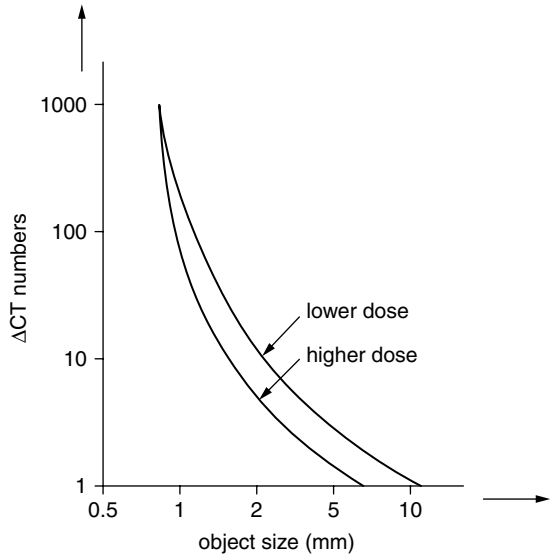


Figure 4.6 Schematic contrast–detail diagram.

influenced by the radiation mechanism, attenuation, and detection efficiency, which are all subject to stochastic deviations. The CT measurement is just at the edge of possibilities, compromising between an acceptable photon noise level, reasonable energetic efficiency of collimation, and tolerable patient dose (noise level decreases about linearly with the square root of dose). The decision on the kind of compromise depends on the radiologist, who also takes into account the diagnostic goal (preference of high-contrast resolution or high geometric resolution). The other sources of data disturbance, namely, thermal noise and other noise types in electronic measuring circuitry, can be kept relatively unimportant.

Scatter of x-ray photons adds some false background to the measured ray intensities; the scatter increases with the attenuation of the measured ray due to object passing. Nevertheless, the scatter influence can be suppressed during postprocessing if the distribution of scattered radiation along the detector field is known. As the scatter is rather omnidirectional, it can reasonably be supposed that the contribution due to scatter is approximately constant over the entire detector field; it is then possible to measure it by detectors placed out of, but near to, the imaging plane.

Polychromaticity of the used x-rays (i.e., wideband character of their energetic spectrum) causes some artifacts in the image that is reconstructed by means of methods based on the assumption of monochromaticity: whitening of soft tissues inside scull bones, streaks in the vicinity of transversal bones, etc. Rigorously taken, a more complex model of imaging and, consequently, more complicated reconstruction methods should be used. Different attenuation properties of tissues at different photon energies E (or wavelengths λ) lead to the phenomenon of *beam hardening*, which should in principle be taken into account by integrating all the above-mentioned monochromatic descriptions and expressions along E in the range of used energies. Nevertheless, this would practically lead to hardly tractable equations, so that only approximations and corresponding corrections are used. The corrections are based on the finding that all soft tissues have similar attenuating (and hardening) properties as water. Thus, the correction can be based on determining the difference between the theoretical linear dependence of monochromatic log attenuation (Equation 4.1) on the thickness of attenuating homogeneous material, and the measured dependence, which is nonlinear due to polychromaticity (Figure 4.7).

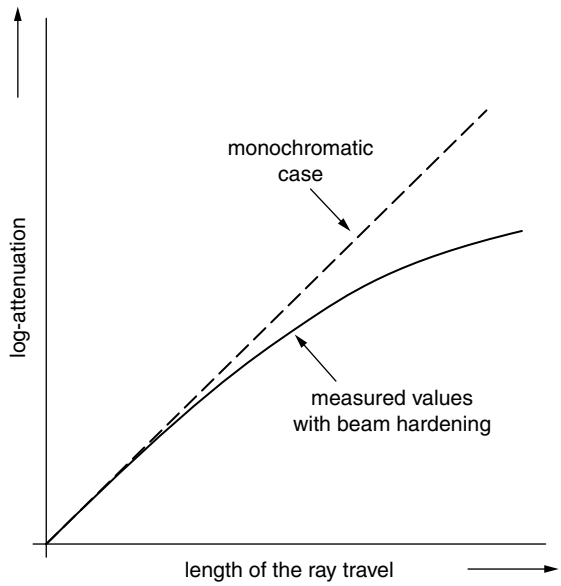


Figure 4.7 Schematic plot of a measured (solid curve) and theoretically expected log attenuation.

Another nonlinear phenomenon is *partial-volume effect*, which appears when more structures of different attenuation are filling a single volume element for which a mean attenuation is to be calculated. The most pronounced situation like this appears when a thick slice is only partially penetrated by a high-contrast structure (Figure 4.8). The measured intensity is then obviously given by the mean value $(aI_1 + bI_2)/(a+b)$ where the weights a, b correspond to relative cross-sections of sub-beams that each, when of full cross-section, would give the intensities I_1, I_2. Nevertheless, the ray integrals are calculated by Equation 4.1, so that the resulting value is proportional to $\log[(aI_1 + bI_2)/(a+b)]$ which is not a linear combination of the two partial-ray integrals, $\log I_1$ and $\log I_2$. Naturally, the same effect would also appear when two or more different materials fill a voxel in a slice, but as the (x-y)-dimension of a voxel is usually small, the artifacts are not that pronounced. The nonlinearity that is in conflict with the reconstruction assumptions influences the whole slice and may again cause rings or streaks. It would be

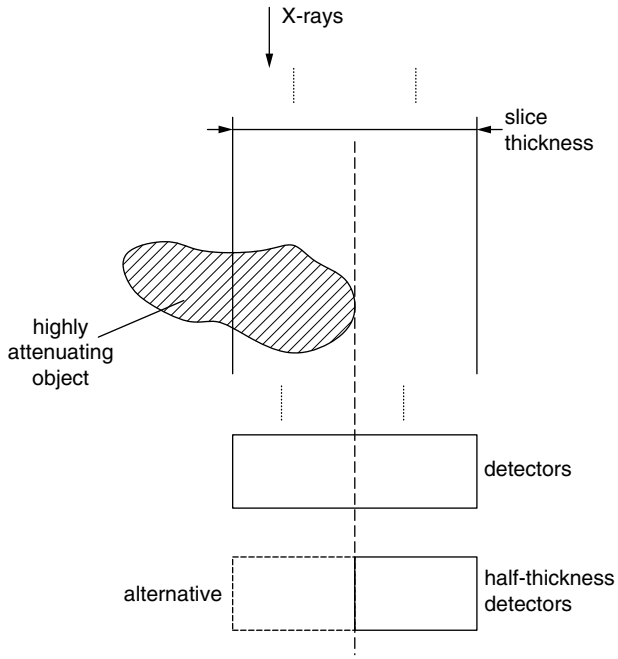


Figure 4.8 Partial-volume effect dependent on slice thickness.

possible to compensate for this nonlinearity by postprocessing, but the knowledge of partial attenuation and of the relative influences would then be needed, which is not easy to provide. Obviously, the effect can be suppressed by reducing the partial volumes, primarily by using thinner slices.

Finally, the artifacts due to partial failure of technical equipment should be mentioned. Failure of a detector in the detection field has different consequences in the third-, fourth- and fifth-generation systems. In the third-generation system, a defective detector means that one ray of all fans is unavailable, thus influencing all projections in a correlated manner, which causes ring artifacts in the reconstructed image. In the fourth- and fifth-generation systems, failure of a detector causes a complete projection to be unavailable, but the others are untouched — no error correlation among projections. It can be shown that the missing projection does not influence the result substantially if the total number of projections per slice is sufficient, as is usually the case nowadays. Similar, though not that apparent, are errors due to uneven sensitivity of detectors. These errors are easy to compensate for by field-homogenizing procedures (Section 12.1.2), providing that calibration data are available. Complementary errors would be caused by nonconstant x-ray source radiation flow: a complete momentary failure of the source would cause a loss of a complete projection in the third generation, and a loss of individual rays in many fans (projections) in the systems with ring type detectors. They could in principle be compensated for in postprocessing if the variations in flow are systematic and known, but this is rarely the case.

From the previous discussion, it is clear that the CT imaging is subject to many influences and parameters that may partly — but mostly cannot — be influenced by the user. The final evaluation of the imaging quality of a particular system must then be done practically by imaging some artificial objects of well-defined properties — *phantoms*. Many different types of phantoms, addressing different aspects of image evaluation, are commercially available. Let us mention only the main image quality parameters, which can be determined using phantoms, schematically depicted in Figure 4.9. The simplest type (a) is formed by a small bead (substantially less than the expected resolution, i.e., ~0.1 to 0.3 mm) of a material with significantly different attenuation, positioned in a surrounding base material; this forms an "impulse" object, to which the response can be considered two-dimensional or, in helical systems, three-dimensional point-spread function (PSF).

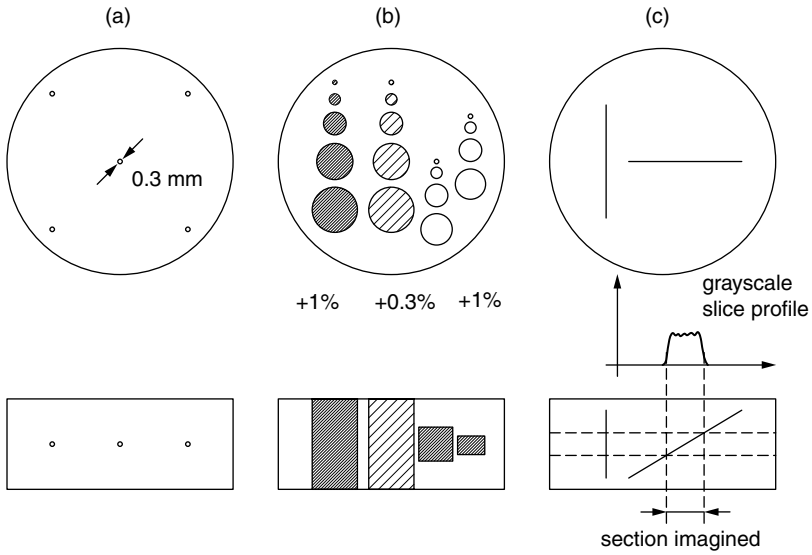


Figure 4.9 Different types of CT phantoms, enabling the determination of (a) point response, (b) contrast vs. size resolution, and (c) slice profile and subslice sensitivity.

Type (b) contains cylindrical objects of different diameters and with different x-ray contrasts with respect to the base material; it enables the checking of the contrast-size resolution dependence. Type (c), where ramp filaments of differing attenuation are placed, enables measurement of the slice profile and corresponding slice thickness; obviously, only the part of the ramp is visible that is inside the slice thickness, while the response intensity informs about the profile.

Still, a very simple phantom that does not need depicting should be mentioned: the water phantom simulating a homogeneous ("average") soft tissue, enabling absolute measurements and calibration of the source–detector field system under more realistic conditions than with air-filled space.

4.4 POSTMEASUREMENT DATA PROCESSING IN COMPUTED TOMOGRAPHY

CT imaging, as the name indicates, is completely based on digital processing of the measured data. The computational *reconstruction from projections* (Chapter 9) is the fundamental step that enables

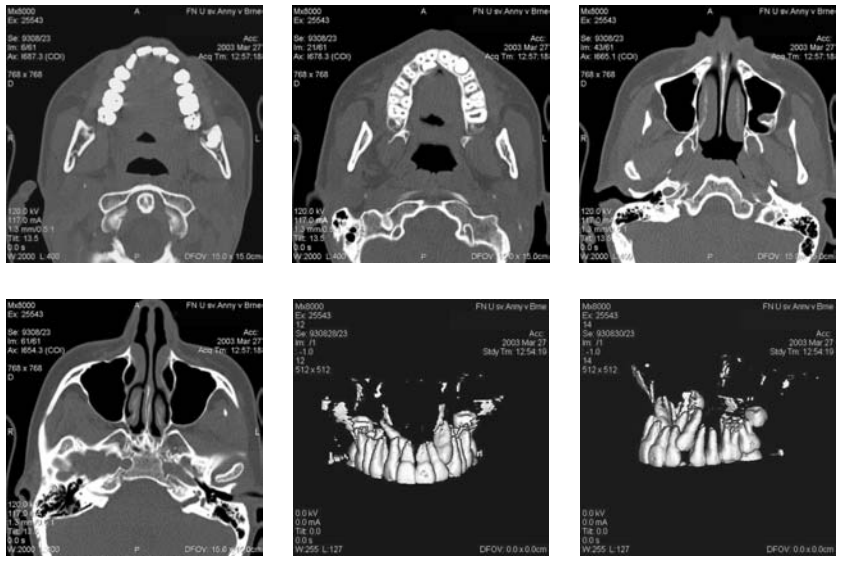


Figure 4.10 Typical CT slice images — selected slices from two sequences and corresponding digital three-dimensional reconstructions (lower right in each part). (Courtesy of the Faculty Hospital of St. Anne Brno, Radiology Clinic, Assoc. Prof. P. Krupa, M.D., Ph.D.)

visualization of the data in a comprehensible form. Several parameters can be chosen during the reconstruction, thus enabling the influencing of the resulting image to a certain extent. Selected slices of two typical series of CT images are shown in Figure 4.10, together with three-dimensional reconstructions based on the two-dimensional slice sets.

Before the image measurement, or with a special arrangement in the CT system even simultaneously with the measurement, *calibration* is performed that may provide data on x-ray irradiation unevenness, detector field inhomogeneity, nonlinearity due to beam hardening, and levels of scattered x-rays. From the image processing viewpoint, it is irrelevant whether these measurements are done regularly, with what intervals, or even just once. All these quantities can be used for correcting the measured raw data — ray integrals — by simple numerical procedures (e.g., field homogenization, lookup table conversions, etc.) before they are submitted to the reconstruction step.

Particularly, the *beam-hardening compensation* deserves special mention. The correction in soft tissue slices is simple when the

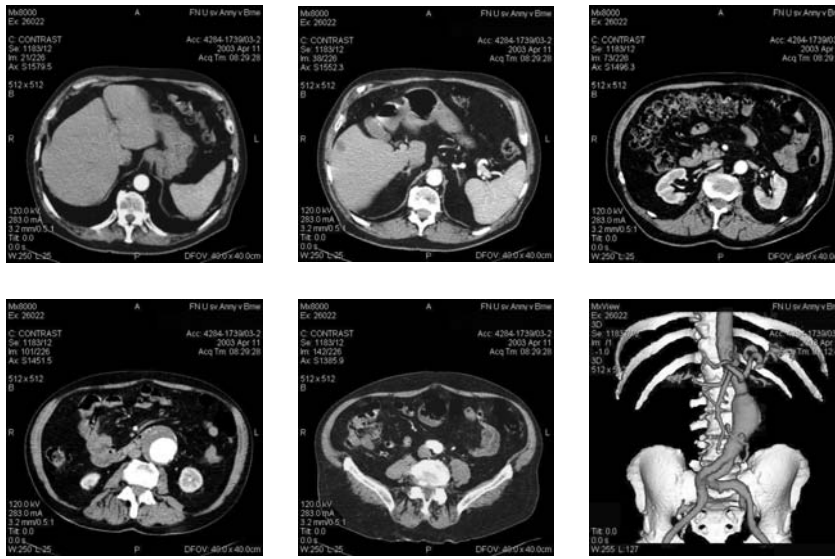


Figure 4.10 (Continued)

correction curve is known, as in Figure 4.7; each ray integral measurement is corrected simply by replacing it with the value expected theoretically. Nevertheless, if bones or other highly attenuating materials occupy an important part of the slice, the assumption of similarity with water is not valid and the technique fails. In this case, every ray integral should be corrected differently, depending on the extent of the bone crossed by the particular ray. To establish this extent, a preliminary reconstruction based on evenly corrected data is provided, correction of each ray integral is improved with respect to the bone extent derived from the provisional image, and, based on improved ray integrals, a new image reconstruction is generated. This procedure may be iteratively repeated.

The images obtained via reconstruction can still become a subject to postprocessing *image enhancement* procedures, namely, contrast adjustment or false color representation to enable better visual evaluation of small-contrast diagnostically important information (Section 11.1). There have been many attempts to improve the image quality, namely, the spatial resolution, by some kind of restoration, but this is probably not very promising when taking into account the complicated process of obtaining the image causing the image to be neither isoplanar nor precisely linear.

The postprocessing thus concentrates rather on image analysis. Of particular interest are the methods of *segmentation* (Section 13.2), enabling the separation of the organs or different types of tissue in individual images (slices).

Three-dimensional representation by maneuverable two-dimensional images is mostly based on previous segmentation of slices providing outlines of anatomic organs, and following reconstruction of space surfaces, which are consecutively converted into the three-dimensional visualization by *surface rendering*. The alternative approach, *volume rendering*, is used as well, though perhaps less frequently, as it may be less flexible when particular inner organs are to be visualized. The higher robustness of volume rendering to image data imperfections need not be utilized in the case of CT image data that are of relatively high quality.

5

Magnetic Resonance Imaging

Magnetic resonance imaging (MRI) is based on a rather complex physical phenomenon of nuclear magnetic resonance (Bloch 1946 [50]; Purcell et al., 1946 [51]), which is basically the exchange of energy between elementary particles placed in a strong magnetic field and the irradiating electromagnetic field of a particular frequency. It is, precisely taken, governed by laws of quantum mechanics and should be described in corresponding terms. Nevertheless, from the viewpoint of medical image data acquisition, processing, and interpretation, the description at the level of individual atoms or their nuclei is, in most respects, not necessary. We will thus present a macroscopic approximation, introduced originally by Bloch (1946), where large sets of nuclei are the subject of observation or measurement. These sets (sc., spin isochromats*) can often be considered to encompass a volume of a voxel — a three-dimensional space element of the discrete image data. In some cases, when some of the parameters influencing the behavior of the group are not homogeneous

*The name is derived from the property that the nuclei in the group are precessing at the same frequency — see below.

throughout the voxel volume, a subvoxel view is necessary; nevertheless, even such a smaller subvolume still contains a huge number of particles, allowing the acceptance of the macroscopic view when analyzing the external magnetic behavior of the subgroup.

MRI is entirely dependent on digital processing of the measured data that have to be converted to the image form by means of algorithms based on the theory underlying this book. While it is impossible here to go into details of many branches of the highly developed MRI field, the purpose of this chapter is to give a consistent and comprehensible explanation of the principles of MRI data acquisition and processing, which are generally considered difficult (and often described in an obscure or vague way). The explanation in Sections 5.1 to 5.5 is thus intended to clarify the image-forming procedures in the manner enabling the user to understand and well interpret the image properties.

Definitely, it should not be considered an in-depth explanation of MRI phenomena or imaging practices. Particularly, the advanced quantum physics of nuclear magnetic resonance is completely omitted, as well as the construction details and technical solutions of MRI systems, and particularities of MRI in specialized diagnostic applications. For a more detailed study or for further reference, see [9–11], [15], [26], [30], [34], [37], [38], [47].

5.1 MAGNETIC RESONANCE PHENOMENA

5.1.1 Magnetization of Nuclei

A single proton (the nucleus of a hydrogen atom or a particle of another atomic nuclei), besides being positively charged, has a property called *spin*, which can be attributed to rotational movement of the proton, to enable easier understanding. The charged rotating particle has a magnetic moment m, like a magnetic dipole. When the particle is situated in an external magnetic field $\mathbf{B}_0$, it can have, according to quantum mechanics, only two values of potential energy, which correspond to either parallel (positive) or antiparallel (negative) orientation of the magnetic dipole relative to the external field. Nevertheless, as the proton of nonzero mass is rotating, the torque acting on the magnetic dipole due to the external field $\mathbf{B}_0$ cannot bring the proton fully to any of the ultimate (parallel or antiparallel) positions, but rather will result in a precession movement — a rotation of the oblique magnetic moment vector around the direction of the field (Figure 5.1). The angle of the tilt

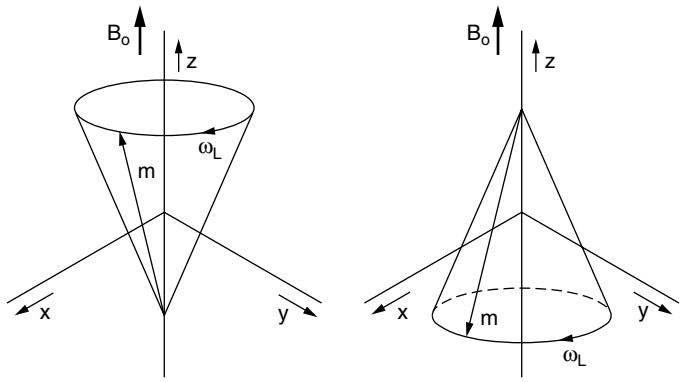


Figure 5.1 Two possibilities of precession of a single-proton magnetization vector. Left: lower-energy state. Right: higher-energy state.

can have only a single absolute value; both cases differ only in the orientation. Nuclei of more complicated atoms behave similarly as far as they have an odd number of protons.

The frequency of the nuclei precession is given by

$$\omega_L = \gamma B_0 , \tag{5.1}$$

where γ is the *gyromagnetic ratio*, a constant specific for a concrete type of nucleus, and B_0 is the magnitude of $\mathbf{B}_0$. The quantity ω_L, called (angular) *Larmor frequency*, is thus determined by the induction of the external magnetic field and by the type of the nucleus. As an example, an isolated proton has $\gamma = 42.58 \times 10^6$ HzT^{-1}, and its Larmor frequency in the magnetic field of, e.g., $B = 1.5$ T, is thus 63.87 MHz.

The mentioned two energy levels of a nucleus differ by $\Delta E = \hbar\omega_L$, where $\hbar$ is Planck's constant; delivering or releasing this energy in the form of a photon of Larmor frequency is connected with switching between both states. Under thermal equilibrium, the ratio of probabilities of both states is given by the *Boltzmann probability factor*

$$\frac{P_{neg}}{P_{pos}} = \exp\left(-\frac{\hbar\omega_L}{k_B T}\right), \tag{5.2}$$

where k_B is a given (Boltzmann) factor and T is the absolute temperature. As this ratio is always less than (though close to) 1 under thermodynamical equilibrium, there will always be, in the equilibrium

state, more positively oriented (energetically lower) than negatively oriented nuclei in a population.

Let us now consider a set of nuclei in a small volume V, where the composition of matter as well as B_0 can be considered constant, e.g., in a voxel or in a part of it. The net effect of the dominance of lower-energy particles is that there will be certain resulting magnetization (per volume) that can be characterized by the vector **M**, given by the sum of all individual particles' magnetic moments m_n normalized by the volume,

$$\mathbf{M} = \frac{\sum_n m_n}{V} \propto \frac{B_0 D_p}{T}. \tag{5.3}$$

The last expression shows that the magnetization is proportional to proton density D_p, one of the important tissue parameters to be measured and imaged. The situation is described by Figure 5.2a, where the spatial coordinates are oriented by convention so that the z-axis has the direction of the magnetic field $\mathbf{B}_0$. In the thermodynamic equilibrium state, **M** obviously has the z-axis orientation, as more particles have the positive orientation. Precessing of individual nuclei has random phases so that transverse magnetization cancels out. The precession frequency is obviously constant in the homogeneous volume—this gives the set its above-mentioned name, *spin isochromat*.

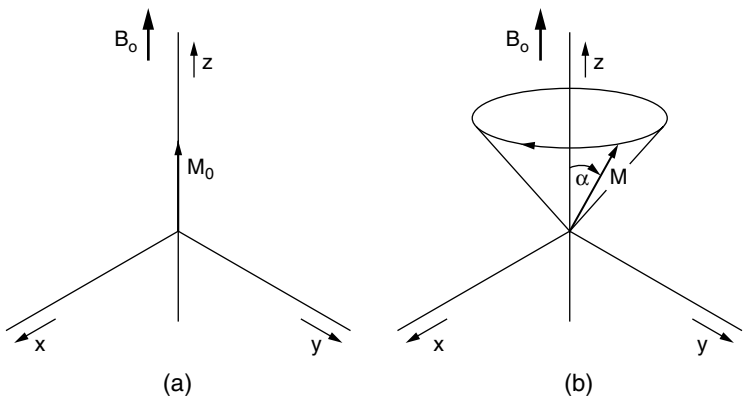


Figure 5.2 Magnetization of an isochromatic set of nuclei: (a) under thermal equilibrium and (b) after irradiation by radio frequency energy.

5.1.2 Stimulated NMR Response and Free Induction Decay

The situation changes if there is a certain degree of phase coherence among particles in the volume V, as shown in Figure 5.2b. Then, the resulting vector $\mathbf{M}$ is inclined with respect to the z-axis by an angle α, and naturally precesses with the same Larmor frequency. The coherence can be achieved by irradiating the object volume by radio frequency (RF) energy oscillating at Larmor frequency. The irradiation also leads to another phenomenon: the population of higher-energy (antiparallel) particles increases at the cost of the lower-energy population. The *flip angle* α may therefore reach values close to 180°, if most particles acquire the higher-energy state. Note that while the precession angle of a single particle has only two possible values, the flip angle α of magnetization of a volume can achieve any value in the continuous range 0° to 180° thanks to the averaging effect of many nuclei in the set*. The value of α is obviously determined by the amount of delivered RF energy, which in turn can be controlled by either the magnitude or duration of the RF pulse, or both. It is then often spoken of as a 90° *pulse* or 180° *pulse*, meaning the RF pulse that leads to the desired value of α.

To explain the macroscopic effect of RF irradiation on the analyzed volume quantitatively, it is useful to introduce the *rotating frame of reference* (x', y', z') besides the stationary (laboratory) coordinates (x, y, z). While the axes z and z' coincide, the (x', y')-plane rotates about the z-axis with the Larmor frequency, so that when the precession of $\mathbf{M}$ is time invariable, its position in the rotating frame is fixed. Let the RF pulse have the form of a transverse circularly polarized wave, i.e., with its magnetic vector $\mathbf{B}_{RF}$ also rotating in the (x, y)-plane of the stationary frame with the Larmor frequency; consequently, it is a fixed vector in the rotating frame, say along the x' direction**. The magnetic field $\mathbf{B}_{RF}$ influences the magnetization $\mathbf{M}$ so that this will precess around $\mathbf{B}_{RF}$ in the same manner, as any magnetic dipole in a magnetic field, as already described.

---

*It should be said that this commonly used explanation is rather rough and does not explain all the macroscopic phenomena — see, e.g., [15]; nevertheless, it is helpful as an introductory approach.

**Note that if a linearly polarized wave is applied, it can be split into two circularly polarized waves, of which one has the described properties. The other component is rotating in the opposite direction and has no net influence due to averaging during a period.

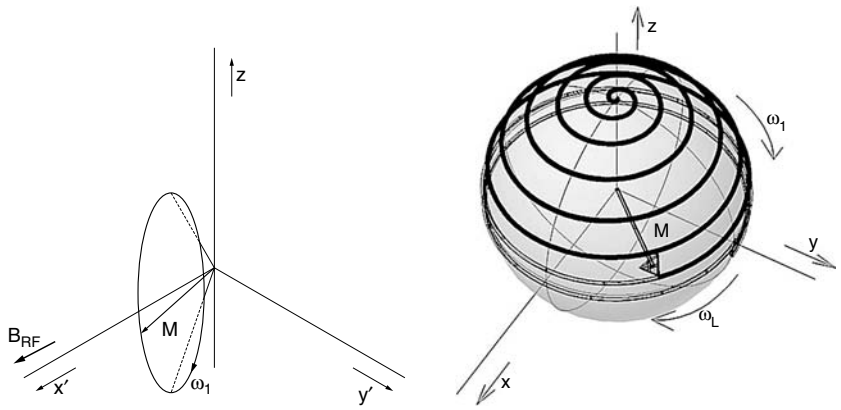


Figure 5.3 Influence of RF irradiation to **M** precession. (a) Isolated effect of $\mathbf{B}_{RF}$ in the rotating frame. (b) Combined effects of both stationary and RF fields in the stationary frame.

This can be well visualized in the rotating coordinates (Figure 5.3a). As the RF magnetic field is several orders weaker than $\mathbf{B}_0$, the precession frequency $\omega_1 = \gamma B_{RF}$ is much lower than ω_L. The combination of fast precession around $\mathbf{B}_0$ and slow precession around $\mathbf{B}_{RF}$ leads to a spiral trajectory of **M** in the stationary coordinates (Figure 5.3b). If the magnitude B_{RF} is time invariable, obviously the 90° pulse duration T_{RF} must be a quarter of the RF precession cycle, $\pi/2\omega_1$, while the duration of the 180° pulse is π/ω_1. Generally, the flip angle provided by a constant field B_{RF} is then $\alpha = \gamma B_{RF} T_{RF}$. In case of time-variable B_{RF}, the flip angle achieved during a time T_{RF} obviously is

$$\alpha = \gamma \int_0^{T_{RF}} B_{RF}(t)\,\mathrm{d}t. \tag{5.4}$$

Irradiation of the object by RF energy is mediated by an *RF coil* (or a couple of coils; see below) supplied by a transmitter working at Larmor frequency.

The measurement of NMR response can be obtained by electromagnetic induction, as seen in Figure 5.4. Let the measuring coil be oriented so that the induced voltage corresponds only to time variations in the y-component of the magnetic field and is insensitive

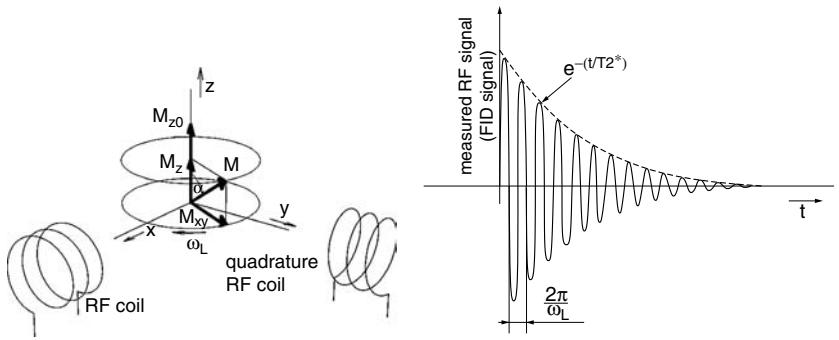


Figure 5.4 Measurement of MR-induced signals. Left: The arrangement. Right: Schematic signal plot.

to the z- and x-components*. The precessing magnetization vector $\mathbf{M}$ (when nonzero) can obviously be decomposed to the axial (longitudinal) component $\mathbf{M}_z$ and the rotating transversal component $\mathbf{M}_{xy}$, which is the only one that generates the measured signal. More concretely, only the time-variable component $\mathbf{M}_y$ is measured, but its amplitude is clearly the same as M_{xy}. Obviously, the geometrical angle of the vector in the rotating (x', y')-plane determines the phase shift of the measured time signal with respect to the phase of previous RF excitation (and to the phase of reference signal in the processing unit of the receiver).

Immediately after the measured volume is irradiated to such an extent that the flip angle is 90°, the measured signal has its maximum amplitude. Then it decays during a relatively short time (several tens of milliseconds)—thus the name *free-induction decay (FID) signal* (right part of Figure 5.4). The initial amplitude of the signal is obviously proportional to the maximum of $M_{xy} = \text{abs}(\mathbf{M}_{xy})$, which in turn corresponds to the magnitude of initial magnetization $\mathbf{M} = \mathbf{M}_z$. Because, according to Equation 5.3, the initial $\mathbf{M}$ is proportional to proton density, the initial signal magnitude describes basically the proton density of the analyzed object volume, one of the important tissue parameters measurable by MR. The time-course of

*The same coil can also be used for irradiating the object when switched alternately between the transmitter and the receiver. Often, a couple of perpendicular (quadrature) coils are used, which serve as either a transmitting or receiving circularly polarized antenna.

the decay (top, Figure 5.5), which carries further substantial information on the object volume, will be analyzed in the following section.

5.1.3 Relaxation

When the analyzed volume is left in field $\mathbf{B}_0$ until the thermodynamic equilibrium is reached (as in Figure 5.2a), the magnetization will have its z-oriented equilibrium value $\mathbf{M}_0$ according to Equation 5.3. At this stage, no signal can be detected, as the individual protons are dephased, and consequently, the transversal magnetization components cancel out (Figure 5.5a).

Immediately after being excited by a 90° RF pulse, the volume has only the transversal magnetization $\mathbf{M} = {}_0\mathbf{M}_{xy}$. The magnitude of ${}_0\mathbf{M}_{xy}$, and thus also of the signal, is maximum just because all the components have been synchronized by the pulse (Figure 5.5b).

The following fast exponential decay of the signal envelope is explained by gradual loss of coherence, i.e., dephasing of precession of proton subsets, as depicted in Figure 5.5c. This causes the exponential decrease in the magnitude of $\mathbf{M}_{xy}(t)$ according to

$$M_{xy}(t) = {}_0M_{xy} \exp\left(-\frac{t}{\tau}\right), \tag{5.5}$$

where τ is a time constant determined by many factors, described below. This process continues until complete incoherence is reached with the partial magnetization vectors of individual subsets of nuclei still in—or near to—the (x', y')-plane (Figure 5.5d). The dephasing is a result of local inhomogeneities of the magnetic field—the Larmor frequencies of the individual subsets thus differ slightly from the mean value, and consequently, the subsets are gradually either gaining or losing in the precession phase.

The inhomogeneities can be divided, according to their cause, into two groups; primarily, they are caused by microscopic magnetic interaction among particles. If there were only this material-related factor, the exponential decay would have a time constant, usually denoted $T2$; this component of the process is therefore called $T2$ *relaxation* (or *spin–spin relaxation*). The values of $T2$, in the range of about 40 to 100 msec, are characteristic for different types of biological tissue; this parameter has relatively good tissue discrimination. Unfortunately, the second group of field inhomogeneities, caused by macroscopic inhomogeneity of $\mathbf{B}_0$ and by differences in magnetic susceptibility in the imaged object, mostly prevails in the resultant

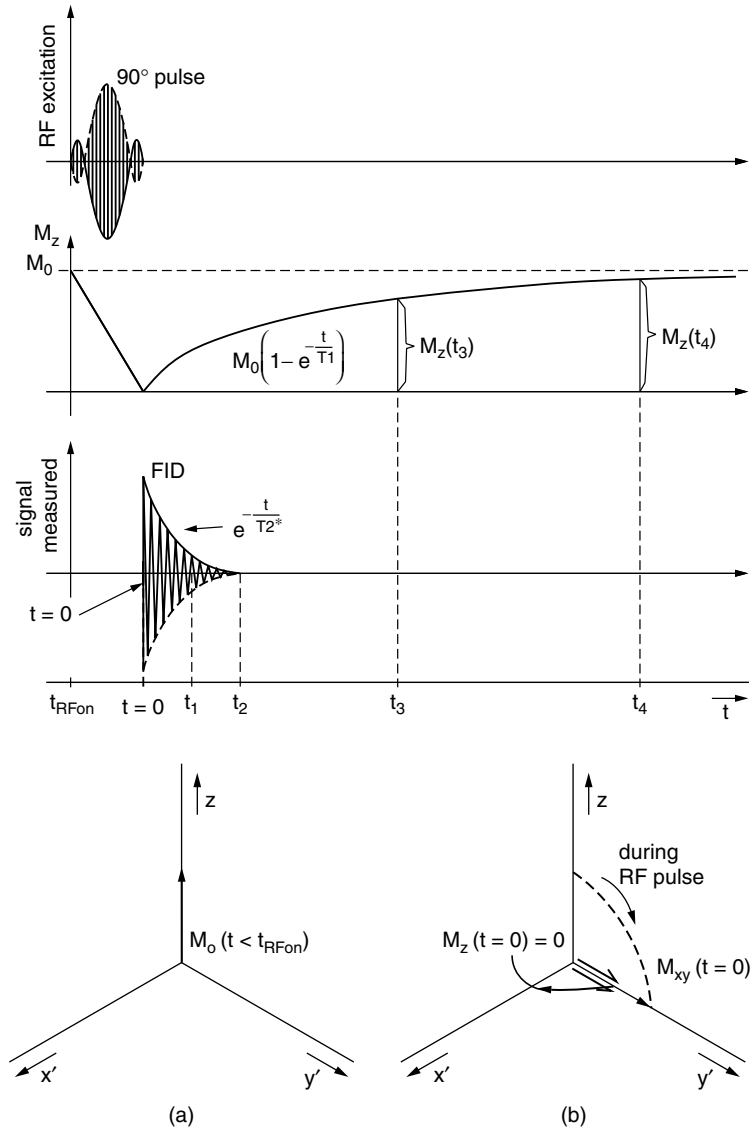


Figure 5.5 Excitation and relaxation phases. Top: timing of the sequence. Bottom: (a) initial equilibrium state, (b) result of a 90° RF impulse, (c) after a partial dephasing, (d) components fully dephased, (e) partial recovery of *z*-component, and (f) almost initial situation.

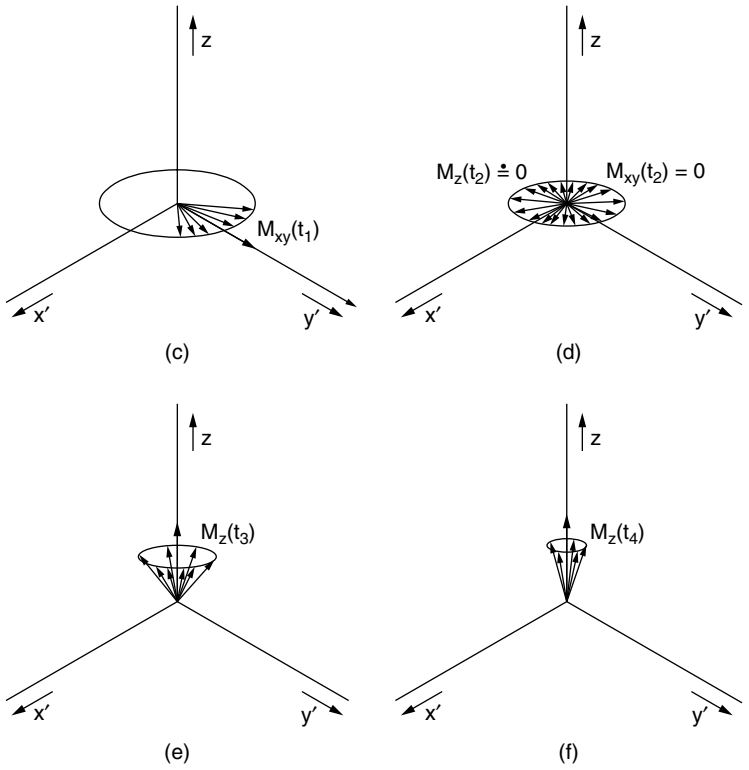


Figure 5.5 (Continued).

decay speed of the signal. The resulting shorter time constant that can be directly determined from the FID signal envelope is denoted $T2^*$. It comprises both mentioned influences according to

$$\frac{1}{T2^*} = \frac{1}{T2} + \gamma \Delta B, \tag{5.6}$$

where ΔB describes the field inhomogeneity having a certain (Lorentzian) expected distribution. Thus, the desired $T2$ value cannot be obtained straight from the FID signal decay, and special means must be employed to separate it from other factors (see below in Section 5.2).

The following, usually longer, phase of total relaxation from the excited state corresponds to finishing the recovery of the original thermodynamic equilibrium, i.e., restoring the original ratio of positively and negatively oriented spins in the analyzed volume.

This is accomplished mostly by interaction with fluctuating magnetic fields of surrounding particles forming the structure of the volume, the lattice*; therefore, it is called *spin–lattice relaxation*, or $T1$ *relaxation*. It is also an exponentially decaying process, of which the time constant $T1$ is again a characteristic parameter of the object matter, discriminating types of tissues rather well. Its values for different biological tissues are in the approximate range of 240 to 810 msec, i.e., about 5 to 10 times longer than $T2$ and much longer than $T2^*$. This is also the reason why $T1$ mechanisms do not influence the FID signal substantially, though they are naturally running during the whole relaxation time since the RF irradiation. The magnetization during the second phase of relaxation is almost purely longitudinal—in the z-direction. It is obvious that direct determination of $T1$— from the FID signal is not possible.

5.1.3.1 Chemical Shift and Flow Influence

According to Equation 5.1, nuclei resonate at the frequency determined by their gyromagnetic ratio γ and the magnitude of the external magnetic field B. Although γ is naturally identical for all hydrogen protons, these resonate at slightly different frequencies when they are a part of molecules. The reason is in the phenomena of magnetic shielding and de-shielding provided by electrons circulating in the molecules and generating a secondary magnetic field so that the net field around a proton is modified. This causes Larmor frequency shifts that are characteristic for concrete chemical compounds, sc., *chemical shifts*. As the secondary fields are proportional to the external field B_0, so are the chemical shifts. They can be discovered and analyzed by detailed spectral analysis of the MR signals. The chemical shifts are what is measured in MR spectroscopy. New directions in MRI—*chemical shift imaging* or *spectroscopic imaging*—include the chemical composition among the imaged parameters, see, e.g., Chapter 13 of [10].

In conventional MR images, the chemical shifts can manifest themselves by undesired localizing artifacts. As the localization is based on frequency encoding (see below in Section 5.4), the small frequency shifts may lead to a corresponding position shift of objects of a certain chemical composition with respect to the background of another material; the amount of the phenomenon depends on data acquisition and processing methods.

Lattice is a traditional term; nevertheless, the $T1$ relaxation runs similarly in liquids or amorphous matter.

Flow in the analyzed object, namely, blood flow, can change the appearance of the respective volumes, depending on the direction of flow (through slice, in slice) and velocity of flow. This may lead to both flow-related enhancement and void of the flow areas, e.g., vessels. The result is also dependent on the method of MR response acquisition. Detailed analysis and applications of these phenomena in MR angiography (MRA) are, however, beyond the scope of this book and should be found in the specialized literature, e.g., Chapter 12 of [10] and Chapter 6.6 of [34].

5.2 RESPONSE MEASUREMENT AND INTERPRETATION

From the previous explanation, it follows that a single FID signal cannot be used directly to determine parameters of the analyzed tissue other than the proton density D_p, which would mostly provide only a low contrast in biomedical images. Hence, it is desirable to base the imaging also on $T1$, $T2$, or $T2^*$ parameters, each of which gives a different tissue contrast. A suitable weighting of these parameters in the resulting image, i.e., combining the image pixel values from all the parameters with the relative weights influenced by the user, may disclose details otherwise invisible. The image, in the contrast of which a concrete parameter p prevails, is called *p-weighted*; thus, the imaging then may be called $T1$-, $T2$-, or density-weighted imaging.

Sophisticated *RF pulse sequences* and corresponding signal analysis must be applied to provide such data for a given tissue; many modifications of several basic approaches are routinely used. The fundamental principle for obtaining a measurable signal difference (and consequently the image contrast) due to differences in $T1$ or $T2^*$ values, respectively, can be seen in Figure 5.6: when a suitable time instant after the excitation is chosen for the measurements, a distinct difference between a slowly recovering (or decaying) function and the faster-changing function appears. The optimal measurement lag for each particular measurement, $t_{opt(T1)}$ or $t_{opt(T2^*)}$ after the RF pulse, can be easily derived for particular values of the respective relaxation time constants of the materials to be distinguished; approximately, it should be about the average of both individual time constants in case of a double-tissue scene. Some of the most common measuring sequences, providing image contrast based on this principle, will be described in the following paragraphs.

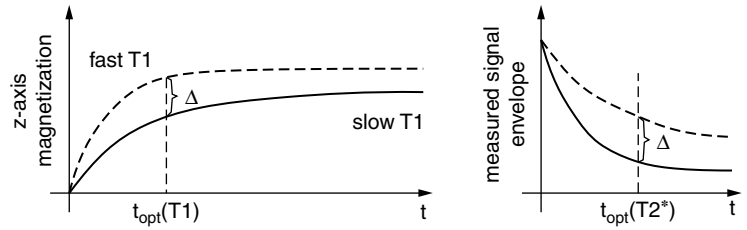


Figure 5.6 Principle of obtaining $T1$ or $T2^*$ contrast by means of delayed measurement: (Left) $T1$ contrast and (Right) $T2^*$ contrast (different timescale in each figure).

It is perhaps useful to repeat here that a 90° RF pulse effectively turns any *instantaneous* z-oriented magnetization $\mathbf{M}_z$ into the (x', y')-plane as a new vector $\mathbf{M}_{xy}$ of the same magnitude, thus enabling the measurement of the immediately preceding magnitude M_z. At the same time, any component of previously existing $\mathbf{M}_{xy}$, perpendicular in the rotating frame to $\mathbf{B}_{RF}$, is turned to the z-axis, so that it ceases to induce any response.

On the other hand, a 180° pulse only reverses the orientation of any z-axis component so that this remains undetectable. It also reverses the perpendicular-to-$\mathbf{B}_{RF}$ component of any existing $\mathbf{M}_{xy}$, thus reversing the angle of $\mathbf{M}_{xy}$ with respect to $\mathbf{B}_{RF}$ orientation to the opposite. Note that such an angle reversal in the rotating frame means change of sign of phase difference between $\mathbf{M}_{xy}$ and $\mathbf{B}_{RF}$ — a lagging vector becomes accelerated and vice versa.

5.2.1 Saturation Recovery (SR) Techniques

Information related to $T1$ can be obtained from periodically repeated FID signals induced by 90° RF pulses*, with the repetition period comparable to $T1$. Each such pulse leads to a FID signal that obviously fully decays before the next pulse. When the repetition period of irradiation T_R is several times longer than the $T1$-relaxation time constant of any of the analyzed tissues, the initial situation of $\mathbf{M}_z$ in

*The periodicity may be utilized, e.g., for improving signal-to-noise ratio by averaging. In principle, a mere pair of pulses can provide the same information, though a stable value of the maximum response is reached only after several periods.

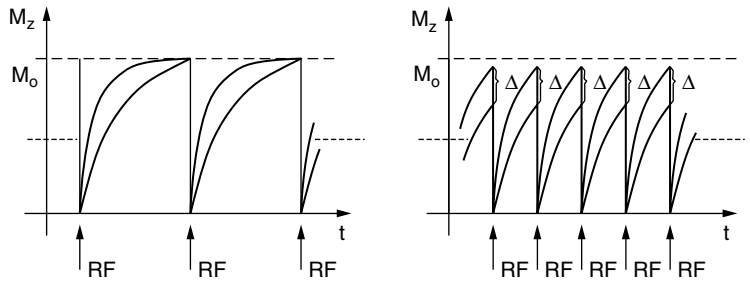


Figure 5.7 Successive application of excitation 90° RF pulses: (Left) repetition period enabling almost full recovery in both compared substances, and (right) optimum-contrast repetition period.

each period is almost entirely restored by thermal equilibrium; there is almost no difference in the measured signals from compared tissues, and consequently, the response amplitude is determined completely by the proton density — see Figure 5.7a, where recovery curves for two tissues differing in $T1$ are plotted. If, nevertheless, the period is shorter, the z-axis magnetization $\mathbf{M}_z$ before the next pulse (and consequently the measured signal amplitude) depends also on $T1$ of the concrete matter (Figure 5.7b). The measured contrast for two voxels differing in $T1$ thus depends on *both* the proton densities of the compared volumes and the $T1$-relaxation times of them. The choice of the repetition period T_R obviously influences the relation between both parameters (weighting). The $T1$-contrast for a particular pair of tissues, obviously proportional to the difference between the reached prepulse z-magnetization in each of both matters, can be optimized by choosing a proper T_R (as already mentioned, it can be shown that the optimum is near to the average of $T1$ values of both tissues). Nevertheless, the techniques analyzing directly the FID signal amplitude, as here, are nowadays rarely used.

A modification of the previous technique (*inversion recovery* (IR); Figure 5.8) starts with a 180° pulse that reverses the instantaneous $\mathbf{M}_z$ into the opposite direction, thus providing twice as large an extent for the recovery (i.e., from $-\mathbf{M}_z$ to $+\mathbf{M}_z$), making the measurement more sensitive to $T1$. After a chosen time t_R, the instantaneous amplitude of $\mathbf{M}_z(t)$ in each tissue type enables estimation of the different rate of recovery. To measure the amplitudes, a 90° pulse is applied, turning the remaining z-axis magnetization into the (x', y')-plane. Note that

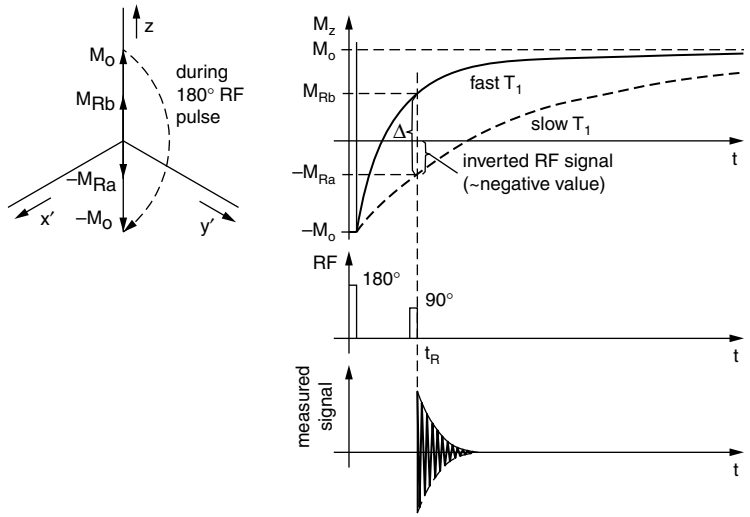


Figure 5.8 Principle of inversion recovery technique to provide $T1$ contrast.

the sign of each individual measured response should be respected; if not, the contrast may be ambiguous.

5.2.2 Spin-Echo Techniques

The RF energy–induced spin-echo technique that exists in many modifications can provide more information on the analyzed volume; primarily, it enables a good estimation of the $T2$ parameter, or it can provide a $T2$-weighted image contrast. Besides that, $T1$ may also be estimated, based on a principle similar to that described in the previous section. The basic signal sequence of this approach is depicted in Figure 5.9. It consists of an excitation 90° RF pulse, followed naturally by the induced FID response, which is usually ignored. After a certain time $TE/2$ of the order of several $T2^*$, but usually less than $T2$, a 180° RF pulse is applied, the purpose of which is to change the phases of $\mathbf{M}_{xy}$ components in such a way that the already dephased components will again acquire the perfect coherence in time TE (see explanation below). This will lead to a new appearance of the already decayed signal around the time TE, with the maximum at this instant. This phenomenon is called *spin echo* (SE).

The explanation of this phenomenon should be followed in Figure 5.10 describing *Hahn's technique* [53]. After the 90° RF pulse with $\mathbf{B}_{RF}$ along, e.g., the x'-axis, the precession of nuclei in partial

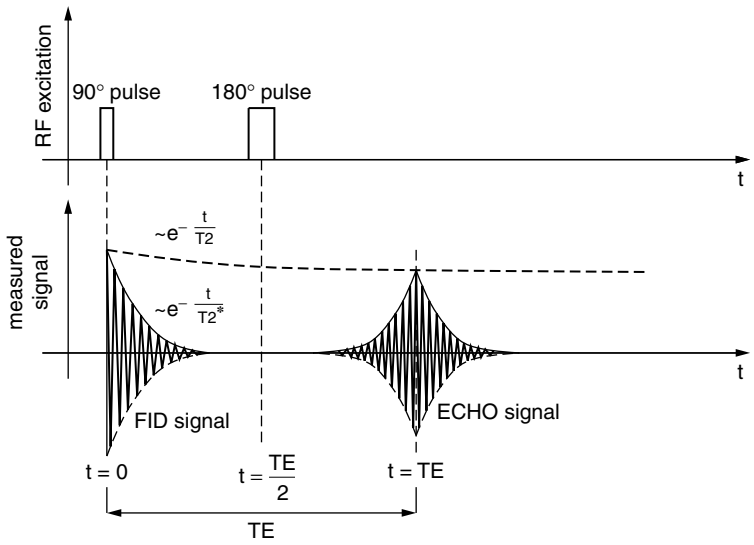


Figure 5.9 Basic spin-echo sequence of signals.

subvolumes* is coherent so that the amplitude M_{xy} of $\mathbf{M}_{xy}$ in the y'-direction is maximum (panel a). Due to inhomogeneities of the magnetic field caused by imperfectly homogeneous $\mathbf{B}_0$ and by space variance of susceptibility, individual subvolume magnetization vectors are precessing at slightly different Larmor frequencies, which leads to the already mentioned fast dephasing, and hence a decrease in M_{xy} and consequential decay of the FID signal (panel b). The following application of a 180° RF pulse along the same x'-axis direction as before reverses the perpendicular (i.e., y'-direction) parts of the $\mathbf{M}_{xy}$ components (panel c) so that the faster vectors now lag and vice versa. On the difference from the previous figure (Figure 5.9), the impulse is applied here before the FID signal completely decays, in order to depict the component reversal clearly. As the speeds of vectors given by the local Larmor frequencies remain unchanged, it is obvious that, after the time interval equal to the dephasing time $TE/2$, all the partial vectors will again reach the state of coherence (i.e., at $t = TE$), when the induced echo signal amplitude will be maximal, consequently decaying again.

*In the literature, it is often spoken of as spins, nuclei, or protons. In light of the previous explanation, the partial magnetization vectors with arbitrary flip angle α cannot belong to individual nuclei, but rather to large sets of them, contained in isochromatic subvolumes.

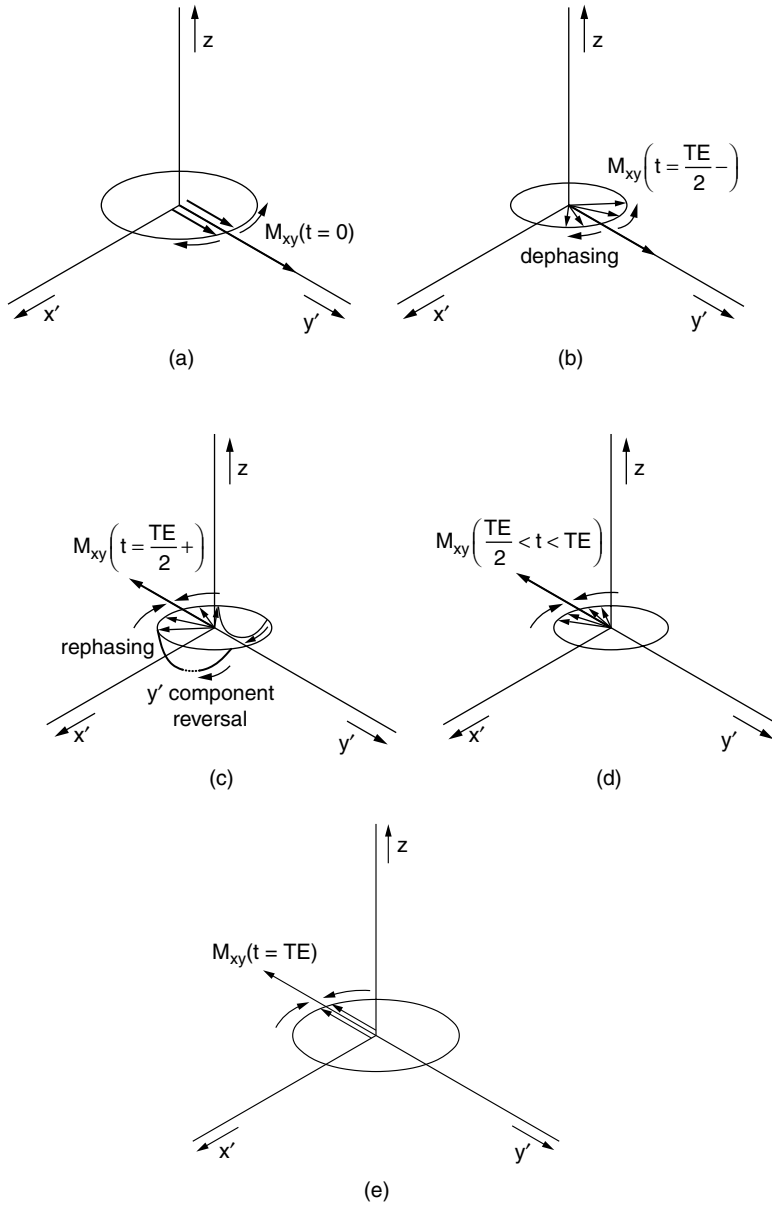


Figure 5.10 Explanation of Hahn spin-echo technique (compare with Figure 5.5): (a) result of a 90° RF impulse, (b) after partial dephasing, (c) result of a 180° RF pulse, (d) partially rephased spins, and (e) maximally rephased situation at $t = TE$.

This mechanism can of course compensate only for the mentioned influences that are deterministic and temporally invariable in the frame of a subvolume during the period of measurement; it cannot compensate for the influence of those $T2^*$ cofactors that are originated by random mechanisms, i.e., $T2$ factors. Consequently, the echo maximal amplitude will be lower than the original maximum of the FID signal due to $T2$ relaxation. The comparison of these two amplitudes may thus provide a clue on $T2$ when the time difference TE is taken into account. Note that dephasing starts after the instant $t = TE$ again, and the signal decays then according to $T2^*$.

Figure 5.11 shows an alternative of the spin-echo approach, sc., *CPMG* technique (Carr–Purcell and Meiboom–Gill). The difference consists of applying the 180° RF pulse perpendicularly to the excitation pulse, i.e., along the y'-axis in the rotating frame, which obviously corresponds to a 90° phase shift in the transmitter current. Now, the x'-components are flipped and, as seen from the figure, the instantaneous coherence is reached similarly as in the previous case, but while the echo signal was reversed (phase shifted by 180°) in Hahn's approach, now it remains in phase. Let us add that the CPMG sequence usually continues with more 180° pulses (see below).

In both above-mentioned SE techniques, the 180° pulse can be repeated several times with the period TE, providing multiple echoes, the peak values of which follow the $T2$ relaxation curve (Figure 5.12), thus enabling a better estimate of the $T2$ time constant, or differently weighted $T2$ contrast. This approach can be generalized: the 180° pulses may be placed on the time axis rather arbitrarily, this way providing the echoes in any desired time instants.

If the $T1$ time constant should be estimated by the SE technique, more excitations per subvolume are necessary, as already seen in the SR technique. Then, the repetition period TR determines the share of $T1$ on the resulting echo contrast. While a very long TR allows a complete recovery of equilibrium and therefore the influence of $T1$ vanishes, a shorter TR, chosen properly (about comparable to the average of the expected values of $T1$), may enable a good contrast in a needed range of $T1$ values.

The time arrangement of the SE techniques is very flexible and allows for differently weighted material parameters (D_p, $T1$, $T2$) expressed in a combination by pixel brightness in the resulting image. Image data may be simply derived from the maximum of echo signal at $t = TE$ (or sometimes also from other higher-order, i.e., later, echoes), but usually the complete course of the echo is recorded in order to localize the response sources (see Section 5.4). Obviously, while a short

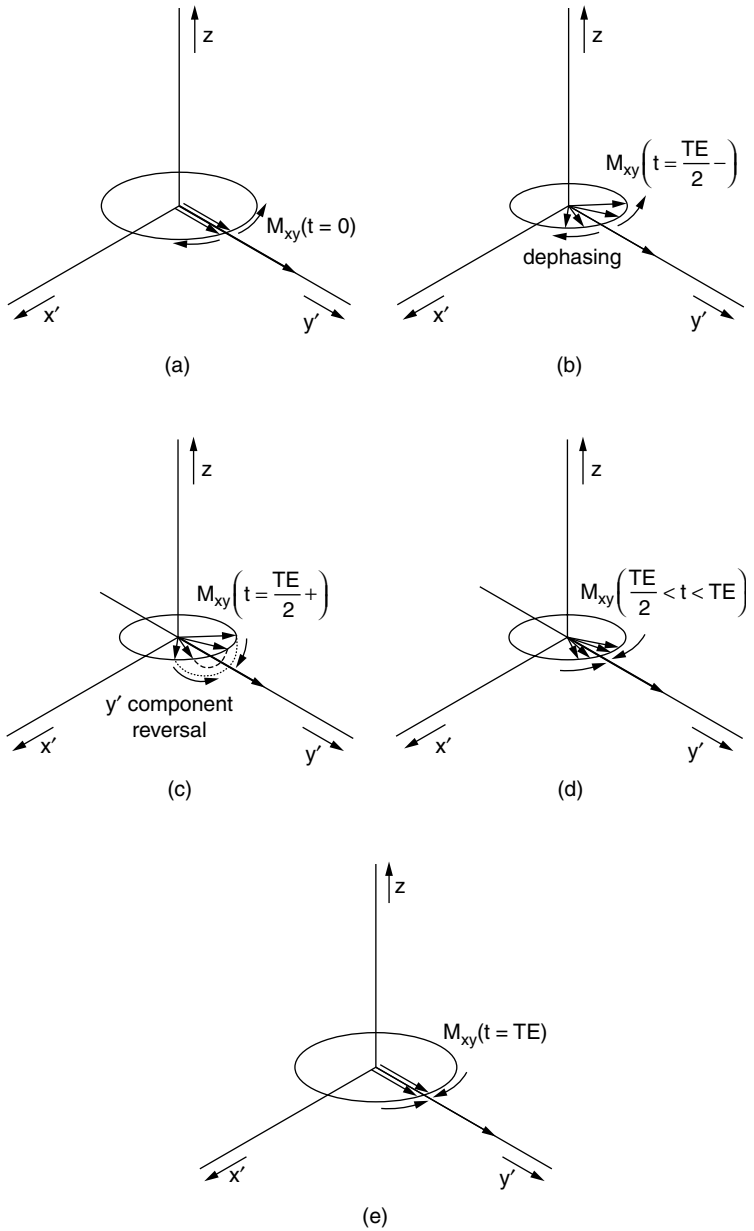


Figure 5.11 Explanation of the CPMG spin-echo technique (compare to Figure 5.10).

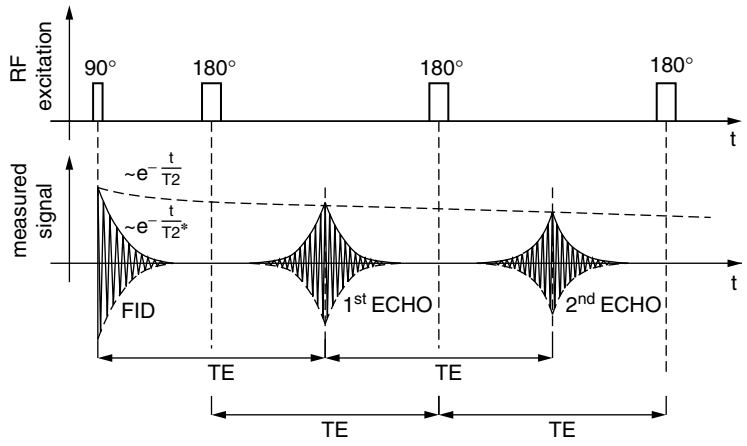


Figure 5.12 Repeated application of 180° pulse revealing the $T2$ decay.

TE will lead to only a small influence of $T2$ on the echo amplitude, thus suppressing the $T2$ contrast, a longer TE emphasizes the $T2$ contrast as the differences in $T2$ attenuation become more pronounced. The overall intensity of any response always depends, of course, linearly on the proton density D_p. We can therefore conclude that:

- The combination of a long TR ($TR >> T1$) and a short TE ($TE << T2$) leads to a contrast reflecting primarily the differences in proton density — resulting in *spin density-weighted images*.
- A long TR ($TR >> T1$) and a proper TE ($TE \approx T2$) emphasize the influence of the $T2$ parameter, which often prevails by large — resulting in *T2-weighted images*.
- On the other hand, a proper TR ($TR \approx T1$) and a short TE ($TE << T2$) accentuate the $T1$ share in the contrast — resulting in *T1-weighted images*.

Naturally, any compromising combination of TR and TE in this most frequently used technique is possible.

An even more flexible modification of this technique using three separated 90° pulses is called *stimulated echo* (see [10]).

5.2.3 Gradient-Echo Techniques

The gradient-echo (GE) techniques should also be mentioned among measuring sequences, though they are closely connected with the

necessity to apply magnetic gradient fields before and during readout of the RF signals; the gradients are otherwise also applied to localize sources of the response components (see Section 5.4). The gradient-echo signal is rather different from the spin echo; namely, it cannot separate the $T2$ value. Its time-course corresponds to an approximately mirrored FID signal, though faster increasing and decaying than in the spontaneous case; consequently, the measuring times may be shorter.

The basic time sequence is depicted in Figure 5.13. After the excitation 90° pulse, an additional spatially linearly increasing

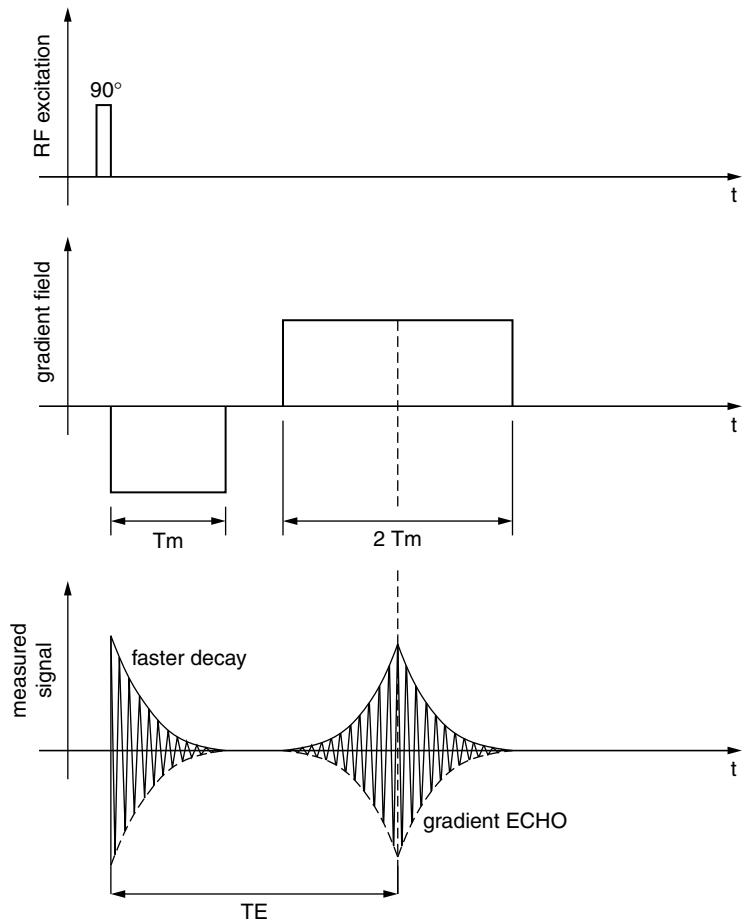


Figure 5.13 Basic principle of the gradient-echo technique.

magnetic field of the same direction as the static field (the *gradient field*) is switched on for a period T_m, causing a fast dephasing of the M_{xy} magnetization due to rather large differences in Larmor frequency throughout the measured subvolumes. Obviously, the stronger the gradient, the faster is the dephasing. When the opposite additional magnetic field of the same strength, forming the *readout gradient*, is switched on consecutively, rephasing occurs with the same speed as the previous dephasing, as the acceleration in ω_L for a particle is replaced by deceleration of the same magnitude and vice versa. Thus, neglecting other sources of dephasing (this is possible thanks to short involved times), the coherency will be regained after the readout gradient has been switched on for the time T_m, when the signal reaches its peak value almost equal to the FID signal maximum. The signal then decays again with the same time constant during the next T_m period, with the readout gradient still on. This principle, providing the fast echo response controlled by the localizing gradient, is indispensable namely in MRI fast imaging (see Section 5.4.7).

5.3 BASIC MRI ARRANGEMENT

To utilize the MR phenomena for imaging, an arrangement (MRI system) must be assembled that provides the necessary magnetic fields and RF excitation, enables response measurement, and also allows spatial localization of the response components. The contemporary MRI systems use interesting high-technology solutions based on long-term physical and technological research and development. For the purpose of this book, we do not need to go into any construction details; all we need is to understand how the involved fields are arranged with respect to the imaged object and to each other, and how the signals are measured. This very basic information can be seen in Figure 5.14 and Figure 5.15.

The most visible feature of every MRI system is the source of the *main static magnetic field* $\mathbf{B}_0$. Technically, it may be an electromagnet, either superconducting or resistive, or even — for lower-intensity fields — a permanent magnet. Its purpose is to provide the field of the needed (high or very high) magnetic induction in the range of about 0.1 to 5 T, extremely homogeneous in the spatial range of the imaged area. The source of such a magnetic field is symbolized in Figure 5.14 by a horseshoe magnet; in Figure 5.15 it is represented more realistically by a couple of big coaxial coils (though a higher number of differently sized coils is used in practice to achieve the desired homogeneity). The static field thus has the

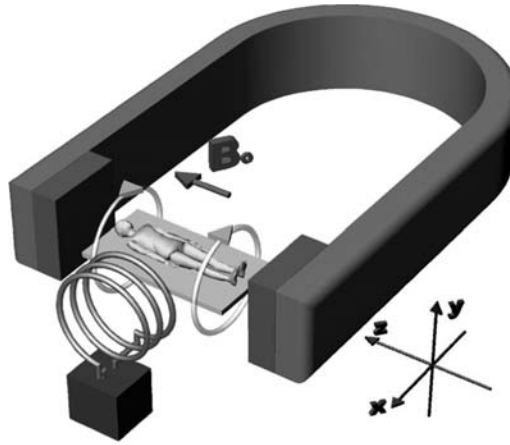


Figure 5.14 Basic principle of the MRI system arrangement.

direction of the coil axis; by convention, this agrees with the z-coordinate of the fixed frame.

Other components of the magnetic field that are time and space variable — so-called *gradient fields* — are produced by gradient coils, of which only the pair of z-gradient coils is depicted in Figure 5.14 and Figure 5.15a. On the difference to the main magnet coils, the z-gradient coils carry the current in opposite directions, thus making the field weakly linearly dependent on z when switched on. The other gradient coils (shown in Figure 5.15b) provide the auxiliary fields dependent on x and y, respectively. It is important to realize that all the gradient fields have the direction of the main field, i.e., in (or against) the z-direction, independently of the coordinate on which they depend; this requirement leads to a relatively complex design of x- and y-gradient coils*. The purpose of the gradient fields and more thorough treatment of their timing is explained in Section 5.4.

Finally, there are also *RF coils* in each MRI system, situated with their axes in a (x,y)-plane transversal to the main field. The coils are often used alternately for transmitting and receiving. The RF

*Strictly taken, the exact solution of the Maxwell equations shows that the desirable gradients are always accompanied by concomitant gradient fields with x- and y-components; analysis of the consequential distortion is beyond the frame of this book (see, e.g., [37]).

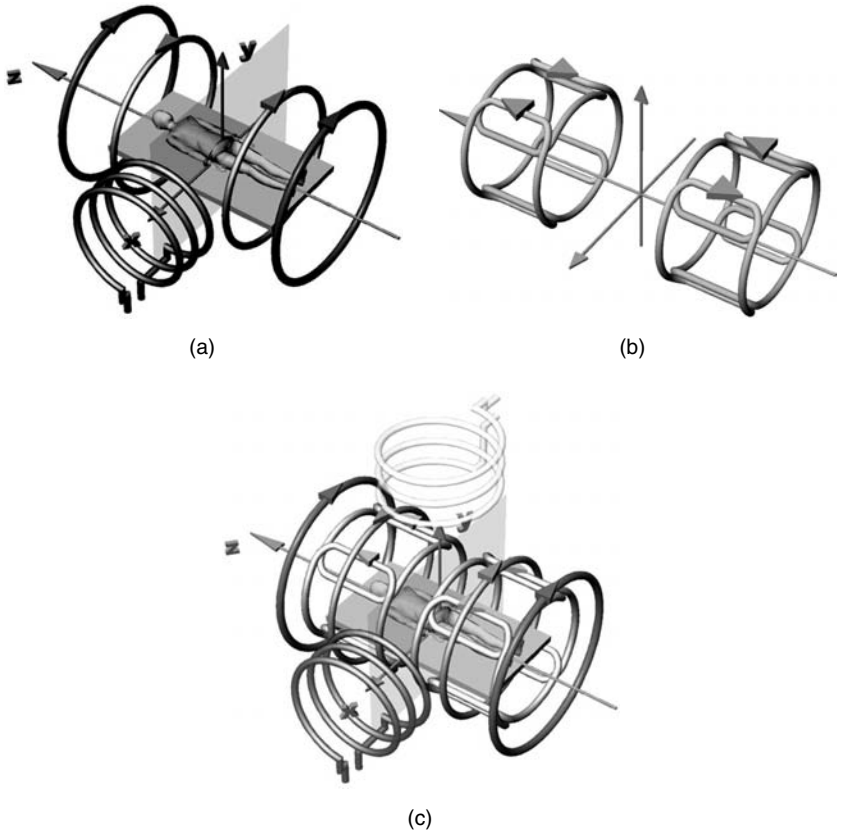


Figure 5.15 More detailed principle of the MRI system: (a) enlarged detail from the previous figure, with outer superconducting coils in place of the magnet; (b) perpendicular couples of x- and y-gradient saddle coils separately; and (c) complete arrangement (schematically).

subsystem usually consists of a couple of perpendicular coils (Figure 5.15c) connected to the RF transmitter/receiver, with one of the coils connected via a 90° phase delay. Each of the coils may be divided into halves symmetrical with respect to the object axis (not shown). This way, the couple serves as a circularly polarized antenna, optimal for both MR excitation and signal reception.

Although exceptions exist, especially in permanent magnet systems, the figure shows the common arrangement where the patient lies on a table with the long axis of his or her body along the z-coordinate. The (x, y)-plane for a particular value of z then

may be the two-dimensional image plane, also called a *slice* (but not necessarily; later we shall see that the slice may be arbitrarily oriented). Multiplanar (multislice) imaging can then supply three-dimensional image data.

5.4 LOCALIZATION AND RECONSTRUCTION OF IMAGE DATA

The RF signal measured in MRI is always acquired from a large volume of the imaged object; it is not feasible to focus the signal acquisition to individual pixels or voxels. It is therefore necessary to use mechanisms that would

- define the *total acquisition volume* as well as possible, and
- provide means for *detailed localization* of the acquired information inside this volume.

Both tasks can be accomplished utilizing the frequency-selective properties of the MR phenomena.

Primarily, the spins or spin sets would only be excited by the RF pulse with a frequency exactly equal to the Larmor frequency of the spins, corresponding to the instantaneous value of the local magnetic field; the spins with a different Larmor frequency remain untouched even when irradiated by the pulse. This fact is utilized for defining the total area of acquisition — no signal would arise from nonexcited parts of the object.

Secondly, once excited, the spin set magnetization precesses at the Larmor frequency, determined by the *instantaneous* value of the local magnetic field, regardless of what has been the excitation frequency. Namely, the precession frequency during the response may be (and usually is) different from the frequency of the excitation RF pulse, because the local magnetic field typically is changed between excitation and measuring periods. This fact is utilized for labeling the space elements by different frequencies of RF responses acquired from the individual volume elements, which are discriminated by differences in the local magnetic field during the measurement. It is then the task of the subsequent data processing — thus closely related to the subject of this book — to localize the sources and to assign each local measured value to a proper position in the image matrix.

5.4.1 Gradient Fields

In order to label the signal from a particular position (x, y, z), the main magnetic field is temporarily modified by auxiliary gradient

fields that can be switched on or off during a measurement (Lauterbur, 1973 [52]). Each of the three gradient fields (often called just gradients) has *the same* direction as the main field — along the *z*-axis — but its amplitude depends on only a single coordinate; this way, the local magnitudes of the static field are modified in a controlled way without changing the direction of **B** (however, see the footnote on page 199).

The effect of the three gradient fields can be linearly combined, as the MR phenomena are linear in *B*. The magnitude of each gradient field is expected to depend linearly on the respective coordinate; this is equivalent to the requirement of geometrical linearity of frequency/phase labeling of voxel positions (see below). Therefore, e.g., for the *x*-gradient field we can write

$$\mathbf{B}_{gx}(x,\ t) = G_x(t)x\mathbf{u}_z,$$

where $G_x(t)$ is the system-controlled time-variable, but space-invariant gradient magnitude (in $\mathrm{T\,m^{-1}}$), and $\mathbf{u}_z$ is the unit vector in the *z*-direction. When denoting the position vector $\mathbf{r} = [x, y, z]^T$ and the vector of magnitudes of partial gradients as $\mathbf{G}(t) = [G_x(t),\ G_y(t),\ G_z(t)]^T$, obviously the total gradient field at $\mathbf{r}$ is

$$\mathbf{G}^T(t)\,\mathbf{r}\,\mathbf{u}_z. \tag{5.7}$$

The magnitude of the total field at the location $\mathbf{r}$ can then be expressed as

$$B_z(t) = B_0 + \mathbf{G}^T(t)\,\mathbf{r}. \tag{5.8}$$

The corresponding local Larmor frequency is then $\omega_L = \gamma B_z$; as far as γ is space invariant, the frequency is obviously a localizing marker. Unfortunately, it is not unique, as in each plane, perpendicular to the gradient vector $\mathbf{G} = G_x\mathbf{u}_x + G_y\mathbf{u}_y + G_z\mathbf{u}_z$, the magnetic induction *B* has a constant value. Nevertheless, by combining time-variable gradients during a measurement and applying subsequent data processing based on advanced signal theory, it is possible to obtain a unique localization for each MR response while still relying only on the Larmor frequency sensitivity to *B*.

The orientation of the resulting gradient **G** may be chosen, according to Equation 5.7, quite arbitrarily, thus enabling different orientations of imaged slices*. For reasons of simplicity and without

*Note, once more, that the orientation of the gradient *field* is always supposed along *z*.

losing on generality, we shall often suppose in the following explanations that the imaged slice is in the (x, y)-plane for a particular z. Nevertheless, it should be understood that the results for an arbitrarily positioned slice or volume could always be easily derived by corresponding rotation of coordinates. MRI systems usually allow easy selection of the slice position and orientation, realized by linearly combining all three gradient fields via the user interface.

5.4.2 Spatially Selective Excitation

In principle, the magnetic field in the area to be excited should have a certain constant magnitude B, differing markedly from the magnetic field of other parts of the object volume. Then, when the object is irradiated by an RF pulse with the frequency $\omega = \gamma B$, only the area of interest becomes correspondingly excited. It would be difficult to generate magnetic fields of the described property for an arbitrarily shaped volume of interest, but it can be easily achieved when a planar slice shape is required.

If a gradient field is applied during a (infinitely) long RF pulse, only a (infinitesimally) thin planar layer of the object is excited, the Larmor frequency of which equals the pulse frequency. Nevertheless, a too thin layer of excited nuclei would lead to a weak MR signal with an unacceptable signal-to-noise ratio (SNR). Thus, excitation of a controlled-thickness slice is needed, which can be achieved by applying an RF pulse of a wider band of frequencies with a constant spectral density inside the band. This way, ideally a slice that is limited by two parallel planes is obtained; the distance of the planes — the thickness of the slice d — is obviously proportional to the signal frequency bandwidth $\Delta\omega$ and inversely proportional to the magnitude of gradient G,

$$d = \Delta\omega / (\gamma G). \tag{5.9}$$

Hence, there are two parameters to control the slice thickness (Figure 5.16). Naturally, only the volume under the influence of the RF coils is excited, but it can be expected that their design guarantees the complete interesting range of the object being covered evenly.

It is known from signal theory that the required band spectrum of the RF pulse corresponds to the original harmonic RF signal modulated (multiplied) by the envelope, which is of the $\sin(at)/(at)$ type, where a is inversely proportional to the width of the frequency band. Practically, this infinite-duration sinc function must be shortened by cutting its side lobes of higher order; usually only three

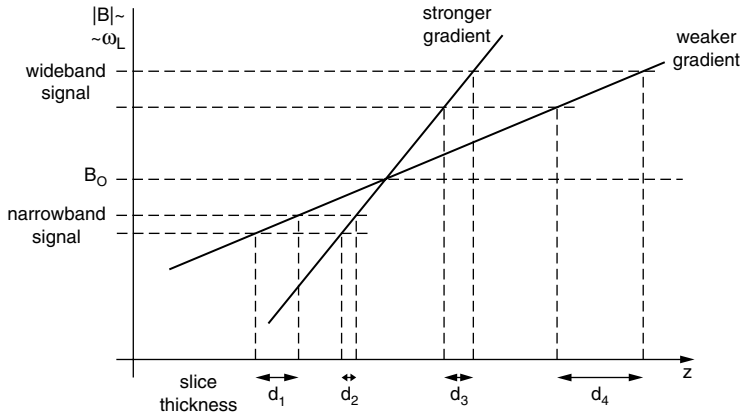


Figure 5.16 Influence of the gradient and frequency band on the slice thickness. Note also the positioning of the slice by the choice of central RF frequency.

to seven lobes are preserved in a pulse (Figure 5.17). Consequently, the frequency band of the signal is slightly uneven and its frequency limits are not quite sharp, but this can be neglected from our point of view. Thus, we may suppose for our explanation that all spin sets

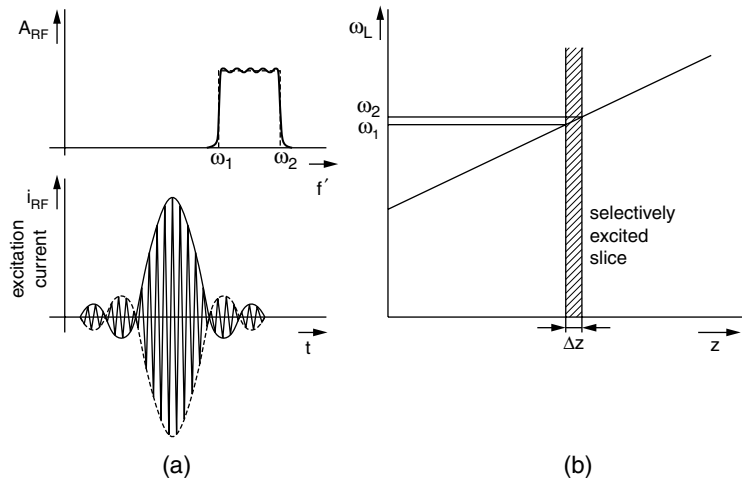


Figure 5.17 (a) Wideband RF impulse in the frequency and time domains, and (b) the defined excited slice of the thickness Δz.

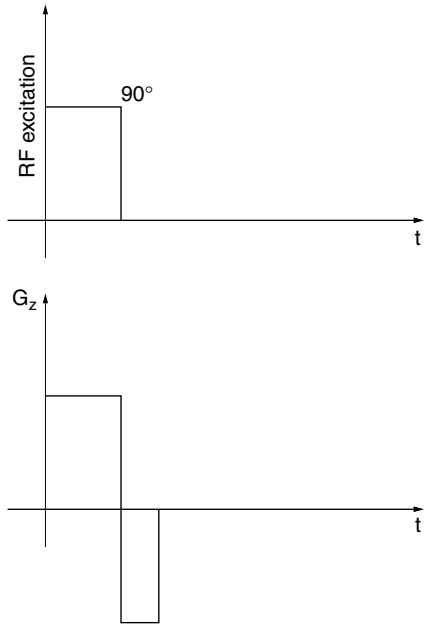


Figure 5.18 Pulse sequence producing an evenly and coherently excited slice.

in the excited volume are precessing with the same amplitude and that there is no overlap with neighboring slices.

Nevertheless, the partial spin sets are losing the phase coherence during the excitation pulse due to differing local Larmor frequencies, as the consequence of the excitation-selective gradient. To restore the full coherence state, the opposite gradient is applied after the RF pulse (Figure 5.18), of either half amplitude or half duration, which accelerates the delayed spins and retards the forwarding ones. After this *refocusing pulse* (sometimes called *balancing pulse*), we can suppose that all the spins in the excited volume are precessing in phase.

This way, it is possible to excite evenly a defined slice-volume. Note that the orientation of the slice border planes may be arbitrary, as it is controlled by the resulting gradient direction according to Equation 5.7. This is an advantage of MRI in comparison with computed tomography (CT) or single-photon emission computed tomography (SPECT) imaging, where the slice orientation is given by the system construction.

Nevertheless, most often, slices in the (x, y)-plane are scanned and therefore excited under the influence of the z-gradient field only.

The thickness of the excited slice determines the geometrical resolution perpendicular to the slice plane in two-dimensional imaging, when no further resolving mechanism in this direction is applied; then the practically used thickness values are in the range of a few millimeters. When the measurement is to provide three-dimensional data, the excited slice thickness may be substantially bigger, which requires a wide frequency band (consequently a short irradiating RF pulse) and a weak gradient during excitation.

5.4.3 RF Signal Model and General Background for Localization

In this section we shall derive a model of the signal received by the RF coil during the receiving period after some previous excitation, i.e., during readout time(s). We shall see that this signal carries the positional information on its components, which are related to magnetization of subvolumes. When arranging the sequence of excitation and gradient pulses properly, it will be possible to reconstruct the two-dimensional or three-dimensional scene from the MR signals, using the multidimensional signal theory.

Let us model the signal under a quite general situation of an arbitrary initial transversal magnetization in consequence of some previous excitation. Thus, the initial imaged scene consists of a three-dimensional distribution of the initial transversal magnetization $\mathbf{M}_{xy0}(\mathbf{r})$, with $\mathbf{r} = (x, y, z)^T$ being the positional vector. Due to previously described precession phenomena, the vectors of magnetization rotate in the (x, y)-plane with the local Larmor frequency, generally different from the excitation frequency ω_L. Starting from the initial state, the magnetization field $\mathbf{M}_{xy}(\mathbf{r}, t)$ is time-variable: besides the mentioned rotation of the vectors, the amplitude $M_{xy}(\mathbf{r}, t) = |\mathbf{M}_{xy}(\mathbf{r}, t)|$ of the local transversal magnetization at $\mathbf{r}$ is time dependent due to $T1$, $T2$, or $T2^*$ relaxation*, or due to inserted excitation, leading possibly to echo phenomena. From the signal model point of view, there is no need to analyze the character of the time variance.

*Note that $T2^*$ includes only dephasing due to inhomogeneity of the main field, but *excludes* the gradient-induced dephasing, the influence of which will be incorporated later in the derivation.

The magnetization is sensed by the RF coil, positioned, say, along the x-axis as in Figure 5.15a. When it is irradiated by the magnetic field oscillating due to MR precession, it produces a signal voltage $s(t)$, proportional to time derivative of the transversal magnetic flux Φ_x through the coil,

$$s(t) \propto \frac{\mathrm{d}}{\mathrm{d}t}\Phi_x(t). \tag{5.10}$$

Therefore, the flux should be determined.

The rotating transversal magnetization vector $\mathbf{M}_{xy}(\mathbf{r}, t)$ can be represented by the phasor*

$$M_{xy}(\mathbf{r}, t)\, e^{-j\varphi(\mathbf{r},t)} = M_{xy}(\mathbf{r}, t)\, e^{-j\omega_0 t} e^{-j\Delta\varphi(\mathbf{r},t)}, \tag{5.11}$$

where the phase $\varphi(\mathbf{r}, t)$ may be interpreted as the instantaneous position angle of $\mathbf{M}_{xy}(\mathbf{r}, t)$ with respect to (x, y)-coordinates, consisting of two components: the linearly fast increasing phase $\omega_0 t$ given by the reference (excitation) frequency ω_0 and the phase difference $\Delta\varphi(\mathbf{r}, t)$. This variable phase difference may be well visualized in the rotating coordinates (x', y', z) as the angle between the vector $\mathbf{M}_{xy}(\mathbf{r}, t)$ and the reference vector, corresponding to the reference signal with frequency ω_0 coherent with the excitation RF pulse. Obviously, the phase $\varphi(\mathbf{r}, t)$ is the total phase accumulated since a time origin $t = 0$, which may be chosen arbitrarily. Once the chosen volume is excited, the spin areas (isochromats) are precessing under the influence of the (generally time-variable) local magnetic field; the instantaneous frequency at the position $\mathbf{r} = (x, y, z)^T$ is, according to Equations 5.1 and 5.8,

$$\omega(\mathbf{r}, t) = \gamma B_z(\mathbf{r}, t) = \gamma\left(B_0 + \mathbf{G}^T(t)\mathbf{r}\right) = \omega_0 + \gamma\mathbf{G}^T(t)\,\mathbf{r}, \tag{5.12}$$

providing that γ is space-invariant. The total phase since the chosen moment $t = 0$ is the integral of frequency,

$$\varphi(\mathbf{r}, t) = \int_0^t \omega(\mathbf{r}, \tau)\,\mathrm{d}\tau = \omega_0 t + \left[\gamma\int_0^t \mathbf{G}^T(\tau)\,\mathrm{d}\tau\right]\mathbf{r}. \tag{5.13}$$

*Real and imaginary components correspond to the x, y-components of the vector.

Here, the quantity in brackets has been introduced by Mansfield (1977) [54] as the so-called $\mathbf{k}(t)$-*term*:

$$\mathbf{k}(t) = \gamma \int_0^t \mathbf{G}^T(\tau)\, \mathrm{d}\tau\,. \tag{5.14}$$

Note that it is generally a three-dimensional quantity in the form of a row (transposed) vector.

Note further that the phasor (Equation 5.11) may also be considered a representation of a real-valued time-variable quantity using the approach, standard in signal theory*. The flux component $\mathrm{d}\Phi_x(t)$, due to excited spins in the infinitesimal volume $\mathrm{d}V_\mathbf{r}$, is a quantity of this kind and, being proportional to the x-component of the transversal magnetization $\mathbf{M}_{xy}(\mathbf{r}, t)$, it may be described as

$$\mathrm{d}\Phi_x(t) \propto M_{xy}(\mathbf{r}, t) e^{-j\varphi(\mathbf{r},t)} \mathrm{d}V_\mathbf{r} = M_{xy}(\mathbf{r}, t)\; e^{-j(\omega_0 t + \Delta\varphi(\mathbf{r},t))} \mathrm{d}V_\mathbf{r}\,. \tag{5.15}$$

The total flux Φ_x through the RF coil is obviously given by the spatial integral in the region V of coil sensitivity (supposedly even in the field of view (FOV)), so that

$$\Phi_x(t) \propto \iiint_V M_{xy}(\mathbf{r}, t)\, e^{-j\varphi(\mathbf{r},t)} \mathrm{d}V_\mathbf{r} \tag{5.16}$$

and according to Equation 5.10, the received signal is**

$$s(t) \propto \frac{\mathrm{d}}{\mathrm{d}t} \iiint_V M_{xy}(\mathbf{r}, t)\, e^{-j\varphi(\mathbf{r},t)} \mathrm{d}V_\mathbf{r}. \tag{5.17}$$

The time variabilities of M_{xy} and of the $\mathbf{k}$-term are orders slower than that of $e^{-j\omega_0 t}$; therefore, the differentiation can be well approximated by a mere multiplication by $-j\omega_0$, which does not change the proportionality and can be omitted. We thus obtain

$$s(t) \propto \iiint_V M_{xy}(\mathbf{r}, t)\, e^{-j\omega_0 t} e^{-j\left[\gamma \int_0^t \mathbf{G}^T(\tau)\, \mathrm{d}\tau\right]\mathbf{r}} \mathrm{d}V_\mathbf{r}. \tag{5.18}$$

*The phasor $e^{-j\omega t}$ may be considered a component of the real signal $(e^{-j\omega t} + e^{+j\omega t})/2$ according to Euler's relation; thus, e.g., $2\mathrm{Re}(e^{-j\omega t})$ corresponds to the real signal.

**The magnitude of the signal depends on the RF coil geometry and other constant factors; nevertheless, the image contrast is fully determined by the *relative* signal dynamics.

Note that it is a complex signal. Though the measured RF signal is usually real-valued, the information on the phase relation with respect to the reference signal is recovered in the following step.

The first step in signal processing is demodulation (i.e., the frequency shift via signal multiplication by $e^{j\omega_0 t}$); the result is, as may be seen, equivalent to measuring the magnetization M_{xy} in the rotating frame. When substituting the **k**-factor for the bracketed term, we have the demodulated signal described by

$$S(t) \propto \iiint_V M_{xy}(\mathbf{r}, t)\, e^{-j\mathbf{k}(t)\,\mathbf{r}} \mathrm{d}V_{\mathbf{r}} \tag{5.19}$$

and simply expanding the scalar product and rewriting the vectors, we finally arrive at

$$S(t) \propto \iiint_V M_{xy}(x, y, z, t)\, e^{-j(k_x(t)x + k_y(t)y + k_z(t)z)} \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\,. \tag{5.20}$$

This expression obviously has the form of three-dimensional Fourier transform (FT) (Section 1.2), where k_n can be interpreted as the circular spatial frequency (in rad/m) in the direction n. The k-space is thus the spectral (spatial-frequency) domain.

The derivation leads to a *conclusion of fundamental importance*: the demodulated MR signal at the time instant t is (disregarding a constant factor) equal to a single value of the three-dimensional Fourier transform of the *instantaneous* magnetization distribution within the excited volume, at space frequencies given by $\mathbf{k}(t)$; hence

$$S(t) \propto \mathrm{FT}_{3D}\{M_{xy}(\mathbf{r}, t)\}\,|\,(\mathbf{k}(t))\,. \tag{5.21}$$

This is a very general signal model that can be applied to any MRI method and any measuring sequence. The formula Equation 5.21 indicates that the signal values should be gradually assigned to the k-space, where they form individual values of the spectral representation of the imaged scene; the trajectory $\mathbf{k}(t)$ in the spectral space is determined by the gradient fields applied during readout. In order to provide sufficient information for the following reconstruction of the image data by the (generally) three-dimensional inverse Fourier transform, the k-space must be filled with the $S(t)$ values rather densely during the measurement, which obviously requires a certain

time. Strictly taken, each imaged scene is time variable to a certain extent so that each value $S(t)$ corresponds to a different scene; thus, each spectral value belongs to a different original function. In other words, the obtained k-space representation is not a spectrum of a single original image, but rather a mixture of spectral values of different images generated during the time development. However, though Equation 5.21 is not limited to static scenes, these are mostly what is to be imaged.

The instantaneous magnetization distribution is given by proton density distribution and its instantaneous relaxation state, and therefore is dependent on the used measurement sequence. By selecting the sequence, differently weighted image contrast can be obtained (see Section 5.2). Of course, a complete spatial distribution of the magnetization $M_{xy}(\mathbf{r}, t)$ for a particular t, thus a static scene, should usually be imaged; there are basically two possibilities to achieve this. The conceptually simpler approach is that the complete measurement is accomplished (after a single excitation) so fast that $M_{xy}(\mathbf{r}, t)$ is practically time invariant during the readout time of the signal. Thus, the complete spectrum is then measured (i.e., the complete k-space scanned) during a very short period, as in echo-planar imaging (EPI) (Section 5.4.7). Naturally, the measurement takes a certain time and the time invariance is then only an approximation. The other, conceptually more complicated but easier-to-implement approach is to repeat the RF excitation in such a way that exactly the same distribution $M_{xy}(\mathbf{r})$ is repeatedly generated before each readout time, during which a different small part of the spectrum is always provided; i.e., new positions of the k-space are written to. Every readout time must be short enough to preserve the time invariance.

In both cases, the previous expression can be simplified to the static image case as

$$S(t) \propto \mathrm{FT}_{3D}\{M_{xy}(\mathbf{r})\}\,|\,(\mathbf{k}(t)) = \breve{M}_{xz}(\mathbf{k}(t)), \tag{5.22}$$

where $\breve{M}_{xy}(\mathbf{k}) = \breve{M}_{xy}(k_x, k_y, k_z)$ is the spectrum (i.e., three-dimensional FT) of the spatial magnetization distribution.

The image contrast used in imaging is given only by the relative range of values, $\Delta M_{xy}/M_{xy\,\max}$, not by the absolute magnitude of M_{xy}, which is why the values of $S(t)$ and of the Fourier transform $\breve{M}_{xy}(\mathbf{k}(t))$ are usually not distinguished, the word *proportional* often being vaguely substituted by *equal* in interpretations of k-space.

For computational reasons, it is necessary to discretize the problem in both original and frequency spaces. The *discrete k-space* of terms $\breve{M}_{xy}(l\Delta k_x + m\Delta k_y + n\Delta k_z)$ is defined by regularly sampling the continuous space $\breve{M}_{xy}(\mathbf{k})$. This k-space is clearly — from a general point of view — nothing other than the discrete spectral domain (Section 2.3). It has to be gradually filled by the sampled $S(t)$ values during the data acquisition process. The order of filling the k-space is obviously irrelevant from the viewpoint of subsequent image reconstruction; many different trajectories $\mathbf{k}(t)$ exist, given by concrete strategies (pulse sequences) used in MRI (see below).

Afterward, the completely filled k-matrix (two-dimensional or three-dimensional, depending on the imaging method) may be transformed by discrete inverse Fourier transform, thus obtaining the desired pixel values of the final discrete two-dimensional image (or three-dimensional image data block). The dimensions of the matrices in original space and in frequency space are identical; the size and dimensions of the final image thus determine the boundary values of k_x, k_y, and k_z, which correspond to maximum space frequencies in a given direction.

Note that the numbers $S(t)$ in the k-space, being Fourier transform values of an asymmetric image, are naturally complex. To acquire complete information, the phase information (time relations) in the measured signal $S(t)$ must be preserved. This is provided for by synchronous demodulation of the measured signal using in principle the auxiliary demodulating (frequency-shifting) signal $e^{j\omega_0 t}$ coherent with the RF irradiation, as mentioned when deriving Equation 5.19.

5.4.4 One-Dimensional Frequency Encoding — Two-Dimensional Reconstruction from Projections

Though this method is not in routine use anymore, it deserves mentioning for several reasons. It is rather straightforward and was probably historically the first method used for MRI. Moreover, it may use the same algorithmic approach as other tomographic methods, like CT, SPECT, or positron emission tomography (PET), so that it is interesting from our point of view, which aims at showing common features of different approaches. Finally, yet importantly, the first phase of this approach — providing one-dimensional spectra of projections — can easily be generalized to more dimensions and hence

forms a good introduction to more sophisticated two-dimensional and three-dimensional localization approaches.

As with other two-dimensional imaging methods, we suppose that a thin planar slice of tissue has been excited by the gradient-controlled RF irradiation. Without losing on generality, we will suppose that the excitation is performed under a z-gradient so that the excited slice lies in an (x,y)-plane. The simplest impulse sequence that can be used for partial localization is shown in Figure 5.19. The 90° RF pulse applied under the z-gradient field excites a thin slice of the thickness Δz; the following negative refocusing impulse provides for coherent precession. For simplicity, let the slice be irradiated evenly and let us neglect any relaxation or dephasing other than that due to readout gradient. After the RF pulse and later the z-gradient are switched off, the perpendicular time-invariant (say, x-) gradient is switched on for the period of measurement. The time dependence of magnetization $M_{xy}(\mathbf{r})$ during readout time is supposed negligible, as explained when deriving Equation 5.22. Equation 5.20 can then be simplified to a two-dimensional integral over the slice area A (i.e., FOV),

$$S(t) \propto \Delta z \iint_A M_{xy}(x,y)\, e^{-j k_x(t)\, x} \mathrm{d}x\, \mathrm{d}y\,, \tag{5.23}$$

as the only frequency difference (and consequently phase difference) is due to the x-gradient. Thus, we have

$$S(t) \propto \Delta z \int_{D_x} \left(\int_{D_y} M_{xy}(x,y)\, \mathrm{d}y \right) e^{-j k_x(t)\, x} \mathrm{d}x\,, \tag{5.24}$$

where D_x and D_y are FOV dimensions. The inner integral is obviously the projection $P_y(x, t)$ of the spatial distribution to be imaged, along the y-axis (compare with Section 9.1). As the applied x-gradient is time-invariant, the expression finally becomes, when taking into account Equation 5.14,

$$S(t) \propto \Delta z \int_{D_x} P_y(x)\, e^{-j\gamma G_x t\, x} \mathrm{d}x\,. \tag{5.25}$$

The integral now has the form of one-dimensional Fourier transform of the projection, where the spatial frequency is $k_x(t) = \gamma G_x t$; in other words, the instantaneous value of $S(t)$ provides the value of the one-dimensional FT at the space frequency $\gamma G_x t$.

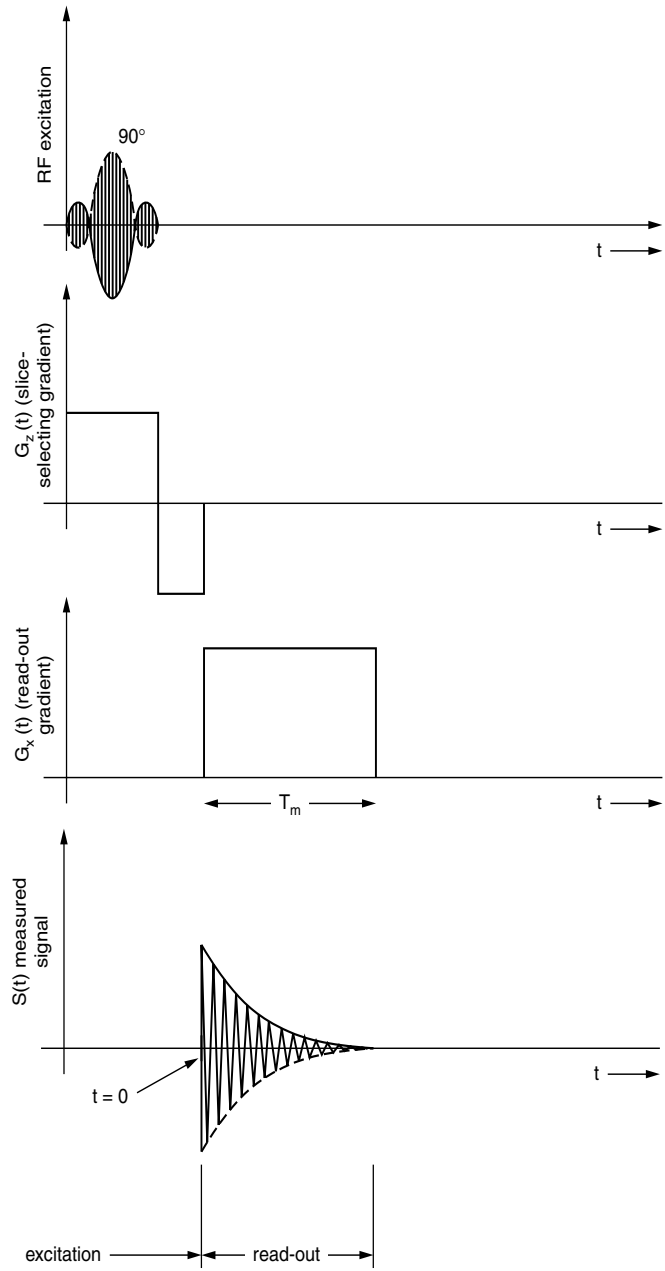


Figure 5.19 The simplest MR measuring sequence with frequency encoding.

The range of the spectrum starts obviously at the spatial frequency $k_x = 0$ (given by $S(0)$) and extends until $S(T_m)$ at

$$k_x = \gamma G_x T_m, \tag{5.26}$$

where T_m is the readout time. Seemingly, arbitrarily high spatial frequencies, hence the corresponding spatial resolution in the image, can be obtained by measuring long enough, but the measurement clearly ends once the signal decays so that SNR becomes insufficient. Consequently, the SNR of the measurement limits the achievable spatial resolution,

$$\Delta x = \pi/(\gamma G_x T_m). \tag{5.27}$$

With the smallest detail size, the image matrix size,

$$N_x = D_x/\Delta x \tag{5.28}$$

can be determined. This in turn determines the necessary density of sampling in the spatial-frequency domain, as the image matrix and the k-space matrix must be of the same size*,

$$\Delta k_x = 2\gamma G_x T_m/N_x \tag{5.29}$$

and consequently, the sampling frequency of $S(t)$ in time-course is

$$f_s = N_x/2T_m. \tag{5.30}$$

By a single RF irradiation, we thus obtain a one-dimensional spectrum of a projection in the direction perpendicular to the gradient. This image projection can easily be recovered by inverse Fourier transform. As the gradient field can be freely combined from x- and y-gradients, $\mathbf{G}_{xy} = G_x\mathbf{u}_x + G_y\mathbf{u}_y$, any other direction of projection can be chosen as well, with all the previous expressions valid after simple rotation of coordinates. By repeating the measurement sequence for different gradient directions, we can obtain a sufficient number of projections, and thus any method of reconstruction from projections (see Section 9.1) can be used to recover the (x, y)-plane image. Under given simplifications, the contrast would be determined basically by proton density.

Note that the first step of a standard method of reconstruction from projections, the method of reconstruction via frequency domain,

*The factor 2 here comes from the necessity to represent also the left-sided part of the spectrum.

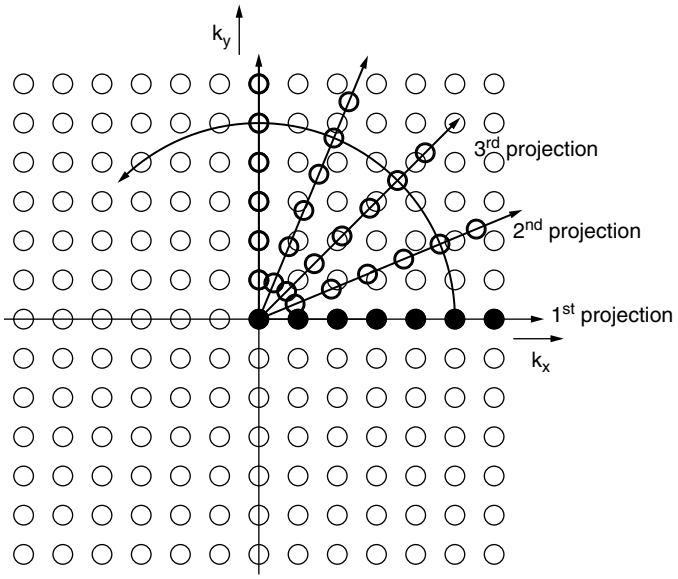


Figure 5.20 Measured signal values in two-dimensional k-space.

is one-dimensional Fourier transforming of all available projections. This obviously has already been done automatically—the MR measurement yields just these one-dimensional spectra of projections. Thus, it is advantageous to apply this method that does not need recovering of the partial projections from $S(t)$ values by inverse one-dimensional Fourier transform.

From this point of view, it is interesting to interpret the obtained values of $S(t)$ in the k-space, which represents the space-frequency domain (Figure 5.20). The values of $S(t)$, provided with readout x-gradient, occupy the positions represented by full dots on the horizontal half line (of the frequency k_x, which corresponds to the positive x-axis in the original domain). The traces of readouts with differently oriented gradients are evidently half lines inclined correspondingly (empty solid dots). To restore the image by means of a standard two-dimensional inverse discrete Fourier transform (2D IDFT) algorithm, we need the k-space values on an equidistant grid (empty thin dots); these must be provided by two-dimensional interpolation. Due to highly uneven distribution of the measured points with respect to the rectangular grid, this may become a source of artifacts.

5.4.5 Two-Dimensional Reconstruction via Frequency and Phase Encoding

To derive another, now commonly used way of filling the k-space, let us start again from Equation 5.20. We again suppose that a thin planar (x, y)-slice of tissue has been evenly excited by RF irradiation under only z-gradient control*, and that the refocusing z-gradient impulse leaves the slice in the state of phase coherence. Any dephasing factors other than the applied x- and y-gradients will be neglected now (these important influences will be considered later). Thus, the time dependence of M_{xy} during readout time is supposed negligible. Repeated irradiation and signal acquisition will be needed, but we shall suppose that the transverse magnetization is identical after each RF pulse. Under these circumstances, the integral can be simplified to

$$S(t) \propto \Delta z \iint_A M_{xy}(x,y)\, e^{-j(k_x(t)x + k_y(t)y)} \mathrm{d}x\, \mathrm{d}y. \tag{5.31}$$

Applying both x- and y-gradients at the same time is equivalent to only a single, though differently oriented, gradient, so that we would naturally arrive again at spectra of correspondingly oriented projections, basically Equations 5.23 and 5.25. Thus, another strategy is needed (Figure 5.21): while we preserve the readout x-gradient resolving the x-position via precession frequency, the time-invariable y-gradient will be applied during a period T_{ph} *before* readout, thus adjusting the $k_y(t)$ value, according to Equation 5.14, to a constant $\gamma G_y T_{ph}$ that remains valid during the signal acquisition period. The integral (Equation 5.31) then becomes

$$S(t) \propto \Delta z \iint_A M_{xy}(x,y)\, e^{-j\gamma(G_x t x + G_y T_{ph} y)} \mathrm{d}x\, \mathrm{d}y. \tag{5.32}$$

Now, the integral obviously has the form of the two-dimensional Fourier transform of M_{xy}, where the space frequencies are $k_x(t) = \gamma G_x t$ and $k_y(t) = k_y = \gamma G_y T_{ph}$. Clearly, an instantaneous value of $S(t)$ provides the two-dimensional Fourier transform value at the spatial-frequency point $(\gamma G_x t,\ \gamma G_y T_{ph})$.

*Note that by combining gradient components, an arbitrary orientation of the slice can easily be provided. The same derivation then still applies after a corresponding rotation of coordinates; nevertheless, the general slice position and corresponding notation would complicate the explanation.

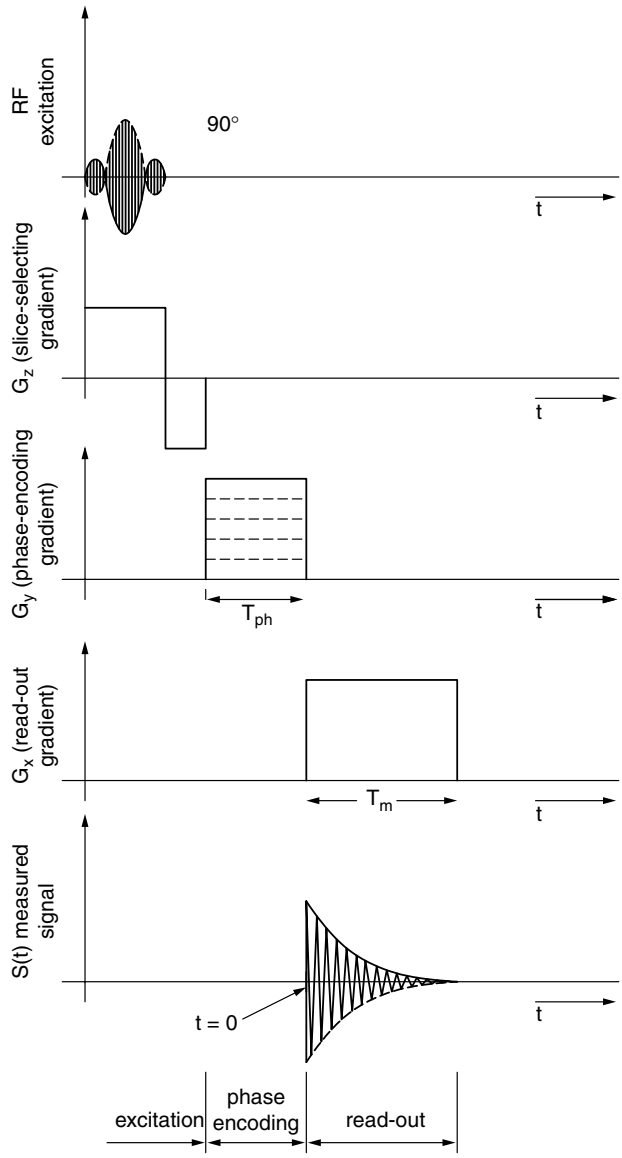


Figure 5.21 A simple phase-encoding sequence.

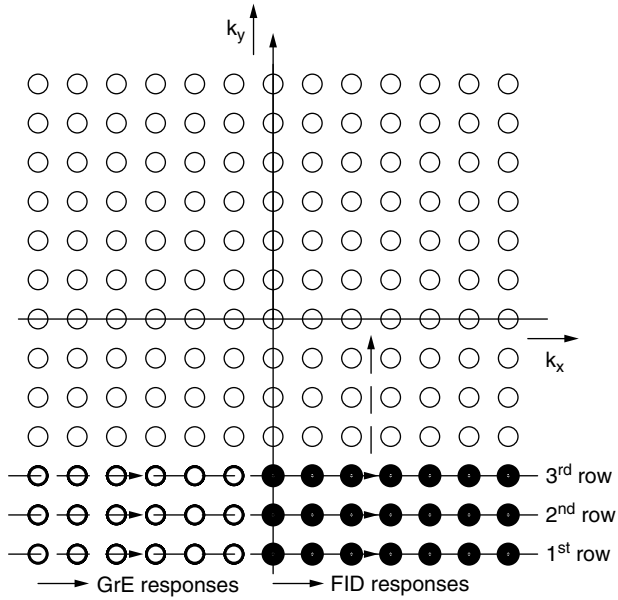


Figure 5.22 k-space filling in by phase-encoding procedure.

Thus, by selecting a particular product $G_y T_{ph}$, we can provide the k-space values on any horizontal half line for the chosen value k_y during a single response to an RF pulse. By changing either the magnitude (as sketched by dotted lines in the figure, possibly even to negative values) or the duration of the y-gradient in each measuring sequence, it is possible to fill in gradually the individual half rows in the k-space, as shown in Figure 5.22, by full dots (note that here t has its origin at the beginning of the readout time). This gradual process is usually called *phase encoding*.

To provide a complete row in the k-space at once (empty solid dots), a different measuring sequence must be used, e.g., such as already mentioned in Section 5.2 under gradient-echo (GE) technique (Figure 5.13). Here, after the excitation sequence, all the spinning is coherent as usual. Then a dephasing x-gradient, $-G_x$, lasting T_m, is applied, causing according to Equation 5.14 a change in $\mathbf{k}(t)$; concretely at the end of this gradient pulse, $k_x = -\gamma G_x T_m$. Consequently, the readout period under the influence of G_x starts, lasting $2T_m$. It is then convenient to set the time origin at the

maximum response time so that the readout period starts at $t = -T_m$, when there is almost no response signal due to the instantaneous state of dephase; during the readout period, k_x obviously increases linearly, $k_x = \gamma G_x t$. Both the degree of coherence and the signal amplitude increase until $t = 0$; after that, the signal decays again due to gradual dephasing until $t = T_m$, when it practically fades away. This way, the signal $S(t)$ is available in the range $t \in \langle -T_m, T_m \rangle$ where T_m is the time needed to reach the minimum acceptable SNR when starting at echo maximum, as explained in the previous section.

A thoughtful reader might object that the obtained signal should be symmetric (or rather conjugate symmetric) with respect to $t = 0$ due to symmetrical rephasing and dephasing, while a row in a two-dimensional image spectrum is generally asymmetric except for a centerline (i.e., with $k_y = \gamma G_y T_{ph} = 0$). This is only a seeming paradox: due to added nonzero phase $\gamma G_y T_{ph} y$ in other rows, the rephasing and dephasing phenomena during readout time are not symmetrical, so that the response will also be asymmetric for any noncenterline.

The GE technique is the fastest way to provide the complete row in the k-space, its speed being controlled by the magnitude of the readout gradient. Nevertheless, it is also possible to use the spin-echo techniques, having the same possibility of two-sided signal acquisition, though at the cost of longer acquisition, as the signal decay is then controlled by $T2^*$. On the other hand, the SE techniques also allow, besides imaging of proton-density distribution, acquisition of $T2$- and $T1$-weighted image data as long as the timing of the SE sequences is chosen properly (see Section 5.2).

Using any echo technique, it is possible to fill in gradually the complete k-space line by line. An example of a simple GE technique-based sequence for frequency- and phase-encoded two-dimensional MRI can be seen in Figure 5.23; a similar sequence for the SE technique is in Figure 5.24.

The frequency range in the x-direction remains the same as in one-dimensional frequency encoding — it is, with double-sided response, $k_x \in \langle -\gamma G_x T_m, \gamma G_x T_m \rangle$, where T_m is the half readout time, limited by the minimum acceptable SNR at $t = -T_m$ and $t = T_m$. Consequently, the pixel size Δx in the reconstructed image is

$$\Delta x = \frac{\pi}{\gamma G_x T_m}. \tag{5.33}$$

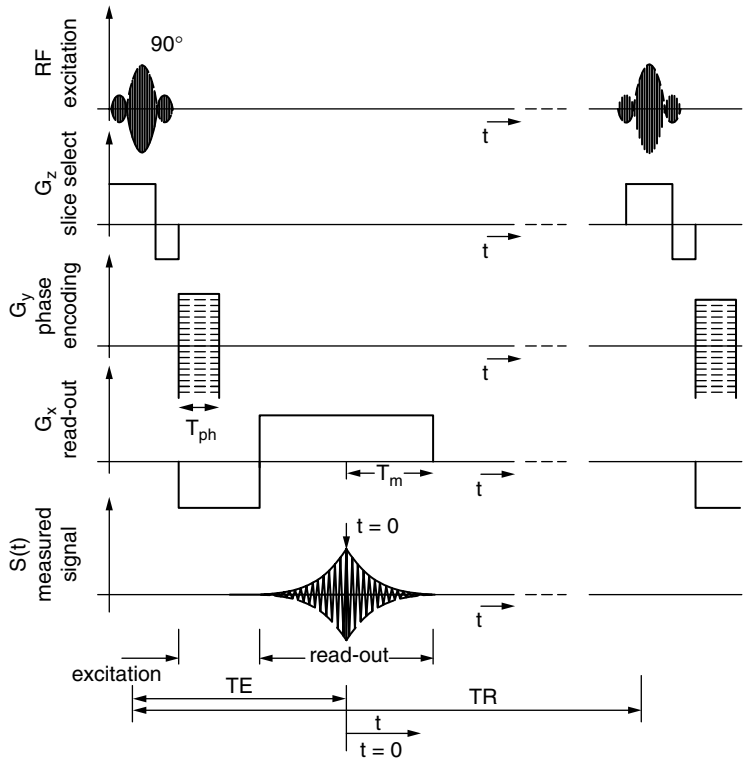


Figure 5.23 A pulse sequence for gradient-echo based two-dimensional MRI using frequency and phase encoding.

The sampling frequency of $S(t)$ in time-course is thus $f_s = N_x/2T_m$. The pixel size Δy in the perpendicular direction is similarly given by the maximum spatial frequency along y,

$$\Delta y = \frac{\pi}{\gamma G_{y\max} T_{ph}}. \tag{5.34}$$

Thereof, the needed range of time-gradient product of the phase-encoding gradient is

$$G_y T_{ph} \in \left\langle -\frac{\pi}{\gamma \Delta y}, \frac{\pi}{\gamma \Delta y} \right\rangle. \tag{5.35}$$

These limits must be observed; otherwise, aliasing appears, causing ringing artifacts.

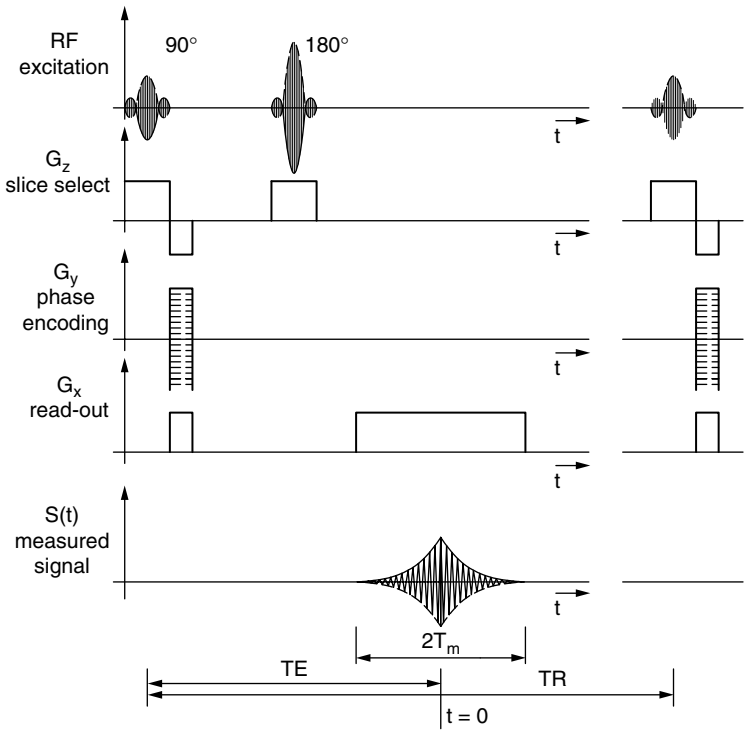


Figure 5.24 A pulse sequence for spin-echo based two-dimensional MRI.

5.4.6 Three-Dimensional Reconstruction via Frequency and Double Phase Encoding

The previous approach can be generalized to three dimensions in a straightforward way. Primarily, the excited volume must include a substantial dimension along the z-axis; this means a weak z-gradient during irradiation, a wideband RF pulse, or both. Alternatively, no gradient may be applied during excitation so that the total object volume is excited, though with a higher risk of wraparound artifact (see Section 5.5).

To derive the method, we again start with Equation 5.20. Now, to encode two dimensions by additional phases, we should follow the approach of the previous section, only adding the third phase term in a similar way as the second one,

$$S(t) \propto \iiint_V M_{xy}(x,y,z)\, e^{-j(G_x t x + G_y T_{ph} y + G_z T_{ph} z)} \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z. \qquad (5.36)$$

In analogy with the previous case, the instantaneous value of $S(t)$ is the value of (now three-dimensional) Fourier transform of the imaged magnetization distribution, evaluated at the three-dimensional spatial frequency ($G_x t$, $G_y T_{ph}$, $G_z T_{ph}$). Each row in the k-space (now three-dimensional matrix) is again described by a single echo. To fill in all the rows, gradually all the combinations of discrete phase factors $G_y T_{ph}$ and $G_z T_{ph}$ must be formed. The *double phase encoding* is provided by auxiliary y- and z-gradients preceding the readout, with either adjustable magnitudes or durations. Both of these gradients may be applied concurrently, as seen in Figure 5.25. It shows the sequence for the GE technique; it is up to the reader to derive the basic sequence for spin-echo imaging by combining properties of this figure with Figure 5.23.

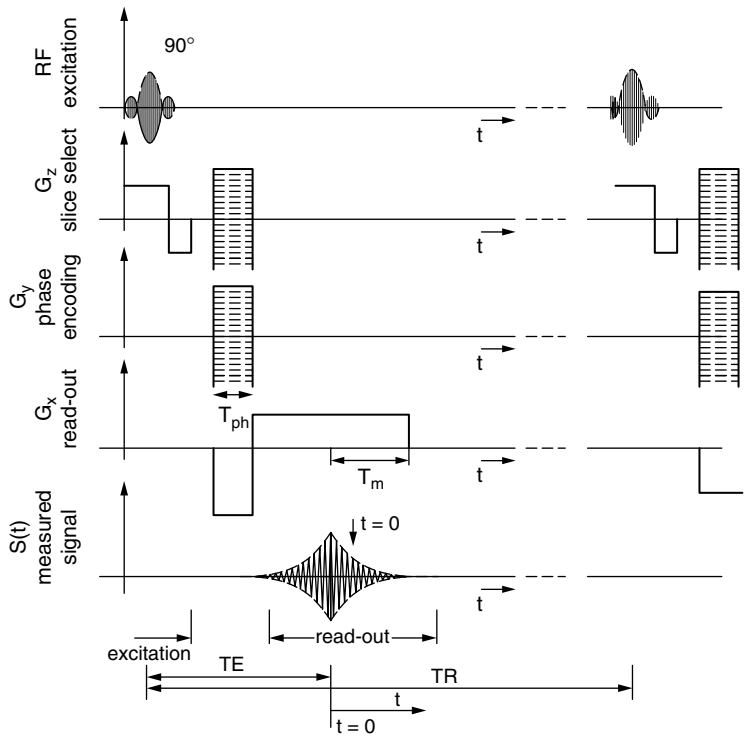


Figure 5.25 Pulse sequence for three-dimensional imaging via frequency- and double-phase encoding.

Using the same argument as in two-dimensional phase encoding, we can derive

$$\Delta z = \frac{\pi}{\gamma G_{z\max} T_{ph}} \qquad \text{and} \qquad G_z T_{ph} \in \left\langle -\frac{\pi}{\gamma \Delta z}, \frac{\pi}{\gamma \Delta z} \right\rangle . \quad (5.37)$$

Three-dimensional imaging may become a lengthy process, as it requires the measurement time $N_y \times N_z \times T_R$. On the other hand, the SNR circumstances are more favorable in three-dimensional imaging than in two-dimensional or one-dimensional imaging, as the signal is acquired from a large volume rather than from a thin slice, and resulting image pixel values will each be calculated by the three-dimensional IDFT from $N_x \times N_y \times N_z$ samples of the k-space, so that the random noise is reduced by weighted averaging. Thanks to this, very thin slices can be calculated — up to a few tenths of a millimeter with still a reasonable SNR.

5.4.7 Fast MRI

The nowadays classical MRI, as described above, requires an RF pulse per line in the k-space. In this standard MR approach, often complete recovery of initial magnetization after each measurement is required (i.e., $TR >> T1$) in order to start the next measurement with the full M_z, thus eliminating the influence of $T1$ and yielding the maximum response. Even in $T1$-weighted imaging, the repetition time TR is still comparable with $T1$, which, for most tissues, is on the order of several hundreds of milliseconds. The total imaging times are then in minutes or (for three-dimensional imaging) even tens of minutes. This is hardly acceptable in many applications, and therefore ways to faster imaging were prospected.

While the above-described imaging and namely localization principles remain completely valid, the additionally used basic notions enabling faster or very fast imaging should be explicitly mentioned here in order to maintain consistency of explanation. The main ideas, which can be combined, may be summarized as follows:

- Multiple-slice imaging.
- Low flip-angle excitation.
- Several or many echoes measured under different phasing after a single RF pulse, i.e., filling in more than a single

line in the k-space per pulse. In *echo-planar imaging*, the complete k-space of a slice per a single pulse is filled in.

- The k-space need not be necessarily filled in by the approach used so far (line by line); other filling strategies are possible.

5.4.7.1 Multiple-Slice Imaging

In principle, multiple-slice imaging means only that in each measuring period TR, the long waiting time after the signal readout, needed to reestablish the initial magnetization in the measured slice, is utilized for exciting and measuring gradually other slices that do not spatially coincide with the relaxing ones. This, under the influence of slice selection gradient fields, also means that their signals are separated in the frequency domain. This way, a kind of time multiplex is introduced, enabling the speeding up of three-dimensional imaging several times, namely, when a complete $T1$ recovery is required, as in $T2$-weighted imaging. In order to prevent any interference, spatial interleaving of two or more alternately imaged slice packages is used, thus keeping sufficient spatial separation among the slices excited simultaneously due to unfinished decay of the previously measured ones.

5.4.7.2 Low Flip-Angle Excitation

So far, we supposed excitation by 90° RF pulses, which leads to a maximum signal response but obviously requires the longest time of $T1$ relaxation. As visible from Figure 5.26, when a weaker RF pulse is applied, leading to a lower flip angle α, the magnitude of transversal magnetization is lower than optimum, $M_{xy} = M_{z0}\sin\alpha$. On the other hand, there remains an axial (longitudinal) magnetization $M_z = M_{z0}\cos\alpha$; for small α, it is close to M_{z0}. This way, the relaxation time of the z-axis magnetization is substantially shorter, while the signal-inducing component M_{xy} may still be sufficient.

For a repetition rate, given by a TR shorter than or comparable with $T1$, a lower angle α provides, paradoxically, a stronger signal. The optimum α (Ernst angle, ~1960), yielding the strongest signal after several TR periods, is

$$\alpha_E \cong \arccos(e^{TR/T1}). \tag{5.38}$$

With small α, short repetition times TR may be used even when $T1$ weighting is required, thus speeding up all the standard acquisition procedures by about a decimal order (TR can be shortened from

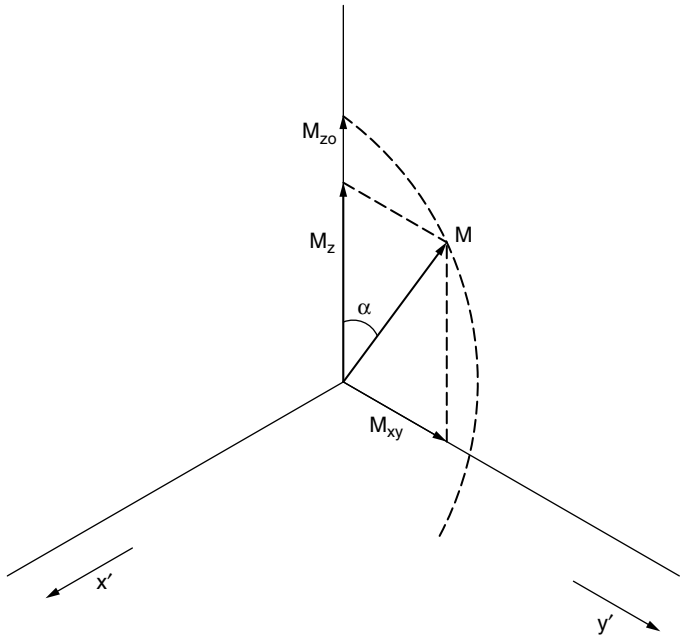


Figure 5.26 Low flip-angle excitation.

hundreds of milliseconds to as little as 10 to 20 msec, when using the GE technique).

5.4.7.3 Multiple-Echo Acquisition

This method is characterized by providing multiple echoes per an RF excitation pulse, on a similar principle like that depicted in Figure 5.12.

Multiple spin echoes may be produced in classical two-dimensional or three-dimensional MRI in the standard way, i.e., all with the same amount of phase encoding, providing data for several autonomous images with differently weighted $T2$ values (Figure 5.27a); naturally, the time separation of the individual echoes must be sufficient to allow a distinct influence of the $T2$ relaxation between neighboring pulses. Thus, each RF 90° excitation is gradually filling equally positioned lines in the k-spaces of all the images. With a standard number of RF excitations, data for *several* differently $T2$-weighted images are thus progressively completed in a single imaging time. The readout x-gradient field has to be applied during each echo,

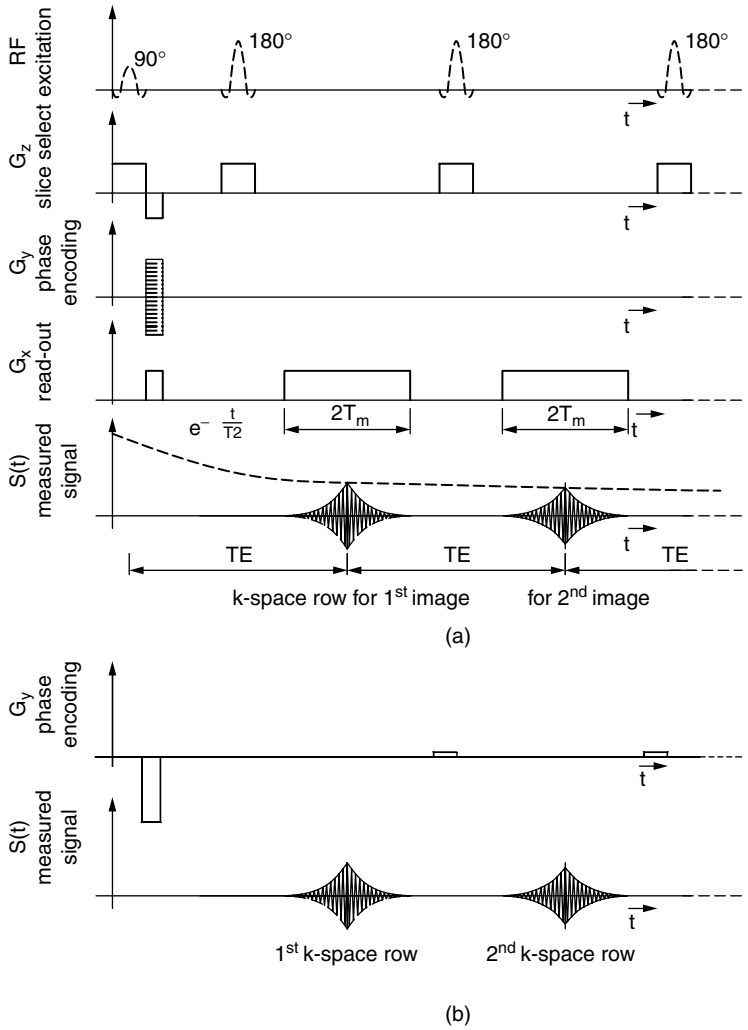


Figure 5.27 Principle of (a) multiple-echo acquisition of the same k-space line in several differently $T2$-weighted images and of (b) multiple k-space lines in a single image with gradual phase encoding, with identical RF, G_z, and G_x pulses as above.

followed by an equally strong opposite field, to compensate for the x-gradient influence on the course of continuing precession.

Another possibility of multiple-echo acquisition is to apply an amount of phase-encoding y-gradient before each echo (Figure 5.27b).

This approach relies on the (approximate) time invariance of spin precession during the measurement time after a single excitation RF pulse; in other words, the complete measurement must be made in a correspondingly short time. The spin sets are then influenced by the time-variable gradient fields, so that they change their frequencies and consequently phases, but the energy of spinning remains (approximately) unchanged. As k_y is accumulated during the whole TR period according to Equation 5.14, each echo is then differently phase encoded and yields values for a corresponding line in the k-space of the image. A single RF pulse may thus cover more lines in the same k-space, thus providing data of an image correspondingly faster. Let us note that, in this case, we suppose that the $T2$-related relaxation is negligible. The number of echoes received after a single RF pulse—up to several tens—is sometimes called *turbo factor*.

5.4.7.4 Echo-Planar Imaging

A similar principle to that of the last section is utilized in *echo-planar imaging*, which can be characterized by very fast data acquisition: complete two-dimensional image data are provided after a single RF pulse. The speed is achieved by using gradient echoes (much faster than spin echoes), which requires strong gradient fields, and by modifying the order of filling the k-space. The last point deserves a more in-depth explanation.

It is seen that the integral (Equation 5.19 or 5.20), as modified for time-invariant distribution M_{xy},

$$S(t) \propto \iiint_V M_{xy}(x,y,z)\, e^{-j(k_x(t)x+k_y(t)y+k_z(t)z)} \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z \qquad (5.39)$$

is symmetrical with respect to all spatial coordinates; there is no inherent notion of "frequency" or "phase encoding". Simply by changing k_x and k_y (and in principle also k_z), values of different locations in the k-space may be obtained from $S(t)$. The final goal is to fill in the complete k-space during a single measurement, but the order is irrelevant*. The time-dependent function $\mathbf{k}(t)$ can then be interpreted as a parametric description of a curve in the k-space,

*Nevertheless, some effects, such as chemical shifts and flow-related phenomena, violate the assumption of time invariance of M_{xy} to a certain extent, and their imaging is therefore partly dependent on the k-space trajectory (see [36]).

the *k-space trajectory*, which depends on the arrangement of the pulse sequence. Naturally, the trajectory should describe the whole k-space sufficiently to enable the following image reconstruction by inverse Fourier transform, and must be completed in a time short enough to allow the assumption of the time invariance of M_{xy}.

The basic pulse sequence and the corresponding k-space trajectory are schematically depicted in Figure 5.28a. After a slice-exciting RF pulse, applied together with the z-gradient (and a refocusing opposite gradient, possibly overlapping with the following step), the short-time negative gradients G_x and G_y adjust the k_x and k_y values, respectively, to their limits, say in the lower left corner of the k-space. Consequently, the readout G_x gradient is applied and the lowest line of k-space is provided by reading the first gradient echo (like in conventional two-dimensional GE MRI). When this is finished, G_y is applied momentarily, incrementing k_y by the required amount Δk_y as needed for the second line. Instead of returning to the left margin of k-space, which would be time- and energy-consuming, the trajectory continues in the opposite direction — from right to left — under the opposite readout gradient G_x. After each line, k_y is incremented by a pulse of G_y and the k-space is gradually filled in a zigzag way. The oscillating x-gradient is the key to the high speed, besides the fast GE technique under strong gradient fields. The data acquisition for a complete image can thus be accomplished after a single RF pulse.

A modification, simplifying G_y generation, consists in replacing the short impulses by a weak time-invariant G_y, the influence of which would accumulate to provide, during a line acquisition time, gradually the increment Δk_y. This leads to a slightly modified trajectory — the horizontal lines are replaced by slightly oblique ones. Obviously, the price for the simplification is that proper k-space samples must be provided by one-dimensional interpolation along k_y to obtain the k-space sample values on the needed rectangular grid.

From the gradient-field generation point of view, it is advantageous to replace the fast-changing rectangular impulses by harmonically oscillating gradients, with lower hardware requirements. When G_x from the first sequence is replaced by a sinusoidal function with the half period equal to echo time, the same zigzag trajectory in Figure 5.28b would appear when G_y is pulsed, only modified by nonequidistant time coverage, which can be compensated for by intentionally uneven sampling of $S(t)$. Alternatively, under the constant weak gradient G_y, a sinusoidal trace like in Figure 5.29a

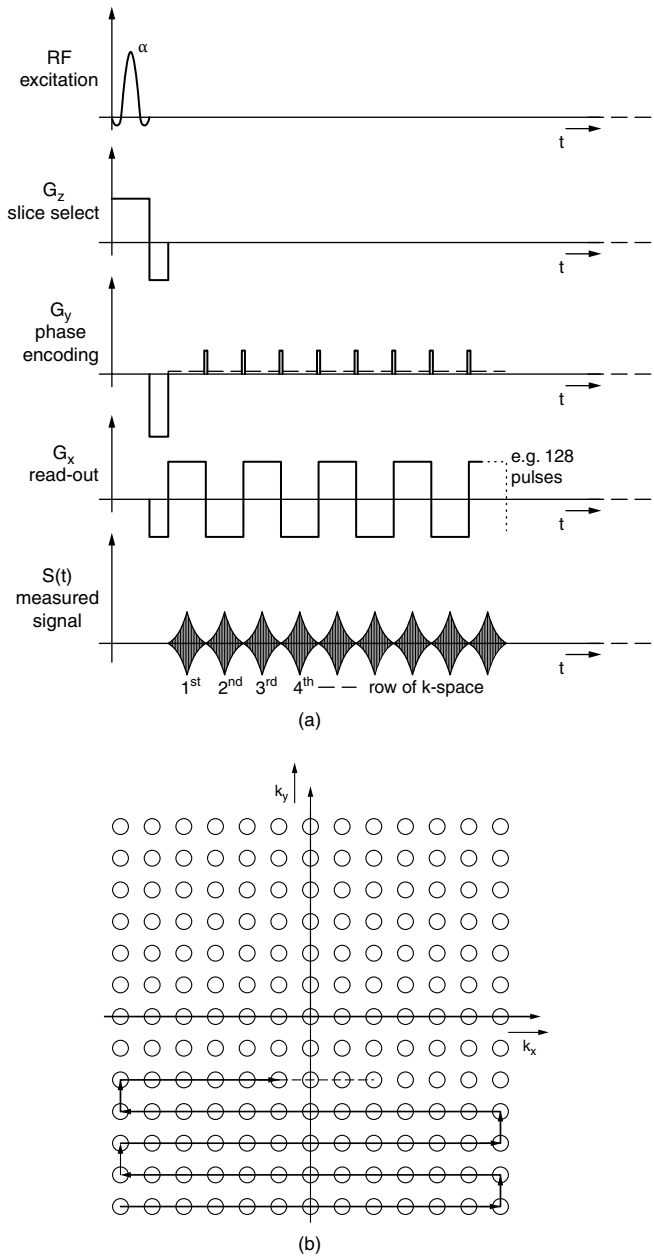


Figure 5.28 (a) Ideal EPI pulse sequence and (b) corresponding *k*-space trajectory.

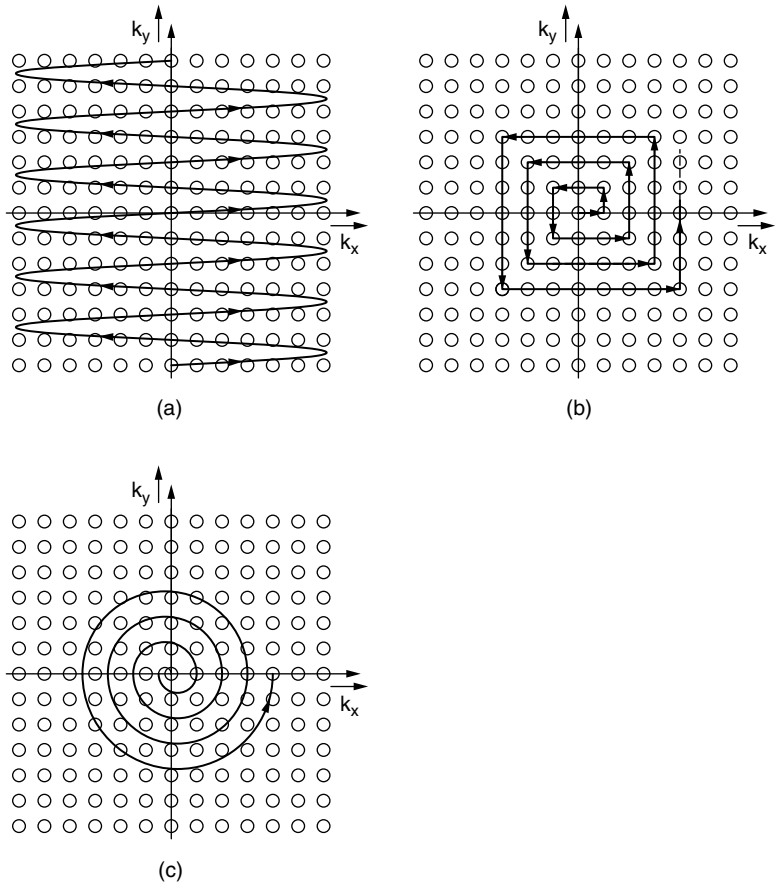


Figure 5.29 Examples of other k-space trajectories.

appears, needing one-dimensional interpolation along k_y, besides the uneven time sampling. The gradients providing the trajectory in Figure 5.29b can be easily derived as a sequence of rectangular pulses of gradually increasing magnitude.

Accepting two-dimensional interpolation in the k-space allows quite general k-space trajectories, as far as they reasonably cover the complete needed range of **k**. This way, trajectories adapted to the special needs of biological activity studies, including flow, diffusion, and chemical composition imaging (or compensation, on the

contrary), can be optimally designed, still with reasonable hardware requirements. A typical example is the spiral trajectory Figure 5.29c that is generated by both gradients G_x, G_y, being sinusoidal with 90° phase shift, and modulated (multiplied) by $t(|k_{\max}|/T_m)$, with $2|k_{\max}|$ being the size of the k-space in either axis. Also, the sampling frequency must clearly be increasing with time to ensure reasonably homogeneous coverage of the k-space. Another advantage of this and similar trajectories is that it covers a circular area of spectrum, thus providing isotropic resolution in the image; this can save about 27% of the measurement time otherwise lost for filling the spectrum corners.

Still another source of time savings is the inherent central symmetry of the two-dimensional Fourier transform, so that theoretically, only one half must be measured. Although slightly more is necessary for the reasons of suppressing noise influence, it allows further cutting of the acquisition time by almost 50%. Specially designed algorithms of two-dimensional inverse discrete Fourier transform are available, enabling reconstruction of the images from such partial data without the need to complement the other half of the k-space.

Finally, it should be mentioned that a sufficiently fast EPI sequence (not including the excitation) could be considered an autonomous block that is capable of converting the *instantaneous* magnetization distribution into the FT of image data *at any time instant* after an RF pulse, or even a sequence of RF pulses. This approach enables provision of not only proton density, but also $T1$- and $T2/T2^*$-weighted images.

5.5 IMAGE QUALITY AND ARTIFACTS

5.5.1 Noise Properties

The basic property of any imaging is the SNR; however, it is far beyond the frame of this book to analyze the noise properties of the MRI process. Such an analysis can be found in specialized literature, e.g., [7], [9], [33], [36], and should investigate in depth the underlying physics at the quantum level, as well as properties of the instrumentation involved (namely, noise generated by RF coils and receiver) and the influence of data processing on the resulting noise level. Nevertheless, it is useful to state, without proof, some basic knowledge that can be partly justified intuitively.

First, let us declare that the SNR can be in principle improved via the two following ways:

- Enhancing the signal strength by adjusting physical conditions of imaging, and the technique and parameters of signal acquisition
- Suppressing noise by averaging of the individual uncorrelated noise components

The first group of possibilities is mainly determined by the system design (e.g., the signal is approximately proportional to the induction of the static field $\mathbf{B}_0$, the signal level may be enhanced by using local RF coils, etc.). It is partly possible to influence the signal strength by the choice of imaging parameters, as mentioned below, which also affect the noise levels at different stages of the imaging process.

The way of image reconstruction and postprocessing may influence the resulting noise level in the image to a great extent, by different explicit or indirect ways of averaging.

The SNR is approximately proportional to the following parameters:

- *Local proton density*, proportionally influencing the signal strength.
- *Static magnetic field induction* $\mathbf{B}_0$ for the same reason, plus some additional effects, thus influencing SNR linearly or even stronger, about in the range $B_0^1 \ldots B_0^{1.5}$. As the mean Larmor frequency ω_{L0} is proportional to $\mathbf{B}_0$, SNR can be alternatively said to depend on ω_{L0}.
- *Voxel size*, codetermining the number of participating nuclei — the higher the requested resolution, the stricter are noise suppression requirements.
- Square root of *number of measurements* taken into account and contributing to averaging. This includes the number of excitations, but also the number of (uncorrelated or slightly correlated) signal acquisitions based on a single excitement, as well as repeated measurements explicitly determined for averaging. Also, the dimensions of the FOV matrix, i.e., the number of elements of the k-matrix, are to be taken into account here, if two-dimensional or three-dimensional Fourier reconstruction is used. The inverse Fourier transform provides every pixel or voxel value as a weighted average of k-space elements, thus suppressing the uncorrelated noise components correspondingly, though not to the full extent, as would be the case if

uniform weights were applied. It can be seen that in this respect, three-dimensional imaging is advantageous compared to two-dimensional imaging, or even one-dimensional projection-based reconstruction. This is the reason why the reconstructed slices may be very thin in three-dimensional imaging without decreasing the SNR to an unacceptable level. Averaging of more images of the same scene with independent realizations of noise also provides SNR improvement proportional to the square root of the number of repetitions, though at the cost of measurement prolongation.

- *Bandwidth* (BW) of the acquired signal $S(t)$ influences SNR by its inverse square root, $1/\sqrt{BW}$, as the noise power in the signal is proportional to BW when the noise can be supposed white in the frequency range of measurement. This can be utilized to improve the SNR by narrowing the signal bandwidth at the cost of slowing down, and consequently prolonging, the measurement under a weaker readout gradient. Note the square rule: improving SNR twice requires fourfold time of signal acquisition. This is an option in some modern systems.
- *System design* parameters, as the quality of RF coils, etc.
- *Acquisition parameters*, as discussed below, influence the SNR in a complex manner, mediated by the above phenomena. To mention just one, often the gap between neighboring slices is presented explicitly among noise-influencing parameters.

5.5.2 Image Parameters

The basic image parameters — field of view (FOV) and matrix size, spatial resolution, and sampling density — are closely related to parameters of the imaging process. Let us summarize the relations, for simplicity still adhering to our above convention concerning axes and slice position, though we realize that any orientation of slices may be obtained simply by rotating coordinates via corresponding combinations of individual gradient fields.

The z-direction resolution depends on the imaging method: in two-dimensional imaging based on selective slice excitation, it is given by the slice thickness that is, according to Equation 5.9, $d = \Delta z = \Delta\omega/(\gamma G_z)$. In three-dimensional imaging, it is given by the same

considerations as resolution in the other two directions (see below), as the description is basically symmetrical with respect to all coordinates.

We shall consider next the more general three-dimensional case; the two-dimensional case description would simply be obtained by omitting all z-coordinate-related quantities. It is visible from Equation 5.20 that the imaging equation is entirely symmetrical with respect to all three coordinates, the only differences being introduced by different ways of encoding the spatial information. Consequently, also the formulae relating the imaging parameters must be in principle symmetrical, with only minor differences reflecting the encoding approach. Often, the imaging parameters (FOV, resolution) are identical in all three directions. Nevertheless, the more general case of different FOV dimensions or resolution along individual coordinates is described.

The extreme values of spatial frequencies in all directions are given by physical and system limits: the maximum usable spatial frequency in each direction is determined by the signal $S(t)$ becoming so weak that the SNR is just at the limit of acceptability. This applies equally in all three directions; thus, according to Equations 5.26, 5.35, and 5.37,

$$k_{x\,\max} = \gamma G_x T_m, \quad k_{y\,\max} = \gamma G_{y\,\max} T_{ph}, \quad k_{z\,\max} = \gamma G_{z\,\max} T_{ph},$$

providing that both phase-encoding gradients are switched on for the same period T_{ph} so that their influence is controlled via magnitudes. Knowing these limit frequencies and choosing the timing T_m, T_{ph}, it is possible to determine the necessary (constant or limit) gradients,

$$G_x = \frac{k_{x\max}}{\gamma T_m}, \quad G_{y\max} = \frac{k_{y\max}}{\gamma T_{ph}}, \quad G_{z\max} = \frac{k_{z\max}}{\gamma T_{ph}}.$$

The maximal frequencies correspond to *spatial resolution*, as they determine the image *pixel (voxel) dimensions* via DFT, as given by Equations 5.33, 5.34, and 5.37,

$$\Delta x = \frac{\pi}{\gamma G_x T_m}, \qquad \Delta y = \frac{\pi}{\gamma G_{y\max} T_{ph}}, \qquad \Delta z = \frac{\pi}{\gamma G_{z\max} T_{ph}}.$$

If the same resolution is requested in all directions, the three maximal frequencies have to be chosen equal — thus $G_x T_m = G_{y\,\max} T_{ph} = G_{z\,\max} T_{ph}$, which is to be provided for by the controlling system.

The rectangular three-dimensional *field of view*, defined by its dimensions D_x, D_y, D_z, can still be freely chosen. Obviously, the size of the image matrix will then be $N_x \times N_y \times N_z$, with $N_x = D_x/\Delta x$, etc. The k-space matrix will be of the same size, from which, in accordance with Equation 5.28, we obtain

$$\Delta k_x = \frac{2\gamma G_x T_m}{N_x} = \frac{2\pi}{D_x}, \quad \Delta k_y = \frac{2\gamma G_{y\max} T_{ph}}{N_y} = \frac{2\pi}{D_y}, \quad \Delta k_z = \frac{2\gamma G_{z\max} T_{ph}}{N_z} = \frac{2\pi}{D_z},$$

so that the increments of variables, controlled by the system, are

$$\Delta t = \frac{\Delta k_x}{\gamma G_x} = \frac{2T_m}{N_x}, \quad \Delta G_y = \frac{\Delta k_y}{\gamma T_{ph}} = \frac{2G_{y\max}}{N_y}, \quad \Delta G_z = \frac{\Delta k_z}{\gamma T_{ph}} = \frac{2G_{z\max}}{N_z}. \tag{5.40}$$

The left equation determines the needed *sampling frequency* of $S(t)$ for the case of regular sampling, in accordance with Equation 5.30:

$$f_s = \frac{N_x}{2T_m} = \frac{\gamma G_x D_x}{2\pi}. \tag{5.41}$$

We see that, given the resolution Δx, the sampling frequency, and consequently the necessary bandwidth of $S(t)$, is proportional to the gradient used and to the size N_x of the image matrix, and thus to the size D_x of the chosen FOV.

The total *scan time* in the three-dimensional case is obviously

$$T_{tot} = N_y \times N_z \times TR,$$

possibly multiplied by the number of scan repetitions if this is needed to improve SNR.

5.5.3 Point-Spread Function

In the analysis so far, we neglected the temporal variability of magnetization distribution in FOV during the readout time.

Under this simplifying assumption, it follows from the preceding derivations that the image data are precisely (with the apparent limitations given by discretization) describing the distribution. This would mean that the *point-spread function* (PSF) of the imaging process is an ideal impulse.

Nevertheless, by a detailed analysis taking into account the temporal variability, it can be found that the imaged information is given rather by convolution* of the imaged distribution with spatially variable PSF, dependent not only on the properties of the excitation and gradient sequence, but also on the local $T2$ relaxation time of the imaged object and on the local inhomogeneity of the main field $\mathbf{B}_0$. Detailed analysis shows that the temporal variances cause widening of the PSF, effectively decreasing the resolution from a single pixel (voxel) to even several pixels; at the same time, the PSF amplitude decreases, causing partial signal loss. These phenomena are particularly severe in fast MRI approaches where more than a single line of k-space or even a complete two-dimensional spectrum is acquired per RF pulse, so that the individual responses are modulated by significantly different instantaneous values of the respective decay exponentials. In EPI, obviously the PSF will depend on the k-space trajectory, as different parts of the trajectory are provided with differently decayed magnetization. When the classical EPI trajectory is used, the phase-encoded y-direction is substantially more affected than the frequency-encoded x-direction, which leads to significantly asymmetrical PSF. On the other hand, spiral EPI (SEPI) provides k-space data affected almost symmetrically with respect to the axes' origin $k = 0$, so that PSF is also rather symmetrical. A deeper analysis of PSF is already beyond the scope of this book; more details can be found in [9], [36] and the references therein.

Attempts to restore the resolution by a kind of deconvolution in postprocessing are complicated primarily by the spatial variance of the PSF, and also by the raise in noise, common to any restoration.

*Strictly, the convolution (convolution integral) is only the operation with an invariable PSF. The operator under discussion is obviously the superposition integral; nevertheless, the notion of space-variant convolution is widely accepted as a colloquial term.

5.5.4 Resolving Power

The geometrical spatial resolution, given by the density of sampling together with the shape of PSF, does not determine completely the diagnostic or *resolving power* of the imaging method. This is also influenced by other factors — image contrast and SNR. The contrast, which should be adapted to the imaged scene, can be considerably influenced by the method and parameters of acquisition ($T1$, $T2$, proton density, or mixed weighting determined via TR and TE selection). The resolving power is to a great extent also affected by the signal-to-noise ratio, which is partly dependent on the system characteristics, but it may be also influenced by the user's choice of voxel volume, and possibly also by repeating the measurements in a way, with subsequent averaging. It is assumed that the resolving power is roughly quadratically dependent on the contrast, as well as on SNR, and finally proportional to the product of both squares. The final practical assessment of the quality of an MRI system should be based on calibrated phantoms containing differently sized and situated objects of diverse MRI-related properties — proton density, $T1$ and $T2/T2^*$, and possibly also chemical composition or intrinsic diffusion (compare with Section 15.8 in [9] on phantoms for CT imaging).

5.5.5 Imaging Artifacts

The imaging artifacts in MRI may be subdivided into three groups:

- Discretization effects
- Effects related to object chemical composition, susceptibility distribution, and time variance
- System-related imperfections

Two phenomena arising from sampling the spectrum and the image are usually mentioned — sc., ringing in the neighborhood of sharp edges and wraparound of the exterior of the imaged volume into the reconstructed image.

Ringing is the truncation artifact caused by too low spatial sampling frequency with respect to a sharp edge inherently containing high-frequency components. Such components with frequencies above the Nyquist limit manifest themselves, due to aliasing, as lower-frequency components generating lines or curves in the neighborhood of the edges and borders. The situation can evidently be improved only by choosing denser sampling (a smaller pixel size) if the SNR and the other acquisition parameters allow.

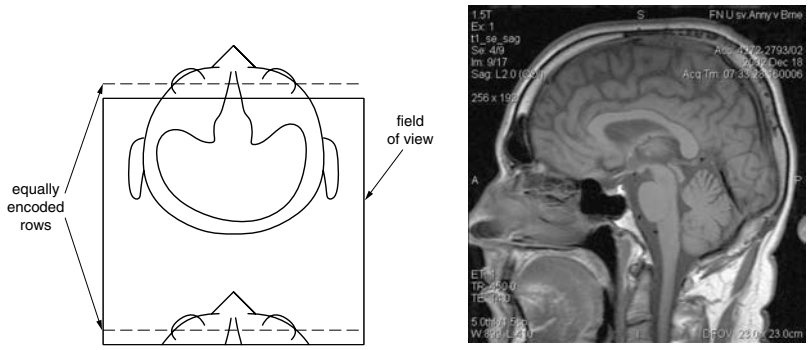


Figure 5.30 Schematic illustration of wraparound effect: schematically (left) and a real MRI image (right). Note the cut nose tip appearing from the right. (Courtesy of the Faculty Hospital of St. Anne Brno, Radiology Clinic, Assoc. Prof. P. Krupa, M.D., Ph.D.)

The *wraparound* artifact is rather specific to MRI with FT-based reconstruction. It manifests itself by imaging the parts of the object that are outside of the FOV, inside the reconstructed image but at the border other than the one to which it is adjoined physically, as schematically depicted in Figure 5.30. This phenomenon can be explained based on the periodicity in the original domain that two-dimensional or three-dimensional DFT imposes on the original domain due to sampling in the frequency domain — the image of FOV is supposed to be periodical in the plane (or space), with the periods D_x, D_y, and, in three-dimensional imaging, also D_z.

An otherwise formulated explanation of the same phenomenon is available: the spatial coordinates of voxels are basically phase encoded (even in the x-direction, as already explained; the difference is only in the way the phase is encoded — see Equation 5.20). As the phase encoding is periodical, with the period given by the dimensions of FOV, the same phase labels appear outside of FOV — the layer just adjoined from outside to, say, the left border of FOV is encoded with the same labels as the first layer of FOV from the right, etc. If the outer parts of the object are not excited, or are empty (no tissue), the wraparound would not appear. However, when the excited area is greater than the FOV, then the signal from the outer structures is combined, in an indistinguishable way, with the signal from identically labeled parts of the FOV, which leads

to incorporating the improper outer structures to the reconstructed image.

The effect can be in principle suppressed by exciting only the area of FOV; this might be of interest in three-dimensional imaging when deciding whether to use a (thick, but limited) slice excitation or a rather spatially unlimited gradientless excitation. Alternatives are to pick up predominantly only the signal from FOV by local (surface) receiving coils or by using *saturation RF pulses* applied outside of the FOV that eliminate signal generation. Still another approach is to increase effectively the dimensions of FOV so that the complete excited area is included; thus, no signal components are received from outside of the FOV. Nevertheless, it may have an adverse effect on either spatial resolution, signal bandwidth, or acquisition time — see the relations among the imaging parameters above. A possibility is to use the *oversampling technique*, sacrificing a part (usually a half) of the reconstructed image and doubling the spectral (k-space) sampling density in the needed direction. This leads to doubling the respective dimension of FOV; should the image resolution remain unchanged, the number of k-space points is obviously also doubled. Fourier transforming the data provides the image with a doubled dimension, of which only the central half is displayed while the outer parts are discarded. In order to preserve the same amount of data, the frequency sampling density in the other direction may be halved, naturally halving the other dimension of FOV.

With respect to the direction of this book, we shall not discuss the artifacts of the second and third groups. Information on the artifact influence and possible remedies can be found in specialized literature as partly listed in the references. Nevertheless, it should be mentioned that the phenomena connected with chemical shift, flow, and other object parameter distributions, considered artifacts in conventional MRI, may, on the other hand, be exploited as a source of important diagnostic information — modern directions of functional MRI or motion and diffusion imaging are based on them. These specialized fields are already beyond the scope of this book; more details and further references can be found in the cited literature.

5.6 POSTMEASUREMENT DATA PROCESSING IN MRI

MRI is obviously fully dependent on postprocessing of the measured data. Primarily, the RF signal has to be processed as described above in order to obtain the k-space data or projection spectra; even this

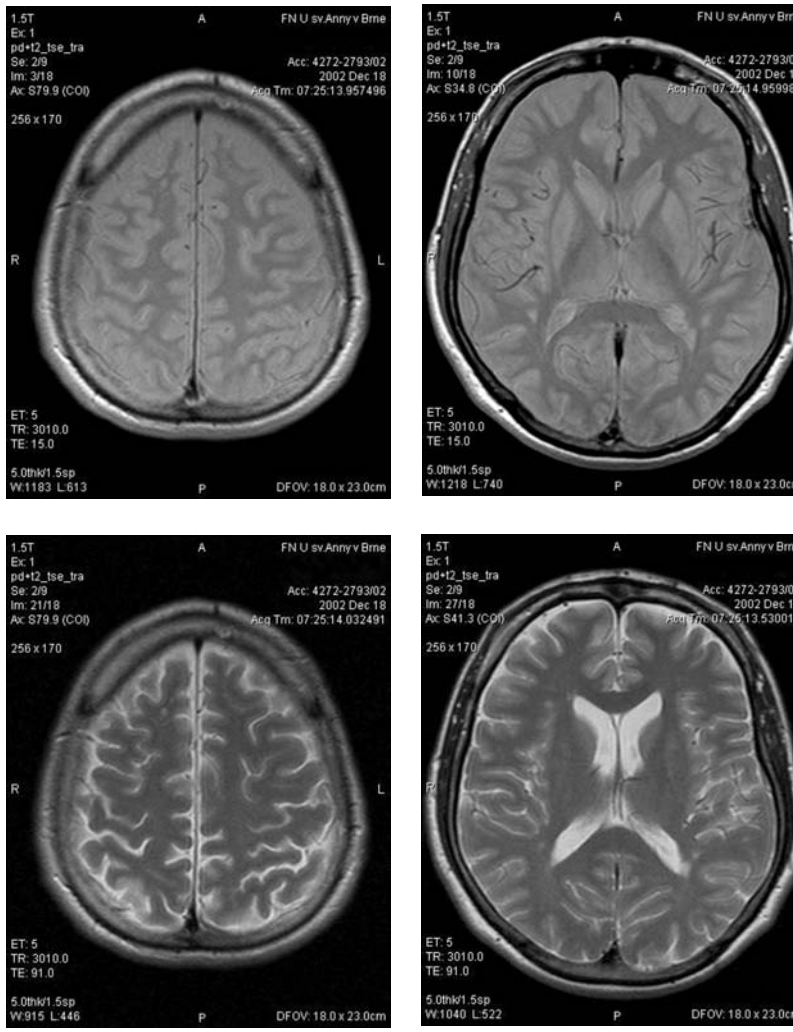


Figure 5.31 (Example MRI images — selected transverse slices of the head. Note the contrast differences in pairs of identical slices provided with different imaging parameters. (Courtesy of the Faculty Hospital at St. Anne Brno, Radiology Clinic, Assoc. Prof. P. Krupa, M.D., Ph.D.)

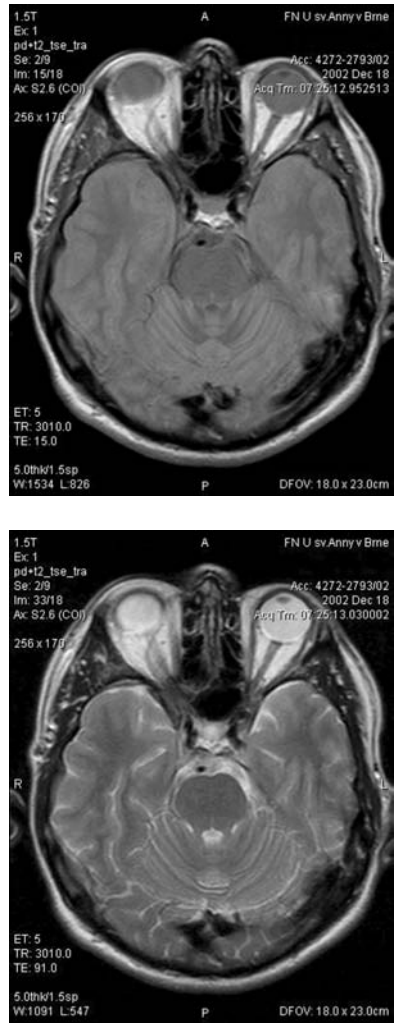


Figure 5.31 (Continued).

stage is nowadays realized largely digitally (except for RF amplification and possibly also demodulation, i.e., frequency shift to the base-band), mostly by means of specialized hardware. When using nonclassical k-space trajectories, as, e.g., in spiral EPI, two-dimensional interpolation (see Section 10.1.2) in the k-space is necessary to provide equidistantly spaced values on the rectangular grid, as needed for the following

transformation. The consequential image reconstruction from the RF responses has to be achieved either by two-dimensional (or three-dimensional) inverse discrete Fourier transform (Section 2.3.2) or by methods of reconstruction from projections (Chapter 9) [34], [38]. Typical examples of MR images are presented in Figure 5.31 and Figure 5.32.

Once the image data are obtained as a two-dimensional matrix or, in case of three-dimensional imaging, a three-dimensional matrix, the standard image processing procedures, described in

Figure 5.32 Example MRI images — selected longitudinal slices: two perpendicular slices of the head, an abdominal slice, and selected slices of a joint series. (Courtesy of the Faculty of Hospital St. Anne Brno, Radiology Clinic, Assoc. Prof. P. Krupa, M.D., Ph.D.)

chapters of Part III, can be applied. A manual or (semi)automatic contrast adjustment and other enhancement procedures (Section 11.1) are standard. In the three-dimensional data, arbitrary slice visualization procedures may be used or three-dimensional rendering applied to provide single-screen or even stereo visualization of the anatomic structures, possibly with a selective suppression or removal of certain tissues. This is naturally also heavily dependent on reliable image segmentation procedures (Section 13.2). When contrast-based imaging is performed, background subtraction techniques may possibly need the image registration (Section 10.3).

6

Nuclear Imaging

Nuclear imaging portrays distribution of radionuclides inside the patient's body by external measurement of γ-rays emanating from the body; this gave the modality its alternative generic name of *γ-imaging* or *gammagraphy* (the latter is usually used only for planar imaging). Radioactive substances of short half-lives in the range of minutes to weeks are administered to patients intravenously, orally, or inhaled. The radiopharmaceuticals then circulate in the organism and concentrate gradually in diagnosed regions, from where they are consequently excreted depending on the activities of the observed regions. The information to be obtained by an external measurement consists of generally time-dependent spatial distribution of the radioactive substance. This way, not only the shape of organs or lesions can be estimated, but also the local activity-dependent time-course of radioactivity density.

Of the radioactive decay products, only high-energy photons forming the γ-component of radiation (ranging in energy from about 50 keV to over 500 keV) are capable of penetrating the surrounding tissue and reaching the external detectors; the α and β radiation only contributes to the patient dose and should be avoided as much as possible. Obviously, the γ-photon energy range is about the same as that of diagnostic

x-rays; physically, the photons are not distinguishable. Nevertheless, the measured intensities of γ-rays, given by the low density of radioactive substances, are generally very weak, being limited by the maximum allowable dose to patients. Thus, the only difference in the character of rays, besides the different mechanism of generation, is in the photon flux density, which is several orders lower in nuclear imaging than in x-ray imaging. At these intensities, the rays do not constitute a continuous flux, but rather they consist of discrete photons that must be individually detected; the intensities are expressed by the counts of detected photons that are stochastically generated. This probabilistic character of the measurements causes problems with a high relative variance of counts and, consequently, a low signal-to-noise ratio (SNR). The photons are generated in all directions, of which only a very small part, on the order of 10^{-3} to 10^{-4}, can be measured in any projection acquisition (see below); this contributes substantially to the very low detected intensities.

It is clear that the photons are subject to the same quantum mechanisms of attenuation and scatter as described in Section 3.1.3. This contributes to complications of the measurement evaluation too: besides the required local density of the applied radionuclide, the measurement is also influenced by the unknown attenuation and scattering, which influences the photon on its way from the point of generation to the detector. It should be understood that the attenuation means a loss of a certain percentage of photons due to interactions with the matter, but the γ-photon, reaching the detector without interacting on the way, always has the exact energy characteristic for a particular used radionuclide. A lower energy of a detected γ-photon means that it is a product of Compton scattering; this is used to discriminate direct photons from the scattered ones. Naturally, if the radionuclide produces γ-photons of two or more energies, a scattered originally high-energy photon may be confused for a direct one with a lower energy.

Two types of radionuclides are used in nuclear imaging. The first type encompasses those for which the radioactive decay directly produces the individual γ-rays (single photons) in the energy range, usually not exceeding 200 keV. The other type is constituted by radionuclides producing primarily positrons, which consequently, when colliding with electrons, produce an annihilation couple of two 511-keV photons traveling in almost exactly opposite directions. The first type is used in single-photon emission imaging that may be either the simpler *planar imaging* (i.e., the two-dimensional projection imaging) or the more sophisticated *single-photon emission computed tomography* (SPECT) based on reconstruction of spatial data

from many planar projections. The positron emission-based imaging utilizes the production of the photon pair to improve substantially the collimation definition of the measured beam. This advanced method is applied namely in *positron emission tomography* (PET). We shall deal with the three mentioned modalities from the image data generation viewpoint in the following sections. For a more detailed study or for further references, see [8–10], [12], [26], [30].

6.1 PLANAR GAMMA IMAGING

Ideally, planar γ-imaging is a parallel projection of the radionuclide spatial density $\rho(\mathbf{r})$ in the patient; the image pixel values are ideally the line integrals of the density,

$$P(x,z)=\int_l \rho(x,y,z)\,\mathrm{d}y \tag{6.1}$$

(with the coordinate orientation depicted in Figure 6.1 — the solid vertical lines represent the ideal integration paths). Similarly as in x-ray planar projection, it is then up to the radiologist to use his *a priori* anatomical knowledge and three-dimensional imagination to evaluate and classify the image.

Unfortunately, in reality, it is not possible to obtain an image corresponding exactly to the plain geometry of Figure 6.1 due to several reasons, which may be subdivided into two groups:

1. *Deterministic phenomena*:
   - The measured counts are given not only by the radioactivity distribution ideally integrated as in Equation 6.1, but also by the spatially dependent attenuation $\mu(\mathbf{r})$ so

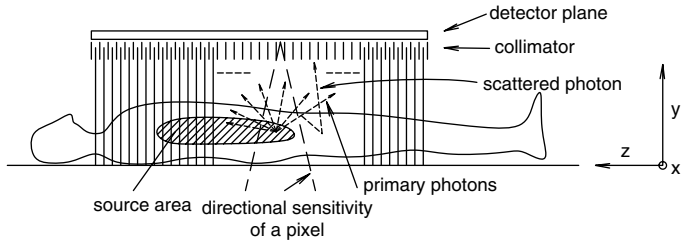


Figure 6.1 Basic geometry of planar gamma imaging. The solid vertical lines are the ideal integration paths.

that, in the simplest case of μ spatially invariant inside the object,

$$P(x,z) = \int_0^b \rho(x,y,z)\, e^{-\mu(b-y)} \mathrm{d}y\,, \tag{6.2}$$

where b is the y-coordinate of the upper surface of the object—the patient's body.

- Imperfect collimation, so that not only the parallel ray integrals of the proper direction are measured, but each image pixel is influenced by a diverging cone of rays (dashed lines), which contributes to a nonpoint, depth-dependent point-spread function (PSF) of the system.
- Interpolation-based localization of a γ-photon interaction on the detector plane (see below) provides only limited s.c. intrinsic resolution, which also contributes to the widening of the total system PSF.

2. *Stochastic phenomena*:

- Radioactive decay is a stochastic (Poisson) process so that the measured pixel values—generally low counts of photon detection, as given by very low detection efficiency—suffer with a high variance, meaning a strong noise component; the image thus appears mottled.
- Due to scatter phenomena, some γ-photons may arrive from other directions than would correspond to the point of origin, though with lower energies (dotted piece-wise straight line as an example in the figure).
- The localization of a detected photon and amplitude response discrimination may be falsified by accidental coincidence of more simultaneously detected photons, by stochastic distribution of emitted light photons among photomultipliers, and by multiple light-emitting locations per a γ-photon due to Compton scattering in the crystal (see below).
- Higher counts may be distorted due to dead time of the detectors, causing some photon interactions to pass undetected.

All the stochastic phenomena are contributing to the error of measurement. While it is possible to predict the mean and variance of the errors, the concrete error components of a pixel value are random, thus contributing to the image noise.

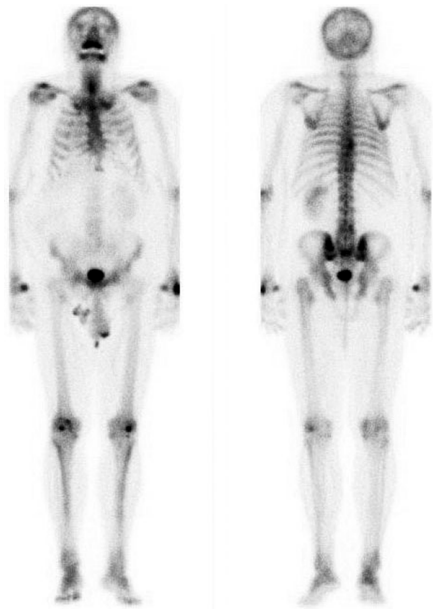


Figure 6.2 An example of a planar gammagram (left, anterior; right, posterior). (Courtesy of the Faculty Hospital Brno-Bohunice, Department of Nuclear Medicine, Assoc. Prof. J. Prasek, M.D., Ph.D.)

Due to all those distorting phenomena, the interpretation of the obtained images is less straightforward than in planar x-ray imaging. Further on, we shall briefly analyze the above-mentioned influences taking into account the principles of the measuring equipment conventional today—the gamma camera, which can produce a digitized two-dimensional output. Examples of typical planar gammagrams are in Figure 6.2. In principle, similar data can be obtained by a single scanning detector, though it is much slower, so that its use for dynamic studies is limited.

6.1.1 Gamma Detectors and Gamma Camera

On the difference from x-ray detectors, gamma detectors deal with an extremely low photon flux that is therefore discretized into individual photons. It is thus necessary to detect the photons individually, which requires extremely sensitive and fast responding equipment—a scintillation detector with a photomultiplier, as in Figure 6.3. The detector consists of a scintillation crystal, about

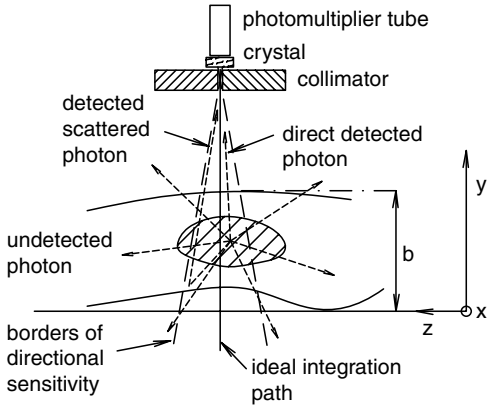


Figure 6.3 Scintillation gamma detector above the measured object.

1 cm thick, which produces a light impulse per detected γ-photon. Of this impulse, only a few photons reach the photocathode of the photomultiplier tube; it is the purpose of the tube to amplify the low number of emitted electrons by several orders, thus producing a measurable current impulse. The height of the current impulse should be proportional to the detected γ-photon energy in order to allow discrimination of the photons and elimination of the unwanted ones (with a lower energy due to scatter) from further analysis. The result of a measurement is the count of incident γ-photons.

In order to detect (approximately) only photons coming from the area of the projection line (the desirable integration path in Equation 6.2), the crystal must be preceded by a collimator—in principle an opaque (lead) block with a hole, of which the diameter and length (thickness of the collimator) determine the spatial characteristic of the detector. As can be seen from the figure, the spatial sensitivity area is approximately a cone with the vertex at the collimator, the apical angle of which depends on the collimator geometry. The width of the corresponding sensitivity in the perpendicular (horizontal) area is thus dependent on the distance (depth) from the detector, instead of being ideally equal (and small) for any depth. Obviously, only a very small portion of the electrons emitted omnidirectionally in the sensitivity area have the direction enabling passage through the collimator, and consequently the detection; this is the reason for the low measured counts. On the other hand, some of the photons generated out of the sensitivity area may also reach the detector due to Compton scatter. These

photons may contribute to the count error when not rejected based on their lower energy and, consequently, lower detector response.

A gamma camera is, in the first approximation, a planar (two-dimensional) set of gamma detectors enabling measurement of a complete projection at once, in parallel instead of by gradual scanning. It is unfortunately not feasible to construct an array of individual detectors dense enough to provide a sufficient geometrical resolution in the image, due to a still relatively large diameter of the photomultiplier tubes. An alternative ingenious principle is used in the assembly as in Figure 6.4 (Anger, 1958[55]): the crystal is not divided into individual detectors; instead, a large planar crystal is used, again about 1 cm thick, and the phenomenon of light scatter in the crystal is utilized to localize the position of the γ-photon interaction, as will be described later. The light generated anywhere in the crystal may be detected by several photomultipliers of a layer (containing up to several tens of multipliers), arranged above the crystal in a dense, nowadays mostly hexagonal, arrangement (Figure 6.4, right). The collimator (now mostly made of shaped lead foil) is in principle a (thick) planar body with transversal holes densely distributed; each identical hole represents its own directional sensitivity, as in Figure 6.3, thus approximating the desirable planar-projection system, as in Figure 6.1. This approximation is better the narrower and longer the holes are, allowing only the photons closely aligned with the projection direction to reach the crystal. Nevertheless, a compromise must be found, as good collimation also means a high rate of γ-photons hitting the septa of the collimator, which are thus lost for the measurement. Good photon effectiveness requires as thin septa as possible, but this may contribute to random penetration of some photons through collimator septa, which increases the measurement error.

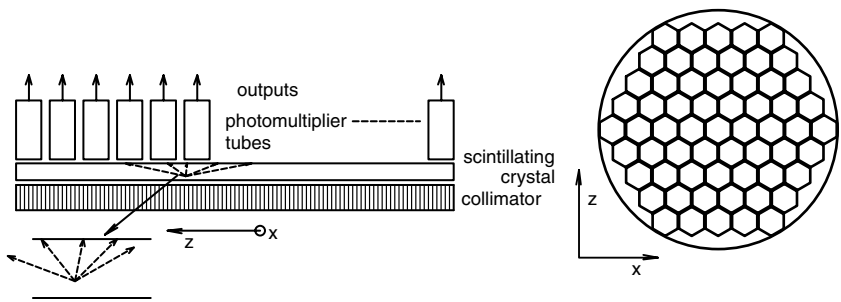


Figure 6.4 Principle of gamma camera: (left) side view and (right) from above.

Often, a selection of several collimators is available for a camera with a different degree of collimation and photon efficiency. It should also be mentioned that collimators exist in which the holes are not parallel but are converging to a focus point in front of the camera, or diverging, i.e., focused behind the camera crystal. They allow either certain enlargement or reduction of the object image on the crystal, dependent on the measuring depth (Figure 6.5), and consequently better coverage of small imaged organs or of a greater part of the body, respectively. Naturally, a converging collimator becomes diverging when reversed. Nevertheless, both types are rarely used, as the magnification varies with depth, making the image interpretation more difficult. The pinhole collimator belongs

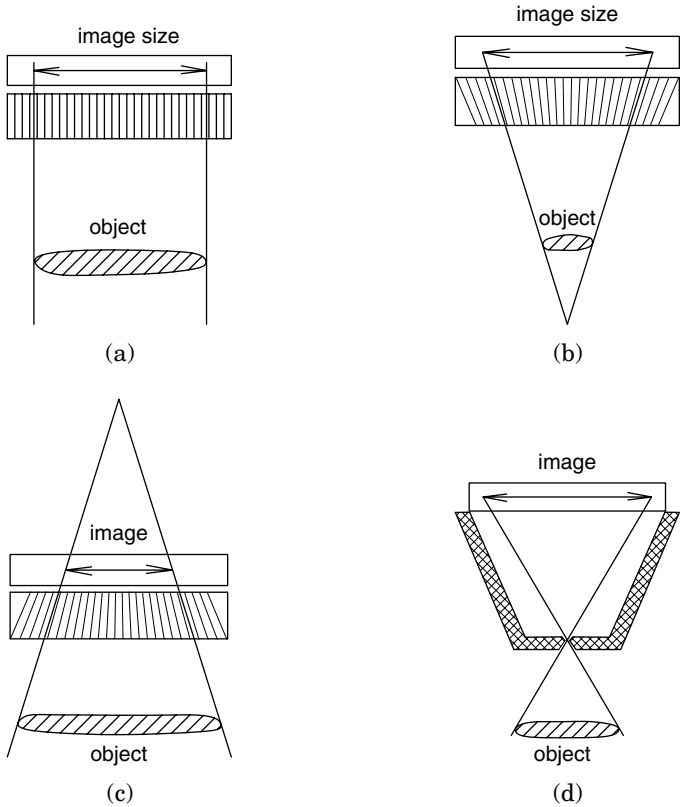


Figure 6.5 Different types of collimators: (a) parallel, (b) converging, (c) diverging, and (d) pinhole.

to the magnifying type; it provides a precise image of the object (again with depth-dependent magnification) with almost point-wise PSF of the collimator, but it has extremely low photon efficiency. Of the nonparallel hole collimators, only the fan-beam collimator, with holes focused in the (x, y)-plane while parallel in any perpendicular plane, is commonly used in SPECT, where it provides better image quality.

An integral part of a gamma camera is a computational block, which discriminates the responses according to the detected γ-photon energy and calculates the localization of the photon incidence on the crystal plane. In modern cameras, the output of individual photomultipliers is digitized and the data then become a subject of digital processing, which must be sufficiently fast to cope with the requirements of individual photon detection and short dead time. The results are then saved in digital form, either as an image matrix, with the pixel values giving the photon counts (sc., *frame mode*), or as a record of individual photon interactions, containing the (x, z)-coordinates for each event, complemented periodically by timing information (sc., *list mode*). The latter is more flexible, but also more demanding, in both memory requirements and computational complexity regarding the necessary reformatting to image matrix form.

When imaging periodically moving organs such as the heart, it may be necessary to provide a sequence of images corresponding to different phases of the movement. If the individual phases do not last long enough to enable acquisition of partial images with a reasonable SNR, it is possible to accumulate a separate image data set for each phase gradually, during many repetitions of the periodic movement. The time axis is then subdivided into short periods corresponding to individual phases of the movement (e.g., in cardiological studies, to heart cycle phases defined by different delays after the R-wave of ECG). The data ((x, z)-positions of count increments), acquired continuously during many cycles, are then distributed in real time to individual image matrices (one for each phase) by a gating mechanism, controlled by the physiological signal. This imaging method, which relies on regularity of the movement, is called *gated mode*.

It is obvious that posterior construction of the gated-mode image sequence is also possible from data recorded in the list mode should the record contain time information and the controlling signal (or derived synchronization data).

6.1.2 Inherent Data Processing and Imaging Properties

6.1.2.1 Data Localization and System Resolution

The resulting PSF of the camera, determining the system resolution, depends on two components:

- The PSF of the collimator
- The PSF of the inner camera (crystal + photomultiplier array), determining the *intrinsic resolution* and given by crystal properties, by arrangement and stability of photomultipliers and also by the localization algorithm (see below)

Ideally, the collimator should provide a precise planar (parallel or focused) projection of the object radioactivity distribution, expressed by local counts of γ-photons reaching the input plane of the crystal. The quality of camera collimation is reflected by the response of the γ-photon image to a point source (the collimator PSF) or a line source (the collimator LSF, of which a perpendicular profile is evaluated). As visible from Figure 6.6, the PSF of the collimator depends on the depth of the source, as it obviously corresponds to the width of the directional pattern of a hole and may also be marginally influenced by septa penetration. Obviously, the spatial resolution and the collimator photon efficiency are conflicting features; the choice of the compromising hole parameters has been discussed above. When neglecting

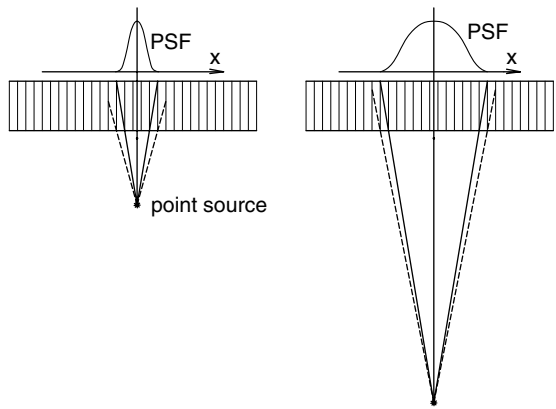


Figure 6.6 Depth-dependent point-spread function of a parallel-hole collimator.

the penetration, the directional pattern and its full-width-at-half-maximum (FWHM) R_c, characterizing the resolution, can easily be calculated based on the geometry of the collimator. The typical values of R_c are in the range of millimeters (1.5 to 15 mm at the depth range of 0 to 20 cm from the collimator).

Because the spatial resolution given by the size of photomultiplier cathodes is definitely insufficient, the position of the light emission (i.e., γ-photon interaction in the crystal) has to be calculated from the simultaneous responses of the neighboring photomultipliers. This localization method is based on the assumption that no more than a single γ-photon is interacting with the crystal at a time. This leads to generation of many (multidirectional) light photons simultaneously reaching a subset of the photomultipliers (Figure 6.4). When the light-spreading pattern from a point source inside the crystal in the (x,y)-plane is known, the response magnitudes of the individual photomultipliers could be predicted, knowing the light source position. By inverting the problem, the position of the γ-photon incidence can be calculated from the known responses, and the photon count of the corresponding pixel then incremented.

Instead of a more precise nonlinear formulation, the weighted average of x- and z-coordinates of centers of individual photomultipliers is often used as the first positional estimate, with the weights given by the individual response amplitudes. Very weak responses of (supposedly) distant photomultipliers are usually given zero weight in the average, in order not to degrade the position estimate by the high amount of noise in those responses. Although the spatial range of involved photomultipliers obviously depends on the depth of the photon interaction in the crystal thickness, the localization result remains basically intact, as the depth influences only the size but not substantially the shape of the light-spread pattern on the output plane of the crystal.

This calculation, historically completed by linear analogue circuitry, can be done more accurately numerically, taking into account a more realistic model of light attenuation and of geometrical relations of the tubes to the source position (distances from the light source, shape, and size of photocathodes). Alternatively, the simple linear position estimate obtained in the above-described way may be consequently improved via lookup tables, based on practical measurements of a phantom with known geometry. The tables are two-dimensional, their input being estimated (x,z)-pairs; two separate tables provide corrections of both coordinates.

The positioning algorithm may be fooled by random Compton scattering of the incident γ-photon in the crystal material, leading to simultaneous multiple light sources, or by simultaneous appearances of two or more original γ-photons. In both cases, the position will be determined as about the center of gravity of both (or more) light scatter patterns, which obviously is not the proper position. These improperly positioned pulses contribute to image noise.

The resulting resolution of the inner camera can be characterized by FWHM R_i of the inner PSF, typifying the intrinsic resolution. The measurement of this parameter is possible when replacing the collimator by an opaque (e.g., lead) thin sheet with a small hole (for PSF) or a narrow linear slit (for LSF), placed on the input plane of the crystal, and by irradiating this arrangement by a sufficiently distant point source of radioactivity. By moving the sheet on the crystal input plane, thus shifting the point or line source, it is possible to check the inner isoplanarity.

Notice that when the sheet is removed, the crystal becomes evenly irradiated and the measurement determines the homogeneity (uniformity) of sensitivity on the image field should the point source of radioactivity be sufficiently distant from the crystal's front side.

The FWHM R_s of the complete system PSF, characterizing the resulting resolution*, can be computationally approximated by

$$R_s = \sqrt{R_c^2 + R_i^2} \ . \tag{6.3}$$

The typical values for modern systems are in the range of several millimeters (about 4 mm at the front of the collimator) to centimeters in the maximum practical depth. It can be measured by imaging a point (or line) radioactive phantom by the complete camera, including the collimator. The way of checking isoplanarity, by moving the phantom in a plane parallel to the camera input plane, is obvious. A more common test method is based on planar radioactive phantoms containing linear, circular, or stripe patterns, enabling the checking of the resolution and often also the geometrical linearity of imaging. The phantoms of point or line sources, formed of radionuclides producing two or three different photon energies, also enable the checking of the proper positioning properties of the camera at different response levels—the component images provided

*Often, R_s (in m) is improperly denoted as resolution; it is obviously a quantity inversely proportional to resolution (in m^{-1}).

at the individual levels should be in register, so that the resulting image would not show multiple points or edges.

6.1.2.2 Total Response Evaluation and Scatter Rejection

As already mentioned, it should be understood that even if the gamma camera had ideal directional properties, it would detect not only the desired direct γ-photons, but also the γ-photons generated by Compton scattering phenomena (Figure 6.1). These photons are approaching the camera at positions other than those corresponding to proper radioactive source locations, and their contribution to a particular position on the detector crystal (and consequently to a concrete image pixel) is thus erroneous. They should therefore be excluded from the image formation, which can be done by energy discrimination via total amplitude of the camera response. The decision on the character of the photon is made based on the amplitude of the total response from all simultaneously responding photomultipliers. Providing linearity of detectors, the total response is calculated as the sum of all synchronous responses. If the total response is in the amplitude band corresponding to a direct γ-photon produced by the applied radionuclide (i.e., in sc. photopeak window), the individual measurement is accepted; otherwise, it is rejected.

The total response may be distorted due to the camera field inhomogeneity (nonuniformity). The position-dependent sensitivity can be compensated by multiplying the total response amplitude (sc., y-coordinate) by an experimentally determined position-dependent factor using another two-dimensional lookup table with the (x, z)-coordinates as inputs. Naturally, the correction should be done before the amplitude tests.

It should be mentioned that an amplitude test might be misled by simultaneous detection of more than a single scattered γ-photon. Two (or more) γ-photons emitted and scattered in a period shorter than the time resolution of the camera may contribute to a seemingly high enough detected energy level, which would increase the count of a pixel erroneously (and in an improper position, as already mentioned). Some radionuclides produce more than one type of γ-photons; in such a case, the total response must be compared with multiple pairs of limits (naturally with a certain risk of confusing a scattered originally high-energy photon with a direct one of lower energy).

6.1.2.3 Data Postprocessing

The quality of gammagrams is limited by the low spatial resolution and high level of noise due to the statistical character of the measured values. One simple way of image correction used is field homogenization (see Section 12.1.2) based on data from a routine uniformity check. If the geometrical linearization is not a part of inner camera processing, it can be done subsequently in postprocessing by the methods of geometrical restitution (Section 12.2) based on parameters provided by imaging, e.g., a rectangular grid phantom.

Contrast transforms (Section 11.1) can be applied to gammagrams as well as to any other modality images. A suitably designed lookup table may provide for background activity suppression (clip-off of pixel values under a certain threshold). Similarly, nonlinearity at high counts caused by dead time in measurement may be approximately corrected by increasing the contrast at the upper end of the pixel value range.

As for application of filtering methods, only smoothing filtering is commonly used, which is intended to suppress the statistical noise, though usually at the cost of further impacting the spatial resolution. To improve the resolution of gammagrams via some kind of inverse filtering (deconvolution) does not seem to be promising due to poor SNR of the measured images, but some relatively successful attempts were made in the use of gamma cameras in SPECT (see below).

During activity studies, image series are provided that carry the information on time development of radioactivity distribution. An important and often used method of this data processing is to derive time profiles of activity at individual pixels or sum/average time-courses in manually or automatically specified areas of interest. Another way to assess the time development is to derive difference images by subtracting the data of the same area, but from different time instants after administering the radionuclide; here, registration of successive images (Section 10.3) may be needed.

6.2 SINGLE-PHOTON EMISSION TOMOGRAPHY

6.2.1 Principle

Gamma imaging is generally based on measuring one-dimensional or two-dimensional projections of three-dimensional radionuclide distribution, in the simplest case according to Equation 6.2. In principle, it is possible to measure a high number of such projections sufficient for consequential reconstruction of the source function—the

radionuclide distribution—by means of the respective methods (Chapter 9). The single-photon emission computed tomography (SPECT), as an individual imaging modality, is based on this idea.

Basically, it is possible to provide the measurements by any arrangement of gamma detectors; even a single detector would do if it is moving in a proper way, thus providing projections as sets of line (ray) integrals—see the principle of x-ray tomography (Chapter 4). A problem is the duration of such an acquisition method, considering the time needed to obtain a reasonable count of photons in each individual line measurement. Therefore, arrangements with multiple detectors are used that allow measurement of many line integrals in parallel. They may be either specially designed multidetector systems or systems based on standard gamma cameras, as described in the previous section. As the cameras have achieved a high degree of sophistication, most contemporary SPECT imaging uses such cameras, either modern single-camera arrangements intended primarily for planar gamma imaging that also allow tomographic positioning, or systems designed primarily for tomographic imaging, with two or more cameras (Figure 6.7). We shall restrict ourselves only to camera-based SPECT imaging; from the image reconstruction point of view, it is not a limitation. Use of a planar gamma camera allows the acquisition of the data with a high parallelism; not only a complete one-dimensional transversal projection is provided, as indicated in Figure 6.8, but the two-dimensional camera provides a parallel set of such projections thanks to its second (axial, z) dimension. Obviously, multislice imaging (providing a series of parallel transversal slices) is possible this way, if all the measured data are properly processed.

If the camera were providing ideal projections, we could stop here and refer to the chapter on x-ray computed tomography, as there is no principal difference. In both cases, line integrals are measured of a spatially distributed parameter—attenuation in CT and radioactivity in SPECT—and from these measurements, the distribution could be reconstructed. Unfortunately, the real situation in SPECT is far from ideal, in contrast with the case of CT.

6.2.2 Deficiencies of SPECT Principle and Possibilities of Cure

The problems of SPECT imaging are partly given by the deficiencies of the used gamma cameras (collimation, scatter influence, nonuniformity of sensitivity, geometric distortion). All the correcting measures mentioned in the previous sections are naturally applied in

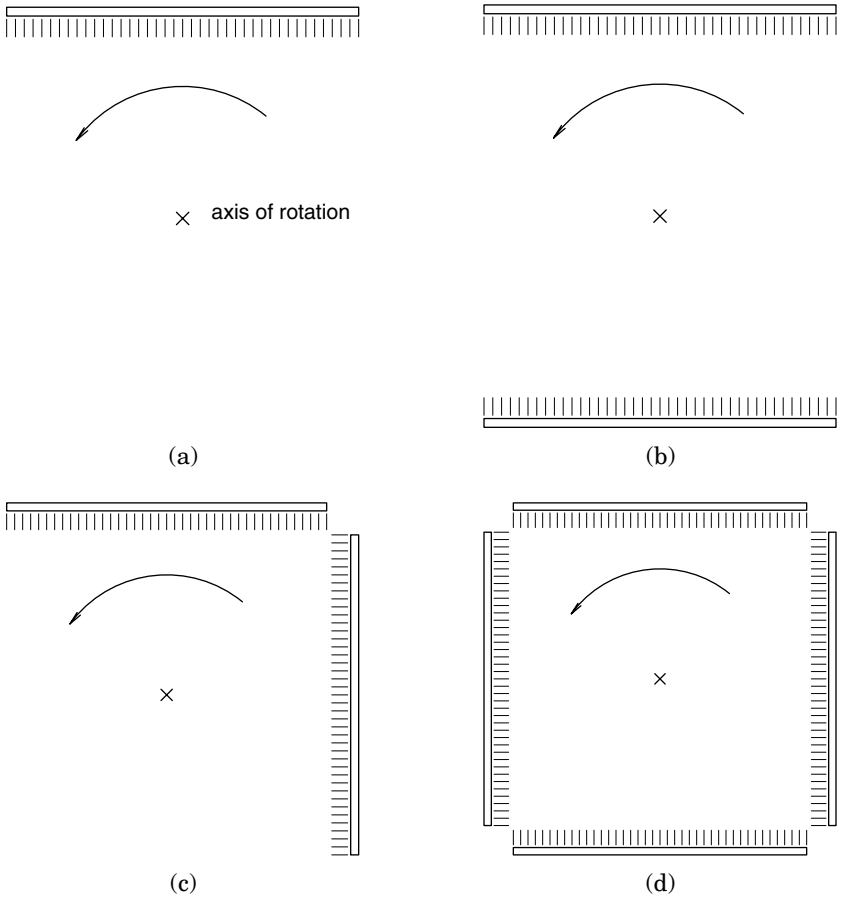


Figure 6.7 Camera-based SPECT systems schematically: (a) single-camera system, (b) dual-camera system with opposite cameras, (c) dual-camera system with perpendicular cameras, and (d) quadruple-camera system.

cameras used for SPECT as well, or even in a more sophisticated manner, required or enabled by way of SPECT data acquisition. Besides that, some phenomena, namely, attenuation, affect substantially the measured data and must be taken into account more thoroughly, even in the sense of modifying the methods of image reconstruction.

The first problem is collimation — the definition of the measuring integration path (or rather volume). As explained in the previous section, the measured volume areas corresponding to individual

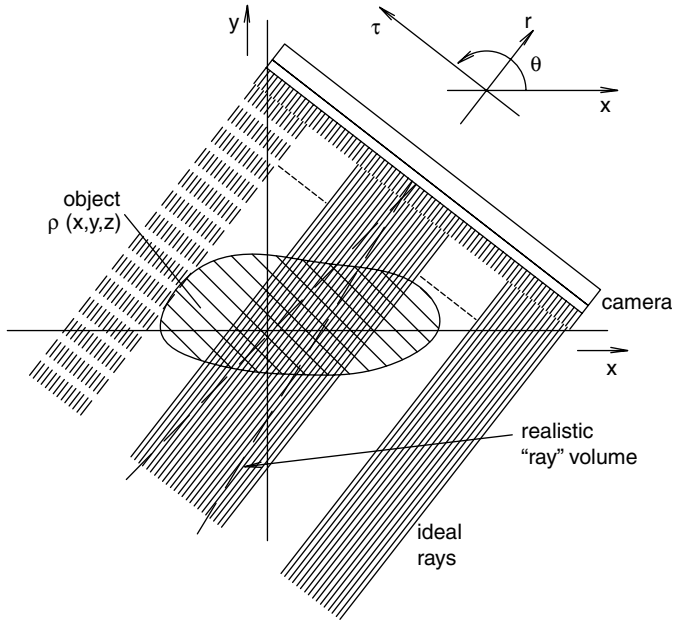


Figure 6.8 Scanning of projections by a gamma camera.

two-dimensional image pixels have the shape of cones with vertices approximately at the detecting crystal surface (dashed lines in Figure 6.8). In order to approximate the structure of parallel integration lines reasonably, high-resolution collimators should be used with long and narrow holes; this unavoidably leads to low photon efficiency. Mostly, parallel-hole collimators are used, but in the case of objects smaller than the camera input area (namely in brain imaging), the fan-beam collimators are advantageous, providing a denser line structure in the object slice and at the same time improving partly the photon efficiency; however, the slices are maintained parallel by the collimator (Figure 6.9). The camera should be kept as close as possible to the body surface, as the absolute lateral resolution is inversely proportional to the distance between the camera and the imaged source; modern systems therefore follow noncircular (body contour) orbits when rotating the camera around the object (Figure 6.10). Theoretically, this does not influence the imaging process; the distance of an ideal parallel-line camera from the object may be chosen arbitrarily for every angle θ, as far as the only contribution to the line integrals comes from the imaged area.

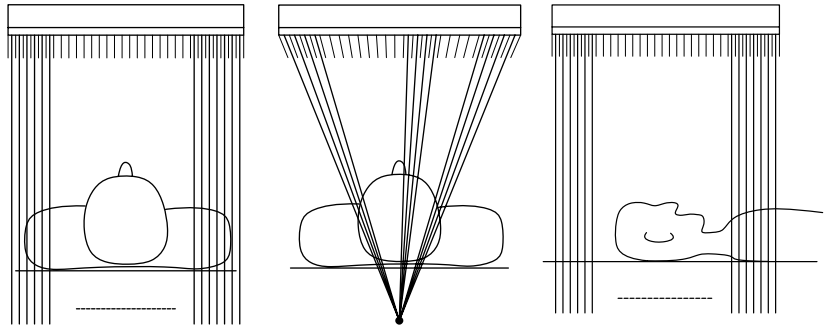


Figure 6.9 Use of (a) parallel-hole and (b) fan-beam collimators-transversal plane view and (c) common axial plane view.

Nevertheless, in reality, the noncircular orbit means uneven integration volumes for different angles, which complicates the attempts to correct the consequences of the nonideal (spatial instead of linear) integration.

The influence of unsatisfactory collimation may be interpreted as an imperfect PSF of the camera: the impulse response should ideally be a point, while in reality it is depth dependent, as mentioned before (Figure 6.6).

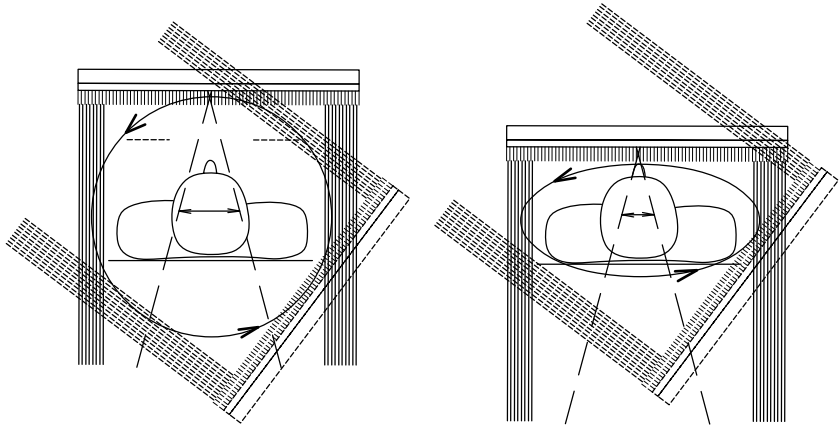


Figure 6.10 Circular vs. body-contour orbits of a gamma camera during SPECT data acquisition.

The second tough problem is the attenuation inside the imaged object. This means that the measurement integral for a vertically positioned camera has a more complicated form than that of Equation 6.1; in the simplified case of homogeneous attenuation, it is Equation 6.2, which must be further modified for more realistic inhomogeneous attenuation into

$$P(x,z)=\int_0^h \rho(x,y,z)\,e^{-A(x,y,z)}\,\mathrm{d}\,y, \tag{6.4}$$

where $A(x, y, z)$ is the total attenuation accumulated on the ray, belonging to the plane (x, y), between the radiation source point (x, y, z) and the height level of the camera input window h,

$$A(x,y,z)=\int_y^h \mu(x,y',z)\;\mathrm{d}y'. \tag{6.5}$$

In the interest of a simple explanation, the previous equations are presented for the particular position of the camera as depicted in Figure 6.3. Naturally, the expressions may easily be converted for any other position of the camera by simple coordinate rotation. Let us formulate it symbolically for a general angle θ of the camera incline with respect to its initial position as the integral along a generic ray $p(\mathbf{t})$,

$$P_\theta(\mathbf{t})=\int_{p(\mathbf{t})} \rho(\mathbf{t},r)\exp\left[-\int_r^{h_r}\mu(\mathbf{t},r')\,\mathrm{d}\,r'\right]\mathrm{d}\,r, \tag{6.6}$$

where $\mathbf{t}$ is the position vector in the plane perpendicular to the projection and r is the coordinate along the projection direction; h_r is the r-coordinate of the camera input window.

Equation 6.6 represents the *generic attenuated projection* for the position angle θ. The inverse transform, providing the radioactivity distribution $\rho(x, y, z)$ based on the measurements of $P_\theta(\mathbf{t})$ for multiple θ, is not known in a closed form. The simplified case of homogeneous attenuation, corresponding to Equation 6.2,

$$P_\theta(\mathbf{t})=\int_{p(\mathbf{t})} \rho(\mathbf{t},r)\exp[-\mu(h_r-r)]\,\mathrm{d}\,r, \tag{6.7}$$

is denoted as *exponential Radon transform*. As we shall see in Chapter 9, this simplification allows the reconstruction of transversal

(x, y)-slices, based on a modified filtered backprojection method, while the more generic approaches permitting inhomogeneous attenuation require iterative algebraic solutions based on some *a priori* estimates of $\rho(x,y,z)$. The attenuation correction is thus postponed to the reconstruction phase of the imaging, where it requires a necessary modification of the reconstruction algorithms.

The attenuation according to Equation 6.6 is naturally different in opposite views (θ differing by 180°); this leads to the requirement to measure the projections in the complete 360° range (different from x-ray CT, where 180° is sufficient). Nevertheless, in some studies, namely in cardiology, the 180° scanning proves better, as the opposite views would lead through the body area with a rather complex attenuation distribution, which is still difficult to take into account even if known (which is often not the case).

It should be mentioned here that the statistical (noisy) character of the measured data also requires some measures that should be applied in the reconstruction process. When filtered backprojection is used as the image reconstruction method, special low-resolution kernels (i.e., impulse responses to be convoluted with the projections) are used, which account for expected average (constant) attenuation in the range of field of view (FOV), as well as for the noisy data. Alternatively, iterative algebraic reconstruction methods are used that enable the inclusion of the realistic attenuation distribution during the computation. The attenuation can be measured in certain SPECT systems by irradiating the imaged object by auxiliary radioactive sources positioned opposite of the camera, as in x-ray CT, their configuration depending on the type of collimator used. This measurement is usually done simultaneously with SPECT data acquisition; the discrimination between both entries is then based on significantly different photon energy of the auxiliary source compared to the radionuclide to be traced in the object. The reconstructed transversal attenuation map (one per slice) is consequently used in the SPECT slice–image reconstruction algorithm. Some more details on the special reconstruction procedures, which take into account the attenuation distribution, are in Section 9.2.1.

Still another source of distortion is scatter of the γ-photons, as explained in Section 6.1.1. Though the cameras eliminate the photons with energy outside of the photopeak window, there is still a part of accepted photons that have been scattered, with the remaining energy falling in the window. If the individual pixel counts of these photons were known, they could easily be subtracted from each

measured value for a complete correction. Though they are not known precisely, they can be estimated based on two-window measurement, the second (scatter) amplitude window being adjusted well distant from the photopeak window; the correction value, though belonging to a different energy range, may be supposed proportional to the measured scatter count. This technique can be further sophisticated based on multiple windows or complete energy spectrum analysis, followed by optimization. Naturally, the correction should be made in individual pixels of projections before the reconstruction; this approach reduces the detrimental influence of spatially variant scatter phenomena on the reconstruction process. The primitive method of subtracting a constant count (or a constant percentage of counts) from all pixel values may be considered a rough approximation of the above method.

A different approach to compensation of scatter is based on the supposition that the scatter deteriorates the PSF of the camera, so that a kind of inverse filtering (Section 12.3) might improve the measured projections.

The resolution limits of SPECT are given by the resolution capabilities of the used gamma camera(s); the potential to influence it by a kind of inverse filtering is similar to that mentioned in the previous section. An example of a set of SPECT slices is in Figure 6.11.

Some image artifacts may be caused by imprecise measuring geometry of the SPECT arrangement, namely, by incorrect alignment of the camera positions with the theoretical axis of rotation, and by a tilt of the camera head with respect to the axis (nonperpendicular integration paths). Conversely, it is possible to identify these imperfections from images of particular phantoms and to use this information for corrections, but the details are already beyond the scope of this book.

6.3 POSITRON EMISSION TOMOGRAPHY

6.3.1 Principles of Measurement

Positron emission tomography (PET) is a sophisticated imaging method enabling the provision of relatively precise two-dimensional or three-dimensional data on spatial distribution of a radionuclide inside the object, e.g., a patient's body. It utilizes radioactive decay of a particular type that involves positron emission.

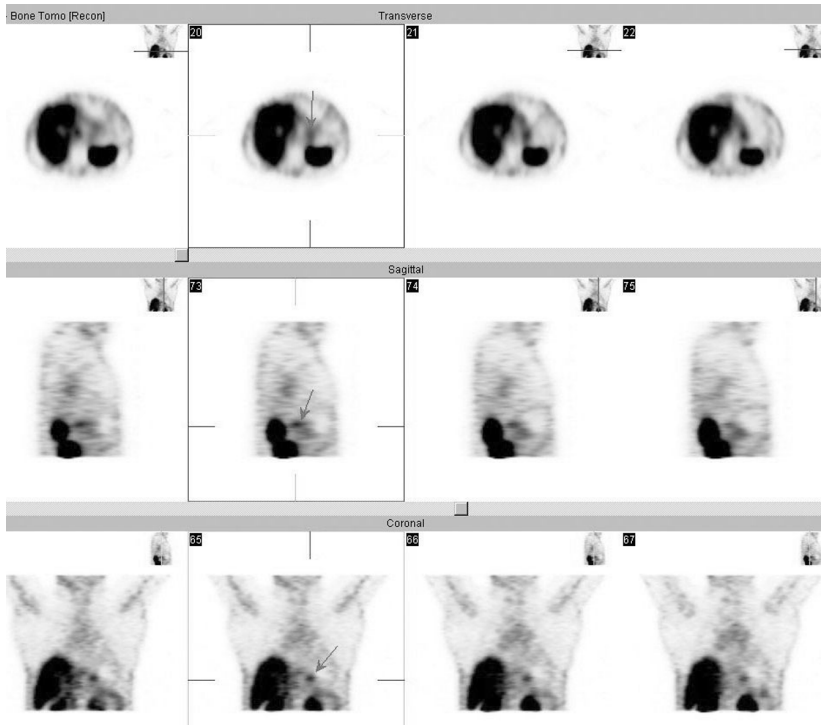


Figure 6.11 Example of a set of SPECT slices (note the slice orientation indicated on miniature torsi). (Courtesy of the Faculty Hospital Brno-Bohunice, Department of Nuclear Medicine, Assoc. Prof. J. Prasek, M.D., Ph.D.)

An individual phenomenon of such decay consists of several phases important from the imaging aspect (Figure 6.12). Initially, a nucleus undergoes a decay producing (besides other decay products) a positron with a certain kinetic energy. Though positrons, as antimatter particles, tend to annihilate in the environment of matter particles, this cannot be accomplished before the positron loses the kinetic energy by colliding with the surrounding medium, thus traveling a certain distance from the place of its origin. The distance is random and its mean value, dependent on the type of the radionuclide as well as on the surrounding matter, is called *mean positron range* (about 1.4 to 4 mm for different radionuclides in water or soft tissue; less in bones, but much more in lungs). Because the position of the decaying nucleus should be imaged while the location of the

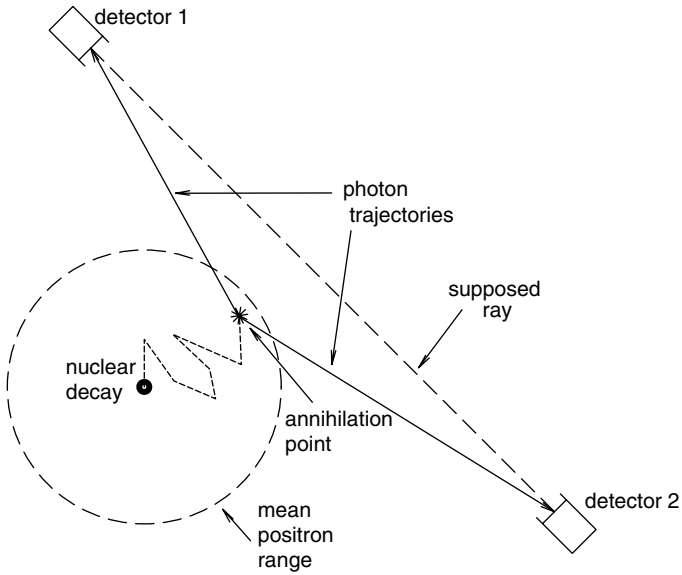


Figure 6.12 Photon-pair emission in a positron-decay radionuclide and its detection.

annihilation is obtained instead, this uncertainty causes a blur in the resulting image denoted as *positron range blurring*.

The positron finally meets with an electron, which leads to annihilation, producing two high-energy (511-keV) photons traveling in exactly opposite directions. The movement of the positron and electron before annihilation leads to the result that, in the imaging coordinate system, the angle between both photon trajectories need not be exactly 180°, but may differ randomly in the range of about ±0.5°. This angle is neither measurable nor predictable; thus, the best estimate is the zero mean, i.e., the assumption that the photon trajectories lie on a single line—a ray—crossing the annihilation point. Nevertheless, the related uncertainty contributes to the blur of the image as *noncollinearity blurring*.

The image data measurement is based, in principle, on coincident detection of both photons by two opposite detectors, as photon travel time of the order of 1 nsec is negligible. Thus, a pair of photons is supposed to be a product of an annihilation event, if the photons are detected at the same time (more precisely, inside a used time window

Δt allowing for inexact equality of both detector responses, that is, on the order of 10 nsec). Besides such pairs forming the measured data—true coincident counts—the detectors respond also to individual photons hitting them nonsimultaneously; these events are not considered in the raw count. Nevertheless, they are detected, contributing to a *single-count rate* for each detector: S_1 and S_2 for detectors D_1 and D_2, respectively. Given the nonzero time window Δt, there will obviously be a certain amount of *random coincident counts*, which appear when two independent photons are registered simultaneously, though they originated from different annihilations. The random count rate R_c is obviously proportional to the time window and both single-count rates,

$$R_c = 2\Delta t S_1 S_2 \tag{6.8}$$

The random counts would add an error to every measured value so that a correction is desirable and usually done based on auxiliary measurements (see below).

The detectors used in PET imaging are basically the same as those used in gamma imaging, including SPECT—scintillating crystals with photomultiplier tubes. Naturally, they must be adapted to a higher energy of photons; i.e., crystals with a higher atomic number are used and their dimensions along the expected photon trajectory are greater to ensure high enough photon efficiency. This is very important: a coincident detection requires that both photons interact with the crystal matter and be detected. As both interactions are independent, the probability of the coincident event being detected is given by squared photon efficiency of a single detector that should therefore be close to 1.

Not only does the detection concern unscattered pairs of collinear photons (as is the ideal case), but also the coincident detection may appear for a pair of photons that, though originating from a single annihilation, underwent a Compton scatter (one or both of them, single or multiple scatter). They are then not moving along the original ray, and the line connecting the detectors does not contain the annihilation point, which leads to false image data—*scattered coincident counts*—increasing the measurement noise. The scattered count rate can be substantially reduced by amplitude-discriminating the detector responses, as in gamma cameras (Section 6.1.1), thus rejecting the photons with energy outside of the 511-keV photopeak window.

The measurement should provide a quantity proportional to the total radioactivity on the line between the detectors, i.e., the

line integral of the radionuclide spatial density, generalized from Equation 6.1, i.e., along the line D connecting both detectors,

$$P = \int_D \rho(r)\,\mathrm{d}r. \tag{6.9}$$

This would be given by the coincident counts if there were no attenuation.

Nevertheless, as in previous sections the emitted γ-photons are attenuated in the sense that there is only a certain probability less than 1 of the photon reaching the respective detector even when traveling originally in the proper direction. This probability is given by the attenuation a on the way d between the annihilation point and the detector input window. For the photon reaching the first detector, D_1, of the respective couple, this is expressed by the line integral of the spatially variant attenuation coefficient $\mu(x,y,z)$;

$$a_1 = \exp\left(-\int_{d_1} \mu(r)\,\mathrm{d}r\right).$$

The probability a_2 of the second photon of the pair reaching the detector D_2 is given by a similar expression, with the integral over the way d_2. As the attenuation events of photons are independent, the probability of both of them reaching the detectors simultaneously is

$$a_1\,a_2 = \exp\left(-\int_{d_1} \mu(r)\mathrm{d}r - \int_{d_2} \mu(r)\mathrm{d}r\right) = \exp\left(-\int_D \mu(r)\mathrm{d}r\right) = a_{\mathrm{ray}} \tag{6.10}$$

substantially lower than 1. This means that the resulting ray count P_{ray} is decreased, with respect to P, proportionally to a_{ray}, which is obviously the total attenuation on the path (ray) between the detectors,

$$P_{\mathrm{ray}} = a_{\mathrm{ray}} \int_D \rho(r)\,\mathrm{d}r. \tag{6.11}$$

We arrived at an important conclusion: the attenuation, while definitely not negligible, is independent of the position of the annihilation on the ray—all the events on the ray are equally influenced by the total ray attenuation. It is then much easier to compensate for the attenuation in PET imaging than in the case of SPECT. Naturally, the attenuation on each ray must be known or determined prior to or after the image data acquisition.

The count is also influenced by the photon effectiveness of used detectors. As the effectiveness with respect to synchronous pair detection is the squared efficiency of a single detector, this individual parameter must be close to 1—much higher than usual for gamma camera detectors. This requirement is in conflict with the need of small size of the detector crystal, as this defines the ray thickness; also, the higher energy of annihilation photons increases the penetration, thus decreasing the probability of interaction with the crystal. Nevertheless, the parameters of detectors are usually given by the system manufacturer and are only indirectly involved in the data processing.

The most important improvement of PET in comparison with SPECT measurement is the much better collimation: the integrating path—the ray—is fully determined by the position of both coincident detectors, and no collimators similar to those used in gamma detectors (Figure 6.3) or gamma cameras are needed. Also, the cross-section of the ray is given only by the size of input windows of the detectors; it does not increase with depth and is thus a much better approximation of the ideal narrow beam.

6.3.2 Imaging Arrangements

In principle, it would be possible to mount the couple of detectors on a shift-rotate mechanism as in Figure 4.1 and to provide, step by step, a sufficient number of parallel projections as sets of gradually shifted ray counts, for different rotation angles. From these one-dimensional projections of a two-dimensional slice, the two-dimensional cross-sectional image of the spatial radioactivity distribution could be reconstructed. Practically, it would be a heavy-going practice requiring a long time, as only a very small fraction of emitted photon pairs is detected; as the direction of a photon pair trajectory is random, most of them leave unobserved.

Several different arrangements have been designed, some of them based on a couple of opposite gamma cameras equipped for coincident detection. As all the arrangements provide the data on the same principle, we shall restrict ourselves to the most common dedicated PET system, the basis of which is a ring of fixed detectors, as schematically depicted in Figure 6.13.

Let us discuss first the simplest case of a single ring formed of densely situated detectors, without any collimators. A ray can be defined by any couple of the detectors, examples being indicated by connecting lines in the figure; each such ray is described by its (τ, θ) coordinates in the Radon space. It is obvious that a set of (almost)

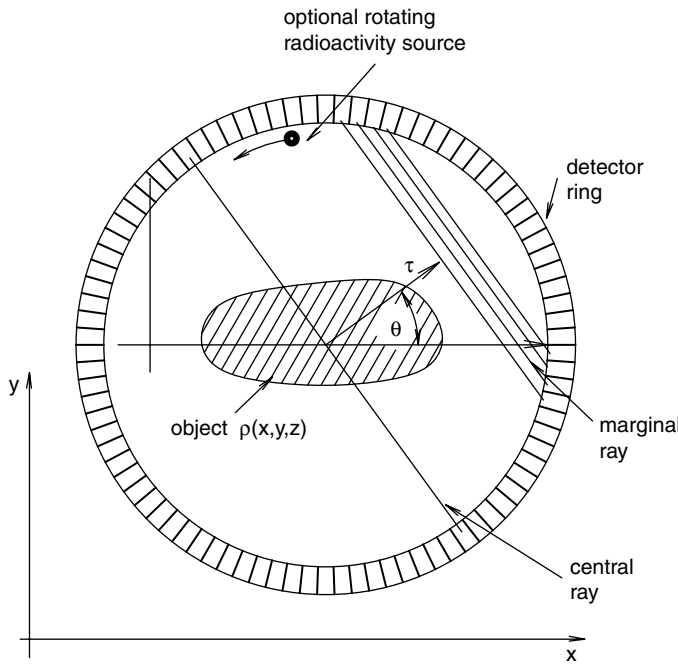


Figure 6.13 PET detector ring schematically.

parallel lines can be found for each of a number of θ angles; the set of counts corresponding to the parallel rays forms a projection. All the detectors in the ring are working simultaneously, but the data are acquired sequentially: a positron decay event in the plane of the circle may generate a photon pair traveling in the plane; such an event is with a high probability detected as a coincidence of two detector responses and should be registered in the system memory. As the directions and positions of the rays belonging to the individual events are random in the measuring sequence, each possible couple of detectors defining a ray has its own bin (location) in the memory that is incremented when the corresponding detection appears. Obviously, the size of the memory is rather large—a ring with $N = 300$ detectors defines theoretically $N^2 = 90{,}000$ rays (though in practice, the number is somewhat lower, as the rays formed by close detectors are not well defined and should be omitted).

The conflicting requirements on the detector properties result in a compromise, as depicted in Figure 6.13 and Figure 6.16: in

order to achieve a good definition of the rays, the number of detectors on the ring must be high and their width therefore low; on the other hand, to achieve a high photon efficiency, the detector crystal should be thick—thus the radial depth of crystals is rather large. Obviously, the efficiency of the detectors with respect to γ-photons with oblique trajectory (arriving from the side) is low and the photon may penetrate several neighboring crystals before being detected. This has consequences, mentioned at the end of this section.

The ring with simultaneously working detectors is obviously much more effective from the viewpoint of photon efficiency than a single detector couple, as all photon pair directions aligned with the slice plane are utilized. Nevertheless, as the photon direction is random, most of the pairs still escape unnoticed. A logical next step in design considerations would be to replace the ring with a spherical (ball) arrangement of detectors, capable of detecting all the ray directions. This is practically unfeasible for obvious reasons, but at least a partial improvement, still enabling good access inside, is possible if more rings are grouped in parallel, thus forming a multilayer detector array (Figure 6.14), today mostly limited to about seven to nine layers. Such a detector array can act primarily as a multislice scanner, when each ring is treated separately; however, it can also be considered an approximation of a partial ball arrangement when interslice rays are

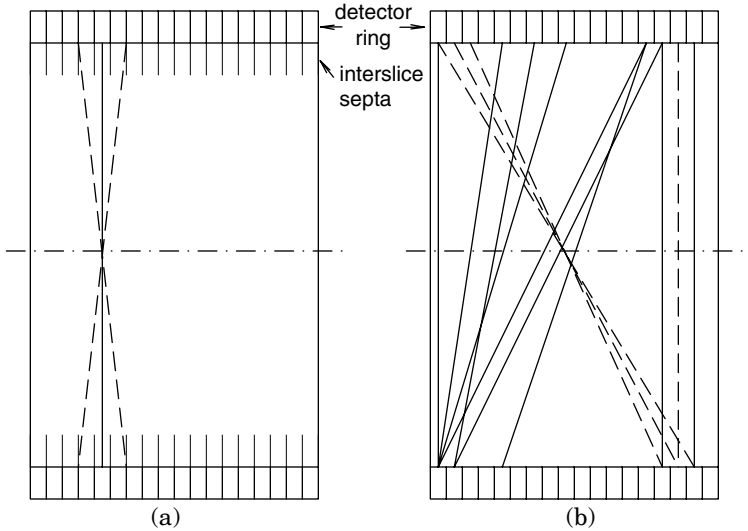


Figure 6.14 Cross-sectional view of a multislice PET detector array with (a) layer-separating collimators and (b) without collimators.

accepted as well, as indicated by skew lines in Figure 6.14b. The total photon effectiveness of the array is higher the greater is covered part of the complete solid angle ($4\pi^2$) which obviously depends on the thickness and number of slices. At the same time, a higher coverage also means more data for true three-dimensional reconstruction, although the measurement is obviously always incomplete and only approximate three-dimensional image reconstruction is possible. On the other hand, this arrangement is naturally more sensitive to scattered photons than an isolated ring, and posterior computational compensation should be applied.

Most multislice PET detector arrays are equipped with removable collimators that, when engaged, separate the individual rings (Figure 6.14a). The septa protect the detectors from oblique cross-rays (with a possible tolerance to rays crossing two immediately neighboring rings) and, namely, from scattered photons that might cause higher dead-time ratios and false detections, thus impairing the measurement accuracy. Obviously, only two-dimensional (though multislice) measurement is then possible, but with a better SNR; two-dimensional PET imaging is therefore considered to provide better (more reliable) quantitative results.

The PET detector arrays may, in principle, consist of individual detectors arranged on a ring or forming a set of parallel rings. Nevertheless, it would have two drawbacks: (1) a detector with its own photomultiplier is too bulky to allow the placing of a high number of detectors along the ring circumference and may also limit the number of rings on a given axial length of a multislice array, and (2) the cost may be prohibitive. Instead, the equivalently acting principle of block detectors is applied. A single crystal, encompassing the area of several tens of detectors, is deeply cut perpendicularly to its input surface, with the cuts filled by opaque material (Figure 6.15), so that the sections act as individual crystals. However, the light generated by γ-photons

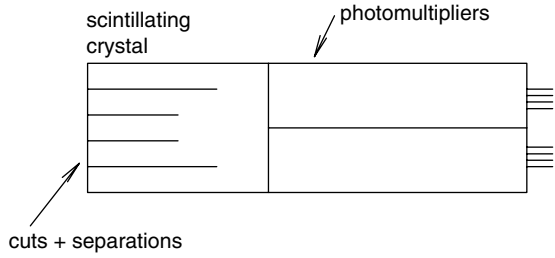


Figure 6.15 A block detector schematically (outputs on the right).

is distributed in the noncut part of the crystal, as in gamma cameras, and can be measured by only a few (four to six) photomultipliers—the position of interaction (the active subcrystal) being located by basically Anger logic. From the image data acquisition and processing aspects, this is equivalent to the above-mentioned arrangement of individual detectors.

6.3.3 Postprocessing of Raw Data and Imaging Properties

The raw data obtained by measurement are imperfect due to several phenomena mentioned in the previous section. Some of the effects can be compensated for, or partly suppressed, by postprocessing of the data. This inherent processing concerns primarily the measured ray counts before the image reconstruction; nevertheless, some of the corrections may be applied to the reconstructed image data, though usually less efficiently. The distorting phenomena can be summarized as follows:

- Attenuation of γ-photons inside the object
- Scatter of γ-photons inside the object
- Random and scatter coincidences
- Dead-time losses
- Resolution limits due to physics of the measurement and also properties of the used image reconstruction algorithm
- Uneven sensitivity of individual detectors

6.3.3.1 Attenuation Correction

It has been mentioned that the attenuation in PET imaging can easily be compensated for according to Equation 6.11, i.e., by dividing each obtained ray count by the respective total attenuation $a_{\text{ray}}(\tau, \theta)$ on this ray. This requires a separate estimation or measurement of the transmission attenuation on each ray under the presence of the object. When relying on the estimates, it is supposed that μ is homogeneous and known inside the body, while outside of the body is negligible. An estimate or approximate knowledge of the object (body) shape is needed. The estimated situation is often not very realistic, but the advantage of this approach is that no random noise is added to the data.

Alternately, the attenuation on individual rays may be measured, using an external source of radiation in the form of a "point" or a thin rod, moving rotationally along the inner surface of the detector array with the collimating septa between slices in place, i.e., as in a CT scanner with a fixed detector array (Figure 4.3). The

ray attenuation is determined, neglecting the air attenuation, via two counts per ray—$C(r, \theta)$ with the object in place and $C_0(r, \theta)$ with the object removed—as

$$a_{\text{ray}}(r,\theta) = \ln\left(\frac{C_0(r,\theta)}{C(r,\theta)}\right). \tag{6.12}$$

For practical reasons, it is preferable to accomplish the attenuation measurement simultaneously with the PET scanning; otherwise, there is a high risk of both measurements being done on the object, positioned differently. Such measurement requires use of the auxiliary radionuclide that emits photons with energy different from the energy of photons emanating from the object; the corresponding difference in attenuation should be estimated and taken into account. The attenuation data must be kept in a memory matrix that is identical in size to the count memory. The attenuation compensation is then realized by simple term-by-term division of both matrices.

It is naturally possible to reconstruct an image from the auxiliary projection data as well; it is an attenuation image resembling that obtained by x-ray CT, though with a lower contrast due to generally higher penetration of the γ-rays. It turns out that when such an image is used as an attenuation map, the noise content should be preliminarily reduced either by spatial smoothing or by segmentation, with the following assignment of average μ values to all pixels of each segmented area. Consequently, the individual ray attenuation values can be calculated based on this map, which gives a better SNR—the random noise in the correction terms is largely eliminated. It should be mentioned that in modern combined CT-PET systems, the attenuation can easily be derived from the CT scans, which are faster and have a better SNR than the attenuation maps derived by means of auxiliary radioactive sources.

6.3.3.2 Random Coincidences

Random coincidences, explained in Section 6.3.1, are corrupting the raw data by improperly increasing the ray counts. As it is visible from Equation 6.8, the false count component, proportional to the random count rate R_c, increases with the square of the radioactivity level to which each of the single-count rates S_1, S_2 is proportional. Thus, while the influence of random coincidences may be unimportant for low activities, it may even exceed the rate of

true coincidences when the radioactivity level is high. It is therefore desirable to subtract this component from the ray counts.

Basically, two approaches are available to this aim. The first one, estimating the quantity indirectly, follows from Equation 6.8: the actual random count $R_c(\tau, \theta)$ for a ray at Radon coordinates (τ, θ) can be determined from the single counts $S_1(\tau, \theta)$ and $S_2(\tau, \theta)$ of the detectors belonging to this ray. The corrected ray count $P_c(\tau, \theta)$, based on the measured count $P_m(\tau, \theta)$, is then

$$P_c(\tau, \theta) = P_m(\tau, \theta) - 2\Delta t\, S_1(\tau, \theta) S_2(\tau, \theta). \tag{6.13}$$

Additional memory bins must be provided to enable counting of single events besides the coincidences, so that other memory locations are needed. Obviously, the posterior correction is done individually for each ray on the basis of the three quantities, once per the complete measurement.

The alternative approach determines the counts of random coincidences by direct measurement. Besides determining the coincidence count via the described method during a time window common to both involved detectors, starting with the response of the first (earlier responding) detector, this approach has another window of the length Δt with a delay $\delta >> \Delta t$ that is always open for the second detector. The first window is called the prompt window to distinguish it from the delayed one. The coincidence inside of the prompt window means an event that may be either a proper photon coincidence or a random coincidence. On the other hand, delayed coincidences, i.e., an impulse being detected by the second detector in the delayed window after an impulse has been detected in the prompt window by the first detector, may only be of the random type. Counting the first type of events would provide the count $P_m(\tau, \theta) = P_c(\tau, \theta) + R_c(\tau, \theta)$, while the count of the second type events is obviously only $R_c(\tau, \theta)$, as a real photon pair cannot be detected in differing time instants. Nevertheless, it is not necessary to have two separate bins for counting both types of events; if the ray bin is incremented by the first type event and decremented by the second type, the resulting count is obviously the desirable $P_c(\tau, \theta) = P_m(\tau, \theta) - R_c(\tau, \theta)$. This way, the resulting quantity is directly provided during the measurement, which has the advantage of saving both the time and computational effort that would be needed for the posterior data correction. It should be noted that the described counting method spoils the Poisson statistics inherent to the proper events counting that forms the theoretical basis for some of the iterative (maximum-likelihood) image reconstruction methods (see Section 9.2.2).

6.3.3.3 Scattered Coincidences

In the count of proper coincidences $P_c(\tau, \theta)$, there is included a certain part of events due to proper pairs of photons, of which one or both have been deflected by Compton scatter. As already explained, the couple of involved detectors then determines an improper ray so that such an event is erratic and should not be counted. If such events are included in the raw data, they contribute to a relatively even background in the reconstructed image, lowering the contrast and reducing the SNR.

Most of the scattered coincidences can be rejected using the principle of amplitude discrimination of detected pulses, as described in Section 6.1.1. Nevertheless, it is less effective than in standard gamma cameras, as the detectors for higher-energy photons have generally lower amplitude resolution. Thus, a non-negligible amount of scattered coincidences remains in the resulting PET ray counts, unless sophisticated methods of multiple amplitude windows are used.

Although many approaches have been suggested to compensate for the remaining scatter influence, this problem is still not considered to be completely solved. One possibility is to postprocess the reconstructed images using standard restoration procedures (Chapter 12); the identification of the distortion may be based on the signal component appearing in the image areas with evidently zero radioactivity, e.g., external parts of the object.

Another approach calculates the scatter component iteratively, usually combining this calculation with iterative image reconstruction algorithms; the principle is as follows. Primarily, the initial image estimate is reconstructed from the measured raw data. Then the following steps are repeated:

- Based on a model of Compton scatter, the scatter component to the raw data is estimated, based on the (so far approximately) reconstructed image, i.e., the radioactivity distribution.
- Estimated ray data are inversely calculated from the available reconstructed image by projection, and consequently corrected by adding the scatter component estimate. If the data thus obtained are sufficiently close to the raw data, the process is terminated; otherwise, it continues.
- A better image reconstruction is calculated based on the raw data corrected by the last estimate of the scatter component; then the process continues by returning to the first step.

6.3.3.4 Dead-Time Influence

As in all nuclear counting devices, each detected event blocks the detector for a certain time. For low activities, the probability of an event appearing during this *dead time* is negligible, but with a growing count rate, it is increasing as well. A high count may then be influenced by a non-negligible number of omitted events that escaped detection due to the relatively high portion of the dead time in the total measuring time. Determining the dead time theoretically is complex and unreliable due to unknown characteristics of all involved elements of the equipment; thus, the correction function (multiplicative factor dependent on the measured value) is usually determined experimentally by a kind of absolute measurement with known sources of radioactivity. An elegant way is to use a radionuclide with a known half-life, thus producing a known curve of the radioactivity decay, which should be fitted with the corrected results of the measurement. The manufacturers of the PET equipment usually provide some information on the correction function under defined circumstances.

The correction is obviously applied as a postmeasurement procedure, modifying the raw data according to the correction function, possibly in the form of a lookup table.

6.3.3.5 Resolution Issues

As already mentioned, the resolution is partly influenced by the physical phenomena involved in the measurement, as explained briefly above: positron range, inexact collinearity in photon pairs, finite size of the detectors, imprecise localization by Anger logic, etc. The influences, given by the system design, can hardly be compensated for—perhaps an exception is a possibility to find a geometrical correction to the Anger logic, similar to the principle already explained in Section 6.1.2. Such correction would most likely be a part of the system firmware.

Another artifact influencing the resolution that is specific to PET should be mentioned. The theoretical ray position, as determined by the join of the input windows of the active detectors (as in Figure 6.13), is well defined for central rays, as the uncertainty due to the size of (tiny) detector windows is small. Nevertheless, positioning of off-center rays may be rather imprecise. As visible from Figure 6.16, the full detector depth, which is needed for high photon efficiency, applies only to near-central rays, while the photons traveling along marginal rays may penetrate several crystals

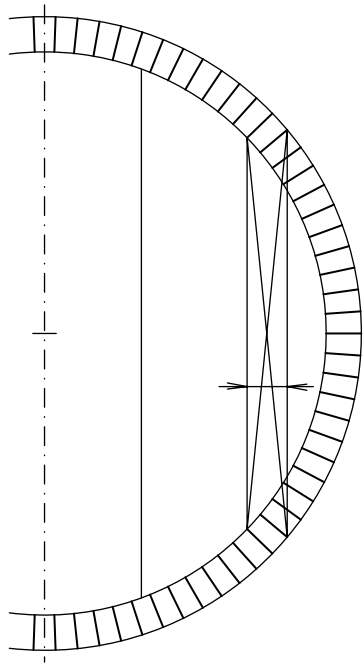


Figure 6.16 Uncertainty in locating an off-center ray due to penetration of photons through more detectors.

before interacting and being detected. As a result, there is uncertainty which of the possible rays (indicated by thin lines) is the correct one. It is obvious that in many cases the recorded position (chosen randomly or, e.g., as the shortest join) is not the right one. This leads to radial blur in the reconstructed image, which increases with the offset from the center.

The final resolution in the image depends also on the used reconstruction algorithm (see Chapter 9).

An improvement in resolution may be, at least theoretically, achieved by image restoration procedures (Sections 12.3 and 12.4) applied to the reconstructed image data. This would require identification of the PSF due to imaging properties, probably on the basis of suitable phantom images. As the image is rather noisy, the properties of the noise should also be taken into account, e.g., its power spectrum, if the noise is sufficiently homogeneous. Due to a rather low SNR and anisoplanar PSF, the problem is quite complicated and a substantial resolution improvement cannot be expected.

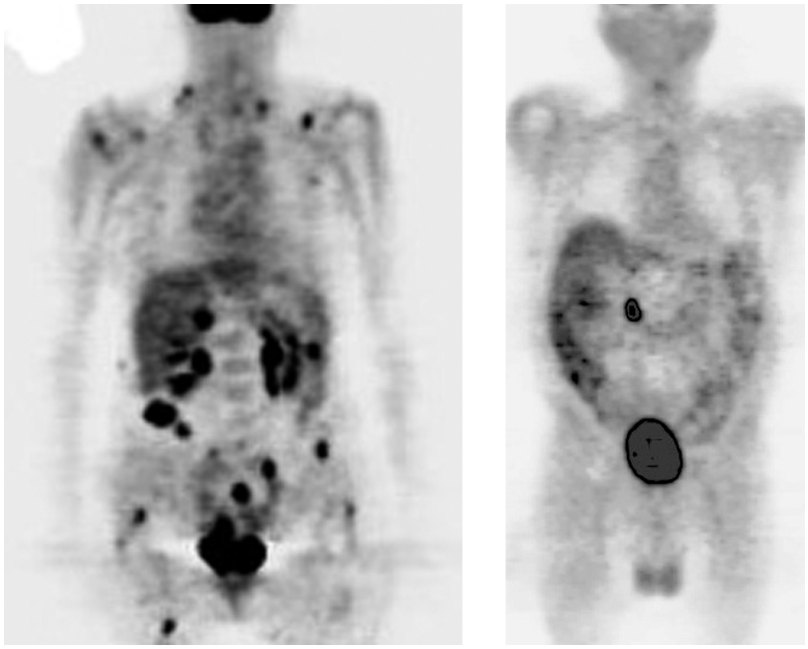


Figure 6.17 Two examples of longitudinal PET tomographic slices. (Courtesy of Masaryk Memorial Cancer Institute, Department of Nuclear Medicine, Karol Bolcak, M.D.)

Examples of tomographic slices provided via PET are demonstrated in Figure 6.17.

6.3.3.6 Ray Normalization

In theoretical considerations, it is supposed that all the detectors in the PET system are equivalent. Obviously, this supposition is not realistic for two reasons. Primarily, it must be admitted that for production and maintenance reasons, the properties of hundreds to thousands of detectors are not uniform. The individual detectors differ in their sensitivities due to different properties and adjustment of the photomultipliers, possibly also due to inhomogeneities in the scintillating crystals, and definitely due to differing positions of the individual (sub)crystals relative to photomultipliers in the detector blocks. Obviously, the amplitude discrimination of detector

responses requires that all the detector sensitivities are equalized or the amplitude windows are adjusted individually, in order that the detectors act equivalently. The related methods are rather specific for different system types and are beyond the scope of this book.

On the other hand, the uniform properties of the system with respect to all the measured rays are immediately connected with the quality of reconstructed images. The reconstruction algorithms expect that all the rays are qualitatively equivalent, i.e., measured with the same efficiency—the same radioactivity integral on the rays should yield identical measured counts. In the raw data, this is not the case: besides the mentioned differences in detector properties, the different rays also have different geometries influencing the effective cross-section of the ray; also, the resulting efficiencies of the involved detectors are affected due to different tilts with respect to the respective rays.

The different ray efficiency must be compensated for by post-processing the raw data, individually multiplying each ray count by a respective preliminarily determined factor. Two possibilities for equalizing the ray measurements exist: either a (periodically repeated) total ray efficiency measurement leading directly to determination of the mentioned factors, or a partially computational approach, in which only the detector sensitivities are repeatedly measured and the geometrical dependencies are determined once forever (experimentally or computationally)—the needed coefficients are combined from all the cofactors. As the latter approach is system-dependent, let us briefly mention only the former method.

The direct determination is based on measurements of individual ray responses to a constant radioactive source of positron type decay (usually a rod, similar to that used in the above-mentioned transmission measurement, moving rotationally around the inner surface of the detector field). Thus, for each position of the rod, say k-th, a number of rays is irradiated, thus providing the individual count $P_k(\tau, \theta)$ for each particular ray. The normalization coefficient is then simply calculated as

$$n(\tau,\theta) = \frac{\dfrac{1}{RT}\displaystyle\sum_{\tau}\sum_{\theta}\sum_{k} P_k(\tau,\theta)}{\displaystyle\sum_{k} P_k(\tau,\theta)}, \tag{6.14}$$

where R is the number of rays across the diameter of the ring (along τ), and T the number of different ray angles θ relative to the x-axis.

Obviously, the numerator in the fraction is the average count per ray over the complete measurement, while the denominator is the (uncorrected) average count of the (τ, θ)-ray. Note, that when multiplying the uncorrected average count by the normalization factor, we obtain the individual average count, which is a constant; hence, all the ray responses are equalized.

6.3.3.7 Comparison of PET and SPECT Modalities

Let us conclude this section with a brief comparison of the imaging properties of SPECT and PET. The most prominent difference between both modalities is in the definition of rays (or better said, of volumes that contribute to ray counts). In SPECT, the ray volume is determined by the collimator properties and basically has the form of a cone widening with the depth; this unavoidably leads to poor resolution inside the object. On the other hand, PET imaging has the ray of an essentially constant cross-section determined by the size of the small input windows of detectors. This leads to the superior spatial resolution in PET (typically 3 to 5 mm), best in the center of image, while SPECT resolution, generally worse (~10 mm), is deteriorating rapidly with the depth and, consequently, is dependent on the shape and size of the camera orbit; the center of the reconstructed image has the worst resolution. On the other hand, the PET range of usable radionuclides is limited to positron-emitting types only, while SPECT can use a wide spectrum of γ-emitters in a broad range of photon energies (with best results for 100 to 200 keV), both single-photon and annihilation couple types. Both modalities are affected by in-object attenuation; the influence is more severe (but easier to exactly compensate for) in PET than in SPECT. The overall imaging properties of PET are favorable (when disregarding the system and data processing complexity, leading to a higher price).

7

Ultrasonography

Ultrasonography (USG) utilizes ultrasonic waves as the information carrier; these are mechanical longitudinal waves of high inaudible frequencies in the approximate range of 1 to 10 MHz, propagating in tissues. They are emitted artificially by a probe that acts usually as both the emitter and, in time multiplex, the receiver of the ultrasonic energy. The ultrasonic imaging may be based on detecting reflected and scattered waves that are responses to the emitted wave (like in radar or sonar systems); then it is called *echo imaging*. Alternatively, it is possible to detect the waves penetrating through the imaged object, in which case it is referred to as *transmission imaging*. The transmission concept, similar in principle to projection tomography, such as computed tomography (CT) or positron emission tomography (PET), enables good specification of the imaged parameter (e.g., ultrasound attenuation or velocity) extracted from the measured data; also, some kinds of artifacts (nonlinear paths due to refraction, reflection or diffusion, or shadows behind highly attenuating tissues) can be better suppressed this way than in the echo mode. The transmission imaging thus has definite advantages over echo imaging—a better possibility of quantitative imaging and the possibility of applying

computational tomographic techniques, at least in principle. However, it is rarely used owing to long imaging times and complicated positioning of probes, with severe practical obstacles in obtaining a reliable acoustic coupling on opposite surfaces of the object.

Echo imaging is practically much simpler to apply, but regarding the physics of the measured signals, it is more complicated. Without going into details, let us state that consequently, most standard echo imaging (traditionally called *B-scan* imaging) is only *qualitative*, describing only the shape and position, but not any concrete tissue parameters, of the anatomic structures. The information, so far utilized in commercial systems, is, besides the spatial coordinates, only the intensity of the detected echo. This intensity, however, is dependent on the imaged scene, on the imaging system properties and adjustment, and on the particular circumstances of the measurement in a very complex manner, so that it cannot be considered to describe a particular tissue parameter.

A special kind of echo-based imaging describes not primarily the structures, but rather the distribution of blood flow on the image slice or in the imaged volume. On the difference to standard B-scan imaging, the flow imaging depicts the blood flow in vessels (and sometimes also tissue perfusion) quantitatively, yielding a clearly defined quantity per pixel. The value, derived by a specialized signal analysis, is mostly the local blood velocity (and perhaps also the degree of turbulence) that can be displayed either in a color scale in the echo image or as separate figures. In case of perfusion evaluation based on contrast imaging, the results are rather qualitative, though there is a certain research effort to quantify the blood content in tissue according to echo intensity or quality.

Although there were many attempts to quantify the ultrasonic echo signals in a way, particularly with the aim of tissue characterization, the main goal of USG remains the spatial two-dimensional or three-dimensional imaging of anatomy via detected tissue interfaces, and of displaying the spatial flow distribution. On the other hand, there is obviously a lot of unused information in the received ultrasonic signals, namely, when going to the raw (unprocessed) radio frequency (RF) signals. It is a subject of a lively research to discover the clinical potential of this so far mostly neglected information.

More detailed information on ultrasonography may be found, e.g., in references [1], [3], [5], [9], [10], [16], [17], [18], [20], [23], [31], [32], [33], [35], [40], [48].

7.1 TWO-DIMENSIONAL ECHO IMAGING

Ultrasonic imaging in the common sense means echo imaging, which is in principle similar to radar or sonar action. The probe emits a short ultrasonic wave impulse that propagates in the medium (e.g., tissue) through the imaged space approximately only along a ray of a certain direction. When touching an object of properties differing from the medium (i.e., an organ of different acoustic impedance), the wave is reflected, refracted, or scattered, depending on the size of the object and the character of the impedance interface. A part of the wave energy returns back to the probe, where it is detected as an impulse. The delay τ of the received impulse with respect to the emission time determines the radial distance r between the probe and the interface,

$$r = \frac{\tau}{2c}, \tag{7.1}$$

where c is the (supposedly constant) velocity of the ultrasound in the tissue (about 1540 m/sec). As the direction of the ray is known thanks to a particular construction of the probe, the spatial coordinates of the point of interaction (i.e., of the object surface) are known and may be utilized when reconstructing the image. The velocity c determines, together with the used frequency f, the wavelength of the ultrasonic energy in the imaged tissue,

$$\lambda = \frac{c}{f}, \tag{7.2}$$

which influences the theoretical resolution limit that is of the same order as λ. Nevertheless, the practical resolution is influenced by other factors too, and generally is considerably worse.

7.1.1 Echo Measurement

7.1.1.1 Principle of Echo Measurement

The basic arrangement for echo measurement is depicted in Figure 7.1. A piezoelectric crystal supplied temporarily by electrical energy from the system transmitter acts as the transducer that generates the impulse of ultrasound. The duration of the impulse in imaging applications is rather short—a few cycles of the used frequency, due to high artificial damping of the crystal. The shape of the real impulse envelope is similar to that depicted: a short rise period followed by a slightly longer decay. When the outgoing

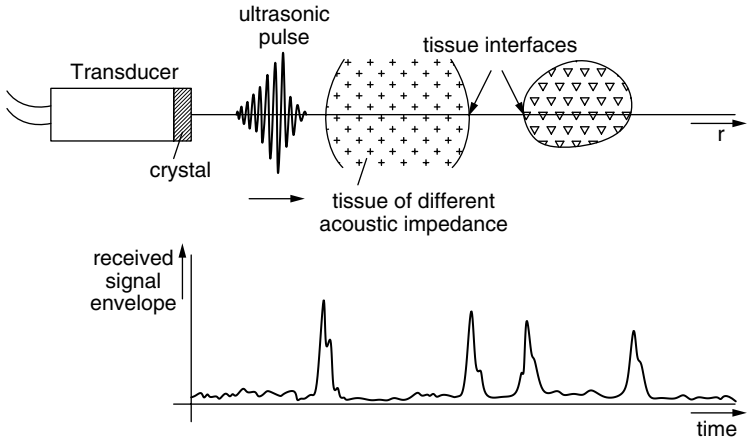


Figure 7.1 Basic single-line ultrasonic measurement. Above: The probe and the ideal ultrasonic ray penetrating through tissue interfaces. Below: Corresponding received (envelope) signal.

impulse has been radiated, the probe is switched over to the input of the system receiver and serves now as the detector of incoming waves; it responds to the mechanical wave oscillations by a proportional RF voltage, the envelope of which is schematically shown in the lower part of the figure. This single-line investigation is historically called *A-mode*; although this mode as such has been practically abandoned, it forms a basis for two- and basically also three-dimensional imaging. Each line of the image is acquired exactly in this way, independently of others so that the line-imaging properties and artifacts directly influence the resulting image quality. Therefore, we shall discuss this principle in a degree of detail.

Obviously, the length of the impulse determines the *radial (axial) resolution* of the measurement; the longer the impulse, the greater must be the difference of radial distances between two targets to be resolvable. As a rule of thumb, when taking into account the nonrectangular shape of the impulse, the radial resolution may be expected to correspond to about one half of the pulse length. As the shorter impulse requires a wider frequency bandwidth, the resolution is directly influenced by the bandwidth of the signal (see below the discussion on bandwidth narrowing due to frequency-dependent attenuation). The radial resolution is related also to the wavelength of the used ultrasound, as only

objects at least comparable in size to λ would cause notable reflection or diffraction, and hence any resolvable echo. Though a higher-frequency f would provide a better resolution, there is a serious limitation in the high attenuation of ultrasound, the attenuation coefficient of which is linearly proportional to frequency, making ultrasonic imaging of tissues in the common sense infeasible with frequencies above 10 MHz. Only ultrasonic microscopy, with imaged depths in millimeters, can use substantially higher frequencies. The most commonly used frequency range in general applications is about 2 to 5 MHz. For abdominal and similar examinations in depths of above 10 cm, the frequencies close to the lower limit of 1 MHz are used; this lower limit is determined by the longest acceptable wavelength.

As already mentioned, the response (signal) from a single ray constitutes data for a line in the final ultrasonic image. The methods of providing multiple lines forming the complete image of a slice, or a three-dimensional data structure, will be described in Sections 7.1.2 and 7.3, respectively.

7.1.1.2 Ultrasonic Transducers

Ultrasonographic probes have a dual purpose: primarily they emit short impulses of ultrasound, and consequently they convert the incoming echoes into measurable voltages.

In the transmit mode, the crystal should emit the ultrasonic energy solely on a thin ray along its axis in order to enable precise localization of the interaction point and to resolve two or more close laterally displaced targets by responding only to one of them. Due to physical limitations, this ideal state can only be approximated. The approximate map of ultrasound intensity in the beam emitted by a simple circular probe of the diameter D is depicted in Figure 7.2. In the *near-field* (Fresnel) *zone* of the length $D^2/(4\lambda)$, the beam is approximately of the same diameter as the probe, slightly narrowing near to the outer border (thanks to s.c. Fresnel focusing); in the *far-field* (Frauenhofer) *zone*, distinct directional lobes are formed. The beam formed by the main lobe is widening in proportion to the distance r from the probe, the beam divergence being given by a constant angle $\beta = \arcsin(1.22\lambda/D)$. The diameter of the main lobe determines the *lateral resolution* of the measurement. Besides the main lobe, there are symmetrical side lobes with substantially lower intensity; however, in the presence of nearby specular reflectors located in the imaged tissues aside of the ray, the side lobes may lead

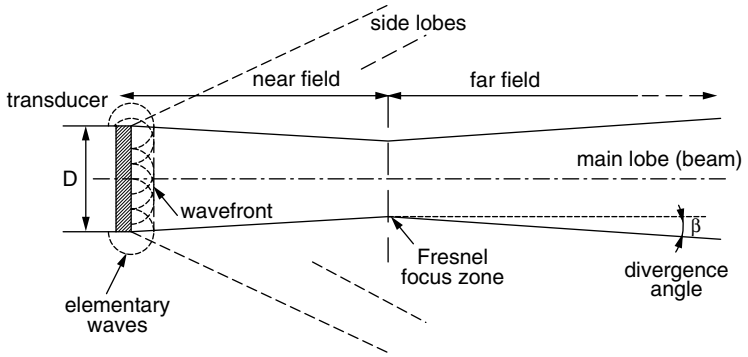


Figure 7.2 Schematic intensity map of the ultrasonic beam.

to artifactual responses, placed in the image on the ray axis. The main lobe is less divergent for greater D, though the ray diameter close to the probe is greater; thus, the larger diameters are advantageous for measurements in greater depths. Conversely, a better lateral resolution in small depths (low r) can be achieved with small transducers, as the diameter of the probe determines the lateral resolution in the near zone.

The described beam pattern is, according to Huygen's principle, the result of superposition of elementary spherical waves emitted by elements of the transducer surface. In case of the planar transducer, the wavefront in the near zone is also approximately planar. When the surface is curved (Figure 7.3 top), the waves are combined differently, which leads to a different directional pattern. A concave spherical surface produces the wavefront that is approximately spherical with curvature decreasing with r in the near zone. Such a transducer effectively focuses the energy to a narrower beam (several mm) in the range of the *focal zone* near the center of curvature where the lateral resolution is improved; unfortunately, the focusing leads to a faster divergence in the far zone so that such a probe is only suitable for examination of correspondingly near objects. An acoustic lens placed in front of the transducer provides a similar effect.

Another method, allowing more flexible beam forming, is to subdivide the planar transducer into small sections (Figure 7.3 bottom), thus obtaining, e.g., a rectangular grid of elementary transducers on the output plane of the probe. Each of the elements is supplied with energy independently, with a controllable delay introduced to each

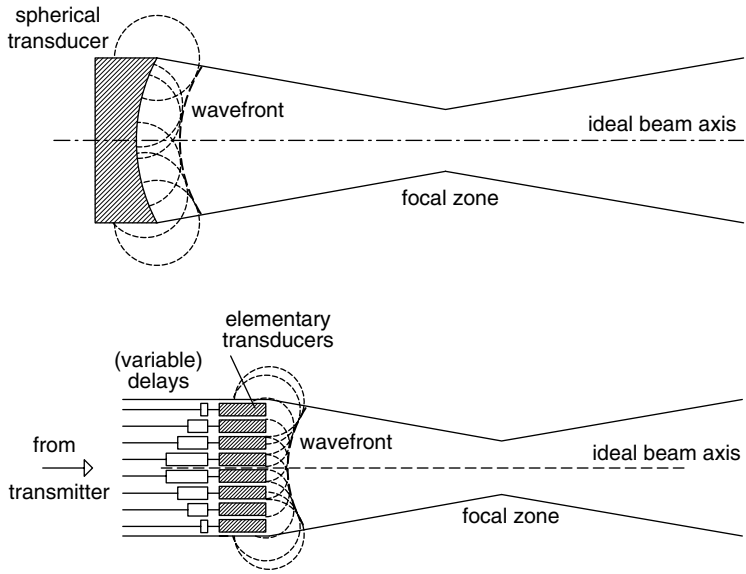


Figure 7.3 Focusing of the ultrasonic beam: (above) by a curved transducer surface and (below) by a phased array.

connection between an element and the transmitter. A delay in the supply of energy with respect to other sections is then equivalent, as for the partial wave superposition, to a spatial (axial) shift relating to the other sections. Such a generally two-dimensional structure in the plane perpendicular to the axis is called the *phased array*; it can mimic any geometrical curvature of the transducer surface and provide the focus distance optimal for a given measurement. When not more than focusing of a circular transducer is needed, the optimal shape of the sections is obviously circular—each annulus with a corresponding delay. Such a transducer is called the *annular array.*

Nevertheless, the possibilities of phased arrays are more generic than just the focusing. By introducing the delays linearly, increasing from one side of the transducer to the other, a planar wavefront can be generated that is inclined with respect to the transducer surface (Figure 7.4). This means that the probe beam is deflected from the axial direction by an angle θ corresponding to the slope of the delay increase, equivalently as if the simple planar probe were inclined respectively. Such beam steering finds a general use in modern B-mode scanners using phased array transducers. The beam steering can be accomplished in combination with focusing;

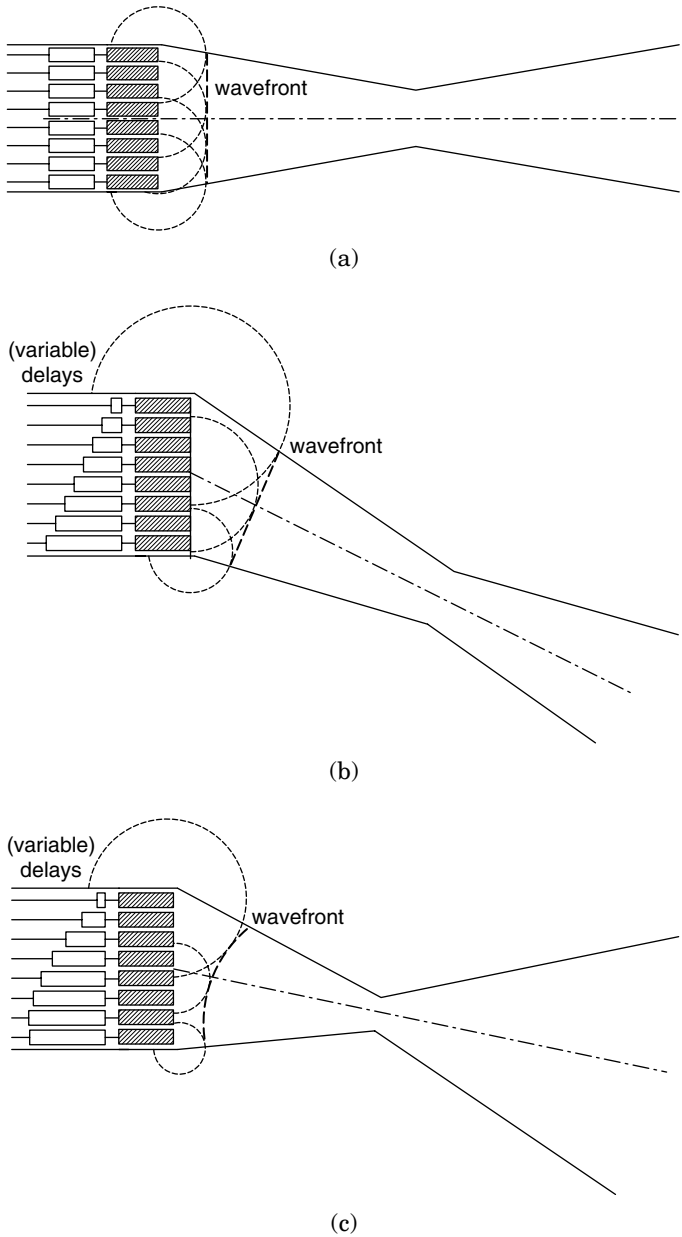


Figure 7.4 Time- and space-variable delay in energy supply used to beam steering in a phased array: (a, b) steering with plane wavefront and (c) steering combined with focusing (the divergence is intentionally exaggerated).

the delays for the individual elements then become sums of individual steering and focusing delays. This way, a true two-dimensional phased array transducer is capable of being flexibly adapted for a particular type of measurement.

Modern two-dimensional phased arrays may even steer (and focus) the ray in two perpendicular angular directions, as needed in three-dimensional ultrasonic imaging (Section 7.3). For technical and cost reasons, phased arrays often consist of only a single row of elements (one-dimensional arrays of 128 to 512 elements); obviously, steering and focusing is then only possible in the plane containing the row (i.e., image slice plane), and the directivity in the perpendicular direction (elevation) is determined by the transverse size and shape of the elements. A certain compromise is the 1.5-dimensional array with a few (five to seven) parallel rows of elements that enable, besides direction steering and focusing in the slice plane, flexible focusing in the transverse direction. The resolution in the direction perpendicular to the tomographic (slice) plane is called *elevation resolution*; when talking about *lateral resolution* in B-mode images, the resolution in the image plane is thought of. Obviously, the elevation resolution determines the effective thickness of the imaged slice.

It should be mentioned that the directional properties of a phased array are partly worsened by grating lobes that are caused by the diffraction effect of separations between neighboring elements in the transducer, acting like a grating put on the active transducer surface. These lobes have similar consequences in the image as the side lobes (see above).

So far, we discussed the transducer as a source of ultrasonic pulses. The same transducer is used as the *receiving* device in the period following the pulse emission, when the same direction of the beam (now expressing the receiving directivity) must obviously be preserved as in the transmitting period. It follows from the duality principle that the directivity of an ordinary (nonphased) transducer is the same in both the transmitting and receiving modes. The effective directivity as to the received signal concerns, which determines the resulting lateral resolution, is then enhanced—the effective width of the beam, as given by cascading the effects of transmitting and receiving beams, is narrowed (the lateral profile at any depth r is obviously the square of the transmitting directivity profile).

In phased arrays, the situation may be more sophisticated. In the receive mode, the signal of each element of the probe array may be processed essentially separately from others. Modern systems are

equipped with separate receiving electronics for each field element, this way ensuring individually controlled delays, besides the needed amplification. Consecutively, all the individually controlled received signals are summed to obtain the transmitter output. While the set of delays, determining the ray deflection, should remain the same in the receiving period as it was during transmission, thus preserving the ray direction, the additional delay components determining the degree of focusing may be changed. In such a case, we have effectively different transducers in the transmitting and receiving modes; nevertheless, the directivity effects in both modes are still cascaded. This means that the lateral profile at any particular depth r is the product of the transmitting and receiving profiles at this depth. It is thus possible to use different focus distances in both modes, which opens a way for *dynamic focusing* during the receiving period. As already explained, the time elapsed from the start of the receiving period is proportional to the radial distance (depth) of each detected target. Thus, it would be advantageous to have a dynamically variable focus during the receiving period, increasing from a short focus distance at the beginning to a longer distance, corresponding to the depth of the field of view (FOV), at the end of the receiving period. As the delays of the individual transducer elements are electronically controlled, this can easily be arranged for by varying the delays during each receiving period correspondingly. Although the transmitted beam can only have a fixed focus, the dynamic focusing of the receiving beam may optimize the combined effective beam, to provide, as a result, enhanced lateral resolution in a greater extent of depth.

The most advanced transducers may be equipped with individual analogue-to-digital (A/D) conversion in each elementary channel, immediately following the signal amplification of each crystal element. In this most flexible (but also most complex) case, the output of the transducer is formed by a set of parallel digital channels for the individual RF signals. This way, each signal can be processed independently, including individually controlled delaying. As the delay means only a shift of the digital RF signal by a certain number of sampling periods, more delays can be applied in parallel to each single-element channel, as schematically depicted in Figure 7.5. This enables the combining of the individual signals with different delays and the provision of several complete transducer output signals at a time, each corresponding to its own receive beam direction and focusing. Such a system finds its use in *multibeam imaging* when the (single) transmitting beam is rather wide (unfocused or possibly properly artificially (de)focused), while combining

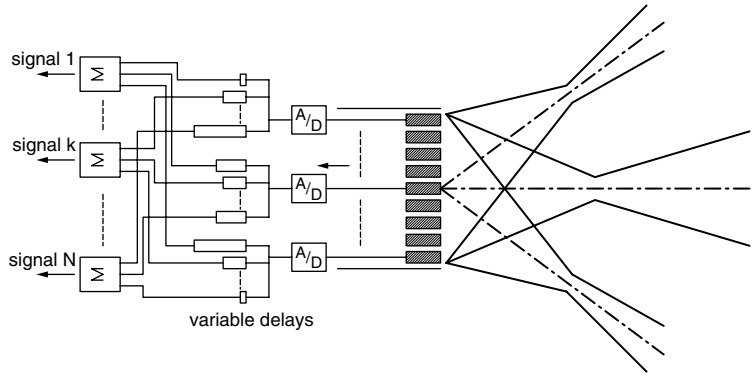


Figure 7.5 Principle of multibeam transducers.

the received differently delayed elementary signals enables the forming of a number of narrow properly focused receiving beams. Obviously, the data acquisition is speeded up by the factor equal to the number of rays acquired in parallel; it is paid for, besides the complexity of the system, by partially decreased lateral resolution, as the transmit focusing cannot contribute to the beam narrowing. On the other hand, this principle enables the increasing of the frame rate substantially, which is particularly important in three-dimensional imaging (see Section 7.3).

Two important parameters of transducers should be mentioned in our context: the central frequency f of mechanical oscillations determining the ultrasound frequency (1 to 10 MHz) and the bandwidth (relatively 50 to 150% of f) given by the probe construction and damping, which influences the bandwidth of the received signal and consequently the time resolution, which determines the radial spatial resolution. Naturally, the transducers are described by many other parameters — efficiency of energy conversion, sensitivity, maximum power, impedance matching to a tissue, etc. — but these are unimportant from the viewpoint of image data processing.

7.1.1.3 Ultrasound Propagation and Interaction with Tissue

The ultrasound, when propagating in tissue, is subject to phenomena common to all wave fields: reflection, refraction, diffraction or scatter, and attenuation. The parameter of the medium, which

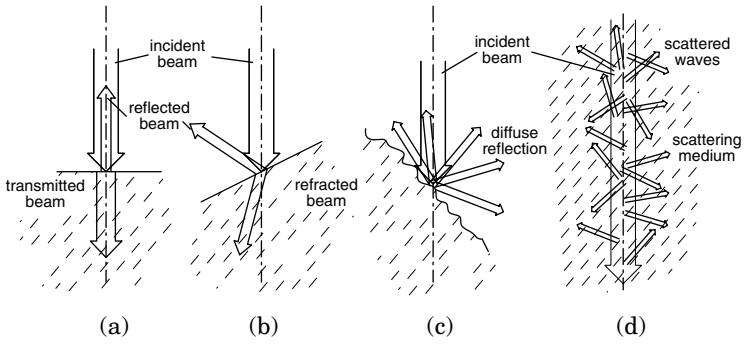


Figure 7.6 Propagation of ultrasound in tissue: (a) specular reflection on a smooth perpendicular interface, (b) reflection and refraction on a smooth slanting interface, (c) diffuse reflection/scatter on a rough interface, and (d) scatter on numerous small object interfaces.

controls the first three phenomena, is the *acoustic impedance* Z, proportional to density ρ of the matter and the ultrasound velocity c,

$$Z = \rho\, c\,. \tag{7.3}$$

All the phenomena appearing at interfaces between two media with different impedances Z_1, Z_2 are depicted in Figure 7.6.

The *specular reflection* intensity depends on the relative difference of impedances, $(Z_2 - Z_1)/(Z_2 + Z_1)$; the transmitted portion, as the remainder, may even be strongly attenuated, which causes a shadow (weaker response) in the deeper part of the image line (Figure 7.6a). Nevertheless, the point of interaction as well as the possible deeper structures are visualized at proper positions. It may not remain true in case of *refractive interaction* on a skew smooth surface (panel b): the reflected part of the signal may not reach the transducer at all, and therefore may not be imaged, while the refracted part propagates farther in a different direction given by Snell's law, $\sin\beta_1/\sin\beta_2 = c_2/c_1$. When interacting later with a structure, situated outside of the original beam, that would return some echo to the transducer, the respective interaction will be improperly placed as an artifact on the original ray in the image, because the image-reconstructing algorithm assumes only the straight-line propagation. The situation is in a sense more beneficial in case of *diffuse reflection* on rough surfaces (panel c), where the ultrasound is reflected to a wide solid angle; in this case, a part of the echo reaches the transducer anyway, though with a

low amplitude, so that the interaction on the interface may be detected and consequently used for image reconstruction. Most of the interfaces in living tissues are of this kind, thus enabling imaging of organ borders along almost the whole circumference, at different angles of the surface to the incident wave (except slopes close to 90°).

The last type of the impedance-related phenomena concerns the interaction of the ultrasound with numerous small objects, of a size comparable with or smaller than λ. In this case, wave diffraction appears, causing *scattering* of the ultrasonic field along the path of propagation. This leads to attenuation of the propagating wave, as well as to massive generation of multiple (though weak) elementary echoes spreading to all directions. Some of them then interfere when reaching the transducer surface. This creates a time-variable signal, instantly weaker or stronger, depending on whether a destructive or constructive interference prevails. This way, an RF signal component is generated that, though being random, is known to have relatively slowly varying amplitude with a tendency to quasiperiodic character. The signal values are highly correlated in a certain extent of time, corresponding to the spatial extent of several millimeters in the distance (depth) r, with possible periodic extension of the autocorrelation function throughout the scattering section. This signal, relatively stationary in the statistical sense inside a macroscopically homogeneous area (e.g., an organ or lesion area), can be decomposed into its mean value and the variable component. While the mean value characterizes the *echogeneity* of the area, represented by the mean gray level of the corresponding image area,* the variable component constitutes a very unpleasant type of noise, difficult to suppress, which causes *speckles* in the reconstructed images (see below). Obviously, the local extremes of this signal have nothing to do with the macroscopic structure of the imaged area; moreover, it has been shown that the character of this signal component is more dependent on the probe and system properties than on the character of the tissue. Nevertheless, certain information on the tissue type is undoubtedly present; it is a subject of many past and running research attempts to isolate it from other factors. These efforts are encouraged by USG practitioners' claims of being able to diagnose certain types of tissue, recognizing a tissue

*The corresponding extremes are denoted as hyperechoic or hypoechoic regions.

"signature" after accumulating enough experience with a particular ultrasonic scanner.

When the ultrasound propagates through a homogeneous tissue, it is attenuated exponentially, so that the local intensity at the distance r is

$$I(r) = I_0 \, e^{-\mu r}, \tag{7.4}$$

I_0 being the initial intensity at $r = 0$, and μ the attenuation coefficient (dB/cm), dependent approximately linearly on the frequency,

$$\mu = \mu_1 \, f \tag{7.5}$$

with μ_1 (dB cm^{-1} MHz^{-1}) being the attenuation coefficient for the particular tissue at $f = 1$ MHz. The value of μ_1 ranges from about 0.2 to 1.8 in soft tissues and may reach over 20 in bones — the extremes being 0.0002 in water and about 40 in air (or lungs). It should be realized that the total attenuation of a received echo corresponds to twice the radial distance to the target (forward and backward travel). The resulting attenuation, even when not counting the low efficiency of wave-returning phenomena — reflection or scatter — reaches rather high values, namely in larger depths and with higher frequencies (e.g., in an average tissue with $\mu_1 = 1$, $r = 10$ cm, and $f = 5$ Mhz, the attenuation would be approximately $2 \times 10 \times 5 \times 1 = 100$ dB).

In the imaging systems, the short transmitted impulse has a broad frequency bandwidth. As the attenuation increases with frequency, the higher-frequency components are attenuated more when traveling in the attenuating tissue; consequently, the mean frequency of the echo signal becomes notably lower than the transmitted mean frequency. This phenomenon is called *frequency shift*; the loss of higher-frequency components may lead to a decrease in radial resolution.

7.1.1.4 Echo Signal Features and Processing

Ultrasound is differently attenuated in the various imaged tissues, and also the portion returned by targets is very different; the signal, as received by the transducer from differently effective and distant targets, thus has an enormous dynamic range, reaching some 100 to 120 dB, with the lowest signal amplitude on the order of 1 μV. The central frequency f is in the mentioned range 1 to 10 MHz — thereof the usually used term *radio frequency signal*. As the first step of processing, the signal is amplified to a level suitable for further processing, usually in the range of volts, in which it can be digitized.

The extreme dynamic range can be diminished by a depth-related amplification compensating for attenuation of echoes coming from more distant targets, so that the echo amplitudes from different depths are made comparable; thus, the deeper structures would not become too dark or invisible in the image. This is done by *time-gain compensation* (TGC), a time-variable amplification, included somewhere in the signal processing chain and adapted approximately (or even more accurately) to the estimated attenuation along the ray. In the simplest form, the amplification is low when the receiving period starts and increases approximately exponentially with the time elapsed until the end of the listening period, this way being adapted to average tissue attenuation, equally on all rays of the image. This fixed amplification curve may be corrected manually to a certain extent (based on the observed image), this way taking partly into account the real situation that may differ substantially from the supposed homogeneous attenuation distribution. However, all the rays are still influenced equally without taking into account the often important differences among attenuation profiles of individual rays.

More sophisticated methods of TGC (still rather experimental) use more realistic models of individual attenuation distribution along each ray, either estimated based on the known anatomy or determined by measurement*. The attenuation distribution in the imaged area may be measured using special equipment, or it may be provided simultaneously with the image data acquisition from the received RF ultrasonic signal utilizing the frequency shift information (see the next paragraph). The design of the amplification curve for each particular ray can then be based on the relevant attenuation profile along this ray. The principal advantage of this approach is obvious: the amplification curve may become the inverse of the actual attenuation profile; also, each ray may use a particular compensation curve. The artifacts caused by improperly compensated attenuation, e.g., shadowed regions behind highly attenuating tissues, or seemingly hyperechoic areas after low-attenuation regions, could thus be suppressed. However, these methods, though interesting from a signal processing viewpoint, are still not widely used and clinically accepted; interested readers may find further references in [5], [28], [31].

*The obtained information may be presented in the two-dimensional form of the *attenuation map* of the imaged area.

The relative bandwidth of the transmitted signal reaches about 1; i.e., it typically occupies the frequency range of <0.5 *f*, 1.5 *f*> or even wider, meaning several megahertz of absolute bandwidth. Nevertheless, reduction of higher-frequency components due to the already mentioned frequency-dependent attenuation narrows the bandwidth of the received RF signal and causes the phenomenon of frequency shift. Obviously, the received signal then carries a signature of the attenuation distribution on the way of ultrasound traveling, which suggests that estimates of attenuation distribution along a ray may be possible based on evaluation of spectral changes in the received signal; this interesting technique has already been investigated. While a frequency-independent TGC mechanism in any form more or less compensates for the spatially variant mean attenuation, it does not provide any correction of the spectral changes characterized by the frequency shift. A kind of RF signal filtering should be used to properly restore the original spectrum. It is a topical theme of research to find a reliable method that would provide for both the estimates of the attenuation profile, based on the spectral changes, and the corresponding compensation of these changes, based in turn on the determined attenuation profile. Anyway, the frequency shift compensation is not presently a routinely used procedure.

Even when the attenuation has been compensated for by a kind of TGC, the resulting signal still has a large dynamic extent due to differences among specular reflections on one side and weak scattering on the other; this range is usually further suppressed by approximately logarithmic amplitude transform. In the older systems, an analogue amplifier provides for this transform, then the signal is demodulated (envelope detected), and the resulting low-dynamics base-band signal, often denoted as *video signal*, is digitized for further processing (say with 8 bits precision and sampling frequency in units of megahertz).

The modern approach is to digitize the unprocessed, only linearly amplified RF signal directly to a digital sequence with 16 to 24 bits precision and with a sampling frequency of about 20 to 40 MHz—let us call the obtained sequence the *digital raw signal*. When the digitizing is fast and precise enough, the complete information contained in the received signal is preserved and the entire signal processing, including the mentioned dynamic range compression (TGC and logarithmic transform), can be done on the digital side, with all the known advantages. Obviously, the complete analysis of the signal leading to image data is then also done digitally.

The next step in the routine signal processing is demodulation, more properly also called *envelope detection*, converting the RF signal to its base-band version, the video signal. More precisely, it should provide the information on the instantaneous amplitude of the RF signal, as the only quantity routinely used in standard ultrasonography. In fact, it is a crude nonlinear step after which not only a substantial part of potentially useful frequency and phase information is lost, but also any analysis based on the linearity of the imaging process is at least challenged, if not made impossible. Commonly, the envelope $S(t)$ of the RF signal $s(t)$ is approximated by the low-pass-filtered absolute value of the RF signal,

$$S(t) = [\,|\,s\,|\,*h\,](t)\,, \tag{7.6}$$

$h(t)$ being the impulse response of the used low-pass filter. Formally more satisfying is the envelope detection based on the analytic version of the RF signal,

$$s_A(t) = s(t) + j\mathrm{H}\{s(t)\}\,, \tag{7.7}$$

where H{…} represents the Hilbert transform of the signal (see, e.g., [7] or [19] of Part I) and j is the imaginary unit. The video signal defined as the envelope of the analytic signal is

$$S(t) = |\,s(t) + j\mathrm{H}\{s(t)\}\,|\,. \tag{7.8}$$

As visible from Figure 7.7, while the filtered absolute value needs a subsequent low-pass filtering with the cutoff frequency not clearly defined, no further filtering must be applied to the magnitude of the analytic signal. This envelope detection method derives definitely better image data, as can be judged by the visibly higher quality of resulting images with higher resolution in spite of smoother appearance.

Another possibility for providing the video signal can be derived from a model of imaging in the form of convolution between the respective RF impulse response (RF PSF) of a particular imaging system and the imaged tissue. In one-dimensional (ray) interpretation, the echo RF signal is composed of the weighted sum of impulse responses to numerous point targets, so that the signal describing the structure profile can, in principle, be derived by deconvolution. Though an attractive idea, it is not easy to implement practically; the identification of the concrete RF (oscillating) impulse response is itself a problem, and the deconvolution is known to be an ill-posed

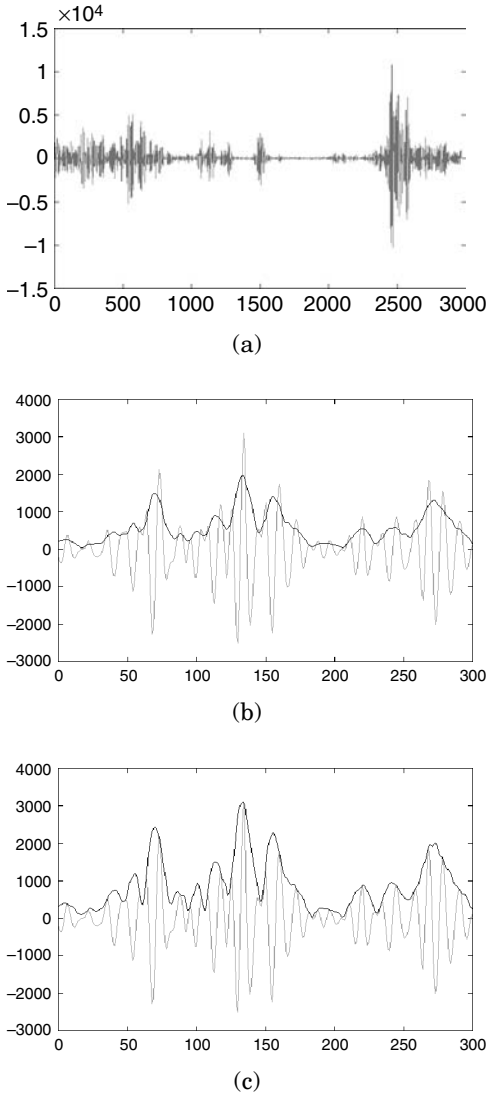


Figure 7.7 RF signal of a part of the ultrasonic line: (a) original RF signal of a complete ray, (b) envelope of a short signal section derived as the filtered absolute value by Equation 7.6, and (c) envelope derived as the magnitude of analytic version of the same signal by Equation 7.8.

problem, here complicated by the involved wave interference of individual point responses. More can be found, e.g., in [43].

The sequence of the derived video signal, which has the same amount of samples as the original RF signal, can be decimated, as its bandwidth is substantially lower, being only in the range of about 1 MHz (rather than several megahertz, as with the RF signal). As the bandwidth of the signal derived according to Equation 7.8 is not clearly defined (the linear theory does not apply after the nonlinear operation of the absolute value), the sampling density may be determined rather from the imaging needs and possibilities; the number of samples on a ray, given by the radial resolution defined physically, is usually in the hundreds. To avoid aliasing, the video signal should be low-pass filtered before the decimation, with the cutoff frequency defined by the used spatial resolution. This way, a conventional one-dimensional data set for an image line is provided that will be consequently processed to provide two-dimensional or three-dimensional image data (see Sections 7.1.2 and 7.3).

Not much new information can be expected to become available by further analysis of the envelope data. On the other hand, when returning to the digital raw RF signal currently accessible in some systems, an attempt to provide clinically useful extra information by further analysis need not be hopeless, as advanced analytic methods, feasible only in digital form, may be applied. However, most modern commercially available systems do not offer such a deep analysis of the RF signal, being basically restricted to providing amplitude (envelope) images, though perhaps based on more sophisticated envelope detection, as described above, and possibly also utilizing other partial improvements, namely, in signal acquisition and transducer control. Naturally, the inner data processing in modern ultrasonic systems, except for initial amplification, is done entirely in digital form and in real time, partly by specialized hardware. Deeper signal analysis is still not included or used routinely, obviously because there is a lack of clinically approved procedures. Hence, the more advanced RF signal analysis and interpretation remains a topical matter of research.

7.1.2 B-Mode Imaging

7.1.2.1 Two-Dimensional Scanning Methods and Transducers

The line (ray) data, provided in the above-described manner, constitute the elements of which the resulting image is formed. The imaged area — in this section a two-dimensional slice through the

object — must be sampled by the rays densely enough to cover the region of interest. This means that the distances between the lines (or, more generally, between the locations of available image data) should not exceed the resolution limits given by the physics of imaging and parameters of the used system. As a standard, several hundreds of lines (typically 128 to 512) are used to form an image; such a set is called a frame. The number of lines N_{line}, together with the imaged depth (radial distance) $r_{\max}$ (in m), obviously determines the maximal available frame rate FR — the number of images scanned in a second. The scanning time of a line is obviously

$$T_{line} = 2\frac{r_{\max}}{c}, \tag{7.9}$$

where c is the velocity of ultrasound (~1540 msec^{-1}); thus, the theoretical maximum frame rate is

$$FR = \frac{1}{N_{line}T_{line}} = \frac{770}{r_{\max}N_{line}}. \tag{7.10}$$

Several arrangements of rays are in use; the most commonly used being the parallel scan and the fan scan, as shown in Figure 7.8. Both have certain benefits and disadvantages. The parallel scan covers a wide rectangular area with a constant lateral density of rays; it is suitable for imaging well-accessible object areas (e.g., abdominal or obstetrical applications). Technically, the sampled image data, as provided by the line acquisitions, are obviously obtained on a rectangular sampling grid and therefore form directly the image matrix in Cartesian coordinates. Such a matrix can be displayed directly (or with only simple separable one-dimensional interpolation and aspect ratio correction) on common rectangular monitors using the standard

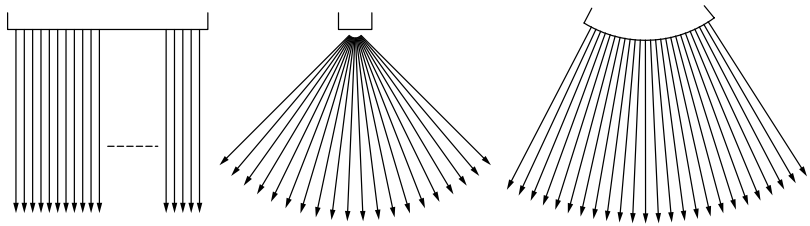


Figure 7.8 Configurations of rays forming the two-dimensional scans: parallel scan, fan scan, curved scan.

hardware. On the other hand, the fan scans suffer from diverging rays so that the lateral coverage density decreases with radial depth; however, this may reasonably correspond to the unavoidable lateral resolution decrease along a ray due to beam widening with depth, thus not impairing the effective resolution. The main advantage of the fan scan is the small interface area needed as the window to the imaged slice, necessary when the approach to the imaged internal organs is restricted by structures unpenetrable by ultrasound (like bones or air in lungs). This is typically the situation in echocardiography, where only a few small windows between ribs are available for the heart imaging. A marginal advantage is that it is easier to provide a reliable acoustic contact on a small window area than on a large window needed for parallel scanning. The data on diverging rays constitute a matrix in polar (r, ϑ)-coordinates and must therefore be converted into Cartesian (x, y)-coordinates before they can be displayed on standard monitors (see below). The curvilinear scan in the third part of the figure is a compromise in a sense, its main advantage being the wider coverage.

In principle, it is unimportant how the lines forming the image are scanned; all common systems provide the data sequentially, ray by ray. The conceptually simplest way is to move a single-ray transducer manually according to the chosen scan geometry as in the early systems.

Nevertheless, modern imaging probes (Figure 7.9) provide for automatic acquisition of complete scans; they can be classified into *linear arrays* (sequential, parallel), *curvilinear arrays*, and *fan-scan probes*; each has its better and worse features, as partly discussed above. Conceptually straightforward are the fan probes, with wobbling or rotating transducers that actually mean a mechanical implementation of the previous principle of a moving single-ray transducer. On the other pole of conceptual simplicity, plain (sequential) linear arrays are found that represent merely a spatially parallel arrangement of independent single-ray transducers, the number of which (up to 512) determines the number of image rays.

Modern linear arrays work in a more sophisticated manner: each transducer element of the array is connected to the system via variable delays, and each beam is formed by the concerted action of several neighboring elements, thus enabling focusing of both the transmitting and receiving beams, as explained in Section 7.1.1, possibly including dynamic focusing. The dynamic focusing may be combined with *dynamic aperture*; it means that while there are only a few elements involved in the receiving action for close targets (thus keeping the beam thin in the narrow-field region), the number

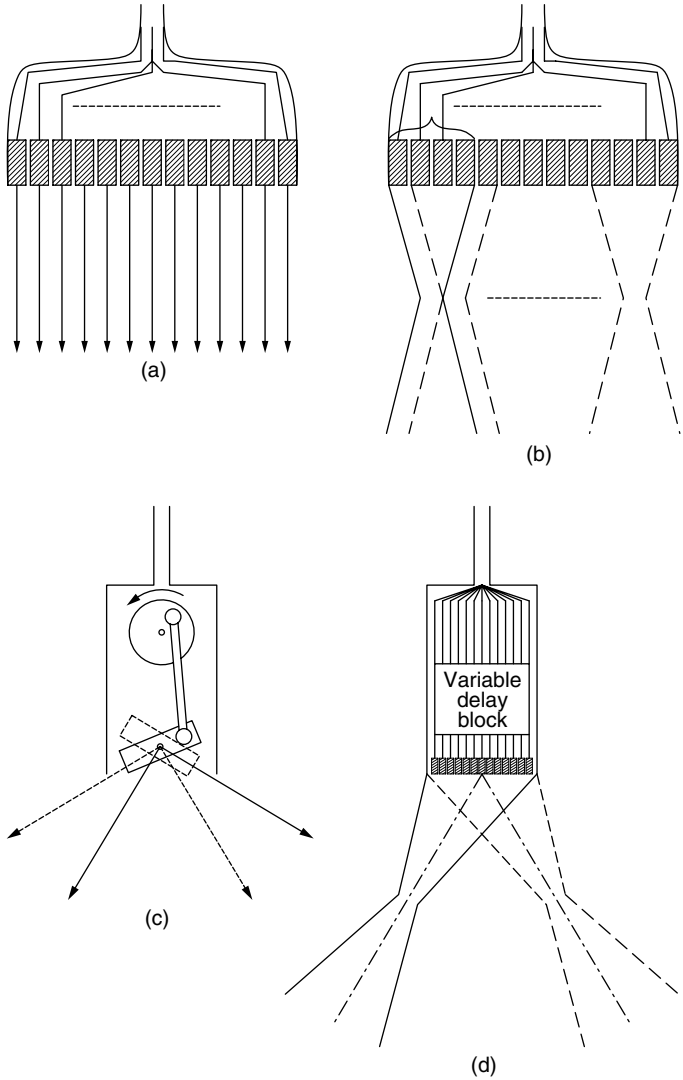


Figure 7.9 Main forms of two-dimensional ultrasonic scan probes: (a) plain linear array, (b) focused linear array, (c) fan probe with wobbling crystal, and (d) phased array.

of elements increases with the target depth, so that the lateral resolution in the distant field does not deteriorate. Different image rays are examined sequentially, by shifting the group of involved elements along the transducer row, one element (or more) at a time. A disadvantage of the linear scanners is the large footprint, the width of which (~10 cm) determines the width of the imaged area.

One-dimensional phased arrays designed for fan scanning have much smaller footprints (with a width of ~2 cm), still with a large number (up to 128) of small transducer elements in the row to enable beam direction steering, perhaps including simultaneous dynamic focusing in the slice plane, as explained in Section 7.1.1. The beam properties in the direction perpendicular to the slice are given by the transversal dimension of elementary transducers and possibly also by their transversal curvature; therefore, they are fixed. This is improved in the 1.5-dimensional phased arrays, which have the transducers, subdivided into several elements also in the transversal direction, so that the basically linear arrays have several rows. The 1.5-dimensional transducers cannot deflect the beam in the transversal direction, but are capable of dynamically focusing in the direction perpendicular to the slice plane. A general disadvantage of phased arrays in comparison with the wobbling crystal probes is that while the latter preserve the same beam parameters exactly for all lateral angles θ, the electronically steered beams vary their properties (namely, cross-sectional profiles as a function of radial distance r) remarkably with θ, thus partly hampering possibilities of advanced image analysis or restoration. On the other hand, the absence of any moving parts, enabled by the ingenious steering principle, is definitely advantageous.

The modern phased arrays with individual signal treatment and digitization for each crystal element enable multibeam acquisition (see Section 7.1.1) via parallel forming of more receiving beams, thus increasing the achievable frame rate substantially.

7.1.2.2 Format Conversion

Most signal and image data processing involved in two-dimensional ultrasonography belongs to one-dimensional ray examination, which has been described in Section 7.1.1. Thus, we shall restrict ourselves to the remaining step of format conversion, the need of which arises from different formats of measured data matrices and of the used display equipment.

The format conversion is elementary in the case of linear (parallel) arrays. The data matrix of a rectangular slice area is provided

column by column, and the only modification that might be needed is a change in pixel size of the displayed image from the acquired one. Then, obviously only two independent one-dimensional backward interpolations (Section 10.1.2) may be needed to obtain the values on a different rectangular grid, possibly with a different aspect ratio, as needed for the realistic display on common monitors or printers. With the inclusion of small necessary delays (in the range of several times that of a column's acquisition time) in order to consider future values in the interpolation, this may easily be implemented in real time.

The situation is more complex in the case of curvilinear arrays (providing a part of a fan scan) or true fan-slice scanners. Here, the image data are provided on a set of lines characterized by certain lateral angles ϑ, so that the natural data arrangement is a matrix with columns corresponding to the individual values of ϑ and rows related to the depth values (radial distances) r. Usually, the data are equidistant in these coordinates. However, when such data were directly interpreted in Cartesian (x, y)-coordinates, as on a display with standard line/frame scanning or any computer monitor with rectangular organization of the display, the image would be heavily distorted, enlarging the lateral dimensions of the image more the closer the respective area is to the probe. Besides that, the circles of constant r would be displayed as horizontal lines. The relations between the polar coordinates (r, ϑ) of the measured data and the proper rectangular coordinates (x, y) of the corresponding display pixel are, according to Figure 7.10,

$$r = \sqrt{x^2 + y^2}, \quad \vartheta = \arctan \frac{x}{y}. \tag{7.11}$$

As the display grid is fixed, the backward interpolation must be used (Section 10.1.2), according to the relations presented. For a real-time format conversion, two matrices — one for each format — are necessary, as the measured data must be completely available to the interpolating algorithm when a pixel value of particular (x, y)-coordinates is requested by the display control, independently of the measurement timing. By inspection, it may be observed that a single displayed image may be formed of data from a number of subsequent frames of ultrasonic data; this is the price paid for fast and simple operation. A more complex interpolation with proper buffering may obviously remove this imperfection. Let us note that the digital format conversion already used in ultrasonic scanners in the 1970s was probably

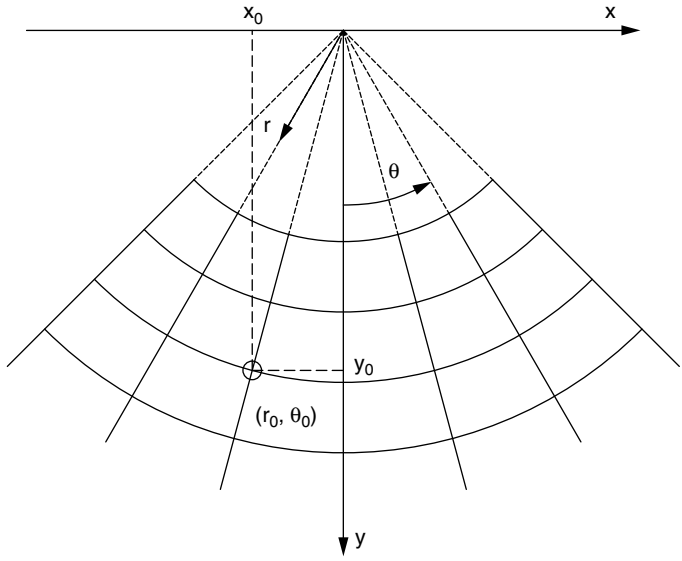


Figure 7.10 Geometry of fan-scan format conversion.

one of the first nontrivial commercial applications of real-time digital signal processing (though most processing in the systems was still analogue at that time).

When the scanner provides the original data in a more complex format, the algorithm must be modified, perhaps even using anisoplanar relations or other approximation techniques, instead of Equation 7.11. The scanning format can be considered a geometrical distortion of the desired image anyway, and the restitution methods described in Section 12.2 apply adequately.

7.1.2.3 Two-Dimensional Image Properties and Processing

The standard USG images of cross-sectional slices of the object can be characterized as qualitative in the sense that the imaged parameter(s) represented by the shade of gray (brightness) in the image are not well defined. As already mentioned, the echo signal to which the brightness corresponds is given by complex and combined phenomena that cannot be easily decomposed into clearly defined influences of particular tissue parameters. Thus, only the spatial description of the

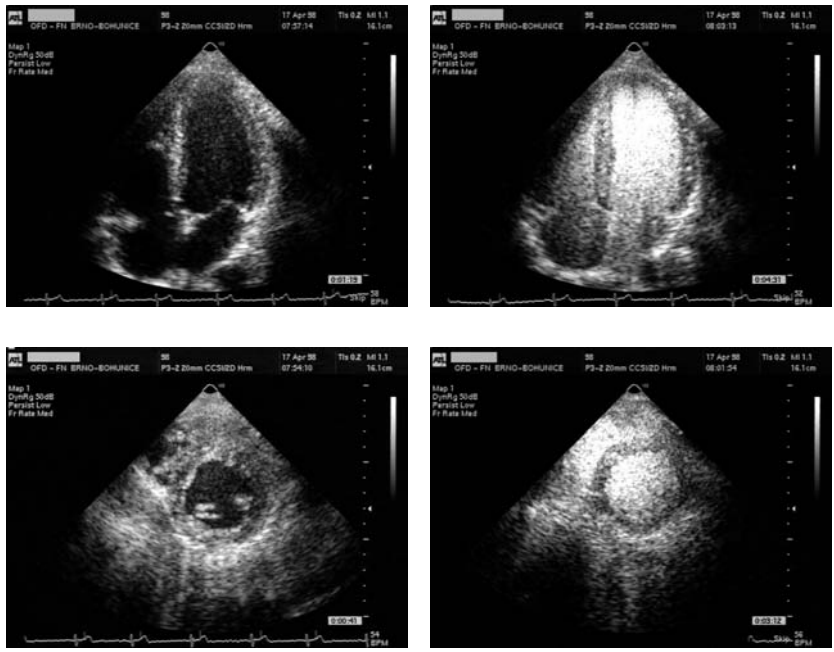


Figure 7.11 Example ultrasonographic cardiological B-scans. Left: Without contrast agent. Right: Chambers filled with contrast marked blood. (Courtesy of the Faculty Hospital Brno-Bohunice, Department of Functional Diagnostics, Assoc. Prof. V. Chaloupka, M.D., Ph.D.)

object structures (tissue areas, organs, lesions) is offered; nevertheless, it may give a clear picture of the concrete anatomy to an experienced ultrasonographer. Typical ultrasonographic images are shown in Figure 7.11.

Generally speaking, the resolution of USG images is still rather poor, particularly in the lateral direction, where it decreases with depth (~2 to 10 mm), while the resolution in the radial (depth) direction is better (about or under 1 mm) and invariant with r. The imaging is thus both anisotropic and anisoplanar. The attempts to improve the resolution by modified inverse filtering based on identification of local PSF have achieved interesting results in cases of separate point targets, but in real images the high level of noise (namely, speckles) complicates achieving a real improvement. Thus, the restoration attempts using formalized restoration methods (Chapter 12), like

modified two-dimensional Wiener filtering utilizing also the spectral information on noise (e.g., [25]), are more successful in the speckle noise suppression without blurring the image than in a substantial resolution improvement.

A characteristic feature of the USG images is their speckled appearance that, on one hand, theoretically carries certain indirect information on the imaged tissue microstructure; on the other hand, this information is rather hidden under influences of transducer parameters and system adjustment. Therefore, any conclusions concerning tissue characterization can be considered relevant only if made in comparison with images provided by the same system under identical adjustment. The methods applied to image areas in order to classify the respective tissues belong mostly to the category of texture analysis (Section 13.1.3); alternatively, the tissue character can also be assessed by analyzing directly the RF echo signal, using detailed models of ultrasonic impulse and mechanical properties of tissue [23].

More frequently, the speckles are considered a disturbing texture that may distort or interrupt boundaries of image areas and hide small details. Many methods have been tried aiming at speckle suppression, from heuristic approaches to highly formalized ones, but so far no really satisfactory solution is available that would smooth the image areas corresponding to homogeneous structures while not spoiling the anatomic spatial information. It should be understood that speckles cannot be considered additive noise; though their character is complex, more adequate is the idea of speckles being a rather multiplicative type of disturbance. This follows even from theoretical considerations showing that the envelope amplitude of the RF signal has a Rayleigh probability distribution with the characteristic feature of a constant ratio (approximately 2:1) of its mean and standard deviation. Hence, increasing intensity of the signal does not change the signal-to-speckle-noise ratio. The multiplicative character of the speckle noise excludes linear processing from speckle suppression procedures; nonlinear processing has to be applied. The so far partially successful approaches are all nonlinear, should they be derivatives of median filters (Section 2.2.2) or more or less heuristically designed procedures basically estimating each speckle region and replacing its value with the mean of the area. The more formalized approach is based on demultiplication via homomorphic filtering (Section 1.3.4), transforming the multiplicative mixture into the additive one by a logarithmic transform, which allows the application of linear filtering as the core procedure. In fact, this was also the case of the above-mentioned

Wiener restoration, as this has been applied on the data, the dynamic range of which had been suppressed by previous point-wise logarithmic transform. Some further details on the methodology can be found in Section 11.3 on noise suppression and in Chapter 12 on image restoration.

Due to speckle noise presence, the task of reliable USG image segmentation is more difficult than in other modalities. Standard approaches of segmentation based on thresholding of brightness or on edge detection fail as the contours found this way are rough, imprecise, and mostly not closed. Thus, more sophisticated methods that implement inherent constraints have to be used (Section 13.2). Promising and partly already routinely used approaches are the methods of deformable contours that may possibly utilize the initial estimates delivered by the operator, deformable models that may utilize *a priori* anatomical knowledge, or a recently suggested method of active shape models based on statistics of a set of images of the relevant area.

Sometimes, though less often, USG images are to be registered, usually in order to allow for noise suppression by averaging images with diversified speckles, or to enable subtraction of images to enhance a differential feature, even less frequently for the sake of comparison with another modality image. Again, the speckled structure practically excludes the methods based on detailed landmarks, and the only suitable matching criterion seems to be the mutual information (Section 10.3). This applies to intramodal (among USG images) registration, and even more to intermodal registration, when combining USG images with those from CT, magnetic resonance imaging (MRI), or gammagraphy.

A routinely used final step in analysis of segmented USG images is the evaluation of geometrical — linear or volume — parameters, usually based on simple geometrical models and still relying on interactive, rather than fully automated, positioning of characteristic points in the image.

7.1.2.4 Contrast Imaging and Harmonic Imaging

Contrast imaging utilizes contrast agents, administered to patients usually intravenously, that enable better tracing of blood flow and delineation of blood-filled volumes, like large arteries or heart cavities. Also, the blood labeled by a contrast agent may fill an organ (e.g., a muscle like myocardium); thus, perfusion of the organ and its time development may be assessed from the developing echogeneity of the corresponding area in the USG image sequence.

The contrast agents consist of microscopic gas bubbles encapsulated in envelopes (e.g., of albumin) that, after administration, are flowing in blood. The bubbles resonate at frequencies used in USG, thus generating a strong echo in blood vessels or in well-perfused tissues. Thus, common two-dimensional imaging, as described so far, can be used to describe tissue structures in the standard way, but with enhanced blood depiction thanks to contrast agents.

Besides that, there are several special modes based on contrast agents that are interesting from the signal processing aspect. Due to the nonlinear character of the mentioned bubble oscillations, i.e., because their instantaneous dimensions are not exactly proportional to the applied acoustic pressure, higher harmonic frequencies of the basic (fundamental) frequency of the incident ultrasound are generated and can be detected in the received signal. With specially equipped USG systems, it is possible to suppress the basic RF frequency band in the response and to let pass only the frequency band of the second harmonic frequency. This signal naturally carries different information than the first harmonic; this way, a new modality called *harmonic imaging* is created. Though theoretically having lower radial resolution due to longer transmitting impulses used, it provides much better contrast of blood.

It has been found that the harmonic signals are partly generated in tissues even without contrast agents, thanks to a certain degree of nonlinearity in tissue properties. Thus, the harmonic imaging becomes a recognized alternative even in noncontrast two-dimensional imaging. Generally, it complements the information of the standard imaging; it has slightly lower radial resolution, but visibly increased lateral resolution (the effective beam thickness contracts, as only the central higher-power part of the beam generates the harmonics). The harmonic imaging, both contrast based and noncontrast, suffers much less from the influence of secondary reflections and scatter, as well as from the effects of directional side lobes of the transducer than the fundamental frequency imaging, because the harmonic signals are not generated by the transducer, but in the imaged objects. The harmonics are produced only when the original ultrasound is strong enough, which is the case almost solely near the main beam axis.

The way of processing harmonics is naturally more complex, requiring wideband transducers, which would cover both the transmitting band around f_0 and the received band around the second harmonic $2f_0$. During receiving, the transducer signal is filtered by a band-pass or a high-pass filter stopping the fundamental frequency band.

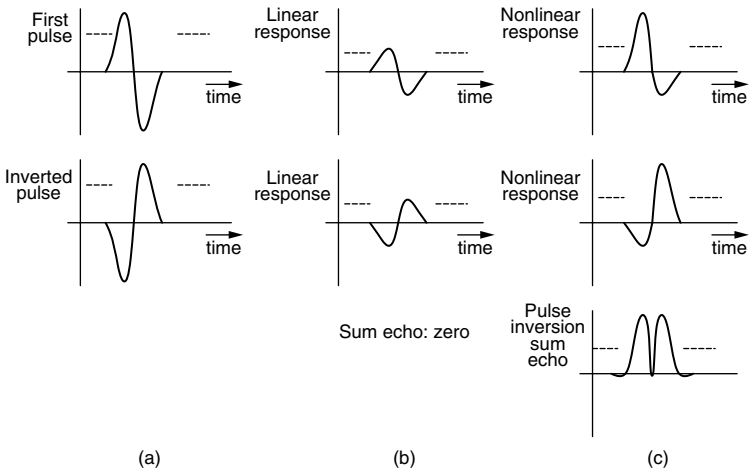


Figure 7.12 Principle of pulse inversion detection of signals produced on nonlinearities: (a) inversed pulses, (b) linear response, and (c) nonlinear response.

Another approach to detect and process the signals generated due to nonlinear behavior of bubbles is *pulse inversion imaging*, basically suppressing standard echo in favor of the differential signal caused by the nonlinearity. It works as depicted in Figure 7.12. Two ultrasonic pulses are sent sequentially, the second being inverted with respect to the first. Should the medium be linear, both responses would be identical but opposite, and after subtraction the result is zero (panel b). However, if there is a nonlinearity due to bubbles, the signals are distorted differently, as seen in panel c, and the resulting difference is the sought signal. The pulse inversion technique obviously doubles the acquisition time and requires more complicated signal processing, but it provides superior contrast of blood-filled volumes.

Quite a different approach to exploitation of contrast agents is *transient disruption imaging*, which detects the strong acoustic emission caused by disruption of bubbles when insonified intensely by an imaging impulse. The detected strong emission arrives at a time corresponding to the target depth, as if it were the normally scattered signal, thus contributing to the image as a special bubble contrast. This is usually further enhanced by subtraction of the same scene image obtained immediately after the disruption when the contrast agent is missing. The difference image emphasizes the blood-filled areas considerably. Still another method is *intermittent*

synchronized power imaging, also using the disruption-caused emission, though in synchrony with a particular phase of the heart cycle and evaluating the flow on the basis of recovery of bubble content after disruption. Image registration and averaging may be involved in all these procedures.

7.2 FLOW IMAGING

7.2.1 Principles of Flow Measurement

7.2.1.1 Doppler Blood Velocity Measurement (Narrowband Approach)

The Doppler effect concerns the change of the observed frequency of a wave field if there is a relative movement between the field and the observer. The frequency f_r measured (received) by the observer is then

$$f_r = f_0 \frac{c + v_0}{c}, \tag{7.12}$$

where f_0 is the frequency of the field as transmitted, c is the velocity of the waves, and v_0 is the relative velocity of the observer with respect to the wave field against the direction of wave propagation. If the observer moves with the velocity v at an angle α to the mentioned direction, then $v_0 = v\cos\alpha$ and, consequently,

$$f_r = f_0 \frac{c + v\cos\alpha}{c}. \tag{7.13}$$

When, on the contrary, the observer is fixed with respect to the medium in which the waves propagate, and the wave source is moving toward the observer with the velocity u (possibly obliquely at an angle β), the observed frequency becomes

$$f_r = f_0 \frac{c}{c - u\cos\beta}. \tag{7.14}$$

When both movements are combined, the received frequency becomes

$$f_r = f_0 \frac{c + v\cos\alpha}{c - u\cos\beta}, \tag{7.15}$$

so that the difference $f_0 - f_r$, called *Doppler frequency*, is

$$f_D = f_0 \left(\frac{c + v\cos\alpha}{c - u\cos\beta} - 1 \right). \tag{7.16}$$

In ultrasonography, the Doppler effect is used for measuring the blood velocity. Both mentioned partial phenomena are involved: the moving blood particles are insonified with a frequency modified due to particle movement in the ultrasonic field of the fixed transducer; consequently, due to scattering, every particle acts as a moving source of this modified frequency field, which is then detected by the transducer. Both effects are combined in synergy, obviously with $u = v$ and $\alpha = \beta$; as the blood velocity is much slower than the speed of ultrasound, Equation 7.16 may be simplified to

$$f_D = 2f_0 \frac{v \cos\alpha}{c}. \tag{7.17}$$

The Doppler frequency is thus proportional to the radial velocity of blood particles with respect to the transducer. As the particles contained in the insonified volume have generally different velocities, a band of Doppler frequencies will typically be observed, possibly including components corresponding also to opposite velocities. Obviously, the Doppler frequencies as defined may become negative; it is therefore necessary to define the Doppler signal in such a way that both radial directions can be resolved.

To describe the signal processing methods that allow deriving of the flow data, let us first consider the case of the transmitted continuous wave (CW) (sc., *CW systems*, when transmitting and receiving transducers are separated, though situated close to each other). In this case, the processing of the signal can be most clearly described in the frequency domain as depicted in Figure 7.13. While the transmitted signal is represented by a single frequency f_0, the received RF signal occupies a certain frequency band, $\langle f_0 - f_{D\max}, f_0 + f_{D\max}\rangle$ where f_{Dmax} corresponds to (±)maximum expected radial velocity (upper part, dashed line). It would be very difficult to separate both side bands corresponding to opposite directions of velocities in the original spectral location, as the necessary filter sharpness would be extremely high. Using complex-valued signal interpretation (see, e.g., [7] in Part I), the digital procedure leading to the separated Doppler signals can be explained in a clear and concise way.

The signal can be digitally shifted to the base-band (demodulated) by multiplication with the complex signal $\exp(-j\omega_0 t)$. The spectrum of the resulting complex signal (after filtering out the high-frequency components around $2f_0$) is then situated around zero frequency (solid line in Figure 7.13), with negative frequencies describing the velocities away from the transmitter, while the positive ones describe those toward it. It is obvious how to separate

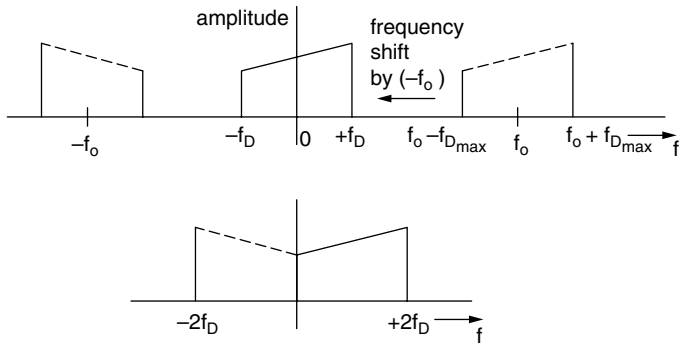


Figure 7.13 Spectra of ultrasonic signal in Doppler measurement by a CW system. Upper line: Forward and backward velocities represented by positive and negative frequencies, respectively. Lower line: Forward and backward velocities represented by high and low frequencies, respectively.

the forward and reverse flow signals: by asymmetrical band-pass filters with complex coefficients, for the upper- and lower-half bands*. The separated channels, both entirely in audible range, may be used for acoustic stereo presentation of the flow. A possibility for obtaining detailed information on the velocity distribution is to analyze the complex signal by flowing short-time discrete Fourier transform (DFT), which provides the time-dependent asymmetric spectrum with frequencies ranging from negative ones for reverse flow to positive ones for direct flow. The sequence of such spectra for one or several heart cycles constitutes a spectrogram — a plot of short-time spectra vs. time, in this application often called a *sonogram*. Besides that, the separate signals of each channel may be reproduced acoustically by stereo speakers, thus representing the velocities on the stereo basis as welcome by clinicians.

Another possibility of reasonable frequency shift is to move the spectrum by only $f_0 - f_{D\max}$ via multiplication of the signal by $\exp(-j(\omega_0 - \omega_{D\max})t)$, thus obtaining the spectrum in the range $\langle 0,\ 2f_{D\max}\rangle$ as seen in the lower part of Figure 7.13 (solid line, plus

*It can be shown that the partial bands can be separated from the complex signal concretely by certain phasing filters or via Hilbert transform, e.g., two complementary analytic filters may in principle serve as the half-band filters.

dotted line after omitting the imaginary part, thus providing a real signal). The spectrum of this signal obviously describes the forward and reverse flow by different frequency bands (reverse flow is under, while forward flow is above $\omega_{D\max}$) and may therefore be used for graphical representation of flow in the *sonograms* as well. Besides that, this signal is still completely in the range of audible frequencies and is thus suitable for a single-channel acoustic representation. However, it should be stressed that the frequencies in this spectrum are not Doppler frequencies ω_D, but rather $\omega_D + \omega_{D\max}$.

The volume of sensitivity of a CW system is given by the total length of the transducer beam, without any possibility of resolving the position of the velocity contributions. Thus, the result is a ray (beam) integral of the velocity distribution.

The needed possibility to localize the flow along the radial distance (depth along the beam) is added in *pulsed-wave (PW) systems*, transmitting the pulsed wave — regularly repeated ultrasonic impulses. These are similar to, but generally longer than, impulses used for B-scan, the length being determined by the necessity to approximate reasonably the single-line spectrum of the CW method. The response arriving with a certain delay after transmission corresponds to a certain depth, like in B-scan imaging, so that when selecting a short received signal segment (the length determined by the considered extent of depth and the length of the transmitted impulse), we obtain the information on the volume corresponding just to the selected depth range. Nevertheless, the Doppler signal analysis, described above, does not directly apply to the short signal segments because they cannot describe the (low) Doppler frequencies sufficiently. Rather, the segments may be considered samples of the complete signal, sampled with the frequency of pulsing — pulse repetition frequency f_{pr}. This repetition frequency is relatively low — in tens of kilohertz — seemingly insufficient to comply with the sampling theorem when the mean ultrasound frequency is in units of megahertz. Nevertheless, the periodicity of the sampled signal spectrum corresponds to f_{pr} (see, e.g., [7] of Part I), and as seen in Figure 7.14, where the original spectrum is plotted by a solid line and the replicas due to sampling by dotted lines, no aliasing appears if the Doppler band of the signal is sufficiently narrow. The remaining processing may clearly be similar to that in CW systems, the only important difference being the possibility of aliasing in case of excessively high velocities. Obviously, the highest velocity $v_{\max}$, which does not cause aliasing, is given by $f_{D\max}$ via

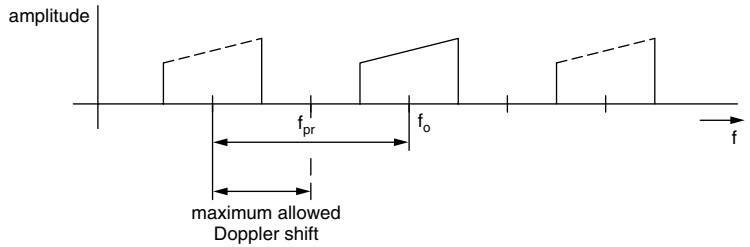


Figure 7.14 Spectra of signals in Doppler measurement by a PW system.

Equation 7.17; the sampling theorem requires that

$$f_{D\max} = 2f_0 \frac{v_{\max} \cos\alpha}{c} < f_{pr}/2\,. \tag{7.18}$$

The maximum pulse rate frequency is naturally given by the inverse of the return time from the deepest location (Equation 7.9). This limitation should be observed; in case of extreme velocities, the eventually aliased spectrum must be interpreted with care.

It is obvious that for a good representation of the Doppler spectra, relatively long sections of the signal are needed; in PW systems, mostly sections of several hundred samples (i.e., the same number of ultrasonic impulses) are involved in a spectrum. While this is fine in dedicated blood flow Doppler equipment, since it corresponds to time resolution in tens of milliseconds, as needed for a good description of changes during a heart cycle, it is prohibitive in the case of flow imaging, where several tens of measuring lines must be provided during a frame with a reasonable frame rate above 10 sec^{-1}. How the evaluation of the signal should be modified sacrificing the spectral details in favor of speed and quantity of measurements will be a subject of the next section.

The PW systems as described are capable of providing the flow information for a particular depth, while the information content of the rest of the signal, refused by the gate, is discarded. Obviously, when more identically working parallel blocks are added, differing only in gating with diverse delays after the transmission impulse, the flow information as a function of depth may be provided. Such systems, providing the time-dependent flow profile along the complete ray, are denoted *PW multigate systems*. However, the parallelism is in fact superfluous, as the signal to individual gated subsystems comes naturally sequentially; hence, basically serial

processing that acts equivalently is possible. The flow profiles are exactly what is needed for constructing flow images, but the time requirements of the mentioned approach are prohibitive for real-time imaging. A modified simpler and fast approach, providing less detailed information, is therefore used in imaging (see Section 7.2.2).

In all blood velocity measurements, it is desirable to suppress the relatively strong Doppler signals of low frequencies caused by slowly moving specular structures, such as artery walls. This is provided for by band-pass filters that allow only the frequency band corresponding to a reasonable blood velocity range, and also suppress the unwanted high-frequency products of the frequency shift operations. When the transmitted impulse signal has a sufficiently narrow band and all the spectral components originating from the original frequency band are removed by the filtering, all the remaining components are due to the Doppler effect and represent the flow. Summing (averaging) the squares of all the passed components then provides the estimate of the *power Doppler signal*, an important quantity, namely in flow imaging.

7.2.1.2 Cross-Correlation Blood Velocity Measurement (Wideband Approach)

A completely different approach to flow estimation is based on the direct radial movement measurement in the space domain. When a target response can be identified on the measured ray in two consequential measurements (A-scans), set apart by a time difference Δt (sometimes called the slow-time difference), in the positions differing by Δr, the radial velocity is obviously

$$v_r = \frac{\Delta r}{\Delta t} = \frac{0.5\,\tau\,c}{\Delta t}, \tag{7.19}$$

where τ is the difference between arrival times of the target responses in the respective A-scan signals. The time difference is usually, but not necessarily, the period between immediately succeeding pulses, $\Delta t = 1/f_{pr}$; in the case of low velocities, i.e., small shifts during a single repetition period, it may be advantageous to take Δt over several repetitions. It is clear that this principle of velocity measurement, which requires short (hence wideband) imaging impulses, is not related to the Doppler effect requiring narrowband excitation, and the use of the Doppler name for the flow imaging based on this method is inappropriate, though common.

The identification of corresponding moving target responses on the signal A-lines $s_1(t)$ and $s_2(t)$ is based on the assumption that, generally, a configuration of a blood cell group does not change significantly during the time Δt; i.e., it is moving basically as a whole. It determines then, via scattering, a particular signature in the received signal, in the form of a characteristic short signal segment that can be found on both A-lines. Nevertheless, complete identity cannot be expected, and a similarity criterion must therefore be defined in order to seek the best correspondence of the segments.

The method relies on fully digital implementation so that the received RF echo signal is digitized immediately after initial amplification. This way, the discrete sequences $s_1(n)$ and $s_2(n)$ are obtained, n being the sequential index, corresponding to sampling along the ray (sc., fast-time sampling). Both sequences are divided into segments of N samples, with N corresponding to the expected signature length. The purpose of the next step is to determine the optimal shift Δn_i of the i-th segment of $s_2(n)$, with respect to the corresponding segment of $s_1(n)$, that would lead to the best match. These shifts, determined for each segment pair, then yield, via known sampling frequency, the delays τ_i and the radial velocities v_{ri} in the individual segments according to Equation 7.19. This way, the complete velocity profile along the investigated ray can be determined.

Several segment similarity criteria have been suggested, obviously the best being the cross-correlation coefficient,

$$R_{12}(\Delta n, i) = \frac{\sum_{n=0}^{N-1} s_1(n + iN)\, s_2(n + iN + \Delta n)}{\sqrt{\sum_{n=0}^{N-1} s_1^2(n + iN) \sum_{n=0}^{N-1} s_2^2(n + iN + \Delta n)}}, \tag{7.20}$$

which is to be calculated for a range of Δn corresponding to the maximum expected velocity v_r. Obviously, the optimal (maximum similarity) shift is not influenced by the first factor inside the square root, which may therefore be omitted from computation. However, even with this simplification, the computation repeated for every segment and for each new pulse is rather intensive, even when taking advantage of realizing the calculations via the frequency domain utilizing the convolution theorem. The obtained best-fit shift $\Delta n_{i,opt}$,

$$\Delta n_{i,opt} : \; R_{12}(\Delta n_{i,opt}, i) = \max\{R_{12}(\Delta n, i)\}, \tag{7.21}$$

determines the sought velocity value $v_{r,i}$ at the respective depth.

This principle has some advantages over the Doppler method: primarily, it does not suffer with the aliasing problem; also, when used for flow imaging, it is better compatible with wideband transducers designed for B-scanning. On the other hand, it is computationally very intensive, exceeding the requirements of the Doppler approach by more than two decimal orders. There is also a possible source of errors in falsely found correspondences, as the criterion in Equation 7.20 does not provide a reliable measure on the quality of the found optimal match. A detailed assessment of both methods can be found in [17]. The demanding realization is probably the reason why the cross-correlation method is still implemented much less in commercially available equipment, and used is in flow-imaging systems more than in quantitative dedicated blood flow meters. Interesting implementation approximations include rounding of the correlated signals to a single-bit precision, i.e., counting only sign products in Equation 7.20.

7.2.2 Color Flow Imaging

Dedicated blood flow measurement usually provides the velocity profile along an ultrasonic ray, or the sonogram for a chosen spatial element on the ray. It turned out to be very advantageous, besides the mentioned quantitative measurement, to combine the more qualitative two-dimensional flow imaging with standard B-scan imaging that provides auxiliary information on the anatomy of the imaged region. Commonly, the gray-scale anatomic image is combined with the color-coded flow image (usually available only on a part of the imaged area, in order to maintain a reasonable frame rate); this mode is generally called *color flow imaging* (CFI). Independently of the method of flow imaging, algorithms must be applied that decide on the level of flow data for a particular pixel, so that the pixel is displayed with either anatomic or flow image content. This way, only the pixels with a certain minimum level of detected flow (e.g., velocity or flow-related power) are displayed in color, while the others, depicting anatomic tissues, remain in gray, providing important support in localization and interpretation of the flow image data.

7.2.2.1 Autocorrelation-Based Doppler Imaging

The flow imaging based on the Doppler effect utilizes the same principles as PW multigate systems, described in the first part of Section 7.2.1. However, the detailed spectral estimates described there are based on numerous samples (several tens or hundreds) per a depth level, which are not available in flow imaging should a reasonable

frame rate be preserved, since each sample means a new impulse to be transmitted. With respect to this limitation, only several (3 to 16) impulses per image line are transmitted; hence, only the same number of samples per depth level is available, not enough for classical spectral estimates. A different approach, suggested by Namekawa et al. [56], limits its interest to only three quantities: the mean local power of the Doppler signal, the mean Doppler frequency, corresponding to the mean local radial velocity, and the variance of the velocity that describes the degree of local turbulence. These three local values can be estimated based on the mere few available samples, at the cost of sacrificing other details. The mentioned localization concerns the depth on the line (ray) expressed by the variable time t_r since the impulse transmission. It should be understood that the following Equations 7.22 to 7.31 all correspond to a particular chosen t_r, but this parameter has been omitted in order to simplify the notation.

The mean Doppler signal power $\bar{P}_D$ is, according to the energy preservation (Parseval) theorem,

$$\bar{P}_D = \int_{B_D} S(\omega)\,\mathrm{d}\omega \tag{7.22}$$

where $S(\omega)$ is the power spectrum of the Doppler signal and B_D is the frequency band occupied by the Doppler frequencies. It is important that only Doppler frequencies of the required type (blood flow) are included in B_D; the high-frequency band of the transmitted signal as well as the low-frequency Doppler signals due to tissue movement must be excluded by preceding filters.

The mean Doppler frequency $\bar{\omega}$ may be obviously defined as

$$\bar{\omega} = \mathrm{E}\{\omega\} = \frac{\int_{\mathrm{B_D}} \omega\, S(\omega)\mathrm{d}\omega}{\int_{\mathrm{B_D}} S(\omega)\mathrm{d}\omega}; \tag{7.23}$$

the variance of the frequency σ_ω^2 is similarly

$$\sigma_\omega^2 = \frac{\int_{\mathrm{B_D}} (\omega - \bar{\omega})^2\, S(\omega)\mathrm{d}\omega}{\int_{\mathrm{B_D}} S(\omega)\mathrm{d}\omega} = \mathrm{E}\{\omega^2\} - (\bar{\omega})^2. \tag{7.24}$$

Providing that the processed RF signal is generated by a stationary and ergodic stochastic process, its autocorrelation function $R(\tau)$ is, according to the Wiener–Khintchin theorem, the inverse Fourier transform of the power spectrum,

$$R(\tau) = \int_{B_D} S(\omega) e^{j\omega\tau} \, \mathrm{d}\omega, \tag{7.25}$$

which, when differentiated with respect to τ, yields

$$\dot{R}(\tau) = j \int_{B_D} \omega S(\omega) e^{j\omega\tau} \, \mathrm{d}\omega \tag{7.26}$$

and

$$\ddot{R}(\tau) = -\int_{B_D} \omega^2 S(\omega) e^{j\omega\tau} \, \mathrm{d}\omega. \tag{7.27}$$

By comparison of previous equations, we obtain

$$P_D = R(0), \tag{7.28}$$

$$\bar{\omega} = -j \frac{\dot{R}(0)}{R(0)} \tag{7.29}$$

and

$$\sigma_\omega^2 = \left(\frac{\dot{R}(0)}{R(0)} \right)^2 - \frac{\ddot{R}(0)}{R(0)}. \tag{7.30}$$

We arrived at an interesting result: all the needed local quantities can be derived from the values of the autocorrelation function at $\tau = 0$, and its first and second derivatives at the same point. As the function is derived from discrete samples, the derivatives can only be approximated using the first and second differences,

$$\dot{R}(0) \cong \frac{1}{T}[R(T) - R(0)], \quad \ddot{R}(0) \cong \frac{1}{T}[\dot{R}(T) - \dot{R}(0)], \tag{7.31}$$

T being the pulse repetition period. To calculate these estimates, the values $R(0)$, $R(T)$, and $R(2T)$ are needed that can be estimated

as simple time averages of pair products of signal values from at least four consecutive samples, as

$$R(0,t_r)=\frac{1}{M-1}\sum_{k=1}^{M-1}{}_k s^2(t_r),\ \ R(T,\ t_r)=\frac{1}{M-2}\sum_{k=1}^{M-2}{}_k s(t_r)\ {}_{k+1}s(t_r),$$
$$R(2T,t_r)=\frac{1}{M-3}\sum_{k=1}^{M-3}{}_k s(t_r)\ {}_{k+2}s(t_r), \tag{7.32}$$

where the variable t_r is the time since the impulse transmission, localizing a particular set of correlation values in a certain depth; M is the number of pulses per ray and k is the sequential index of the pulse. The necessary number of pulses per line (ray) is thus reduced to a minimum of four; when more samples are available, the estimate variance is lowered. By investigating the properties of the complex autocorrelation function $R(\tau)$ derived from the complex RF signal, it has been shown that the three desirable quantities could be provided knowing only the complex values $R(0)$ and $R(T)$, so that the minimum number of samples can be further decreased to three. The details can be found in [16] or [17], where some alternative, less frequently used approaches are also discussed.

As several pulses per line are needed, the frame rate decreases correspondingly in comparison with the plain B-scan imaging. In order to maintain a reasonable frame rate, the user-adjustable area, where either color flow imaging or power Doppler imaging (see below) is applied, is usually smaller than the complete image; the repeated and longer impulses are limited only to lines forming this limited area.

The obtained values may be used for CFI representing the direction of flow in a color scale (forward, red shades; reverse, blue shades) and roughly also the velocity of flow by coloring the pixels in the interpolated color scale. Only pixels with the power of the Doppler signal in excess of a chosen threshold are colored. The variance of the frequency is usually added to the displayed information in the form of additional yellow tone, the intensity of which indicates the local turbulence by changing the resulting colors to either orange–yellow on the forward side or cyan–yellow on the reverse side of velocities.

Another possibility is to represent only the Doppler signal power in a color scale (usually dark blue or brown via orange to yellow); then it is called *power Doppler imaging* (PDI). This mode, resigning any velocity or directional information, obviously needs only $R(0)$ to be evaluated. This can be estimated using the maximum

of available sample pairs, obviously with a lower variance than $R(T)$ and $R(2T)$, and further improved by time averaging as the imaged information is less time variable than velocity. Thus, this mode generally has a better SNR and provides more stable images, enabling also imaging of perfused areas of tissue with small blood vessels, including the vessels perpendicular to the rays, thanks to the phenomenon of spectrum broadening. These possibilities are achieved at the cost of sacrificing the information on the magnitude and direction of velocity; the Doppler power corresponds to the volume of moving blood, rather than to its velocity.

7.2.2.2 Movement Estimation Imaging

The color flow imaging based on movement estimation in the original domain of the digitized RF echo signal uses the concept described in the second part of Section 7.2.1. The cross-correlation principle yields the optimum time delay differences corresponding to the spatial shifts of blood clusters during the time between successive pulses. The time delay finally determines the sought radial velocities (including direction). An important advantage of this mode is that it does not suffer from aliasing due to insufficient pulse repetition frequency, which is often difficult to prevent (or to interpret properly) in Doppler CFI of more complicated anatomic structures. The power of the velocity signal is not readily available via correlation, so that the power Doppler imaging should be arranged on the spectral basis (as the name indicates), i.e., using properly filtered signal, as in the previous section.

A note concerning the correlation interpolation: the optimum shift, calculated by Equation 7.21, can only correspond to an integer number of sampling distances along the ray. If this is too rough, it is possible to interpolate among the neighboring similarity criterion values, this way providing the subsample precision of $\Delta n_{i,opt}$ and hence more precise velocity determination.

7.2.2.3 Contrast-Based Flow Imaging

Using contrast agents ([3]; see also the last part of Section 7.1.2) in the flow imaging is generally profitable, as it increases enormously the Doppler components of the echo signals. This leads to a notable improvement in SNR and, consequently, to suppression of ambiguities in decisions on whether to display the color flow information. Combined with special techniques, such as, e.g., disruption imaging, it may give results not available otherwise.

7.2.2.4 Postprocessing of Flow Images

Flow imaging suffers from several typical phenomena observed in the final images, mainly due to the rather stochastic character of the flow signal. Primarily, it leads to dropouts of colored values in flow-containing areas as well as to some colored pixels in stationary tissues, both randomly changing from frame to frame. Generally, the appearance of the flow image, especially of the velocity (CFI) image, is influenced by the stochastic changes between following frames, thus causing both disagreeable flicker and ambiguities as to decisions on whether the gray-scale B-image or colored flow information should be displayed. This in turn has the consequence of frayed time-variable borders, which are rather unpleasant and disturbing to the observer.

These phenomena can be partly suppressed by spatial in-frame processing (Section 2.2); local linear averaging or median filtering in small neighborhoods of processed pixels, or similar *ad hoc* procedures, would suppress the randomness to an extent.

Another possibility for suppressing the noise and ambiguities, and consequently the flickering, is to use some kind of temporal (interframe) processing. A partial suppression of the flickering of individual pixels may be expected from temporal uniform or exponential averaging or median filtering (Section 11.3). On the other side of the line of complexity, sophisticated, mostly intuitively based, nonlinear procedures have been designed that take into account the time development and decide, based on circumstances, whether to accept or reject certain partial operations; such procedures react differently according to the local and instantaneous situation. These rather specialized methods (as listed, e.g., in [17]), adapted to a particular type of imaging, system properties, or a type of examination, are beyond the frame of this book.

7.3 THREE-DIMENSIONAL ULTRASONOGRAPHY

Ultrasonography provides basically tomographic data, which can be arranged into three-dimensional data blocks describing the interior of the imaged object, so that the analysis and display methods used for three-dimensional imaging in other modalities may be applied adequately. In principle, data may be collected ray by ray and compiled into a three-dimensional data block taking into account the instantaneous position of the transducer and the image data format determining the ray geometry.

However, several problems appear when trying to apply this simple principle:

- Long acquisition times, as given by physical limits of the USG modality
- Precision of transducer probe localization, and consequently proper positioning of measured data in the set
- Anisotropic and inhomogeneous image data as provided by the ultrasonic modality, which nevertheless should be combined in a consistent data block
- Unequally dense space sampling, which is connected with the methods of ray positioning, and hence also with the transducer scan format
- Consequently, the need of rather complicated interpolation with resampling
- Low signal-to-noise ratio due to strong three-dimensional speckle texture

7.3.1 Three-Dimensional Data Acquisition

Even though the three-dimensional data acquisition based on a single-ray transducer would be in principle possible, the procedure would be obviously too heavy going, and thus impractical. The three-dimensional ultrasonic imaging may therefore be reasonably based either on sequentially collected two-dimensional scans with gradually repositioned slices (tomographic planes) or on providing directly the three-dimensional data by means of specially designed three-dimensional transducers.

7.3.1.1 Two-Dimensional Scan–Based Data Acquisition

Diverse possibilities of three-dimensional data acquisition based on two-dimensional scans (slices) are schematically depicted in Figure 7.15. If the imaged scene is stationary, the third dimension can be gradually investigated by many types of systematic slice movement: the linear shift perpendicular to the slice planes (panel a), which may be combined (panel c) with an in-plane movement that effectively increases the cross-sectional (slice) area (panel b), or an inclination providing a pyramidal block of slices (panel d), or a rotation along the longitudinal axis of the transducer (panel e). All of these movements can be provided by mechanical supports guaranteeing via electronic control identical distance or angle differences among the slices, thus simplifying

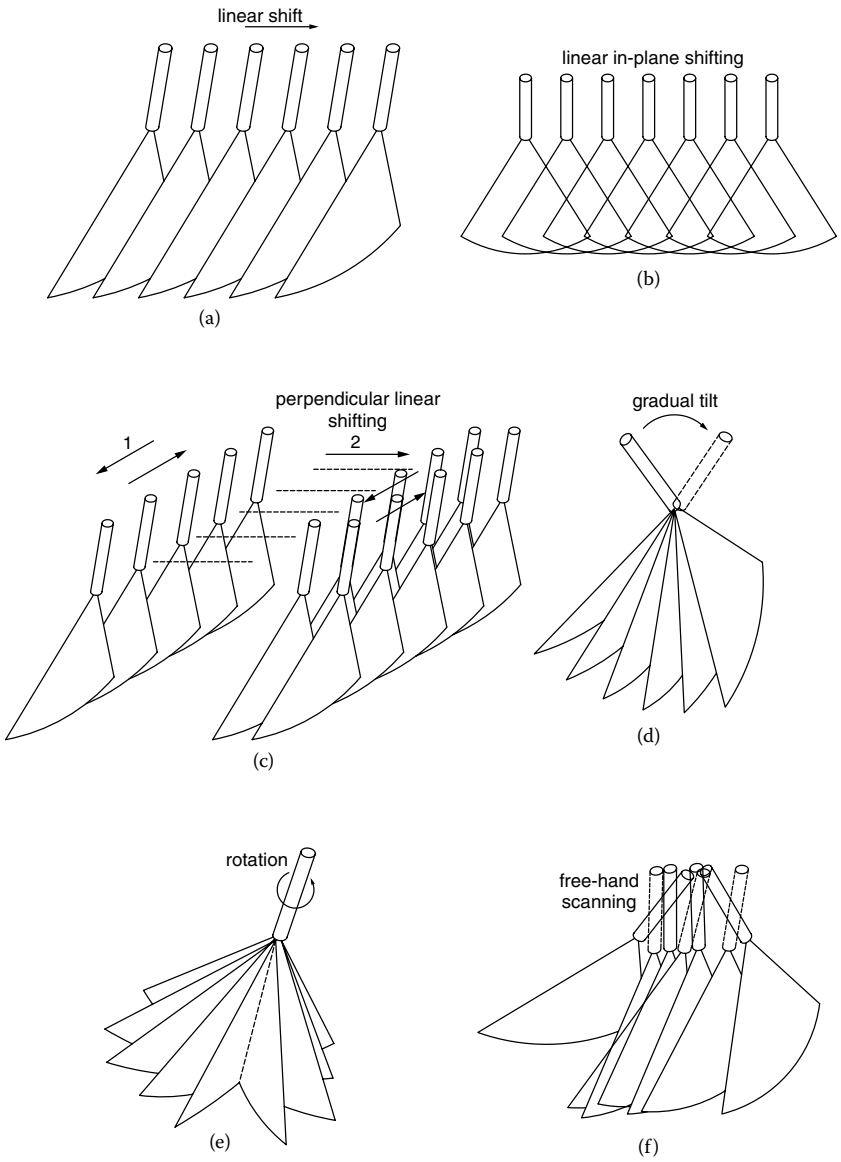


Figure 7.15 Schematic arrangements of two-dimensional scans as three-dimensional data sources: (a) regular parallel scans, (b) in-plane scans, (c) combined in-plane and shift, (d) tilt scanning, (e) rotational scanning, and (f) irregular (freehand) scanning.

the following data processing. Endoscopic probes may offer additional (not depicted) possibilities by means of their shift and rotation.

Panel (f) of the figure depicts an example of a random scanning as provided typically by *freehand scanning*. This is perhaps the most acceptable method for an experienced echoradiologist, as such a three-dimensional examination does not differ in principle from a standard procedure: the radiologist collects (and evaluates) the best available images by moving the probe (possibly taking care of certain regularity in the movement). During the examination, the image data are stored together with the transducer position data and gradually compiled into the three-dimensional block. Afterwards, i.e., offline, the data are processed to obtain a consistent data block, from which differently oriented slices or reconstructed three-dimensional views may be generated. Providing the precise information on the instantaneous position of the probe (six parameters—three spatial coordinates plus three spatial angles of rotation/swing) is a self-standing problem. Several physical principles are routinely used for this purpose: magnetic, acoustic, or optical positional systems are commercially available. The most interesting approach from the image processing point of view relies on decoding the spatial information contained in the degree of two-dimensional or even three-dimensional correlations between the image data of neighboring scans. Deriving the local spatial shift by exploiting the known spatial variability of anisotropic correlation, and this way determining even the global shift, tilt, and rotation of the neighboring scan planes, might limit or eliminate the need of the special localizing equipment.

In the simplest cases of a reasonably regular freehand movement (mere incline or shift), the position information need not be necessarily recorded; the image data are then supposed to be collected in a prescribed regular manner and situated correspondingly in the block. Naturally, such imaging may not be considered reliable enough with respect to absolute geometrical dimensions.

Cardiological three-dimensional echo imaging poses a particular problem, as the imaged structures are quickly moving so that the image data do not describe a stationary scene as supposed so far. In this case, four-dimensional data must be provided, with time as an additional parameter of every image data item. Nevertheless, due to limited capabilities of the three-dimensional imaging via two-dimensional scans, it is impossible to provide sufficiently dense data for any particular heartbeat phase during a single cardiac cycle. The data must be collected throughout several cycles (that should be similar, with exclusion of possible extrasystoles), together with

the timing related to the last QRS complex of the ECG signal, indicating a fixed initial phase. This naturally complicates the acquisition process substantially; the unpleasant consequence is that the examination is prolonged. Nevertheless, still there seems to be clinical interest in developing the relevant methods further.

7.3.1.2 Three-Dimensional Transducer Principles

In principle, any transducer providing two-dimensional scans can be converted to a three-dimensional image data transducer by providing a means of, e.g., inclining it with respect to the probe holder. Naturally, the axis of inclination must lie in the plane of each slice, thus forming the line of intersection of all the slice planes, as in Figure 7.15d. Probes for three-dimensional imaging, using two-dimensional sector scanners (as, e.g., one-dimensional phased arrays) that would be periodically inclined or rotated along such an axis mechanically, are imaginable.

Nevertheless, the development aims at probes without mechanically moving parts, i.e., *true two-dimensional phased arrays* that may orient the ray to any direction in the frame of a relatively wide solid angle, on the same principle as explained for two-dimensional imaging in Section 7.1.2. The technological development solved the problems with producing the two-dimensional phased arrays containing numerous evenly sensitive subcrystals; thus, the main obstacle remains the relatively long time needed for the acquisition of a single three-dimensional image. If the ray-by-ray principle is used, the acquisition time of a single three-dimensional data set by the two-dimensional phased array is equivalent to that of the concept with mechanically inclined standard B-scan transducer: with the usual frame rate of tens per second, the acquisition time per three-dimensional set of about a hundred slices requires several seconds. While this may be acceptable for static anatomic structure imaging, it is too slow for dynamic real-time imaging of any moving organ. It is definitely insufficient for cardiological applications, unless synchronization with a chosen heart cycle phase is provided; this extremely prolongs the examination times.

The acquisition time can be substantially shortened by using the principle of *multibeam imaging*, as explained in Figure 7.5. Naturally, in the case of three-dimensional imaging, the number of simultaneously measured rays should be as high as possible in order to speed up the acquisition sufficiently. Obviously, the extreme speed reaching the physical limit is achieved when a single transmitting impulse irradiates the whole solid angle of the intended FOV. In this case, the

insonification would not provide any lateral or transversal (elevation) selection; all the directional information in the acquired signals is then based entirely on the beam deflecting (and focusing) during the receive phase. The data for all the rays forming the three-dimensional set can be in principle collected simultaneously. The digital RF signal of each subcrystal channel is in that case delayed in parallel by as many generally different delays as there are rays in the set (M, of the order 10^4 or more); for each ray, an adder is provided, summing the adequately delayed signals from all the subcrystals. This way, the digital signals for each needed ray direction are provided in parallel instead of sequentially as usual, and the achievable frame rate would be given by the inverse of a single return time (Equation 7.9).

However, such a two-dimensional phased array transducer becomes extremely complicated: should the two-dimensional array have the square shape with N (tens to hundreds) subcrystals on each side, indexed by i, k, then N^2 (i.e., thousands to tens of thousands) individual subtransducers and as many signal channels, including the input amplifiers and A/D converters, are needed. Each of those signals must be delayed in parallel by M different delays $\Delta t_{i,k;m}$, each being specific for individual sums forming the output signals $s_m(t)$; each of these M (thousands to tens of thousands) signals describes an individual ray. Such a system becomes feasible only thanks to modern large-scale integration circuitry that may be integrated into the transducer; also, it requires a very high degree of parallelism in the data processing needed for image reconstruction and display.

Less complicated would be the three-dimensional transducer using the same phased array, but forming simultaneously only rays of a single B-scan; the individual B-scans are acquired sequentially by adequately modifying the delays. In this case, the number of different delays per subcrystal, equal to the number of adders, is reduced to the number of rays (R) in a slice (about a hundred). The receiving hardware and algorithms are simplified correspondingly. Besides that, it is in principle possible to improve the transversal (elevation) resolution by thinning the slice thickness via suitable transmit focusing that would concentrate most of the ultrasonic energy in the slice volume. It is obvious that the maximum frame rate decreases n_s times (n_s being the number of B-scans in the three-dimensional data set), thus substantially in comparison with the purely parallel acquisition. However, the resulting three-dimensional frame rate is still comparable with two-dimensional frame rates of the common B-scan imaging, and therefore allows real-time three-dimensional imaging. Note that providing subpyramids of the pyramidal FOV, instead of planar fan

frames, may provide similar results, while forming the transmission cones (instead of slices) might be easier.

Three-dimensional USG imaging provided by such special probes is very promising. It is still in the experimental phase, both technologically and clinically; however, the first commercial systems are already available.

7.3.2 Three-Dimensional and Four-Dimensional Data Postprocessing and Display

7.3.2.1 Data Block Compilation

The obtained USG data can be processed in basically two ways. It is possible to analyze each two-dimensional scan independently, thus identifying and localizing particular features or borders; these would in turn be used in three-dimensional space (matrix) to build a spatial structure describing more or less roughly the anatomy. Obviously, the information contained in the spatial relations among the scan data is not utilized then, which not only is a definite loss anyway, but also may lead to inconsistencies in the reconstructed scene.

Hence, to compile all the acquired data into a three-dimensional (or four-dimensional, with time development) block of data, and to base any subsequent analysis on this complete spatial information appears to be a better approach. It turns out that this seemingly elementary step of compiling the data structure from numerous B-scans, is a rather demanding action, the quality of which is decisive for the success of the final reconstruction. There are several problems involved in this procedure.

Proper positioning of the measured data into the block requires good localizing entry provided either by means of an external localization device or by correlation analysis of the image data itself—a separate, not completely solved, problem. With any localizing equipment, even if the coordinates of the probe were exact (not always the case), the localization of an individual data item might be distorted by ultrasound wave propagation phenomena (reflection, refraction, scatter), as explained in Section 7.1.1.3. Uneven sound velocity leads to errors in radial distance, whereas the refraction phenomena cause lateral and elevation errors; moreover, the secondary reflections and scatter may introduce artifacts in the form of grossly displaced and distorted ghost objects. While in B-scan imaging these errors lead only to shape distortion that is usually unimportant or easily recognizable from the diagnostic point of view, in three-dimensional imaging it may

cause data to become inconsistent: the data concerning the same location provided from different B-scans are mutually displaced. Consequently, there is a possibility of falsely associating unrelated data in the block. Thus, a careful analysis of the data similarity from scans to be associated is required, and possibly *localization correction* based on principles of image registration (see Section 10.3) may be needed.

Another problem is that the data on a scan plane are inhomogeneous and anisotropic. In principle, they could be restored by methods of formalized inverse filtering (Sections 12.3 and 12.4), but this may be computationally too demanding. Nevertheless, associating uncorrected data from different scans, which come from very different positions in the scans, leads inevitably to loss of three-dimensional data quality. Therefore, certain *homogenization* of the measured data is recommendable, before the data are incorporated in the three-dimensional block.

The measured scan data are randomly positioned with respect to the nodes of the three-dimensional rectangular sampling grid. A kind of *image data interpolation* is thus needed (see Section 10.1.2), which may be done in a single process with the data homogenization. The interpolation is complicated by the fact that the resulting samples at the grid nodes will be obviously derived from a variable number of randomly positioned input data with a different quality. The quality may be taken into account by weighting the input data correspondingly when interpolating; however, determining the quality is itself a difficult problem related to homogenization.

Finally, in case of moving object imaging (e.g., cardiological applications), a longer series of measurements is needed, filling the four-dimensional space by data covering the complete space volume of FOV and the time extent covering the complete movement (e.g., a heart cycle). As it may be difficult or impossible to cover a particular phase of the movement completely by the spatial data in a single cycle of the movement, several cycles should be measured and the data *synchronized* with the cycle phase. Above the high acquisition time and difficulty of measurement, inclusion of the fourth coordinate also adds to the complexity of the homogenization and interpolation task.

Most of these problems are alleviated or suppressed when using a true three-dimensional probe with a two-dimensional phased array transducer. Depending on the type of beam deflection control and on echo signal processing (see above), a compromise results. A shorter three-dimensional frame acquisition time with lower spatial resolution may be achieved when using a multibeam technique (possibly up to the extreme of a single spatially wide

transmitted pulse per three-dimensional frame), or the single narrow-beam transmit technique is used, allowing the highest spatial resolution at the cost of a (much) lower frame rate.

7.3.2.2 Display of Three-Dimensional Data

Once the three-dimensional (or four-dimensional) data block is provided, all the techniques used in the display of tomographic data can in principle be used, although the low ultrasonic image data quality complicates the situation in comparison with such modalities as x-ray CT or MRI. This does not impose many restrictions on the two-dimensional (multiplane) viewing, consisting of synthesizing arbitrarily positioned tomographic planes where the noise, primarily speckles, is an expected feature. However, in the true three-dimensional volume reconstruction (though projected on two-dimensional display), the speckles may be rather disturbing, as they severely complicate proper segmentation of the imaged spatial objects. As already mentioned, this may be the reason why methods of volume rendering based on spatial ray tracing are more commonly used than the surface-based methods, requiring the preliminary spatial segmentation. Moreover, the ray-tracing methods provide images in which the speckles are partly suppressed by spatial averaging. An example of a typical three-dimensional reconstruction from ultrasonographic data can be seen in Figure 7.16.

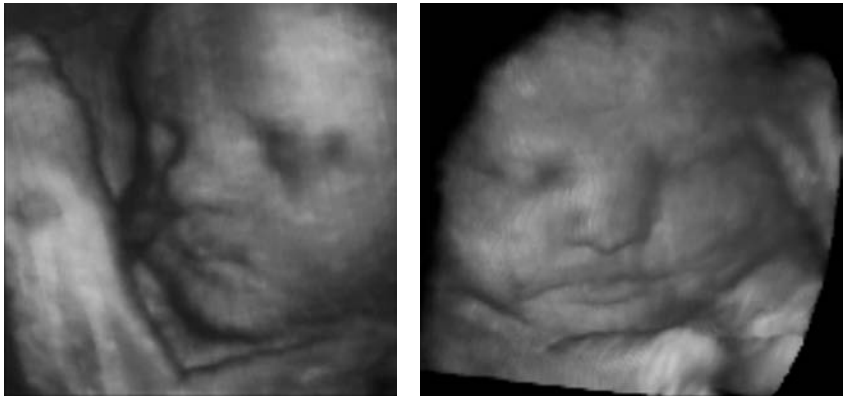


Figure 7.16 Three-dimensional reconstructions of fetal faces and hands from ultrasonographic data. (Courtesy of the Faculty Hospital Brno-Bohunice, Department of Obstetrics, Assist. Prof. R. Gerychova, M.D.)

An often used and perhaps preferred display mode in three-dimensional USG is a combination of multiplane viewing with three-dimensional rendering of a polyhedron, the walls of which are formed by the two-dimensional tomographic images. This way, the simplicity of the multiplane imaging is combined with the visually comprehensible (often revolving) spatial display of the polyhedron, which can be properly oriented in space.

8

Other Modalities

8.1 OPTICAL AND INFRARED IMAGING

Optical imaging as a modality providing medical diagnostic information is perhaps generally considered somewhat marginal and often not listed among the standard modalities of medical imaging. The reason may be that the information carrier — visible or infrared light — cannot provide information on deep tissue structures because the penetration of light is limited to a rather thin layer in most tissues.

Nevertheless, several important areas of medical optical imaging deserve mentioning. It is primarily optical microscopy that provides the image information on thin tissue layers, nowadays even as three-dimensional data. Optical imaging naturally plays a significant role in ophthalmology, where the diagnosed structures are transparent or translucent. Dermatology is an obvious application field of optical surface imaging. Also, modern laparoscopic, cystoscopic, gastroscopic, rectoscopic, and other endoscopic probes mediate optical images, in this case of internal structures. Finally, infrared imaging should be mentioned as providing the information on tiny differences of body surface temperature that may indicate internal changes. In almost all those cases, the image is formed on the sensor plane (or on the retina

of the observer) by means of optical elements—lenses, mirrors, and prisms that transform the light field. The approximate analysis of such an optical system may be based on well-known geometrical optics or, with a greater degree of realism, on wave (Fourier) optics. Further reading in this direction is, e.g., [24], [39]. From the signal processing aspect, it is interesting to realize that the information in an optical system has alternately the form of the image (original-domain signal) and its two-dimensional Fourier transform (spectral representation) when going along the light path from one focal plane to the next. As for the result influencing the subsequent processing, we can describe the imaging quality in terms of the system point-spread function (PSF) and its (an)isoplanarity, field of view (FOV) and scale (magnification), geometrical distortions, image dynamic extent and contrast suppression, and signal nonlinearity (if nonlinear elements such as film or camera-screen subsystems are included). Isoplanar systems are often described by the *frequency transfer function* (FTF) as the Fourier transform of the PSF, or less completely by the *modulation transfer function* (MTF)—absolute value of FTF. From our point of view, these characteristics cannot be influenced and either are given by the optical system manufacturer or should be identified before an attempt is made to correct the image imperfections on the digital side (see Chapter 12).

In a great part, the optical information is still directly observed and evaluated without being digitized, but there is an obvious tendency toward digitizing and archiving even the images based on optical imaging; the digital camera becomes a standard part of optical systems. From the aspect of this book, the direct conversion of an optical image provided by an optical system into the image data via a solid-state matrix sensor (CCD or more modern CMOS light-sensing matrix structure), followed by an analogue-to-digital (A/D) converter, may be considered standard. This way, an image data matrix is provided, the size of which is determined by the sensor structure. Also, the other parameters, like the bit depth of the data, signal-to-noise ratio (SNR), possibly color resolution in color imaging, etc., are determined primarily by the sensing element, but are also influenced by the properties of the preceding optical system, illumination, use of contrast agents, etc. It is beyond the scope of this book to describe the underlying physical mechanisms and internal structure of the sensors; more details can be found, e.g., in Section II of [24]. The reader may realize the similarity of principle with the flat panels used as matrix detectors of x-rays (Section 3.1.4). The modern matrix light sensors may have up to tens of millions of pixels, thus providing resolution comparable to

that with large-format photography; the standard pixel count used by most current equipment is between 0.5 and 10 Mpixels. The bit depth of the obtained data is usually between 8 and 16 bits per pixel in gray-scale images, or three times that in color images.

The data provided by the matrix sensors are immediately available for image processing procedures; thus, there is no need to describe the internal mechanisms and intrinsic processing. The procedures still often used to correct the original data are sensitivity homogenization (Section 12.1.2) and possibly also linearization, though with decreasing importance, as the modern elements show a large dynamic extent with good linearity. We shall only mention two modalities that are not that standard and are interesting from the data acquisition point of view.

8.1.1 Three-Dimensional Confocal Imaging

The confocal imaging is a tomographic modality determined for imaging of (perhaps partially) transparent objects, like eye structures or microscopic cell preparations (Petran et al., 1968 [57]). It provides the tomographic slices perpendicular to the optical axis of the imager, utilizing optical projection by a lens system and a certain kind of sequential scanning [42].

The basic principle is depicted in Figure 8.1. Here on the left we see a telecentric couple of projection lenses positioned at the distance $d = f_1 + f_2$ (f_1, f_2 being the focus lengths of the lenses) so that their inner focal planes coincide. This optical system forms an image of the object plane on the image plane via magnifying projection with the (lateral) magnification m. While the beams of rays originating from points on the object plane are focused on the image plane, the beams coming from points under or above the object plane are focused below or above the image plane, so that their traces on the image plane are defocused wide patches. It follows from the optical system analysis that while lateral magnification is m, the axial magnification — the ratio of the axial focus shift in the image space and that in the object space — is m^2, hence exaggerated and leading to important defocus for already small axial differences (thus meaning a low depth of focus).

This (otherwise unfavorable) phenomenon is utilized in confocal imaging, as visible on the right side of the figure. The object is illuminated only point-wise, by a light source (often, but not necessarily, a laser), the light beam of which is focused first onto the illumination diaphragm, situated on the common focus plane of the

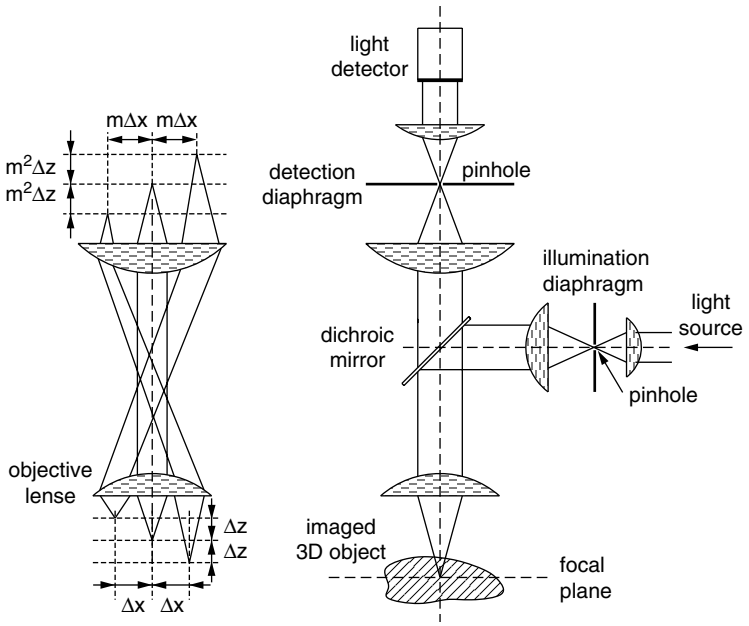


Figure 8.1 Principle of confocal imaging: (left) axial vs. lateral magnification and (right) concept of a scanning confocal microscope.

collimator lenses, with a pinhole that nevertheless allows unrestricted passage of the focused beam. The beam is then collimated by the following lens and reflected by a dichroic mirror into the objective lens, which focuses the light on the image plane. The light scattered, reflected, or emitted from the illuminated spot (point) of the object into the objective lens is then normally processed by the projection system (when neglecting the partial loss of light on the dichroic mirror) and focused onto the proper point of the image plane. Here, another (sc., detection) diaphragm is placed, again with a pinhole that allows passing of all the available light energy, which is consequently measured by a light detector. The output of the detector thus describes intensity of the light originated in the object space at the focus position. If the light-emitting or -scattering object spot lies above or below the focused object plane, its trace on the detection diaphragm is wide and only a small part of the available light passes through the pinhole; the light part is smaller the greater the axial shift. Hence, the light components due to structures not on the image plane are substantially suppressed in the measurement.

In order to obtain a complete image, a two-dimensional scanning mechanism must be provided (is not depicted in detail). The conceptually simplest approach is to move both the diaphragms in their planes mechanically in synchrony, thus covering gradually the complete image area of interest. Another possibility is to use tilting mirrors to provide the same effect. Anyway, scanning the complete image takes time in the range of a fraction of a second to minutes, because there is obviously a principally low light efficiency, and achieving a reasonable SNR requires accumulating enough photons per pixel. Direct observation of the image is only possible with sufficiently fast scanning conditioned by bright illumination; alternatively, the detector signal would be digitized and data gradually filled into a memory matrix.

By shifting the complete object with respect to the system axially, the focused plane would be moved to another plane of the object; this way, another slice will be scanned. By gradually shifting the object like this (or by refocusing the system correspondingly), it is possible to collect the data for multiple slices, thus providing three-dimensional data on the object.

It can be inferred that the resulting lateral and axial resolution of the confocal imaging may be greater than that of the same imaging system without the point-wise illumination, since the resulting PSF is narrowed by the concerted action of both illuminating and imaging focusing. Details on resolution analysis, as well as on other imaging parameters, can be found, e.g., in [42].

It is obvious that the principle of confocal imaging applies only to (semi)transparent objects. On the other hand, the use of the principle is not limited to microscopy; the same principle can obviously also be used for macroscopic imaging, e.g., in ophtalmology.

8.1.2 Infrared Imaging

Infrared imaging (IR) serves, in the given medical context, primarily for estimation of spatial distribution of temperature on the body surface. It is the only passive medical imaging modality, utilizing radiation energy produced by the body itself. It has been found that a skin surface behaves, in the used wavelength range $\lambda = 2 \ldots 10$ μm, approximately as a black body (emissivity ~ 0.98 to 0.99), i.e., radiating almost optimally. It is then possible to interpret the obtained intensity image in terms of surface temperature and, consequently, to localize the sources of heat, like blood vessels or pathologic (e.g., inflammatory) areas close to the body surface. The interpretation is

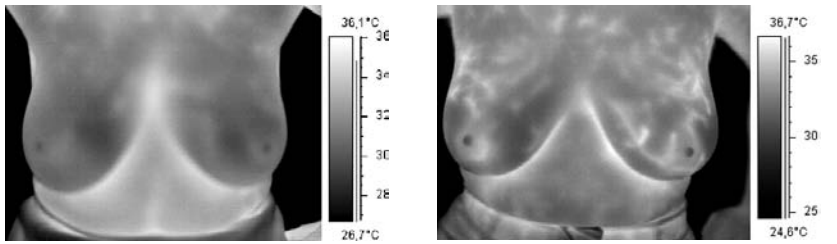


Figure 8.2 Examples of IR breast images, normal and pathological finding. (Courtesy of the Brno University of Technology Assoc. Prof. A. Drastich, Ph.D. and Faculty Hospital Brno-Bohunice, Department of Obstetrics.)

unfortunately complicated by the complex heat transfer in tissues, which leads to convoluted surface temperature patterns due to a combination of many factors: influence of multiple sources, external temperature and uneven external cooling of the body surface, non-Lambertinian character of the body radiator, partial skin tissue transparency at the shorter end of the wavelength range, etc. The attempts to obtain more definite images by blind deconvolution and similar approaches were not particularly successful, and so far infrared imaging is not widely clinically used. Experimental images aiming at evaluation of IR imaging capabilities for early detection of mamma carcinoma are shown in Figure 8.2.

As the body temperature is rather low and the measured differences small, rather sensitive and stable equipment as well as a precisely air-conditioned environment are needed. The IR camera may be based on the scanning principle using a single detector (either of bolometric type or a semiconductor photoresistive or photovoltaic detector) with oscillating or rotating mirrors or prisms to cover the FOV by shifting the IR image two-dimensionally in front of the detector; these systems often need intensive cooling by, e.g., liquid nitrogen. The advantage of the single-detector cameras is the easier calibration, as well as the capability of absolute IR radiation evaluation based on comparative measurement using a known reference radiator. Alternatively, an integrated line of semiconductor detectors may be used; then only one-dimensional scanning is needed. The most modern IR cameras use two-dimensional integrated detector fields (e.g., 320×240 pixels), thus achieving almost the same degree of flexibility as visible-light equipment, although with a lower resolution due to both optical system properties and lower pixel number. The basic parameters of the imaging system, besides the spatial resolution, are spectral sensitivity and temperature resolution; they should be specified by technical parameters

guaranteed by the manufacturer, or they are to be identified experimentally before any interpretation or analysis of the measured image data is performed. Further details can be found, e.g., in [14].

8.2 ELECTRON MICROSCOPY

Electron microscopy (EM) appeared in about 1930 (Glaser and Scherzer, Knoll, and Ruska, see [58]) as a logical extension of light microscopy, which reached its physical limits of spatial resolution (fraction of a micrometer) as determined by the wavelength of the visible light. The information carriers in EM, the electrons of practicable energies, have substantially lower effective wavelengths on the order of 10^{-9} m, and the electron-based imaging can thus in principle achieve many orders higher resolution. Practically, the resolution is determined by the quality of the used electron optics, particularly by the objective lens aperture; the practicable resolution may reach the nanometer range (up to about 0.2 nm). It should be mentioned that EM is used also in low-magnification applications where the resolution of otherwise simpler light microscopy would suffice; besides the possibility of imaging different parameters, the incomparably larger depth of focus is often highly appreciated.

Electron microscopy has two distinctive forms: *transmission* EM (TEM), applicable to very thin specimens transparent (or translucent) for the electron beam, and *nontransmission* EM, which analyzes basically surfaces of arbitrarily thick opaque specimens. While transmission EM visualizes the information carried by the electron beam, modified by passing through the specimen, nontransmission EM is based on detection and analysis of (different) products of electron beam interaction with the specimen surface.

The transmission principle allows illumination of the complete FOV at once and formation of the magnified image directly in electron representation by means of an electron-optics system, consequently converting the electron image into a visible image or image data. On the other hand, the interaction products in nontransmission EM cannot be directly focused and projected. Hence, the imaging must be based on scanning the surface with a highly focused (point) electron beam and detecting its interaction products sequentially (von Ardenne, ~1930; Oatley, ~1960, see [58]). The scanning principle is not limited to nontransmission EM; obviously, it may be used for the transmission EM as well, actually with some advantages.

As the scanning principle is used in transmission EM rather exceptionally (and then abbreviated STEM), the abbreviation TEM

is implicitly understood as full-field transmission imaging, while *scanning electron microscopy* (SEM) denotes the nontransmission mode. In the following comments, we shall limit ourselves to those two common modes. Further details can be found in, e.g., [22], [41].

8.2.1 Scattering Phenomena in the Specimen Volume

When the electrons are passing the specimen volume, they interact with the matter and become scattered. There are basically two mechanisms of scattering events: elastic and nonelastic.

The elastic scattering is marked out by a relatively low loss of electron energy. Depending on the distance of the original electron path with respect to the interacting atom, different mechanisms appear and the electrons are more or less scattered. Rutherford scattering leads to large scatter angles, and some of the electrons may even be *backscattered* (important for SEM; see below). These electrons are largely lost for TEM. Nevertheless, most electrons travel far enough from the atoms and are forward scattered by rather small deflection angles (up to several degrees); their use in TEM depends on the objective aperture.

Inelastic scattering results when the interaction concerns the electron shell of the object atoms. When an inner-shell electron (core electron) is removed, obtaining the energy in tens or hundreds of electronvolts, the incident electron loses the corresponding energy and is slightly scattered (under 1°). This event also leads to a consequent recombination that produces an x-ray photon, the wavelength of which is characteristic to the matter (see energy-dispersive x-ray (EDX) analysis in SEM), or it may generate an electron that may escape from the matter (sc., Auger electrons). The interaction with outer electrons of the atomic shells (conduction band electrons) decreases the energy of the incident electron only a little, and consequently, the scatter angle is rather small (under 0.1°). The freed electron from the outer shell may obtain energy up to tens of electronvolts and may be emitted from the specimen material as a *secondary electron* (of primary importance in SEM). The inelastically scattered electrons may obviously contribute to the postspecimen beam processed in TEM, as their scatter angles are in the range accepted by the projection system.

The scatter phenomena lead to a very short mean free path of electrons in the specimen material. Thus, the thickness of the TEM specimen must be correspondingly small to enable passage of a

majority of electrons — under 1 μm, up to units of nanometers. The same reason — small free path in the air — leads to the requirement of well evacuating the microscope tube, naturally with adverse influence on, namely, biological specimens.

8.2.2 Transmission Electron Microscopy

The arrangement of the *transmission electron microscope* is schematically depicted in Figure 8.3. The electron gun produces electrons accelerated by a high voltage, usually in the range 50 to 150 kV, but with rare exceptions on both margins: high-resolution microscopes use up to 3 MV, while low voltages of up to several hundred volts are also possible (though more in SEM). Hence, the electrons gain the corresponding energy in the range of 10^2 to 10^6 eV. The electron beam generated by the electron gun is collimated by the condenser using magnetic lenses, thus ideally providing a constant (possibly — with some arrangements — even coherent) illumination of the specimen.

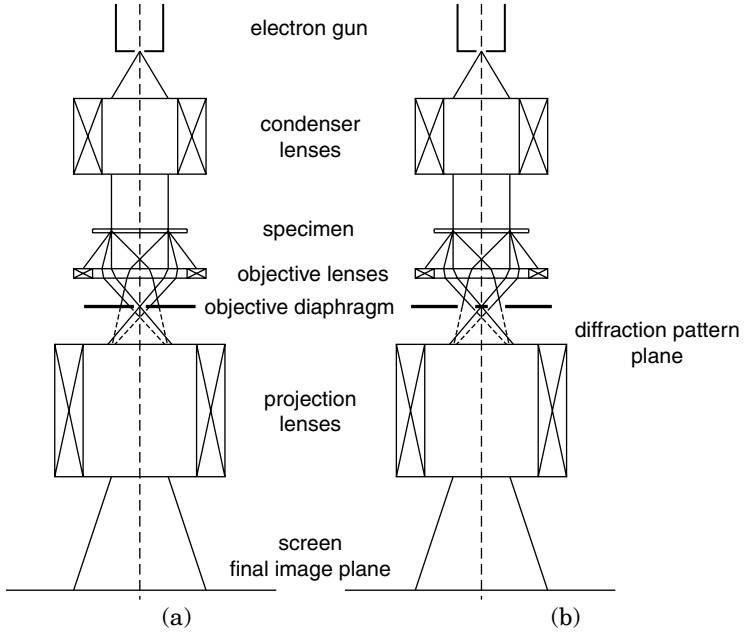


Figure 8.3 Scheme of a transmission electron microscope: (a) bright-field imaging and (b) dark-field imaging.

The incident beam is influenced by passing the specimen, which may be described in terms of particle physics as scattering of individual electrons (see above) so that the output (postspecimen) beam consists of electrons traveling in differing directions with uneven velocity. Whether the scattered electron is utilized in forming the image by the following projection system depends on its individual scatter angle with respect to the projection objective aperture. As the magnetic lenses would have too large aberrations with wider apertures, the typical aperture value is rather low, 0.3 to 1.5° (5×10^{-3} to 25×10^{-3} rad). Thus only electrons scattered less than this will be processed, the other being refused by the diaphragm placed in the objective back-focus plane (see Figure 8.3a). The image contrast is then explained as being a consequence of locally different degrees of scatter in the specimen, leading to smaller or greater loss of part of the electrons stopped by the limited aperture (scattering contrast).

Another view of the same phenomenon is possible when using the wave interpretation: the incident electron beam may be considered a plane wave, which is locally modified in its intensity and phase by the specimen, so that a (complex-valued) image is formed on the output plane of the specimen. Every point in the plane then becomes, according to Huygen's principle, an elementary source of the spherical wave, which is transferred to the other focal plane of the objective lens and interferes there with other elementary waves, thus generating the Fourier transform of the image. This two-dimensional spectral function is further processed by the following lens system to provide finally the output image. The image contrast is then due to interference of the elementary waves, which is dependent on the phase relations (phase contrast, which can be shown to depend on certain defocus of the projection). Let us note that by changing the focusing of the projection system, either the described image formation is realized or the spectral function may be formed in the output plane, its diffraction pattern providing the information on the three-dimensional spatial structure of the specimen, informative especially in case of a crystalline type specimen.

The equivalent wavelength of the wave function of accelerated electron is

$$\lambda = \frac{h}{\sqrt{2\,vem_0\left(\frac{ve}{2m_0c}+1\right)}} \approx \frac{1.2261_{10}-9}{\sqrt{v}} \quad [\mathrm{m}], \tag{8.1}$$

where h is the Planck constant, v (V) the accelerating voltage, e (C) the electron charge, m_0 (kg) the rest mass of the electron, and c (msec^{-1}) the light velocity; the second expression is an approximation for nonrelativistic velocities (under about 100 kV). The wavelength is very small — about 0.004 nm with the standard 100 kV, and still about 0.06 nm with mere 500 V. This does not mean a resolution limit; practically, the resolution is much worse, due to numerous imperfections with which the real TEM imaging suffers — primarily lens aberrations (namely, spherical) and, consequently, the resulting necessity to limit the lens apertures, both contributing to nonpoint PSF of the imaging. Particularly, the resolution limit due to diffraction on the finite aperture α (rad) of the objective lens may be described by the effective diameter d of the diffraction pattern,

$$d = \frac{0.61\lambda}{\sin\alpha} \; [m] \tag{8.2}$$

giving, for $\alpha = 5 \times 10^{-3}$ rad, about 0.5 nm, with acceleration by 100 kV, and some 7 nm, with 500 V. The combined resolution-limiting phenomena should be taken into account and perhaps (partly) compensated for in postprocessing by a kind of (approximate) inverse filtering.

On the other hand, the small apertures provide for an excellent depth of focus D relative to the resolution limit R,

$$D = \frac{R}{\tan\alpha} \; [m]; \tag{8.3}$$

by orders greater than in light microscopy (relatively, with respect to resolution — for the resolution of 10 nm and $\alpha = 5 \times 10^{-3}$ rad, the focus depth D is about 2 μm). As a rule of thumb, it can be said that the focus depth in EM may be comparable with (though mostly less than) the width of the image.

The described processing is the standard mode, denoted as the *bright-field imaging*. A complementary mode, called *dark-field imaging*, is based primarily on electrons with greater scatter angles, while the little scattered electrons are refused; the discrimination is provided for by a wider but centrally opaque diaphragm (Figure 8.3b). The alternative modality may provide additional information, e.g., by emphasizing the highly scattering areas.

The output image in TEM, historically visualized by a fluorescent screen or registered on photographic emulsion or another

medium, can nowadays be directly discretized by a (often cooled) semiconductor light detector field (CCD or MOS type) with the resolution of several megapixels and a rather high dynamic range of about 10^5. The input plane of the sensor is separated from the harmful electron beam image by a scintillation layer converting the electron energy into light photons (and unfortunately, also partly lessening the resolution due to lateral leakage of light in the layer). The embedded (or external) A/D converter provides the data that can be stored in the standard matrix form.

8.2.3 Scanning Electron Microscopy

The principle arrangement of a scanning electron microscope is depicted in Figure 8.4. The electron gun provides the beam of electrons accelerated to the energy of about 10 to 50 keV, which is focused by the condenser lens system, in this case to the plane of the object (specimen). It is an important property of the focused beam that the convergence angle is very small (on the order of 1 mrad = 0.05°), so that the lateral beam width is almost constantly narrow (up to 1 to 3 nm) in a relatively large range of axial depth (~100 times the beam diameter or more). The quality of focusing the electron beam is critical, as it determines the resolution and depth of focus of the microscope.

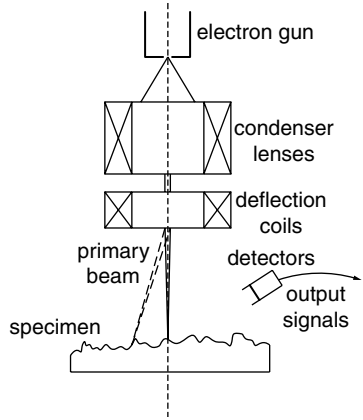


Figure 8.4 Principle arrangement of a scanning electron microscope.

The beam is deflected by the auxilliary magnetic field of the deflection coils, which provides for two-dimensional scanning of the chosen area on the specimen surface, usually in the noninterlaced rectangular manner (line by line across the FOV, several hundreds of lines, each sampled at up to ~1000 points). The instantaneous response, measured by the detectors positioned above the specimen, thus characterizes a tiny sample area — the just illuminated spot, approximately of the cross-sectional size of the beam. The acquisition of data on the chosen parameter (see below) is thus sequential and must be synchronized with the currents in the deflection coils. The depicted conventional type of SE microscope suffers from a high background of disturbing secondary and backscattered electrons generated outside of the illuminated spot or even outside of the specimen, due to secondary and tertiary effects. This can be prevented in the *in-lens* type of the SE microscope, where the specimen is situated between the pole pieces of the immersion lens, while the detectors are located above the lens.

As mentioned above, the detected quantity may be the count rate of secondary electrons (*secondary electron imaging* (SEI)), of backscattered electrons (*backscattered electron imaging* (BEI)), or x-ray photons (possibly with energy discrimination, allowing chemical element determination based on spectral analysis of the characteristic radiation — see below); other less frequently used quantities are the Auger electron rate, photoemission, or resulting current through the specimen. The image reconstructed from these data is highly dependent on the type of imaged parameter, not only as to the image appearance (contrast) concerns, but also with respect to lateral (x,y) resolution and spatial (depth) interpretation of the results. The last aspect needs explanation: the different measured quantities describe different volume areas belonging to the same illuminated spot. As schematically shown in Figure 8.5, the best lateral resolution (up to units of nanometers) may be achieved in SEI or Auger electron imaging; these two differ in the effective depth of the volume sample. The BEI mode is characterized by a substantially larger (up to units of micrometers) and deeper active sample volume, limiting the lateral resolution and describing also subsurface structures. Even deeper is the effective sample volume generating the characteristic x-rays, while the braking radiation (*bremsstrahlung*) comes from still more deep structures. These facts should be considered when interpreting the obtained images.

The scintillation detectors of electrons, similar in principle to those mentioned in Chapter 6, may serve both SEI and BEI modes;

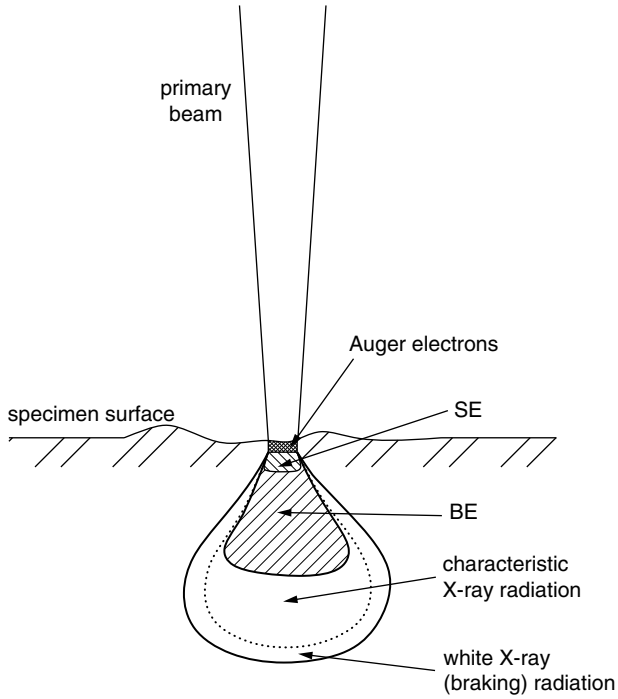


Figure 8.5 Schematic representation [22] of the SEM sample volume.

the choice of mode is then arranged for by selectable electrical potential applied to the detector input window. However, the high-energy backscattered electrons may also be detected separately by a semiconductor detector. The detector of x-ray photons should be capable of wavelength recognition, based either on the less discriminative, but faster *energy dispersive x-ray analysis* (EDX), i.e., on amplitude classification of detection pulses, or on diffraction of x-rays on a suitable crystal followed by a variably positioned counting detector (more sensitive *wave-dispersive analysis* (WDX)).

The extent of the sample volume may be better defined, especially in low-atomic-number and low-conducting specimens, by thin metallic coating (1 to 2 nm) that localizes the effective volume at the surface and also prevents the deteriorative charging effects. This is a matter of specimen preparation that is generally rather demanding in EM and may also include decontamination, polishing, drying, etc.

8.2.4 Postprocessing of EM Images

Modern electron microscopes are completely digitized and provide image data in the standard matrix form. Examples of typical SEM images are shown in Figure 8.6. Of the spectrum of applicable image processing approaches, a few deserve particular mentioning.

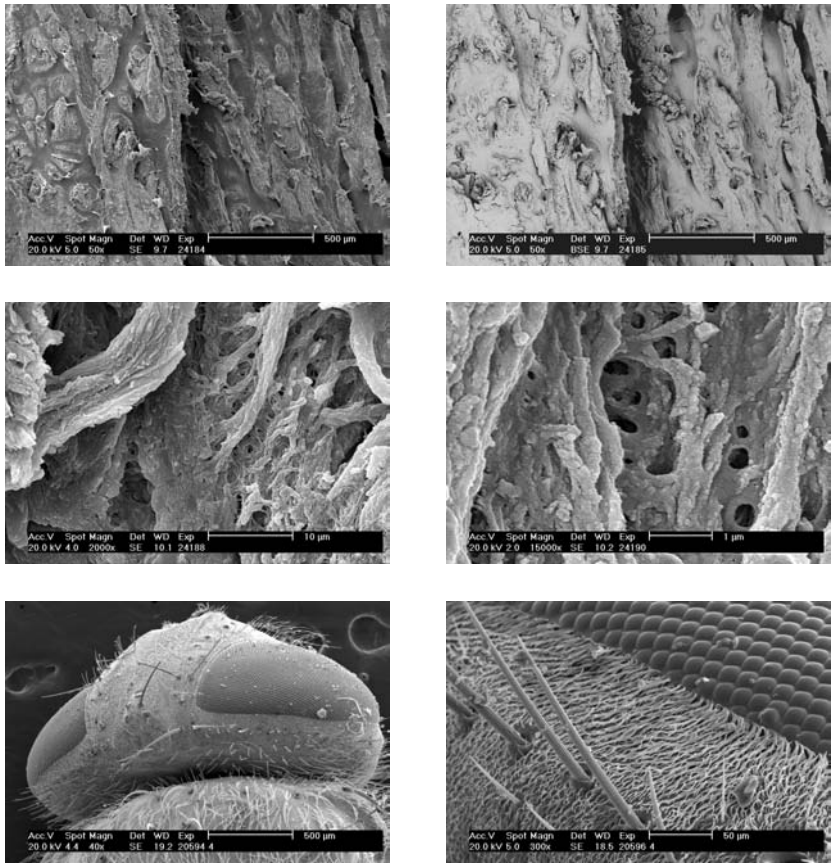


Figure 8.6 Examples of SEM images. Upper line: Surface of a bone splinter as imaged in (left) SEI and (right) BEI modalities at a low magnification (50×). Middle line: The same sample as above in higher magnification (SEI, 2000×, 15000×). Lower line: SEI images of an insect eye in different magnification (40×, 300×); note the extraordinary depth of focus. (Courtesy of Institute of Material Science, Faculty of Mechanical Engineering, Brno University of Technology, D. Janova, M.Sc.)

As the images may be rather noisy, especially when using a narrow-beam spot and low current or fast scanning in SEM, averaging of images (Sections 10.4.1 and 11.3) may be used to increase the SNR. If the consequential images are not well aligned due to any instabilities in imaging, the step of image registration (Section 10.3) may be needed, perhaps even of a complicated flexible type requiring a complex identification of the registration transform. When periodic structures are imaged, it is possible to average over the periodic patterns in the frame of a single image; it requires the individual pattern identification and registration via correlation analysis (Sections 2.4.2 and 10.3). A similar, though more complicated, task is to derive the disparity map(s) between two or more couples of images (see Section 10.2). When the aberrations of concrete EM imaging are identified, restoration of image data may be attempted (Chapter 12), aiming at either resolution improvement and noise suppression, or dynamics linearization, geometrical restitution, etc. In TEM, it may be useful to acquire, besides the image, its Fourier transform (the diffraction pattern), which may complement in a sense the available information and enable a more satisfactory restoration.

An interesting task of SEM is to recover three-dimensional information on the specimen. The three-dimensional *surface reconstruction* (Section 10.4.5) is based on a pair (or a series) of images of the specimen, provided by differently tilting the specimen with respect to the optical axis. The reconstruction algorithm, providing the z-coordinate for each (x, y)-pixel, utilizes the information from the above-mentioned disparity maps (Section 10.2). In some types of specimen that may be considered (semi)transparent in the used imaging mode, a series of images of gradually tilted specimen may be considered a set of projections, thus allowing the obtaining of the complete three-dimensional structure of the specimen by a method of image data reconstruction from projections (Section 9.1).

Naturally, all image-enhancing methods (e.g., contrast enhancement, false color representation, background suppression, etc.—Chapter 11) are routinely used as well as analytic procedures (e.g., segmentation, analysis of texture, morphological transforms, determination of particle size, count, orientation, etc.—Chapter 13).

8.3 ELECTRICAL IMPEDANCE TOMOGRAPHY

Electrical impedance tomography (EIT) is a modality aiming at imaging the spatial distribution of electrical conductivity $\sigma(x, y, z)$ (S m^{-1}) in tissue. Though it is seemingly a clearly defined physical

parameter of tissue, the complex microscopic structure makes its realization less straightforward, see, e.g., [2].

Because of disturbing polarization phenomena, the measurement cannot be done by means of direct current (DC), and therefore, the measured quantities are derived from (frequency-dependent) impedances **Z** at the measuring frequency f of the used alternating current (AC), rather than from the pure resistances R. However, due to stray capacitances in the measuring arrangement, it is infeasible to measure precisely the imaginary part of **Z**, and hence only the real part is usually measured via synchronous detection. Even these quantities, though determined by the resistivity $\rho = 1/\sigma$, are frequency dependent, as is the conductivity itself due to complicated conductive mechanisms. The known frequency dependence may, on the other hand, be utilized for multiple measurements, enabling partial suppression of the measurement errors (see difference imaging below).

The conductivity is usually physically defined as a scalar quantity, thus isotropic in the sense that the current of any direction is conducted (or impeded) equally. This may be right on the microscopic level of tissues, not available to EIT measurement, but on the measured macroscopic level, there is important anisotropy (e.g., obvious difference between the conductivities of muscles along and across fibers that are significantly different). Should this be taken into account, the imaged parameter is no more scalar, but rather a vector quantity. So far considered a limitation, it may carry a potential for improving the diagnostic value of the images by a more complex measurement, enabling reconstruction of the vector-valued image with the orientation and degree of directivity presented in false color (color EIT). Nevertheless, in the following outline we shall neglect this phenomenon.

The measuring arrangement for two-dimensional (slice) imaging is depicted in Figure 8.7. A set of N electrodes (usually $N = 16$ or 32) is positioned around the boundary of the imaged object. To some of them, in the simplest case to a couple of adjacent ones, the current sources are connected, causing current flow inside the object. This in turn generates a field of electrical potential $\varphi(x, y, z)$ that can be sensed via the voltages measured at the electrodes. The same electrodes may be in principle used both for introducing the driving current and for the voltage measurement; nevertheless, the unknown resistances between the electrodes and tissue prevent the measurement on the electrodes just being used as driving ones. Although completely separate sets of driving and measuring electrodes may be used with some advantage, it is not common.

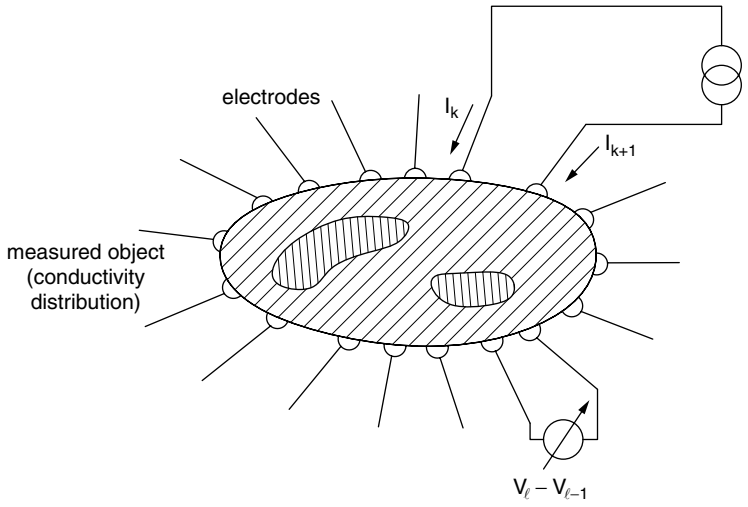


Figure 8.7 Principle of EIT measuring arrangement.

On a set of N electrodes, we thus generally define a vector of externally controlled input currents $\mathbf{I} = [I_1, I_2, \ldots, I_N]^T$, obviously with

$$\sum_{k=1}^{N} I_k = 0, \tag{8.4}$$

and a vector of measured voltages $\mathbf{V} = [V_1, V_2, \ldots, V_N]^T$.* Assuming linearity, we obtain

$$\mathbf{V} = \mathbf{A}(\sigma(x, y, z))\,\mathbf{I}, \tag{8.5}$$

where the elements of the $N \times N$ matrix $\mathbf{A}$ are determined by the conductivity distribution. As $\sigma(x, y, z)$ is to be determined, the matrix $\mathbf{A}$ is initially unknown. In order to determine $\mathbf{A}$, many independent measurements must be made with different current patterns controlled by external supply. When denoting the individual

*With respect to the previous limitation, only the voltages V_k: $I_k = 0$ may be available, but theoretically, when the contact resistances can be neglected, all the voltages can be measured.

couples of measurement vectors by ($\mathbf{V}_n$, $\mathbf{I}_n$), $n = 1,\ldots, M$, Equation 8.5 may be extended into the linear equation system

$$\bar{\bar{\mathbf{V}}} = \mathbf{A}(\sigma(x, y, z))\,\bar{\bar{\mathbf{I}}}, \quad \bar{\bar{\mathbf{V}}} = [\mathbf{V}_1, \mathbf{V}_2, \ldots, \mathbf{V}_M], \quad \bar{\bar{\mathbf{I}}} = [\mathbf{I}_1, \mathbf{I}_2, \ldots, \mathbf{I}_M]. \tag{8.6}$$

As the measurements cannot be considered completely independent, M must be high enough to enable determination of $\mathbf{A}$ by pseudoinversion. Nevertheless, even if $\mathbf{A}$ were well estimated, the problem of deriving the corresponding conductivity distribution $\sigma(x, y, z)$ remains.

The current density distribution $i(x, y, z)$ is governed by the set of partial differential equations, linking the electrical potential field $\varphi(x, y, z)$ with the current field,

$$i(x, y, z) = \sigma(x, y, z)\,\mathrm{grad}\varphi(x, y, z), \tag{8.7}$$

$$\mathrm{div}\, i(x, y, z) = 0, \tag{8.8}$$

which can be solved numerically. It is then possible to solve the so-called forward problem, i.e., to determine the current and voltage distribution if the conductivity distribution and driving currents are known. The finite-element approach is used; i.e., basically, the spatial distributions are discretized into small spatial elements indexed i, j (for the two-dimensional solution), each with a constant $\sigma_{i,j}$, thus forming the matrix $\boldsymbol{\sigma}$. The forward problem then consists of finding the forward operator F,

$$\varphi(x, y, z) = \mathrm{F}\{\mathbf{I}, \boldsymbol{\sigma}\}, \tag{8.9}$$

whence the set of electrode voltages can be determined as the differences of the potential with respect to the reference electrode, whose potential is defined as $\phi_0 = 0$.

The inverse problem—finding σ when knowing $\mathbf{I}$ and ϕ (only approximately, in a few discrete points)—cannot be directly solved. An iterative procedure is therefore used to reconstruct a reasonable approximation of $\boldsymbol{\sigma}(x, y, z)$ based on the measurements. The operator F can be expressed in the matrix form,

$$\mathbf{V} = \mathbf{B}(\mathbf{I}, \boldsymbol{\sigma})\,\bar{\sigma}, \tag{8.10}$$

where $\bar{\sigma}$ is the vector of conductivity values (elements of $\boldsymbol{\sigma}$) and components of $\mathbf{V}$ are $V_i = \varphi_i - \varphi_0$. The elements of the $\mathbf{B}$ matrix can be shown to depend on conductivity distribution; thus the equation is obviously nonlinear with respect to σ values. This again indicates that direct inversion of Equation 8.10 is impossible.

After the measurement, the pairs of $\mathbf{I}$ and $\mathbf{V}$ are known and the voltage vector $\mathbf{V}_{\text{pred}}$ predicted by Equation 8.10, based on a certain previous estimate $\bar{\sigma}_j$, can be compared with the measurement. This produces the error vector

$$\Delta\mathbf{V} = \mathbf{V} - \mathbf{V}_{pred}, \tag{8.11}$$

which is used in a formula, based on the sensitivity matrix of $\mathbf{B}$ with respect to $\bar{\sigma}$, for determining the correction vector $\Delta\bar{\sigma}$. This way, a new estimate of $\bar{\sigma}$ becomes

$$\bar{\sigma}_{j+1} = \bar{\sigma}_j + \Delta\bar{\sigma}. \tag{8.12}$$

This iteration continues, using repeatedly all the measurement data, unless the correction may be considered negligible. In each iteration step, the partial differential equation system (Equations 8.7 and 8.8) must be in principle solved; hence, the procedure is rather complex.

The EIT modality is interesting from the theoretical point of view; namely, the signal acquisition and processing, including the image reconstruction, pose challenging problems. The information on the conductivity distribution might form an interesting complement to other diagnostic data, including images in other modalities, especially as EIT imaging is entirely noninvasive. However, the images so far obtained are of rather poor quality, and the clinical experience is limited to the experimental phase. A promising approach is the already mentioned differential imaging when only differences in two conductivity images, caused either by time development (e.g., by breathing) or by using different measuring frequencies, are imaged. In this case, the influence of disturbing phenomena mostly cancels out by subtracting almost equal consequences. Three-dimensional EIT imaging represents another direction of possible development, although so far it is even more experimental than the two-dimensional impedance tomography. More details and further references may be found in [2], [46].

REFERENCES for Part II

[1] Albert, W. and Pandit, M. Processing of ultrasound images in medical diagnosis. In *Handbook of Computer Vision and Applications*, Vol. 1, *Sensors and Imaging*, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[2] Barber, D.C. Electrical impedance tomography. In *Biomedical Engineering Handbook*, 2nd ed., Vol. 1, Bronzino, J.D. (Ed.). CRC Press/IEEE Press, Boca Raton, FL, 2000.

[3] Bechr, H. and Burns, P.N. *Handbook of Contrast Echocardiography.* Springer-Verlag, Heidelberg, 2000.

[4] Beutel, J., Kundel, H.L., and Van Metter, R.L. (Eds.). *Handbook of Medical Imaging*, 2nd ed., Vol. 1. SPIE Press, Washington, DC, 2000.

[5] Bijnens, B. Exploiting Radiofrequency Information in Echocardiography. Ph.D. dissertation, Catholic University of Leuven, Belgium, 1997.

[6] Boone, J.M. X-ray production, interaction and detection in diagnostic imaging. In *Handbook of Medical Imaging*, Vol. I, Beutel, J., Kundel, H.L., van Metter, R.L. (Eds.). SPIE Press, Washington, DC, 2000.

[7] Bronzino, J.D. *Biomedical Engineering Handbook*. CRC Press/IEEE Press, Boca Raton, FL, 1995.

[8] Budinger, T.F. and VanBrocklin, H.F. Positron emission tomography. In *Biomedical Engineering Handbook*, Bronzino, J.D. (Ed.). CRC Press, Boca Raton, FL, 1995.

[9] Bushberg, J.T., Seibert, J.A., Leidholdt, E.M., Jr., and Boone, J.M. *The Essential Physics of Medical Imaging*. Lippincott Williams & Wilkins, Baltimore, MD, 2002.

[10] Cho, Z.H., Jones, J.P., and Singh, M. *Foundations of Medical Imaging*. John Wiley & Sons, New York, 1993.

[11] Conoly, S., Macovsky, A., Pauly, J., Schenk, J., Kwonk, K.K., Chesler, D.A., Hu, X., Chen, W., Patel, M., and Ugurbil, K. Magnetic resonance imaging. In *Biomedical Engineering Handbook*, Bronzino, J.D. (Ed.). CRC Press, Boca Raton, FL, 1995.

[12] Croft, B.Y. and Tsui, B.M.W. Nuclear medicine. In *Biomedical Engineering Handbook*, Bronzino, J.D. (Ed.). CRC Press, Boca Raton, FL, 1995.

[13] Dawant, B.M. and Zijdenbos, A.P. Image segmentation. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[14] Dereniak, E.L. and Boreman, G.D. *Infrared Detectors and Systems*. John Wiley & Sons, New York, 1996.

[15] Elster, A.D. *Magnetic Resonance Imaging, Questions and Answers*. Mosby-Year Book, 1994.

[16] Evans, D.H., McDicken, W.N., Skidmore, R., and Woodcock, J.P. *Doppler Ultrasound: Physics, Instrumentation, and Clinical Applications*. John Wiley & Sons, New York, 1989.

[17] Evans, D.H. and McDicken, W.N. *Doppler Ultrasound: Physics, Instrumentation and Signal Processing*, 2nd ed. John Wiley & Sons, New York, 2000.

[18] Fenster, A. and Downey, D.B. Three-dimensional ultrasound imaging. In *Handbook of Medical Imaging*, Vol. I, Beutel, J., Kundel, H.L., van Metter, R.L. (Eds.). SPIE Press, Washington, DC, 2000.

[19] Fitzpatrick, J.M., Hill, D.L.G., and Maurer, C.R., Jr. Image registration. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[20] Goldberg, R.L, Smith, S.W., Mottley, J.G., and Ferrara, K.W. Ultrasound. In *Biomedical Engineering Handbook*, 2nd ed., Vol. 1, Bronzino, J.D. (Ed.). CRC Press/IEEE Press, Boca Raton, FL, 2000.

[21] Goodenough, D.J. Tomographic imaging. In *Handbook of Medical Imaging*, Vol. I, Beutel, J., Kundel, H.L., van Metter, R.L. (Eds.). SPIE Press, Washington, DC, 2000.

[22] Hrivnák, I. *Electron Microscopy of Steels* [in Slovak]. Veda, Slovakia Bratislava, 1986.

[23] Insana, M.F., Myers, K.J., and Grossman, L.W. Signal modelling for tissue characterisation. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[24] Jahne, B., Haussecker, H., and Geissler, P. (Eds.). *Handbook of Computer Vision and Applications*, Vol. 1. Academic Press, New York, 1999.

[25] Jan, J. and Kilian, P. Modified Wiener approach to restoration of ultrasonic scans via frequency domain. In *Proceedings of the 9th Scandinavian Conference on Image Analysis*, Uppsala, Sweden, 1995, pp. 1173–1180.

[26] Kak, A.C. and Slaney, M. Principles of Computerized Tomographic Imaging. Paper presented at SIAM Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[27] Kalender, W.A. *Computed Tomography*. Publicis MCD Verlag, Munich, 2000.

[28] Kilian, P., Jan, J., and Bijnens, B. Dynamic filtering of ultrasonic responses to compensate for attenuation and frequency shift in tissues. In P*roceedings of the 15th EURASIP Conference BIOSIGNAL*, Brno, Czech Republic, 2000, pp. 261–263.

[29] Koeppe, R.A. Data analysis and image processing. In *Principles and Practice of Positron Emission Tomography*, Wahl, R.L. (Ed.). Lippincott Williams & Wilkins, Baltimore, 2002.

[30] Krestel, E. (Ed.). *Imaging Systems for Medical Diagnostics*. Siemens Aktiengesellschaft, Berlin, Germany, 1990.

[31] Lovstrom, B. Ultrasonic Signal Processing: Attenuation Estimation and Image Enhancement. Ph.D. dissertation, University of Lund, Sweden, 1992.

[32] Nelson, T.R. and Pretorius, D.H. Three-dimensional ultrasound imaging. *Ultrasound Med. Biol.*, 24, 1243–1270, 1998.

[33] Palmer, P.E.S. (Ed.). *Manual of Diagnostic Ultrasound*. World Health Organization, Geneva, 1995.

[34] Pickens, D. Magnetic resonance imaging. In *Handbook of Medical Imaging*, Vol. I, Beutel, J., Kundel, H.L., van Metter, R.L. (Eds.). SPIE Press, Washington, DC, 2000.

[35] Reiber, J.H.C. et al. Angiography and intravascular ultrasound. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[36] Rowlands, J.A. and Yorkston, J. Flat panel detectors for digital radiography. In *Handbook of Medical Imaging*, Vol. I, Beutel, J., Kundel, H.L., van Metter, R.L. (Eds.). SPIE Press, Washington, DC, 2000.

[37] Schmitt, F., Stehling, M.K., and Turner, R. *Echo-Planar Imaging: Theory, Technique and Application*. Springer, Heidelberg, 1998.

[38] Schreiber, W.G., and Brix, G. Magnetic resonance imaging in medicine. In *Handbook of Computer Vision and Applications*, Vol. 1, *Sensors and Imaging*, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[39] Scott, C. *Introduction to Optics and Optical Imaging*. IEEE Press, Washington, DC, 1998.

[40] Sheehan, F., Wilson, D.C., Shavelle, D., and Geiser, E.A. Echocardiography. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M. and Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[41] Stegmann, H., Wepf, R., and Schroder, R.R. Electron microscopic image acquisition. In *Handbook of Computer Vision and Applications*, Vol. 1, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[42] Stelzer, E.H.K. Three-dimensional light microscopy. In *Handbook of Computer Vision and Applications*, Vol. 1, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[43] Taxt, T. and Jirik, J. Superresolution of ultrasound images using the 1st and 2nd harmonic. In *Proc. IEEE Trans. Ultrason. Ferroelec. Freq. Control*, in print.

[44] Thompson, C.J. Instrumentation. In *Principles and Practice of Positron Emission Tomography*, Wahl, R.L. (Ed.). Lippincott Williams & Wilkins, Baltimore, 2002.

[45] Wahl, R.L. (Ed.). *Principles and Practice of Positron Emission Tomography.* Lippincott Williams & Wilkins, Baltimore, 2002.

[46] Webster, J.G. (Ed.). *Electrical Impedance Tomography.* Adam Hilger, New York, 1990.

[47] Woodward, P. (Ed.). *MRI for Technologists*, 2nd ed. McGraw-Hill, New York, 1995.

[48] Xu, C., Pham, D.L., and Prince, J.L. Image Segmentation Using Deformable Models. In *Handbook of Medical Imaging,* Vol. 2, *Medical Image Processing and Analysis,* SonKa, M. and Fitzpatrick, S.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA.

[49] Yaffe, M.J. Digital mammography. In *Handbook of Medical Imaging*, Vol. I, Beutel, J., Kundel, H.L., van Metter, R.L. (Eds.). SPIE Press, Washington, DC, 2000

[50] Bloch, F. Nuclear induction. *Phys. Rev.*, 70, 460–474, 1946.

[51] Purcell, E.M., Torrey, H.C., and Pound, R.V. Resonance absorption by buclear magnetic resonance in a solid. *Phys. Rev.,* 69, 37–38, 1946.

[52] Lauterbur, P. Image formation by induced local interactions: Examples employing nuclear magnetic resonancce. *Nature*, 242, 190–191, 1973.

[53] Hahn, E. Spin echos. *Phys. Rev.*, 20(4), 580–594, 1950.

[54] Mansfield, P. Multi-planar image formation using NMR spin echoes. *J. Physics*, C10, L55, 1977.

[55] Anger, H.O. Scintillation camera. *Rev. Sci. Instrum*. 29, 27, 1958.

[56] Namekava, K., Kasai, C., Tsukamoto, M., and Koyano, A. Real time blood flow imaging system utilizing autocorrelation techniques. In *Ultrasound '82,* Lerski, R.A., Morley, P. (Eds.). Pergamon, New York pp. 203–208, 1982.

[57] Petrán, M., Hadravsky M., Egger M.D., and Galambos, R. Tandem scanning reflected-light microscope. *J. Opt. Soc. Am*. 58, 661–664, 1968.

[58] Hawkes, P. (ed.). The beginnings of electron microscopy. *Adv. Electronics and Electron Phys.*, Suppl. 16, 1985.

[59] Hounsfield, G.N.: Computerized transverse axial scanning (tomography): Part I. *Brit. J. Radiol.* 46, 1016–1022, 1973.

Part III

Image Processing and Analysis

The third part of this book is devoted to the concepts of more commonly used methods of biomedical image processing and analysis. Again, the emphasis is put more on the principles and explanation of the philosophy behind the methods than on the implementation details specific to concrete image processing tasks. The final goal is the *insight* into the approaches and methods, so that the reader will be able to choose appropriate methods and use them with understanding, possibly properly adjusting their parameters or modifying partial features. Understanding of the concepts should also enable correct interpretation of the obtained results from the physical point of view, even when the user is more inclined toward concrete medical applications. The treatment of the material is adapted to these goals: the necessary mathematical formalism is supported by references to physical and intuitive interpretation in order that the explanation would be perceived (hopefully) as consistent and meaningful. When something is exceptionally introduced without (at least indicated) derivation, it is presented with a notice; in other cases, it should be possible to follow the text as a continuous development.

For those who intend to go further in developing the image processing theory and methods, this part of the book should be considered introductory. Details on individual methods and deeper underlying mathematical theory can be found in the literature, partly listed in the reference section. The following chapters are intended to provide a lucid way through the fundamental terms, concepts, and methods of medical image processing, thus providing the knowledge necessary for a meaningful study of the specialized monographs and original papers.

The third part of the book is basically divided into three segments:

- Methods of *obtaining new quality* from measured data, fundamental in today's medical imaging: the *reconstruction of tomographic images* (Chapter 9), forming the computational engine of tomographic modalities like computed tomography (CT), single-photon emission computed tomography (SPECT), and positron emission tomography (PET) (magnetic resonance imaging (MRI) and ultrasonography will also be mentioned, though the reconstruction may be differently formulated), and *image fusion* (Chapter 10) — the area which provides qualitatively higher information by combining, in a broad sense, image data from different sources, obtained at different times, or from different aspects (image registration, comparison and fusion, disparity and flow analysis).
- *Image processing and restoration* (Chapters 11 and 12) defined as computationally deriving an output *image* (hopefully better, in a sense) based on input image data; the methods are rather generic and apply to most imaging modalities, though with modality-specific particularities and modifications.
- *Image analysis* (Chapter 13), including image *segmentation* as the usual first step of analysis, and further analytic approaches providing a *description* of the input image (or images)—a vector of parameters and/or features describing the relevant properties of the image that can be used in higher-level processing—classification, and diagnostics, either automatic or based on interaction with a human operator. The output vector may have the form of a parametric image, often acquiring only a few different values, denoting the character of differently sized areas (as in segmentation or tissue characterization) up to the pixel resolution (e.g., in edge representation).

The concluding Chapter 14 comments on the image processing environment and briefly mentions some aspects of medical image data handling that go beyond the chosen frame of image processing and analysis, as, e.g., the *image compression* concepts needed in archiving and communication of medical image information essential in contemporary health care applications.

The literature specific to individual topics in Part III is cited in the respective chapters. Other sources used but not cited are [8], [19], [20], [27], [28], [30], [31], [33], [38], [46], [49], [56], [58], [65], [67]–[69], [73]–[75], [77], and [78]. As for the topical information on development in a particular area of methodology, the reader should refer to numerous international journals that deal regularly with novel methods of medical image processing and analysis, often in the frame of special issues devoted to problem areas determined by a unifying aspect of the application, as, e.g., [29]. The Internet became a fast source of preliminary information, including that on the wide choice and contents of journals.

9

Reconstructing Tomographic Images

As we have seen in Part II, tomographic data are mostly provided in the form of projections of the desirable images. As the projection data are totally incomprehensible, it is necessary to convert them into an informationally equivalent but understandable form of images. This procedure is called image *reconstruction from projections*. As only sample values of the projections are available in practical imaging, the reconstruction can only yield approximate images even under the most favorable circumstances.

The basic reconstruction principles are treated first (Section 9.1). We shall deal here with different methods of the approximate reconstruction from projections under the ideal situation when all the side effects can be neglected. This should primarily provide the insight into the different approaches to the reconstruction problem; however, the methods are directly applicable, namely, in x-ray computed tomography (CT), which under certain conditions approaches the ideal situation. The fact that the realistic measurement of a ray integral concerns not the ideal line path, but rather a certain volume—in the best case a thin "pencil" beam—is partly taken into account in the discrete formulation.

In Part II of the book, we noticed that the measured values in practical imaging may be to a lesser or greater extent influenced by, besides the chosen measured parameter, other parameters of the investigated object, or affected by some disturbing physical phenomena, like scatter, attenuation, and quantum noise. The nonideal situation is the subject of Section 9.2. In Section 9.2.1, the model of data measured with attenuation is formulated, and some methods that enable taking attenuation into account during the reconstruction process will be mentioned.

Measured values in some modalities, namely, in gamma tomography (SPECT, PET) are rather imprecise due to quantum noise, as the available measured intensities are extremely weak, and therefore represented by individual photon counts with large variances. In this case, the situation is substantially complicated by the necessity to take into account the stochastic model of the measured values. This problem will be briefly discussed in Section 9.2.2.

Two common tomographic modalities — MRI and ultrasonography (USG) — use (mostly) other approaches to reconstruct the image data from the measurements. Nevertheless, both belong to the area of tomographic reconstruction; the approaches used will therefore be briefly discussed in Section 9.3.

The explanation of the reconstruction methods is based on theoretically simpler parallel projections. The practically important case of technically advantageous fan projections will be mentioned as an extension of the basic methods. Of the literature cited in the references to Part III, the most relevant sources to this chapter are [26], [37], [8], and also [20], [23], [51], [58], [66], [71], where further references can also be found.

9.1 RECONSTRUCTION FROM NEAR-IDEAL PROJECTIONS

9.1.1 Representation of Images by Projections

Let us have a two-dimensional function $f(x, y)$ (Figure 9.1), the physical meaning of which is arbitrary (to be concrete, it may be interpreted, e.g., as a distribution of a certain tissue parameter $f(x, y, z)$ on a spatial slice for $z = z_0$). To make the depiction simpler, the function value is supposed nonzero only inside the indicated contour. Auxiliary coordinates τ, r are introduced that are inclined with respect to x, y by the angle θ_0. A line R, parallel to the axis r, is

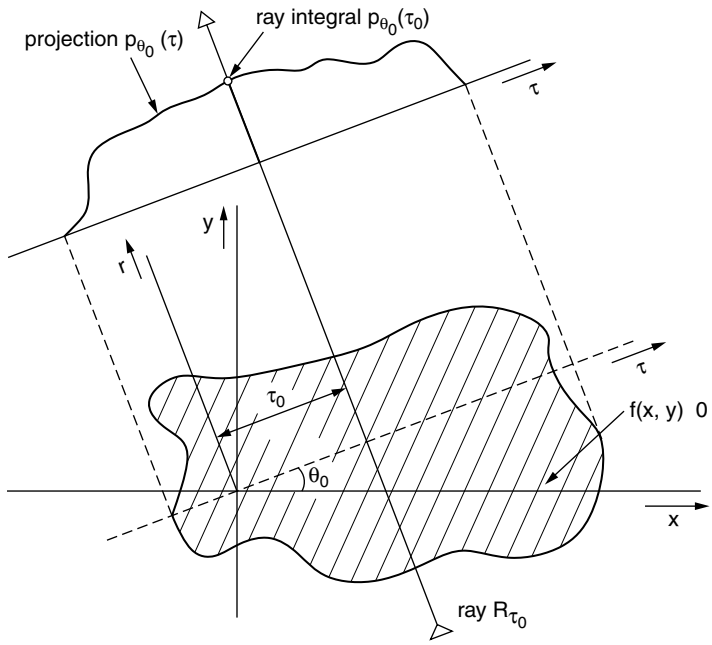


Figure 9.1 Ray integrals and a projection representation of an image. (Modified from Jan, J., *Digital Signal Filtering, Analysis and Restoration*, IEE, London, 2000. Courtesy IEE Publ., London.)

chosen at $\tau = \tau_0$; this line will be called a *ray*. The integral of the function $f(x, y)$ along this line,

$$p_{\theta_0}(\tau_0) = \int_R f(x, y)\,\mathrm{d}r, \tag{9.1}$$

is then called the *ray integral* at (τ_0, θ_0). The integral characterizes the distribution of the measured parameter on the ray by a single overall value; naturally, it is not possible to determine from where the individual contributions to the total value have come.

The choice of τ_0 was arbitrary; the ray integral may therefore be evaluated for any τ, thus becoming a function of τ,

$$p_{\theta_0}(\tau) = \int_{R_\tau} f(x, y)\,\mathrm{d}r, \tag{9.2}$$

which can be plotted as depicted in the upper part of the figure. This one-dimensional function is called the *projection* of $f(x,y)$ under the angle θ_0. Because the angle has also been chosen arbitrarily, the projections $p_\theta(\tau)$ under different angles θ are possible. The ensemble of the projections, infinite in the continuous formulation, then becomes a two-dimensional function in the (τ,θ)-coordinates,

$$p_\theta(\tau) = p(\tau,\theta) = \int_{R_{\tau,\theta}} f(x,y)\,\mathrm{d}r, \quad \tau \in (-\infty,\infty), \quad \theta \in (0,2\pi), \tag{9.3}$$

which is called the *Radon transform* (RT) of $f(x, y)$. In 1917 Radon proved that this transform is inversible, so that the information content in the transform function $p(\tau,\theta)$ is the same as in the original image $f(x,y)$. Hence, it is possible to denote the Radon transform as the (ideal) *projection representation* of the original function (image) $f(x, y)$.

The RT may be expressed by some alternative formulae. The following form expresses the one-dimensional integral via the two-dimensional integral where the linear integration path is determined by the (one-dimensional) Dirac impulse being nonzero only when its argument is zero,

$$p(\tau,\theta) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,\delta(x\cos\theta + y\sin\theta - \tau)\,\mathrm{d}x\,\mathrm{d}y. \tag{9.4}$$

When realizing that the integration path $R_{\tau,\theta}$ is a line $\tau = \text{const}$, and utilizing the rotation of coordinates

$$\begin{bmatrix} x \\ y \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix} \begin{bmatrix} \tau \\ r \end{bmatrix}, \tag{9.5}$$

the generic projection can obviously also be expressed as

$$p(\tau,\theta) = \int_{R_{\tau,\theta}} f(x,y)\,\mathrm{d}r = \int_{-\infty}^{\infty} f(\tau\cos\theta - r\sin\theta,\ \tau\sin\theta + r\cos\theta)\,\mathrm{d}r. \tag{9.6}$$

The formula of the *inverse Radon transform*,

$$f(x,y) = \int_0^{\pi}\int_{-\infty}^{\infty} \frac{\partial p(\tau,\theta)}{\partial\tau} \frac{1}{(x\cos\theta + y\sin\theta - \tau)}\,\mathrm{d}\tau\,\mathrm{d}\theta, \tag{9.7}$$

is presented here without proof for the present; we shall return to it later, when dealing with reconstruction from projections by filtered back-projection.

In the context of image reconstruction from projections, the important (plain) *back-projection operator* should also be mentioned, converting the transform function $p(\tau,\theta)$ into the image

$$b(x,y)=\int_0^{\pi} p(\tau(x,y,\theta),\theta)\,\mathrm{d}\theta=\int_0^{\pi} p(x\cos\theta+y\sin\theta,\theta)\,\mathrm{d}\theta. \qquad (9.8)$$

The image reconstructed this way may be intuitively supposed to approximate $f(x,y)$, but more importantly, the back-projection principle also forms an element of theoretically exact reconstruction algorithms. Every point (x,y) in b is here obviously contributed to by a single value from each projection (at every θ), this value being the ray integral along the ray crossing that point. Conversely said, for a constant pair (τ,θ) defining a ray, the value of the corresponding ray integral contributes evenly to (and only to) all points of this ray in the (x,y)-plane (i.e., on the line $x\cos\theta+y\sin\theta=\tau$). This may be visualized as "smearing" each projection $p(\tau,\theta)$, $\forall\tau$ over the total reconstruction plane (x,y) in the same direction as was originally used for projecting (see Figure 9.2). The resulting back-projected image is thus, according to Equation 9.8, composed of all those smeared projections.

The plain back-projection is not the inverse of the RT. When considering Figure 9.2e, the image reconstructed by the plain back-projection will be obviously heavily distorted. This may be quantitatively visualized when deriving the response of the cascade, formed by the RT, followed by the back-projection, to an image consisting of a single impulse. Both RT and back-projection are linear operators, and it can be shown that their impulse responses are space invariant; thus, the derived response is the PSF of the chain. As it may be intuitively expected, the response is strongest at the position of the original impulse, but is nonzero off this center; the derivation shows the decay with $1/d$, where d is the distance from the center. Thus, instead of the properly reconstructed image $f(x,y)$, we obtain its blurred version $b(x,y)=(1/d)*f|(x,y)$. However, we shall see later that back-projection of *modified* projections may allow the exact reconstruction.

In Part II, we met several types of imaging modalities that provide data approximating the projection representation. Naturally,

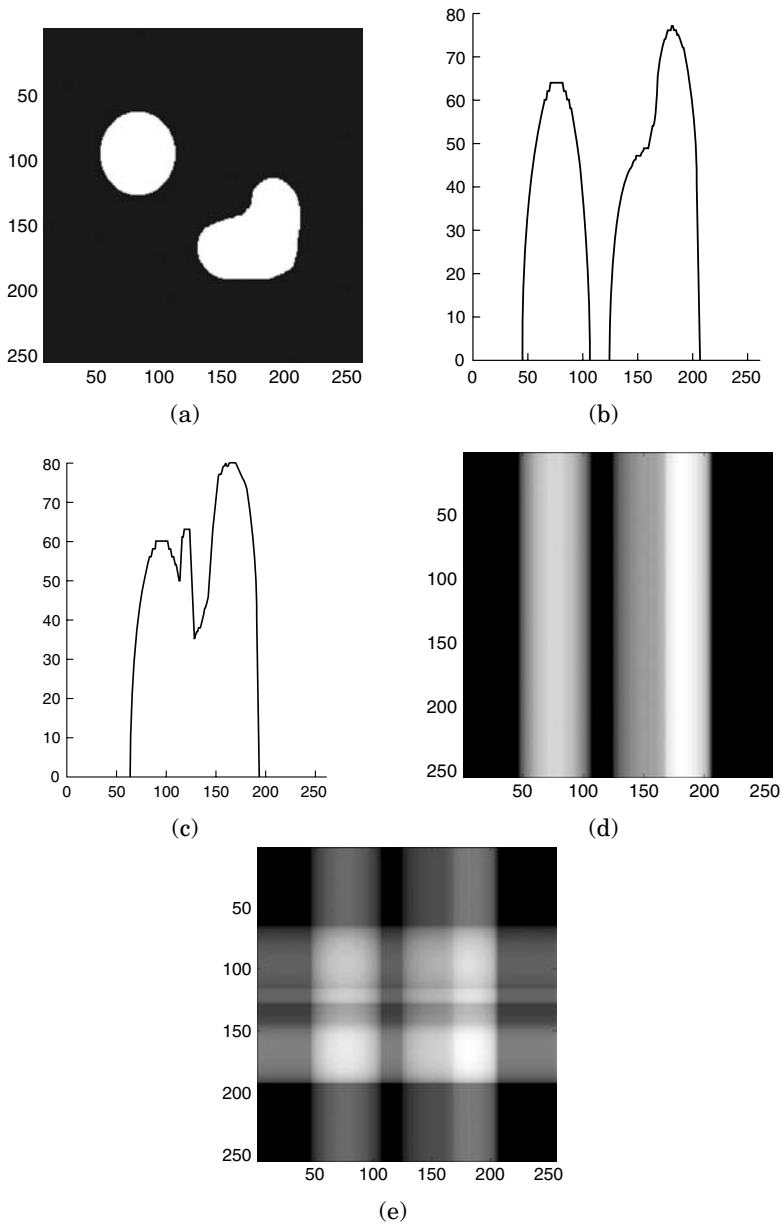


Figure 9.2 Plain back-projection reconstruction: (a) a simple image, (b, c) its vertical and horizontal projections, (d) smeared vertical projection, and (e) partial back-projection reconstruction containing only these two projections.

only a finite number of projections are practicable, and each projection can be described by only a finite number of samples—ray integral values. The measurements thus provide sample sets of the Radon transform*; the sampling densities in both coordinates τ, θ must be high enough to prevent aliasing phenomena and, consequently, complicated artifacts in the reconstructed image. The practical values are usually in the hundreds of projections per image, each consisting of several hundred ray integrals. Since only this limited data set is available, instead of the continuous two-dimensional RT (2D RT), the task of reconstructing the original function $f(x, y)$ (or rather its approximation) from the sample set may be denoted as the *approximate inverse Radon transform*. Although the exact integral, Equation 9.7, for the inverse RT exists, it cannot be directly applied due to the discrete character of the measurement data. The practical methods must therefore take into account the discrete character of the available data either from the very beginning (this leads to algebraic reconstructions) or in the course of implementation (reconstruction via frequency domain, or via filtered back-projection).

The described situation is partly idealized, as it is supposed that the projection values really reflect the ideal ray integrals of the chosen measured parameter $f(x, y)$, which is not entirely true in practice. Of the common modalities, the x-ray CT most approaches the ideal case, when the scatter phenomena are neglected (or compensated for) and the measured values are log-transformed in order to obtain the ray integrals of the attenuation coefficient. Nevertheless, the same principles, though partially modified, are the basis for reconstruction in other modalities as well.

The definition of the Radon transform comes out of the described parallel projections, and consequently, the basic methods of approximate inversion are based on this projection mode. However, most of the modern imaging systems provide fan projections instead, because they are technically more convenient and less invasive to the patients. The modifications of the methods needed to handle the fan data will be mentioned where needed.

*It is tempting to call this sample set the discrete Radon transform. Obviously, it would be improper, as the projection samples are still measured as line integrals of a continuous function.

9.1.2 Algebraic Methods of Reconstruction

9.1.2.1 Discrete Formulation of the Reconstruction Problem

The algebraic methods are based on discretization of the projection scene. The continuous image function $f(x, y)$ is then approximately expressed by means of its samples $f_{i,k}$ via interpolation as

$$f(x,y) = \sum_{i=1}^{I} \sum_{k=1}^{K} f_{i,k}\, g_{i,k}(x,y), \tag{9.9}$$

where the functions $g_{i,k}$ are suitably chosen to form the interpolation basis. When the 0th order interpolation is used (mostly in practice), the function is "staircase" approximated by pixel regions of constant values $f_{i,k}$, as in Figure 9.3. Let us reorder the image matrix into

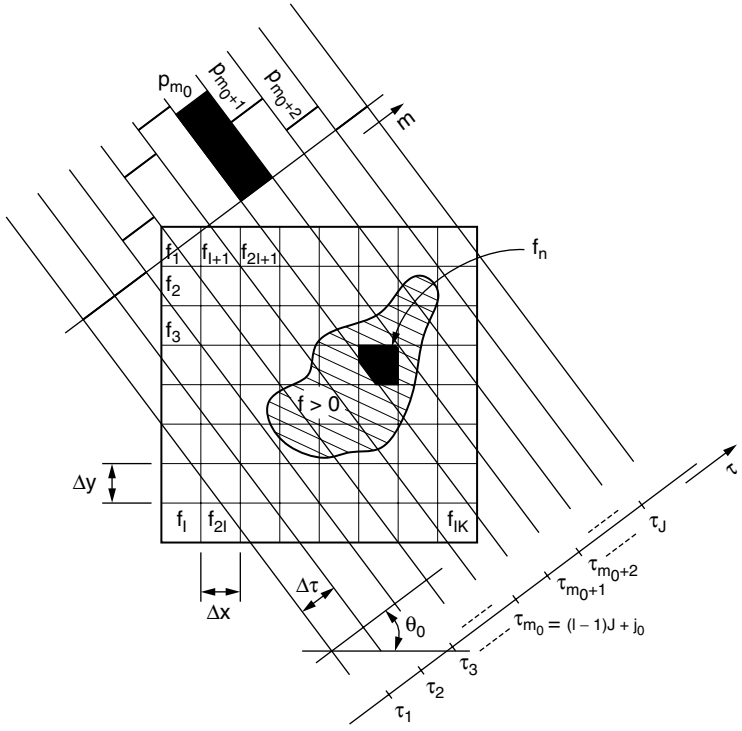


Figure 9.3 Discretization of the projection scene with parallel projection.

the corresponding vector $\mathbf{f} = [f_n]$, $n = 1, 2, \ldots, N = IK$ by column scanning, as in the figure.

Each projection $p(\tau,\theta)$ for a fixed θ is discretized as well, to equidistant samples for τ_j, $j = 1, 2, \ldots, J$, in the range of τ corresponding to the maximum projection size of the image. The sampling corresponds to the physical fact that the projection values (ideally ray integrals) are influenced by certain sensitivity volumes of finite cross-sectional dimensions (corresponding in the best case to narrow pencil beams). The depicted skew stripes in the two-dimensional slice may be understood as approximate representation of the measuring beams replacing the ideal rays for a particular θ; the corresponding projection values $p(\tau_i, \theta)$ are then the measured individual detector outputs. Also, the projection angle is discretized equiangularly, usually in the range $\theta \in <0, \pi>$, providing the set of θ_l, $l = 1, 2, \ldots, L$, the angle increments being determined by the tomographic system construction (see Part II). Hence, $M = JL$ "rays" (beams) are measured, yielding the projection values $p_{j,l}$, which again may be arranged into a vector $\mathbf{p} = [p_m]$, $m = (l - 1)J + j = 1, 2, \ldots, M$.

To be able to calculate projections in the discrete approximation, let us define the projection value (the *stripe sum* approximating a ray integral) as

$$p_m = \sum_{n=1}^{N} w_{m,n} f_n = \mathbf{w}_m \mathbf{f}, \tag{9.10}$$

where the weights $w_{m,n} \in <0,1>$ are given by the intersection area of the m-th stripe with the n-th pixel region divided by the total pixel area $\Delta x \Delta y$. Obviously, most weights are zero, so that the matrix $\mathbf{W} = [w_{m,n}]$ is sparse. Equation 9.10 is a model of projection measurement that may be considered true to the degree in which the discretization is dense enough to describe the measured scene appropriately. As the equation is valid for $\forall m, \forall n$, it represents a set of linear equations

$$\mathbf{Wf} = \mathbf{p}, \tag{9.11}$$

in which elements of $\mathbf{W}$ may be determined based on the geometry of measurement according to Figure 9.3. As far as the projection model (Equation 9.10) may be considered correct, this system also describes the measurements — $\mathbf{p}$ is then the known vector of measured beam ("ray") integrals. Should the matrix $\mathbf{W}$ be square (i.e., when $JL = IK$,

which may always be arranged) and regular, the system may be inverted, yielding the vector of reconstructed pixel values,

$$\mathbf{f} = \mathbf{W}^{-1}\mathbf{p}. \tag{9.12}$$

This way, the problem of reconstruction from discrete projections has been transformed into the known problem of solving a system of linear equations. It is theoretically satisfactory and probably the easiest approach to show that the problem is solvable, at least in principle.

The algebraic formulation leads to computationally intensive algorithms (see below); on the other hand, it has two important advantages:

- Any imaging geometry may be expressed by the weight matrix **W**; the approach is thus, e.g., equally suitable for both parallel and fan projections (see Section 9.1.5, Figure 9.15).
- When implemented via iteration (see below), *a priori* knowledge may be utilized in the form of constraints — e.g., a known finite size of the object (i.e., a part of pixels *a priori* set to zero), non-negativity of pixel values, corrections of nonideal imaging, etc.

9.1.2.2 Iterative Solution

Equation 9.12 is correct but has two problems. The size of the system is enormous: because I, K, J, L are each of the order 10^2 to 10^3, the number of equations is of the order 10^4 to 10^6 and the size of **W** may then reach some 10^8 to 10^{12} elements. Obviously, such a system cannot be solved by standard procedures (matrix inversion, lower–upper decomposition, etc.) with respect to numerical stability problems and prohibitive time consumption, nor would it fit reasonably in a contemporary hardware. The other problem concerns the regularity of **W**: the neighboring equations have similar coefficients, and are therefore close to linear dependence; thus, the matrix is near-singular. It is known that such systems are sensitive to small errors in the right-hand-side vector so that the obtained solution, if any, may suffer extreme errors. This problem is alleviated by increasing the number of measurements so that $M > N$; the overdetermined system should then be solved by the least mean square error method (by *pseudoinversion*) or in a similar sense.

As an alternative to the mentioned one-step methods, iterative procedures may be used for solving such large systems. Let us describe

the Kaczmarz method (1937), which will also have an interesting interpretation in terms of the iterative reprojection improvement (see below).

Each of the equations described in Equation 9.10, $\mathbf{w}_m \mathbf{f} = p_m$, represents a hyperplane in the N-dimensional vector space of $\mathbf{f}$. Should $N = M$, all the hyperplanes intersect at the proper solution point of the reconstructed image $\underline{\mathbf{f}}$. When $N < M$, the hyperplanes would not intersect due to measurement errors; however, there will be a region of reasonable solutions close to $\underline{\mathbf{f}}$, where partial intersections will be situated. The iterative procedure seeks the intersection (or the intersection region) via consecutive approximations. The iteration starts from an arbitrary initial estimate $\mathbf{f}_0$, e.g., a constant gray field. Each iterative step starts from the previous estimate ${}^{t-1}\mathbf{f}$ and provides a new estimate ${}^{t}\mathbf{f}$ by projecting the old one perpendicularly to the t-th hyperplane as

$$ {}^{t}\mathbf{f} = {}^{t-1}\mathbf{f} + \frac{\left(p_{\{t\}} - \mathbf{w}_{\{t\}}^{T}\, {}^{t-1}\mathbf{f}\right)}{\mathbf{w}_{\{t\}}^{T}\mathbf{w}_{\{t\}}}\mathbf{w}_{\{t\}}. \tag{9.13} $$

Note that $p_m / |\mathbf{w}_m|$ is the distance of the m-th hyperplane from the origin and $\mathbf{w}_m$ is a vector perpendicular to it. Thus, the fraction is $|{}^{t-1}\mathbf{f} - {}^{t}\mathbf{f}| / |\mathbf{w}_{\{t\}}|$ so that the correction term is the vector $\Delta^{t}\mathbf{f} = {}^{t}\mathbf{f} - {}^{t-1}\mathbf{f}$, perpendicular to the t-th hyperplane. Usually, the number of available equations (M) is not sufficient to reach the solution (M iterative steps are insufficient), so that when the iteration index t exceeds M, the equations are reused, the index of the equation used in the t-th iteration becoming $\{t\} = t \bmod M$. It has been proved that the iteration converges to $\underline{\mathbf{f}}$, as illustrated for $M = N = 2$ in Figure 9.4. When $M > N$, the iteration reaches the region of partial intersections and ends in a limit cycle; each of the reappearing final approximations (or, e.g., their average) may be used as the estimate of the solution (Figure 9.5).

9.1.2.3 Reprojection Interpretation of the Iteration

The vector equation (Equation 9.11) relates the original image $\mathbf{f}$ (or its reconstruction) to the measurement vector $\mathbf{p}$. The weight matrix $\mathbf{W}$, however, may be utilized in the same manner to also derive the projection vector ${}^{t}\mathbf{q}$ from any of the intermediate approximate solutions ${}^{t}\mathbf{f}$,

$$ {}^{t}\mathbf{q} = \mathbf{W}\, {}^{t}\mathbf{f}, \qquad \text{i.e.,} \qquad {}^{t}q_m = \sum_{n=1}^{N} w_{m,n}\, {}^{t}f_n = \mathbf{w}_m^{T}\, {}^{t}\mathbf{f}. \tag{9.14} $$

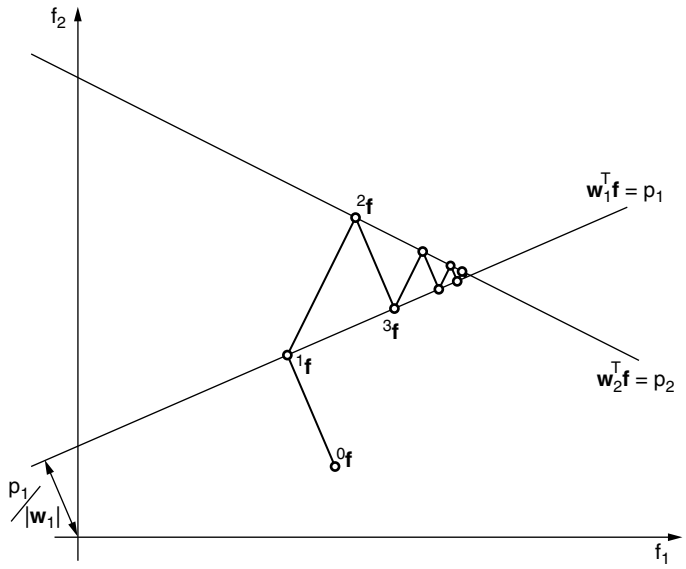


Figure 9.4 Projective iteration for $M = N = 2$.

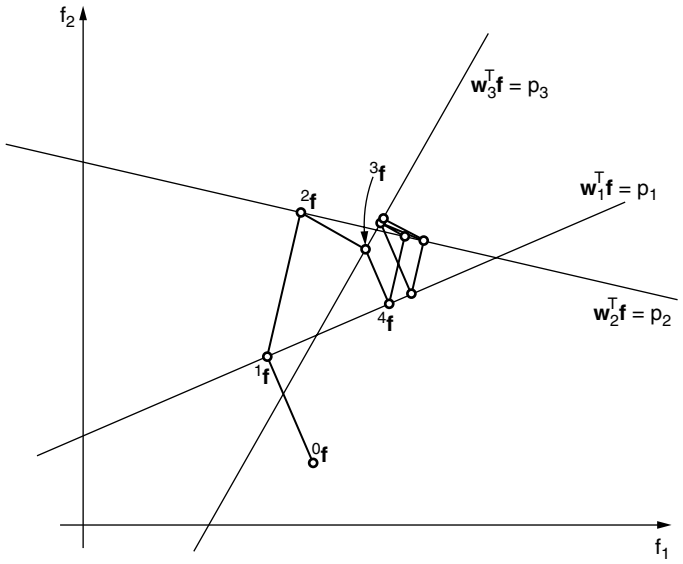


Figure 9.5 Projective iteration in the case $M = 3, N = 2$.

The iterative correction (Equation 9.13) expressed for an individual pixel gives

$$ {}^{t}f_n = {}^{t-1}f_n + \frac{\left(p_{\{t\}} - \mathbf{w}_{\{t\}}^{T}\,{}^{t-1}\mathbf{f}\right)}{\mathbf{w}_{\{t\}}^{T}\mathbf{w}_{\{t\}}} w_{\{t\},n}\,, \tag{9.15} $$

and when substituting for both dot products and realizing that in the t-th iteration, $\{t\} = m$, we obtain

$$ \Delta\,{}^{t}f_n = {}^{t}f_n - {}^{t-1}f_n = \left(p_m - {}^{t-1}q_m\right) \frac{w_{m,n}}{\sum_{l=1}^{N} w_{m,l}^2}\,. \tag{9.16} $$

This is an interesting result: the correction of the particular n-th pixel is determined by the projection residuum—the difference between the real beam measurement and its estimate from the current approximation of $\mathbf{f}$. The residuum $(p_m - {}^{t-1}q_m)$ is distributed among all the pixels that intersect the m-th stripe, in proportion to the respective pixel-stripe weight $w_{m,n}$ divided by the square length of $\mathbf{w}_m$. As all the stripes belong to L different projections with J stripes each, the algorithm may be better visualized as depicted in Figure 9.6. For each projection angle θ, the measured projection is compared with the reprojection based on the last approximation of $\mathbf{f}$. The residual differences are used to correct all pixels of the respective stripes; the correction of an individual pixel is proportional to its relative contribution to the active area of the stripe. Interpreted this way, the algorithm contains elements of back-projection, though not the measured projections but rather the corrections are back-projected.

Different algorithms diversified by computational details are based on (or approximating) the iteration in Equation 9.16; all of them may be denoted as *algebraic reconstruction techniques* (ART algorithms), although this abbreviation is primarily used for a group of simplified algorithms (see later).

Doing this for all J stripes of a projection is obviously equivalent to one step of the iteration (Equation 9.13), providing that the complete reprojection was calculated beforehand, after completing corrections due to residua of the previous projection. There is an obvious advantage of the residual algorithm over the direct implementation of Equation 9.13: the possibility of running a check of the improvement during iteration—the residua should, at least on average, decrease monotonously. The algorithm is terminated when the residua, and consequently the corrections, become negligible.

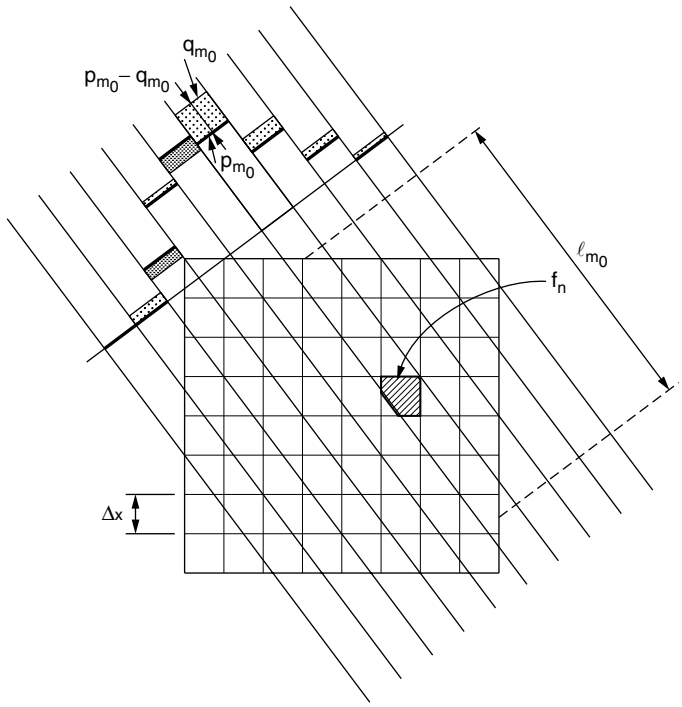


Figure 9.6 Residual of a projection distributed among pixels.

Correcting strategies other than the one mentioned above is possible, with the extremes, on one hand, correcting each stripe independently and, on the other hand, correcting all the pixels simultaneously only once per a complete cycle of M stripes (i.e., after processing all the projections). The first (*immediate*) strategy may introduce artifacts due to frequently correcting the same individual pixels so that the consecutively used equations may become rather inconsistent; this phenomenon is particularly serious in simplified calculations (see below). The other (*accumulating*) approach consists of accumulating the corrections $\Delta^t f_n$ for each (n-th) pixel calculated from all M stripes in an iteration cycle (for $t \in \langle (T-1)M, TM-1 \rangle$, i.e., $\forall m = t \bmod M$), during which $\mathbf{f}$ remains constant. The resulting correction is then the average

$$\Delta f_n = \frac{1}{M} \sum_{t=(T-1)M}^{TM-1} \Delta^t f_n, \qquad \text{i.e.,} \qquad \Delta \mathbf{f} = \frac{1}{M} \sum_t \Delta^t \mathbf{f}. \tag{9.17}$$

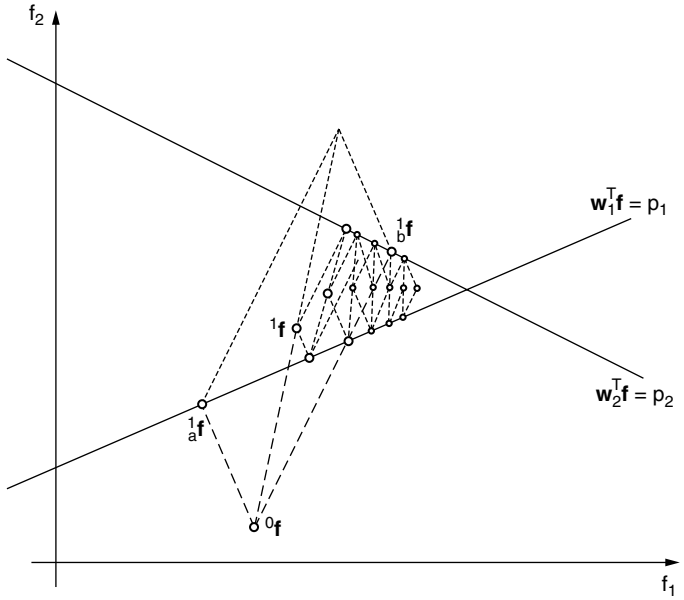


Figure 9.7 Averaged-correction strategy in an iteration for $M = N = 2$.

Thanks to averaging, the errors in the partial corrections tend to cancel even in the case of simplified calculations, as mentioned below. The averaged-correction strategy in a two-dimensional case is visualized in Figure 9.7, where the two correction vectors corresponding to Equation 9.13 under the fixed **f** are averaged in each cycle; obviously, the consecutive approximations then do not lay on the hyperplanes.

9.1.2.4 Simplified Reprojection Iteration

Though the **W** matrix is sparse and, when using the reprojection iteration, only weights concerning the just processed stripe are currently needed, still the computational load due to exact calculation of the weights based on the cross-sectional areas is enormous. In order to cut down the requirements, the matrix **W** may be approximated by a binary incidence matrix: the weight $w_{m,n}$ is 1 when the n-th pixel center lies in the m-th stripe; otherwise, it is zero. The decision can be done with much less effort than calculation of the cross-sectional area, so that it is even possible to calculate the just needed weights during the iteration, thus also releasing the extreme memory requirements.

Obviously, in this case the correction (Equation 9.16) is approximated as

$$\Delta {}^{t}f_n \approx \frac{1}{L_m}(p_m - {}^{t-1}q_m) \tag{9.18}$$

where L_m is the number of incident pixels on the m-th stripe. All incident pixels are thus corrected equally — the (normalized) projection residuum is "smeared" over the matrix of previous estimate ${}^{t-1}\mathbf{f}$ and added to the current values of the incident pixels.

When choosing the immediate correcting strategy (see above), the resulting images may suffer with impulse noise caused by inconsistency of equations, in which the sums are imprecise due to the discontinuous character of incidences used instead of smoothly changing weights. In the narrower sense, the above-mentioned abbreviation ART is used for algorithms of this group. The algorithms, denoted as SIRT (*simultaneous iterative reconstruction techniques*), may use the same simplified correction formula, but differ in using the accumulating correcting strategy with the aim to suppress the mentioned noise due to the step-wise inconsistencies by averaging the individual corrections. The algorithms of the SART group (*simultaneous* ART) use the accumulating correction strategy as well, but besides other differing details (see [37]), the stripe sums are calculated based on improved interpolation representation of the reconstructed image, this way enabling faster convergence.

9.1.2.5 Other Iterative Reprojection Approaches

The iterative reprojection method, described in detail above, is just an example from a rich set of similarly based approaches. The common idea is to provide an initial reconstruction by any method, possibly under simplified conditions, neglecting some of the disturbing phenomena (e.g., spatially dependent attenuation; see later in Section 9.2.1). Based on a known estimate of the reconstructed image, reprojections are provided in the *feedback step* as precisely as possible and compared with the measured projections. The differences are then used in the following *forward step* to provide corrections that (hopefully) lead to an improved image reconstruction by modifying the previous estimate. The forward and feedback steps are cyclically repeated, while the reconstructed image should improve monotonously and converge to a proper reconstruction, if the algorithm is well designed. The cycle repeats until the differences are negligible (or tolerable).

It should be noted that the mentioned criterion — a good agreement of the reprojections and originally measured projections — is a necessary, but not automatically sufficient, condition of a good reconstruction.

As mentioned above, the iterative reconstruction algorithms, most of which are based on the forward-feedback scheme, have the advantage of easily implementing additional constraints that are too difficult to be included in the analytic solutions, e.g., non-negativity, attenuation influence in emission tomography (see Section 9.2.1), etc. That is why they are gaining attractiveness, namely, when the main previous obstacle of their practical exploitation — lack of computing power needed for the demanding iterative procedures — is overcome by the hardware development.

9.1.3 Reconstruction via Frequency Domain

9.1.3.1 Projection Slice Theorem

The methods of reconstruction via the frequency domain are based on the projection slice theorem, which formulates the relation between the two-dimensional spectrum of a two-dimensional image,

$$F(u,v) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-j(ux+vy)}\mathrm{d}x\,\mathrm{d}y, \tag{9.19}$$

and the one-dimensional spectrum of a one-dimensional projection of this image under the angle θ,

$$P_\theta(w) = \int_{-\infty}^{\infty} p_\theta(\tau)e^{-jw\tau}\mathrm{d}\tau. \tag{9.20}$$

Substituting for $p_\theta(\tau)$ from Equation 9.6, we obtain

$$P_\theta(w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\tau\cos\theta - r\sin\theta, \tau\sin\theta + r\cos\theta)e^{-jw\tau}\mathrm{d}r\,\mathrm{d}\tau \tag{9.21}$$

and, transforming the coordinates according to Equation 9.5, finally

$$P_\theta(w) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)e^{-jw(x\cos\theta + y\sin\theta)}\mathrm{d}x\,\mathrm{d}y, \tag{9.22}$$

as the determinant of the transform Jacobian is 1. The last expression is obviously the two-dimensional FT of $f(x, y)$ with $u = w \cos \theta$ and $v = w \sin \theta$. Hence, the value of the one-dimensional spectrum for the spatial frequency w is equal to the two-dimensional spectrum value at the absolute spatial frequency w, located on the line, crossing the origin of (u, v) and inclined by the angle θ relative to the u-axis.

This result—the *projection slice theorem*—thus states that the one-dimensional spectrum $P_\theta(w)$ of a projection $p(\tau, \theta)$ obtained from the image $f(x, y)$ is equal to the central slice, at the angle θ, of the two-dimensional image spectrum $F(u, v)$.

9.1.3.2 Frequency-Domain Reconstruction

Based on this theorem, a conceptually transparent approach to approximate image reconstruction from a finite number of parallel projections is formulated (Figure 9.8). In principle, having measured an ensemble of projections, each projection is Fourier transformed.

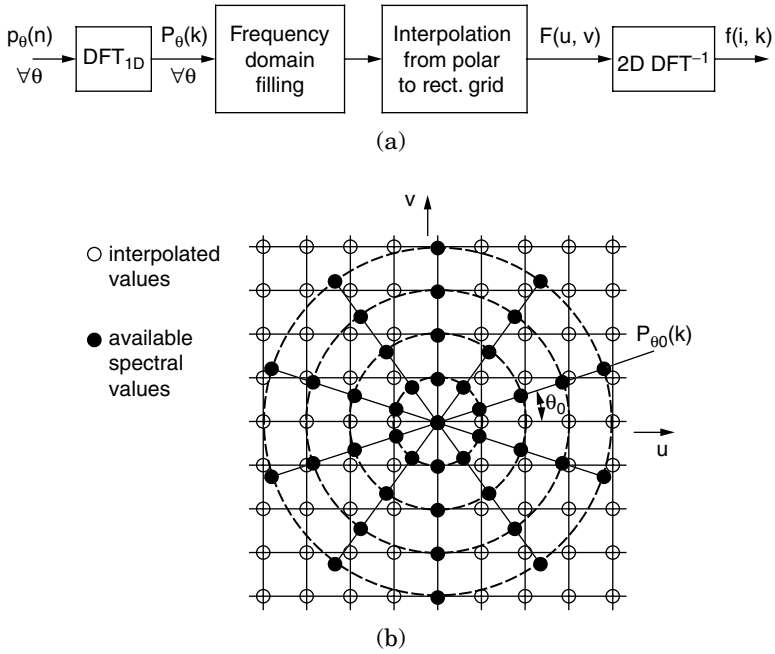


Figure 9.8 Fourier domain reconstruction from projections: (a) the concept and (b) sampling in the u, v domain.

The obtained one-dimensional spectra provide, according to the slice theorem, the values of the two-dimensional image spectrum on the corresponding central lines, so that an (angularly sampled) description of the two-dimensional spectrum on the line set is obtained. If a sufficient number of lines is available, the missing spectral values may be interpolated, thus obtaining the complete (though approximate) two-dimensional spectrum of the image. The approximation of the image can then be obtained by the inverse Fourier transform.

The practical tomographic measurements provide only sampled projections that are then transformed by one-dimensional DFT yielding sampled one-dimensional spectra. Consequently, the two-dimensional spectrum is described only discretely, by sampled lines that altogether form a polar raster of samples equidistant in w and equiangular in θ (Figure 9.8b). The samples are therefore dense in the neighborhood of (u, v)-origin, but rather scarce in the region of high frequencies. As the fast algorithms of two-dimensional DFT require spectral samples on the equidistant rectangular grid of u,v, interpolation is necessary that, due to scarceness of the high frequency coverage, may lead to inaccuracies manifesting themselves as artifacts in the reconstructed image. The sampling densities (including sampling in θ that determines the number of taken projections) must therefore be appropriately chosen with respect to the character of the imaged scene. When substituting the integral FT and FT^{-1} by the corresponding discrete transforms, the imposed periodicity implicit to DFT must be observed and appropriate measures like zero padding and windowing used.

9.1.4 Reconstruction from Parallel Projections by Filtered Back-Projection

9.1.4.1 Underlying Theory

Let us primarily show why the continuous-space plain back-projection (Equation 9.8) cannot provide the proper image reconstruction, and how to modify it so that the distortion is removed. Then we shall prove the continuous inverse Radon transform formula (Equation 9.7).

The image can be obtained from its spectrum via inverse FT,

$$f(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} F(u,v)e^{j(ux+vy)}\mathrm{d}u\,\mathrm{d}v, \tag{9.23}$$

which can be expressed in polar coordinates, $w = \sqrt{(u^2 + v^2)}$, $\theta = \arctan(v/u)$, as

$$f(x,y) = \int_0^{2\pi} \int_0^{\infty} \tilde{F}(w,\theta) e^{jw(x\cos\theta + y\sin\theta)} w \, \mathrm{d}w \, \mathrm{d}\theta, \tag{9.24}$$

where $\tilde{F}(w,\theta) = F(w\cos\theta, w\sin\theta)$ is the two-dimensional spectrum expressed in polar coordinates. Utilizing the equivalence of positions $(w, \theta + \pi)$ and $(-w, \theta)$, the last double integral may be rewritten as

$$f(x,y) = \int_0^{\pi} \int_{-\infty}^{\infty} \tilde{F}(w,\theta) e^{jw(x\cos\theta + y\sin\theta)} |w| \, \mathrm{d}w \, \mathrm{d}\theta, \tag{9.25}$$

where the absolute value of the Jacobian, implicit already in Equation 9.24, must be explicitly mentioned when w may become negative.

The inner integral in Equation 9.25, in which θ is a parameter, may be expressed as

$$\int_{-\infty}^{\infty} \tilde{F}(w,\theta) e^{jw(x\cos\theta + y\sin\theta)} |w| \mathrm{d}w = \int_{-\infty}^{\infty} P_\theta(w) |w| \, e^{jw(x\cos\theta + y\sin\theta)} \mathrm{d}w, \tag{9.26}$$

using the projection slice theorem (for a constant θ, the two-dimensional spectrum equals the one-dimensional spectrum of the respective projection). Note that the last integral is the one-dimensional inverse FT of the spectral function $Q_\theta(w) = |w| \; P_\theta(w)$, which may be considered the projection spectrum $P_\theta(w)$ modified by linear filtering having the frequency response $|w|$. Considering the coordinate transform (Equation 9.5) and applying the convolution theorem, we may write

$$\begin{aligned} q_\theta(\tau) &= \int_{-\infty}^{\infty} Q_\theta(w) e^{jw\tau} \mathrm{d}w = \mathrm{FT}^{-1}\{P_\theta(w)|w|\} \\ &= p_\theta * h|(\tau) = \int_{-\infty}^{\infty} p_\theta(s) \, h(\tau - s) \, ds, \end{aligned} \tag{9.27}$$

where $h(\tau) = \mathrm{FT}^{-1}\{|w|\}$ is the ideal impulse response of the mentioned filter (see the comment below), the frequency response of which is even and linearly increasing with frequency without limit

(Figure 9.10a). The original-domain function $q_\theta(\tau) = \mathrm{FT}^{-1}\{Q_\theta(w)\}$ is denoted as the *filtered projection* at the angle θ.

When substituting the filtered projection for the inner integral in Equation 9.25, we finally obtain, using, as in Equation 9.3, both alternative notations, i.e., $q_\theta(\tau)$ and $q(\tau, \theta)$,

$$f(x,y) = \int_0^\pi q_\theta(\tau)\,\mathrm{d}\theta = \int_0^\pi q(x\cos\theta + y\sin\theta, \theta)\,\mathrm{d}\theta\,. \qquad (9.28)$$

In comparison with the plain back-projection formula (Equation 9.8), we see that the image reconstruction is given by the back-projection of a (continuous) set of projections *modified* by the filter with the frequency response $|w|$. This is a fundamental result — the (continuous) inverse Radon transform formulation, called *filtered back-projection* (FBP), enabling the attainment of a theoretically exact reconstruction from projections. It also explains why the plain back-projection of unfiltered projections yields imperfect reconstructions. The discrete approximations of Equations 9.27 and 9.28 are the basis for the most commonly used reconstruction algorithms.

Before leaving the theoretical considerations, let us discuss the properties of the filter with the frequency response $|w| = w$ sign (w). As this function is not absolutely integrable, it is impossible to derive directly the corresponding impulse response by FT^{-1}. However, any projection $p_\theta(\tau)$ to be filtered is automatically band-limited due to imperfect resolution of imaging so that the inverse FT in Equation 9.27 exists. Consequently, the filter's frequency response may be limited to the extent of the existing signal band by a suitable finite weighting window; then the reconstructed image remains unaffected. Such a filter is realizable and its impulse response can be determined once the applied frequency window is chosen (see the discussion below on common filters in discrete reconstruction implementations).

In Figure 9.9, corresponding to Figure 9.2, both filtered projections are plotted (panels b and c) and smeared over the image area (panels d and e); the partial filtered back-projection (panel e) consists of only two filtered projections, and is thus very incomplete. However, it may be estimated that the back-projection of filtered projections would lead to higher-quality reconstruction thanks to the negative peaks that compensate for the disturbing traces outside of the objects. Naturally, only when back-projecting a high number of differently oriented filtered projections do we obtain a reasonable approximation of the original image.

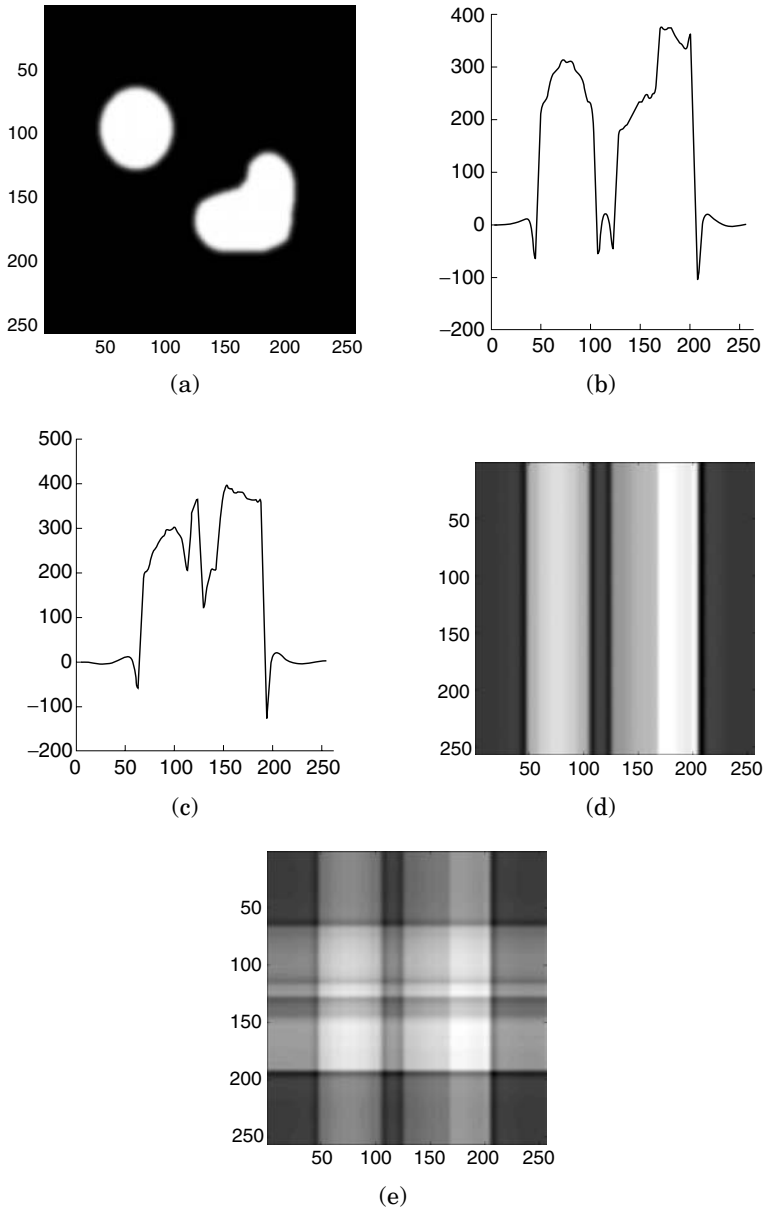


Figure 9.9 Filtered back-projection (compare with Figure 9.2): (a) original image, (b) filtered vertical projection, (c) filtered horizontal projection, (d) smeared vertical filtered projection, and (e) partial filtered back-projection reconstruction consisting of a mere two filtered projections.

In order not to limit our discussion to a particular windowed filter, we shall use an alternative approach of deriving the inverse transform of the complete modified spectrum $Q_\theta(w) = |w|\, P_\theta(w)$. Taking into consideration the basic FT properties, $\mathrm{FT}\{\frac{\mathrm{d}}{\mathrm{d}\tau}\, p(\tau)\} = j\, w\, P(w)$, and $\mathrm{FT}\{-\frac{1}{\tau}\} = j\ \mathrm{sgn}(w)$, we can write

$$q_\theta(\tau) = \mathrm{FT}^{-1}\{w\, P_\theta(w)\ \mathrm{sgn}(w)\} = \mathrm{FT}^{-1}\{w\, P_\theta(w)\} * \mathrm{FT}^{-1}\{\mathrm{sgn}(w)\}$$

$$= \frac{\mathrm{d}p_\theta(\tau)}{\mathrm{d}\tau} * \frac{1}{\tau} = \int_{-\infty}^{\infty} \frac{\partial p(s,\theta)}{\partial s} \frac{1}{\tau - s}\, \mathrm{d}s \tag{9.29}$$

Substituting this result into Equation 9.28 leads to the inverse Radon transform formula (Equation 9.7); it has been thus proved. The formula is rather of theoretical interest, as the integrand in Equation 9.29 and consequently in Equation 9.7 has a singularity so that the integral is to be evaluated analytically in the sense of principal value.

9.1.4.2 Practical Aspects

The above formulae may serve as the basis for a design of practical reconstruction algorithms using a finite ensemble of discrete sampled projections. Obviously, it is necessary to provide reasonable discrete approximations of the formulae (on one-dimensional signal processing, see, e.g., [30] or [51]). The last stage of all the algorithms is the back-projection procedure (Equation 9.28); the three basic types of algorithms differ only in the way of providing the sampled filtered projections $q_\theta(\tau)$ based on the measured (noisy) data.

First, let us consider Equation 9.29 to filter the projections. Obviously, the theoretical procedure consists of differentiating, followed by convolving the differentiated projections with the function $1/\tau$. The algorithm should first approximately differentiate the projections, using finite differences, which is generally not a recommended operation on noisy data. Further, the cumulation that would approximate the convolutional integration (of a function with singularity) is also uneasy to realize with an adequate degree of numerical stability. The convolution, which represents the Hilbert transform,* may be alternatively better accomplished by a discrete filter approximating the Hilbert transformer; however, it is also a demanding task

*The convolution $f(x)*1/x$, i.e., $\psi(x) = \mathrm{HT}\{f(x)\} = \int_{-\infty}^{\infty} f(t) \frac{1}{x - t}\, dt$, is denoted the *continuous Hilbert transform* of $f(x)$.

when accuracy is required. Consequently, Equation 9.29 is not particularly convenient for practical projection filtering.

The two remaining practical methods thus rely on Equation 9.27 in one of its forms, which means filtering either in the frequency domain (upper formula, realization via DFT) or in the original domain (lower formula, finite impulse response (FIR) realization). In both cases, the theoretical filter has to be modified as already discussed above; i.e., its frequency band must be limited on the side of high frequencies. This is primarily enabled by the restricted spatial resolution of imaging, which consequently limits the bandwidth of measured projections; this in turn enables replacing the continuous projections by their sampled versions without any informational loss. Considering the sampling theorem, the upper limit of the filter bandwidth must not exceed the Nyquist spatial frequency w_N, so that the minimally modified filter should have the frequency response

$$H(w) = \begin{cases} |w|, & w \in \langle -w_N, w_N \rangle \\ 0 & \text{elsewhere} \end{cases}, \tag{9.30}$$

as depicted in Figure 9.10b. The corresponding (infinite) impulse response $h(\tau)$ is a rather complicated second-order polynomial in $\mathrm{sinc}(\sim w_N \tau)$, the sampled version of which is approximately depicted in Figure 9.10e.

The limitation of the maximum frequency is equivalent to multiplying the theoretical frequency response $|w|$ by a rectangular unit window of the width $2w_N$ that corresponds to convolving the filtered signal with a sinc function, a rather undesirable operation in the presence of noise. Thus, smoother windows are commonly applied that also partly suppress the high frequencies below w_N, thus expectantly limiting the noise. The choice of window influences the quality of reconstruction; comparisons of reconstruction results obtained with different filters can be found in many references, e.g., in [66], [26], [64], [37].

The filter selection also influences the artifacts due to discretization. The cause of the primary artifact — elimination of the spectral component at $w = 0$ — is seemingly present already in the continuous formulation (Equation 9.27). While the continuous inverse FT is not influenced (the integral value is insensitive to isolated integrand values), in the discrete formulation not only a single frequency, but also a certain band around $w = 0$ is lost. This leads to a change in the mean value (overall brightness) and modifies slow brightness trends in the reconstructed image. A possible cure is to limit the width of the neglected band; this may

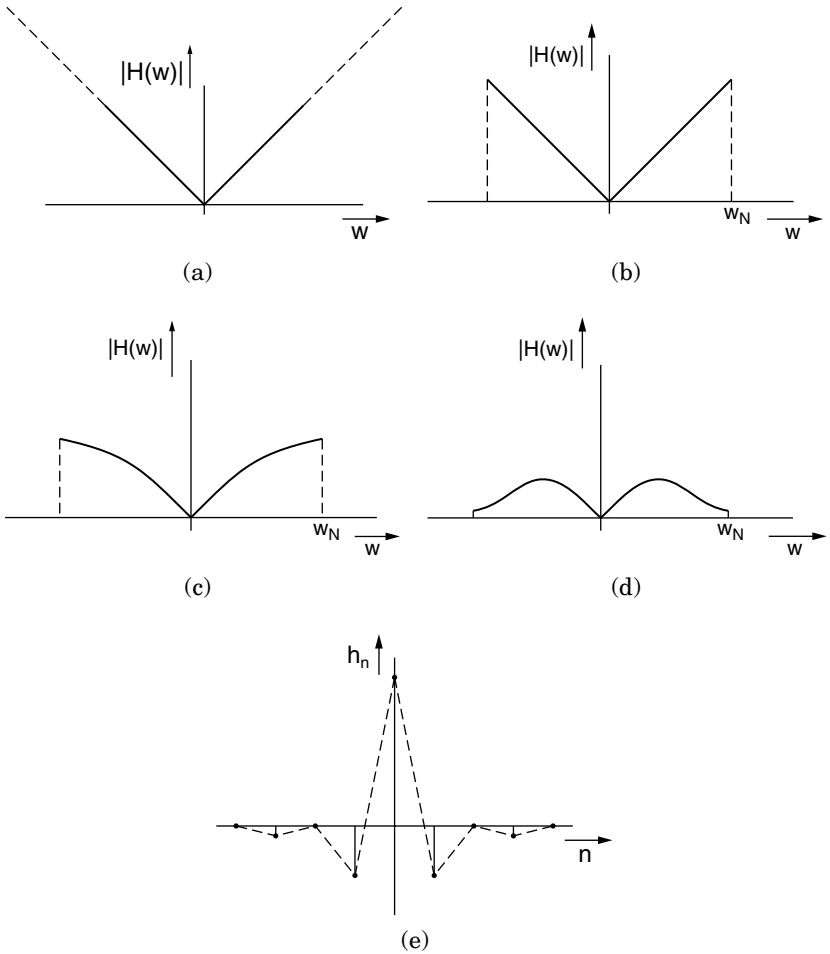


Figure 9.10 Frequency responses of filters in FBP reconstruction: (a) theoretical unlimited filter, (b) minimally modified band-limited filter, (c, d) windowed filters (Shepp–Logan and Hamming windowed), and (e) impulse response of FIR approximation of the filter in panel b.

be achieved by sampling more densely in the frequency domain. As the number of projection samples is given and fixed, this is provided for by zero-padding the projection data, thus doubling or multiplying the data extent. Alternately, the filter frequency response may be heuristically raised at $w = 0$ to a small nonzero value.

The filter is usually implemented as an FIR realization; then effects due to shortening the exact impulse response may appear that should be treated by choosing appropriate weighting, together with a proper length of the filter. These FIR filters are commonly realized via frequency domain, which implies replacing the requested linear convolution in Equation 9.27 or 9.29 by circular convolution. Then it is necessary to prevent interperiod interference artifacts, again by proper zero padding of projections. The direct implementation of the FIR filters in the original domain is an alternative, but usually slower, possibility.

Once the projections are filtered, the back-projection integral (Equation 9.28) must be discretely approximated, e.g., by the rectangular rule,

$$f(x,y) \approx \frac{\pi}{L} \sum_{l=1}^{L} q_{\theta_l}(x \cos\theta_l + y \sin\theta_l, \theta), \tag{9.31}$$

where L is the number of measured projections in 180°. The contribution to pixel values $f(x, y)$ from a particular projection at θ_l is

$$\frac{\pi}{L} q_{\theta_l}(x \cos\theta_l + y \sin\theta_l) = \frac{\pi}{L} q_{\theta_l}(\tau); \tag{9.32}$$

thus identical to all points (x, y) on the line $\tau = x \cos \theta_l + y \sin \theta_l = \text{const}$. The complete projection is thus smeared in the respective direction of projecting and its values are added adequately to all touched pixels — this procedure is denoted a back-projection of an individual filtered projection. The reconstructed image is formed of sums of these contributions from all projections, ideally each position (x, y) receiving just a single contribution from each of numerous individual projections.

So far, we considered only the values of projection angles as discrete. However, the projections are sampled as well, so that only discrete values $\frac{\pi}{L} q_{\theta_l}(\tau_j)$, $j = 1, 2, \ldots, J$ are available that may be understood as being stripe results from a discretized image, similarly as explained in Section 9.1.2. An individual rectangular pixel then obtains, from each filtered-projection value, a contribution proportional to the intersection area of the pixel and the stripe, normalized by the total intersection of the stripe with the image (Figure 9.11). This way, the projection values are distributed similarly as the residual values in the algebraic reconstruction; a pixel thus may (and usually does) obtain partial contributions from more than a single stripe of a projection.

An obvious advantage of the back-projecting methods is the immediate use of each individual projection once it is available

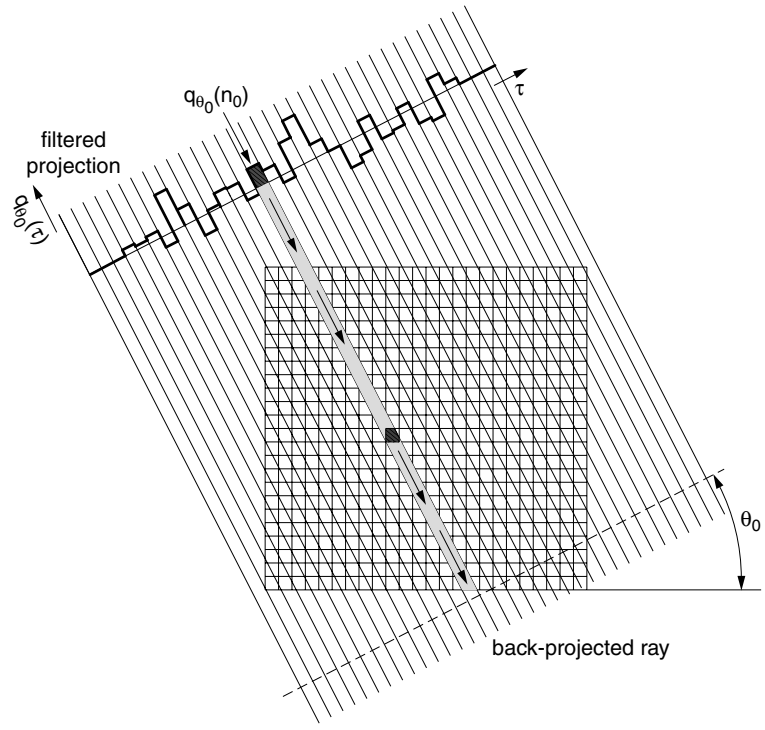


Figure 9.11 Discrete filtered back-projection.

during measurement; it is then filtered and back-projected into the resulting image. In principle, there is no need to keep the projection saved any further, so that the memory requirements are lower than with other methods.

9.1.5 Reconstruction from Fan Projections

Modern imaging systems mostly provide the measurement data from fan projections, with acquisition of data of a complete projection at a time in parallel (see Part II). It is then possible either to reorganize the measurement data so that parallel projection data are obtained and the above described reconstruction procedures may be applied without modifications, or to derive algorithms specially designed for the fan projections. We shall briefly mention both approaches, limiting the discussion to the common case of equiangular fans.

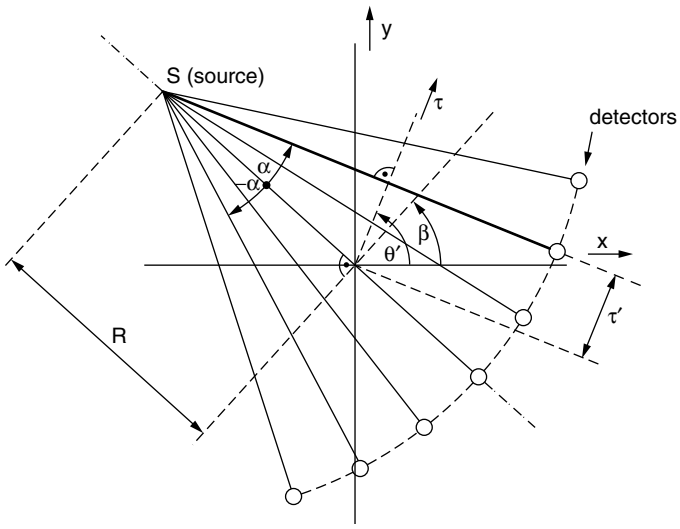


Figure 9.12 Geometry of fan projection measurement.

The measuring geometry is the same in both cases (Figure 9.12). The measuring fan is *equiangular*—i.e., the angular differences between neighboring rays are constant. The detectors are here situated on a circular segment with identical distances to the vertex where a radiation source may be situated (as, e.g., in CT modality); however, the distance is essentially unimportant for most tomographic modalities as long as the ray length differences are in air. The explicit source may be missing, as in convergent SPECT, or another detector may be considered instead of the source, as in PET.

The position of a ray in the fan is defined by the angle α between the ray and the symmetry axis of the fan. The fan is step by step turning around the origin of coordinates to provide successively a set of projections; the angle β between the fan axis and y-axis defines a concrete fan projection. The distance R from the center to the vertex S, about half of the ray length, determines the geometry of measurement.

From the reconstruction point of view, each ray of the fan is independent, defining its ray integral. We may thus consider each ray integral either as belonging to the fan projection $r_\beta(\alpha)$ or as a sample of a parallel projection $p_\theta(\tau)$. As an example, let us consider the ray indicated by the bold line in the figure, which defines both the ray integral $r_\beta(\alpha')$ of the fan positioned at β, and also the ray

integral $p_{\theta'}(\tau')$ from the parallel projection at θ'. It follows from the figure geometry that, for any ray,

$$\tau = R\sin\alpha, \qquad \theta = \alpha + \beta \tag{9.33}$$

and consequently,

$$r_\beta(\alpha) = p_{(\alpha+\beta)}(R\sin\alpha) \tag{9.34}$$

9.1.5.1 Rebinning and Interpolation

The fan projections $r_\beta(\alpha)$ are provided by the imaging system on a two-dimensional discrete grid $\{\alpha_n, \beta_m\}$. The same ray integrals may be considered to also constitute a set of parallel projections $p_\theta(\tau)$ on the grid $\{\tau_n = R \sin \alpha_n, \theta_k = \beta_m + \alpha_n\}$, however unsorted. It is nevertheless possible to re-sort the projections according to $\theta_k = \beta_m + \alpha_n$, thus obtaining sets of discrete parallel projections, although sampled irregularly. This process is denoted as *rebinning*. The needed equidistantly sampled parallel projections may be provided by interpolation along both θ and τ.

The situation is simplified when the fan measurement is equiangular in both α and β, with the increments $\Delta\alpha = \Delta\beta = \Delta$, so that $\alpha_n = n\Delta$, $\beta_m = m\Delta$, and consequently

$$p_{(m+n)\Delta}(R\sin n\Delta) = r_{m\Delta}(n\Delta). \tag{9.35}$$

Thus, the $j = (m + n)$-th parallel projection consists of the ray integrals from all fan projections, the n-th sample from the $m = (j - n)$-th fan projection becoming the n-th ray integral $p_{j\Delta}(n\Delta)$. The obtained parallel projections are equiangularly distributed in $\theta \in < 0, \pi)$, but sampling of projections is obviously not equidistant because of the sin function involvement; interpolation on the τ axis is thus still necessary.

The rebinning and interpolation approach is often applied in practice as it enables utilization of the established algorithms (and possibly also specialized hardware) for reconstruction from parallel projections. However, the method is rather memory-intensive, as the complete set of measurement data from all projections must be concurrently kept in memory.

9.1.5.2 Weighted Filtered Back-Projection

When the storage requirements of rebinning are difficult to fulfill or even prohibitive, the direct reconstruction from fan projections

may be preferable. Cases of different measuring geometry are discussed in detail in [37]. Specifically, for the geometry according to Figure 9.12, a formula of modified back-projection, suitable for discrete realization of direct reconstruction from fan projections, may be briefly derived as follows, paraphrasing [37].

Starting from Equation 9.28 and substituting for $q(\ldots)$ from the last expression of Equation 9.27, we obtain

$$f(x,y) = \int_0^{\pi} \int_{-\infty}^{\infty} p_\theta(\tau)\, h(\tau' - \tau)\, \mathrm{d}\tau\, \mathrm{d}\theta, \tag{9.36}$$

where we interpret $h(\tau)$ as the impulse response of the band-limited filter (Equation 9.30); $\tau' = x\cos\theta + y\sin\theta$ is the distance of the ray, belonging to $p_\theta(\tau)$ and crossing (x,y), from the origin of coordinates — see Figure 9.13. Using polar coordinates ρ, φ of the point (x, y), i.e., $x = \rho \cos\varphi$, $y = \rho \sin\varphi$, τ' can be expressed as $\tau' = \rho\cos(\theta - \varphi)$. Transforming

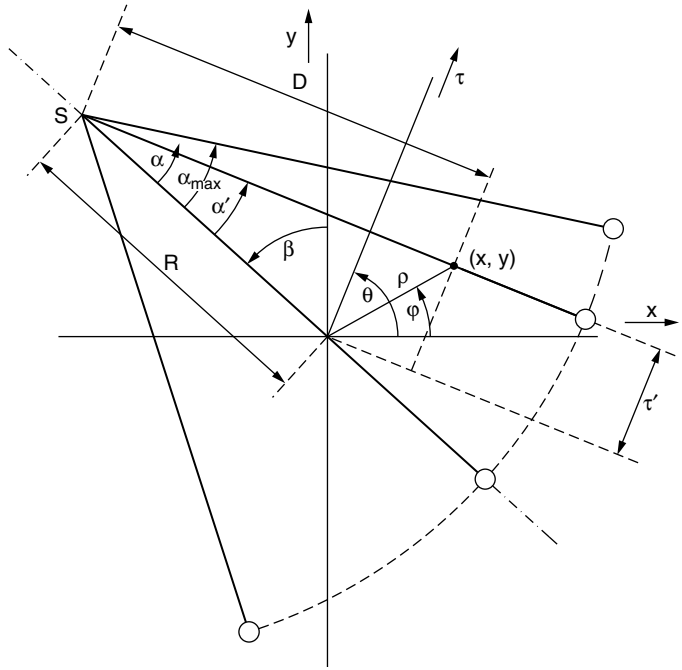


Figure 9.13 Geometry of fan projection — details.

the integral with the use of Equation 9.33, into the coordinates α, β (note that $\mathrm{d}\tau\mathrm{d}\theta = R\cos\alpha\,\mathrm{d}\alpha\mathrm{d}\beta$) yields

$$f(x,y)=\int_0^{\pi}\int_{-\alpha_{\max}}^{\alpha_{\max}} p_{\alpha+\beta}(R\sin\alpha)\; h(\rho\cos(\alpha+\beta-\varphi)-R\sin\alpha)\,R\cos\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta. \tag{9.37}$$

Instead of infinite α-limits, we introduced some finite $\pm\alpha_{\max}$, guaranteeing that any contribution outside of these limits is zero. Such limits, practically respected by the fan width, always exist thanks to finite size of the imaged object. The limits of the outer integral need not be shifted, as the function under the integral is periodical with the period π.

Utilizing Equation 9.34 and the relation $\rho\cos(\alpha+\beta-\varphi) - R\sin\alpha = D\sin(\alpha'-\alpha)$ that can be derived from Figure 9.13*, we obtain

$$f(x,y)=\int_0^{\pi}\int_{-\alpha_{\max}}^{\alpha_{\max}} r_{\beta}(\alpha)\; h(D\sin(\alpha'-\alpha))\,R\cos\alpha\,\mathrm{d}\alpha\,\mathrm{d}\beta. \tag{9.38}$$

Considering that $h(\tau)=\mathrm{FT}^{-1}\{H(w)\}$, and substituting $w' = wD(x,y,\beta)\,(\sin\alpha)/\alpha$ into the inverse FT integral, it can be proved that

$$h(D\sin\alpha)=\frac{1}{D^2(x,y,\beta)}\left(\frac{\alpha}{\sin\alpha}\right)^2 h(\alpha). \tag{9.39}$$

Denoting $\tilde{r}_{\beta}(\alpha)=r_{\beta}(\alpha)\,R\cos\alpha$ and $\tilde{h}(\alpha)=(\frac{\alpha}{\sin\alpha})^2 h(\alpha)$, we finally obtain

$$\begin{aligned} f(x,y)=\tilde{f}(D,\alpha') &= \int_0^{\pi}\frac{1}{D^2}\int_{-\alpha_{\max}}^{\alpha_{\max}}\tilde{r}_{\beta}(\alpha)\;\tilde{h}(\alpha'-\alpha)\,\mathrm{d}\alpha\,\mathrm{d}\beta \\ &= \int_0^{\pi}\frac{\tilde{q}_{\beta}(\alpha'(x,y,\beta))}{D^2(x,y,\beta)}\,\mathrm{d}\beta, \end{aligned} \tag{9.40}$$

*Note that (D, α') are the polar coordinates of the point (x, y) relative to vertex S; both D and α' are determined by (x, y) and β.

where the *modified and filtered projections* $\tilde{q}_\beta(\alpha)$ are

$$\tilde{q}_\beta(\alpha) = \tilde{r}_\beta * \tilde{h}\,|(\alpha) = R\cos(\alpha)\,.\,r_\beta(\alpha) * \tilde{h}\,|(\alpha). \tag{9.41}$$

The algorithm of the discrete *weighted filtered back-projection* is similar to, though somewhat more complicated than, the filtered back-projection for parallel projections (Section 9.1.4). It may be formulated as follows (if $\Delta\alpha = \Delta\beta = \Delta$):

- Each discrete fan projection is modified as $\tilde{r}_\beta(n\Delta) = r_\beta(n\Delta)R\cos n\Delta$.
- The modified projection is filtered, via DFT or direct convolution, by the modified filter according to Equation 9.41, thus obtaining $\tilde{q}_\beta(n\Delta)$.
- The projection is back-projected according to Equation 9.40, with the integral approximated by the sum

$$f(x,y) = \tilde{f}(D,\alpha') \approx \Delta\sum_{m=1}^{L} \frac{1}{D^2(x,y,\beta)}\tilde{q}_\beta(n\Delta). \tag{9.42}$$

We considered here only points (x, y) laying on rays $\alpha' = n\Delta$, $n = 1, 2, \ldots, J$.

The contribution $\Delta\frac{1}{D^2}\tilde{q}_\beta(n\Delta)$ from a single modified and filtered fan-projection to all points laying on such an n-th ray is identical along the ray, except that it is obviously weighted by the inverse square distance of (x,y) from S. Thus, the back-projection means smearing the projection values along the diverging rays of the fan while weighting them with $1/D^2$. However, the image itself is discretized; then the contributions must be redistributed among individual square pixels incident with the diverging stripes that approximate the rays, as roughly indicated in Figure 9.14. The distribution may be organized similarly as in back-projection of parallel projections, based on relative intersecting areas of individual pixels with the stripes. Alternatively, the position (x,y) of an individual pixel may determine the angle α' in the fan projection and the needed $\tilde{q}_\beta(\alpha')$ may then be found by interpolation among the values $\tilde{q}_\beta(n\Delta)$.

The weighted filtered back-projection has the same advantage as the filtered back-projection (Section 9.1.4), consisting of the immediate use of each individual projection once it is available, though its processing is more complicated than in the case of parallel projection imaging. Again, there is no need to keep the projections saved any

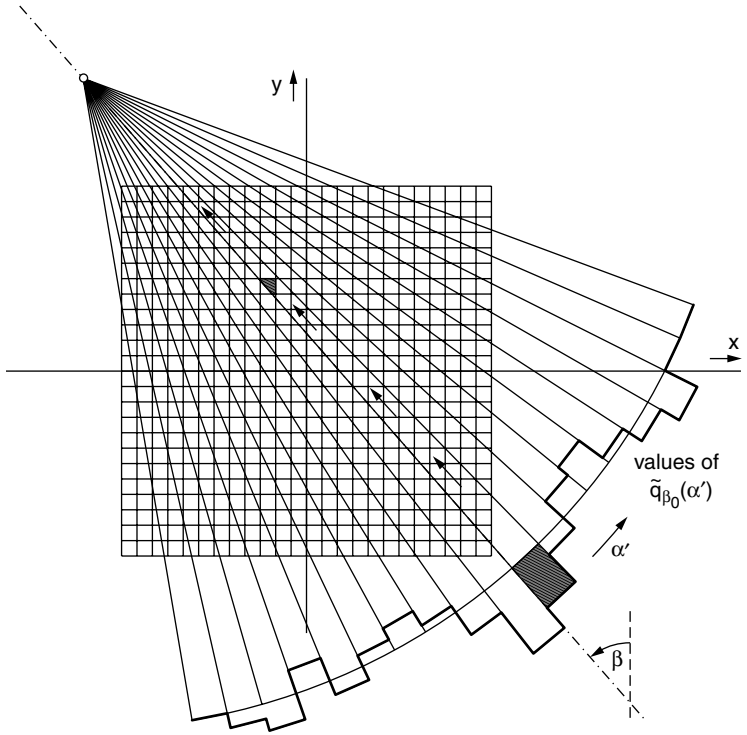


Figure 9.14 Discrete weighted back-projection of modified and filtered fan projections.

further once they have been exploited, so that the memory requirements are lower than with other methods.

9.1.5.3 Algebraic Methods of Reconstruction

As already mentioned in Section 9.1.2, the algebraic methods of reconstruction from projections can be applied to the fan projections as well, without conceptual changes. The fan measurement geometry will only change the weight coefficients, given by the intersection areas of the discrete square pixels, with the now diverging stripes approximating the diverging rays (Figure 9.15). With the system matrix **W** modified this way, the reconstruction method remains in principle the same as in Section 9.1.2. The same iterative methods to solve the equation system may be used,

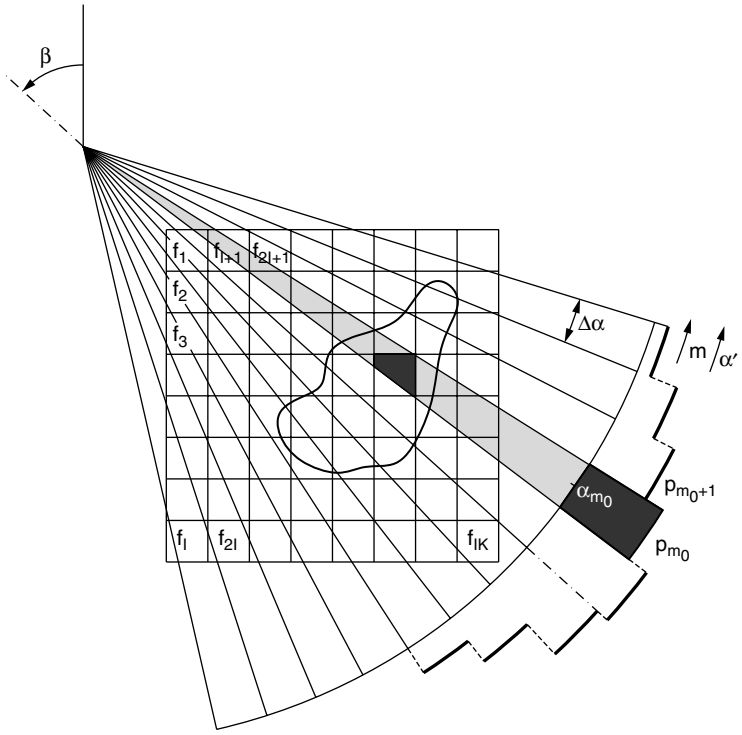


Figure 9.15 Discretization of the projection scene with fan projection.

including the iterative reprojection-based correction according to Equation 9.16.

The algebraic methods for the fan projection case have similar advantages and disadvantages in comparison with the weighted and filtered back-projection, as mentioned for the algebraic reconstruction from parallel projections vs. the filtered back-projection.

9.2 RECONSTRUCTION FROM NONIDEAL PROJECTIONS

9.2.1 Reconstruction under Nonzero Attenuation

In some of the tomographic modalities, the attenuation influencing the measured ray integrals must be considered, as is the case with SPECT (Section 6.2, Equations 6.4 and 6.5) or PET (Section 6.3).

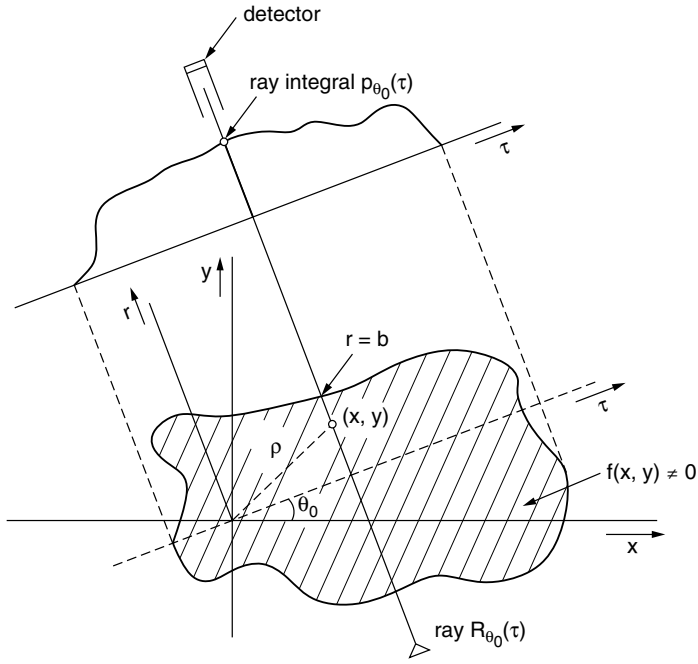


Figure 9.16 Geometry of projection measurement under general attenuation.

In both cases, the spatially dependent parameter, of which the distribution $f(x, y)$ is to be imaged, is the radioactivity distribution. The emanating radiation is attenuated on the way to the detecting device due to interaction with the object matter; this is now a disturbing phenomenon (on the difference from x-ray CT, where the attenuation is the imaged parameter). Let the spatial distribution of the attenuation coefficient on the imaged slice be denoted $\tilde{\mu}(x,y) = \mu(r,\tau,\theta)$, see Figure 6.8 and Figure 9.16.

Although physically similar, the two gammagraphic modalities represent different situations from the reconstruction point of view. In SPECT, a single detector situated at one end of the measured ray detects the integral activity on the ray, so that contributions from more distant locations are more attenuated when passing through the imaged object between the source and the detector. PET measurement also yields the integral ray activity, but thanks to a

different measuring methodology, the contribution from any position on the ray is attenuated equally, by the constant attenuation of the total ray path inside the body.

9.2.1.1 SPECT Type Imaging

The more complicated case (corresponding to SPECT) can be modeled as in Figure 9.16. An elementary source of radiation is supposed to be at $(x, y) = (\rho, \varphi)$. The ray $R_\theta(\tau)$ crossing this point ends in the ray detector. Thus, the radiation from the source $f(x, y)$ is attenuated on the path between the source position (x,y) and the detector; however, only the partial path between (x,y) and the object surface at $r = b(\tau,\theta)$ matters, as the attenuation outside of the object (in air) may be neglected. The r-coordinate of the point (x, y) is, according to Equation 9.5, $r_{xy} = -x\sin\theta + y\cos\theta$; the total attenuation exponent of the contribution is then

$$a(\tau,\theta,r_{x,y}) = \int_{r_{xy}}^{b(\tau,\theta)} \mu(r,\tau,\theta)\,\mathrm{d}r. \tag{9.43}$$

The actual output of the detector is obtained by adding the attenuation factor, in which the lower limit of the integral is obviously r, to Equation 9.6:

$$\begin{aligned} p_\theta(\tau) &= \int_{-\infty}^{b} f(\tau\cos\theta - r\sin\theta,\ \tau\sin\theta + r\cos\theta)\ e^{-a(\tau,\theta,r)}\mathrm{d}r \\ &= \int_{-\infty}^{b} f(\tau\cos\theta - r\sin\theta,\ \tau\sin\theta + r\cos\theta)\ \exp\left[-\int_{r}^{b(\tau,\theta)} \mu(r',\tau,\theta)\,\mathrm{d}r'\right]\mathrm{d}r. \end{aligned} \tag{9.44}$$

The derivation does not require convexity of the object contour; in the nonconvex case, $\mu(r)$ is zero outside of the contour. This is the model of projection under generic attenuation, clearly much more complicated than in the plain-projection case. The complications are of twofold origin: primarily, the formula with the double integral is itself very difficult to invert; moreover, the function $\mu(x, y)$ is usually unknown and would have to be measured separately, in fact via transmission imaging. The determination of $\mu(x,y)$ also yields the surface description $b(\tau, \theta)$, which nevertheless can be determined in a simpler way.

It is evident that the opposite projections (for θ differing by 180°) would differ due to the attenuation involvement. Although theoretically it should still be possible to reconstruct the image from 180° projections, it is customary to provide full-rotation scans, as this may increase the signal-to-noise ratio (SNR) in the resulting image substantially (besides possibly improving the lateral resolution due to better definition of the rays). Heuristically, the opposite projections have been used to produce arithmetic or geometric means as supposedly better estimates of attenuation-corrected projections; obviously, attempts like this work only approximately.

Exact inversion of Equation 9.44 is unknown in a closed form. Thus, only a simplified case of uniform attenuation $\mu(x, y) = \mu$, leading to

$$p_\theta(\tau) = \int_{-\infty}^{b} f(\tau\cos\theta - r\sin\theta,\ \tau\sin\theta + r\cos\theta)\ e^{-\mu[b(\tau,\theta)-r]}\,\mathrm{d}r, \quad (9.45)$$

has been analytically investigated; this expression is called the *exponential Radon transform*. The contour of the object in the slice has to be convex, to guarantee that $(b(\tau, \theta) - r)$ represents a continuous section of the ray. The derivation of the inverting formula is beyond the scope of this text (see, e.g., [37] and references therein). It has again the back-projection form of filtered and modified projections,

$$f(x,y) = \int_0^{2\pi} m(x,y,\theta,\mu)[p_\theta(\tau) * h(\tau,\mu)]\mathrm{d}\theta, \quad (9.46)$$

where both the filter $h(\ldots)$ and the exponential modification factor $m(\ldots)$, applied to each measured projection before back-projecting, are dependent on the attenuation. Even in this simplified case, it is necessary to know, as an input, the border function $b(\tau, \theta)$ that may be determined, e.g., from the initial estimate of $f(x, y)$ neglecting attenuation. The validity of the reconstruction with the assumption of uniform attenuation is naturally limited to situations where it can be justified (e.g., in abdominal imaging); in more complicated cases (e.g., cardiac SPECT), the nonuniformity of attenuation should be taken into consideration.

Conceptually simpler is the attenuation compensation when reconstructing the image by iterative discrete methods, basically similar to those described in Section 9.1.2. Discretization of the

image is identical (see Figure 9.3); the projection model taking into account the attenuation is obtained by modifying each term in the sum (Equation 9.10) with a factor representing the attenuation, similarly as in the continuous formulation (Equation 9.44)

$$p_m = \sum_{n=1}^{N} \exp(-a_{mn}) w_{m,n} f_n = \widetilde{\mathbf{w}}_m \mathbf{f}. \tag{9.47}$$

Here, a_{mn} is the total attenuation exponent for the m-th pixel seen from the m-th direction (direction of the m-th projection integral)

$$a_{m,n} = \sum_{k \in K_m} w_{m,k} \mu_k \tag{9.48}$$

where K_m is the pixel set containing the half ray between the n-th pixel and the detector and not containing the other half ray, and μ_k is the attenuation coefficient in the range of the k-th pixel. Obviously, there is no limit on the character of the distribution $\tilde{\mu}(x, y)$.

The attenuation map must be known in advance as an input to the procedure. Then it is possible to calculate the coefficients $a_{m,n}$ at the same algorithm stage as the weights $w_{m,n}$; it is evidently a computationally intensive procedure that, should all the coefficients $\exp(-a_{m,n})\, w_{m,n}$ be saved, becomes as memory-intensive as the algorithms discussed in Section 9.1.2. Thus, the iterative reprojection algorithm mentioned in the last part of Section 9.1.2 may offer an alternative to the direct inversion of the linear system, modified from Equation 9.11.

When the attenuation map is unknown beforehand, it is possible to start the reconstruction procedure by first estimating an approximate reconstruction providing $\mu = 0$. Consequently, based on the obtained approximate image, assigning estimated values of μ according to expected types of tissue is possible, followed by repeated reconstruction taking into account this estimated attenuation map. The obtained image may become the base for further iterative improvement. This approach is particularly suitable, when the attenuation may be expected uniform on, and around, the slice of the imaged object. The constant attenuation coefficient is then assigned to all pixels of the auxiliary image that correspond to inner pixels of the object in the preliminarily reconstructed image $f(x, y)$.

9.2.1.2 PET Type Imaging

The compensation of attenuation influence in PET images is, at least conceptually, substantially simpler than in the previous case.

As was derived in Section 6.3.3, the attenuation phenomenon influences every contribution to a ray integral by the *total* attenuation along the ray, independently of the position of the elementary radiation source on the ray. All the contributions to a ray integral are therefore to be corrected equally. The attenuation along all the rays used in the imaging process has to be known beforehand, as in the previous case; however, this can be provided in advance by simple measurement without the need to reconstruct the attenuation map. As the needed attenuation values are only ray (and not pixel) specific, only the resulting ray integrals (and not the integrands along the rays) have to be individually corrected with the corresponding factors before the reconstruction from projections starts. The reconstruction is then based on the corrected near-ideal projections and all the reconstruction approaches of Section 9.1 apply.

9.2.2 Reconstruction from Stochastic Projections

So far, the projections were considered deterministic functions, derived algorithmically from the function $f(x, y)$. Though there is always a certain level of noise (e.g., thermal, measuring circuitry noise, etc.), with the modalities like CT the model of deterministically derived projections containing additive noise of a low level is adequate. Although the measured quantities are, exactly taken, formed by flow of individual photons, the obtained measurements approximate very well the mean values of flow thanks to very high numbers of photons involved.

The situation is substantially different in modalities like SPECT or PET, where the measurement of ray integrals consists of counting individual photons and the counts remain in low numbers. Under such conditions, the relative variance of the measured values is high and the obtained projections must be considered random signals — realizations of a one-dimensional stochastic process. Reconstruction from such noisy projections by classical methods based on the mentioned deterministic model would yield low-quality images, if at all possible; the equations obtained from such measurements are usually rather inconsistent, so that the straightforward solution may even fail. In this section, we shall describe the concept of reconstruction from randomly valued projections, considering a stochastic model of the measured signals and incorporating this approach to the reconstruction procedure. In both mentioned modalities, the model is based on Poisson distribution that governs photon emission and transmission;

we shall concentrate on this case, following mainly the ideas of [8], [26], [44], [58]. However, the principle of maximum-likelihood solution that will be briefly formulated applies much more generally.

9.2.2.1 Stochastic Models of Projections

The number of photons (n), emitted from a radioactive source in a (short) period toward the detector, is a random variable with Poisson distribution, giving the probability P of just n photons appearing as

$$P(n) = \frac{\bar{n}^n \, e^{-\bar{n}}}{n!} \tag{9.49}$$

where $\bar{n} = \mathsf{E}(n)$ denotes the mean value of n. The variance of this distribution is known to be $\sigma^2 = \bar{n}$, which means a high relative variance. When there is an attenuating medium between the source and the detector, only a part of the photons emitted in the direction of the detector reach the detector and are registered, as given on average by the attenuation law

$$\frac{\bar{n}_d}{\bar{n}} = e^{-a}, \quad a = \int_R \mu(r)\mathrm{d}r. \tag{9.50}$$

Here, $\bar{n}_d$ is the mean of the count of detected photons and a is the total attenuation on the ray R between the source and the detector. However, the event of an individual photon reaching the detector is again stochastic. It can be shown that the resulting probability distribution of the number of detected photons n_d is again Poisson, with a certain mean $\bar{n}_d$. The value expected to be measured is just $\bar{n}_d$; however, the real measured quantity is random and may differ substantially from the mean. Note that in the previous sections of this chapter we counted on having access to the precise values of $\bar{n}_d$.

The ideal detector output—the *mean value* of the ray integral along the ray R—is, according to Equation 9.44,

$$\bar{p}_R = \int_R f(r) \exp\left[-\int_r^{r_D} \mu(r')\,\mathrm{d}r' \right] \mathrm{d}r, \tag{9.51}$$

where $f(\ldots)$ and $\mu(\ldots)$ are the source and attenuation distributions, respectively, and r_D is the detector coordinate on the ray. Both cases of transmission and emission tomography are included: transmission

tomography has a concentrated source outside of the imaged object at $r = r_S$, while in emission tomography the sources are distributed inside the object.

Although there is basically no difference in the measurement model between transmission tomography (like CT, though the randomness is important there only for very low-dose measurements) and emission tomography (SPECT, PET), the obtained data are processed differently in order to obtain different reconstructed images of $\mu(x, y)$ or $f(x, y)$, respectively. In transmission tomography, the outer integral simplifies so that the ray integral is

$$\bar{p}_R = c_R \exp\left[-\int_{r_S}^{r_D} \mu(r)\mathrm{d}r\right]. \tag{9.52}$$

On the other hand, in emission tomography, the attenuation effects are neglected in the first approximation (and compensated for later — see Section 9.2.1) so that the ray integral becomes

$$\bar{p}_R = c_R \int_R f(r)\mathrm{d}r. \tag{9.53}$$

In both cases, there is a ray-specific constant c_R, which encompasses the influence of the detector efficiency, measurement duration, and — in transmission measurement — also the source intensity. It may be supposed that these constants are determined ahead of the measurement (or simultaneously) by calibration; they are therefore known when processing the ray measurements. Comparison of Equations 9.52 and 9.53 shows that while the emission tomography provides directly the ray integrals in the linear form expected by reconstruction algorithms, log-transforming of each ray integral is needed in transmission tomography.

However, the measurements available in modalities suffering with quantum noise are far from the preceding ideal formulae. Primarily, instead of measuring the mean value $\bar{p}_R$, we always obtain only a realization $\tilde{p}_R$ of the random variable, which may substantially differ from the mean. Also, the measurements in both cases are suffering with a systematic (and possibly ray-specific) error e_R due to disturbing phenomena (scatter, random coincidences in PET, etc.); this error is non-negative and adds to the proper value. Fortunately, these errors may be supposed to be also known from a calibration preceding the tomographic measurement, so that they can be subtracted from the measurements before further

processing. Thus, in the transmission case, the estimated (log-transformed) ray integral as needed for the reconstruction is

$$p_R = -\log \frac{\tilde{p}_R - e_R}{c_R}, \tag{9.54}$$

while in emission tomography it is

$$p_R = \frac{\tilde{p}_R - e_R}{c_R}. \tag{9.55}$$

A result becomes physically unacceptable when the difference is nonpositive due to the presence of large e_R, which may happen frequently due to large variance of $\tilde{p}_R$. In such a case, the obtained values should be replaced by substitutes, e.g., interpolated from neighboring rays.

9.2.2.2 Principle of Maximum-Likelihood Reconstruction

Maximum-likelihood (ML) estimation concerns the situation, when a realization $\tilde{y}$ of a (possibly vector-valued) stochastic variable y is known. The shape (function type) of the probability distribution, which controls the variable, is supposed known except for a (possibly vector-valued) parameter θ of the distribution. For each value of θ, the probability function determines the probability of the particular realization $\tilde{y}$; thus, a *likelihood function* $L(\tilde{y},\theta)$ may be formulated that yields the probability of the given realization $\tilde{y}$ for each θ. The task is to find the value of θ corresponding to the maximum of the function $L(\theta)$; this optimum parameter value is called the *ML estimate*. Obviously, we can also reformulate the definition: the ML estimate maximizes the conditional probability $P(\tilde{y}\,|\,\theta)$ by suitably choosing θ.

When $L(\tilde{y},\theta)$ is differentiable with respect to θ, the maximum (not at the boundary) is at θ_{opt}, fulfilling

$$\left.\frac{\partial L(\tilde{y},\theta)}{\partial\theta}\right|_{\theta_{\text{opt}}} = 0. \tag{9.56}$$

As $L(\tilde{y},\theta)$ is non-negative and the logarithm function is strictly monotonous, the same maximizer is obtained by

$$\left.\frac{\partial \log L(\tilde{y},\theta)}{\partial\theta}\right|_{\theta_{\text{opt}}} = 0, \tag{9.57}$$

which often simplifies the expressions.

Estimating the image vector $\mathbf{f}$ based on the stochastic projection vector $\tilde{\mathbf{p}}$ is a problem of this kind.

When discretizing the scene as in Figure 9.3, the ray integral is approximated as in Equation 9.10, $p_m = \Sigma_{n=1}^{N} w_{m,n} f_n = \mathbf{w}_m^T \mathbf{f}$, and the problem may be described as in Equation 9.11, $\mathbf{Wf} = \mathbf{p}$ in both variants of modalities. The substantial distinction in comparison with the previous sections is that the right-hand-side measurement vector cannot be considered exact, but rather a realization of a random vector, with a large variance. Instead of using the iteration algorithms, discussed above, which would lead to a very noisy image, the problem should be defined and treated differently.

The problem can be stated as follows: let the spatial distribution of the imaged parameter, $\mu(x, y)$ or $f(x, y)$, be denoted $\eta(x, y)$. In the discrete approximation, it is represented by the vector $\boldsymbol{\eta}$ to be determined. The measurement has yielded the stochastic vector $\tilde{\mathbf{p}}$ consisting of $\tilde{p}_R, \forall R$, each with the Poisson distribution. In the sense of ML method, the estimate $\hat{\boldsymbol{\eta}}$ of the image vector $\boldsymbol{\eta}$ is sought, which maximizes the probability of measuring the $\tilde{\mathbf{p}}$ given $\boldsymbol{\eta}$, i.e.,

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} L(\boldsymbol{\eta}), \quad \boldsymbol{\eta}_n \geq 0, \forall_n \tag{9.58}$$

where $L(\boldsymbol{\eta})$ is the likelihood function

$$L(\boldsymbol{\eta}) = \log P\{\tilde{\mathbf{p}} \mid \boldsymbol{\eta}\}. \tag{9.59}$$

The non-negativity constraint is physically obvious.

The elements $\tilde{p}_R$ of $\tilde{\mathbf{p}}$ are mutually independent so that, when each of them is Poissonian,

$$P\{\tilde{\mathbf{p}} \mid \boldsymbol{\eta}\} = \prod_{r=1}^{M} P\{\tilde{p}_r \mid \boldsymbol{\eta}\} = \prod_{r=1}^{M} \frac{[\bar{p}_r(\boldsymbol{\eta})]^{\tilde{p}_r} \; e^{-\bar{p}_r(\boldsymbol{\eta})}}{\tilde{p}_r!} \tag{9.60}$$

and consequently,

$$L(\boldsymbol{\eta}) = \log P\{\tilde{\mathbf{p}} \mid \boldsymbol{\eta}\} = \sum_{r=1}^{M} -\bar{p}_r(\boldsymbol{\eta}) + \tilde{p}_r \log \bar{p}_r(\boldsymbol{\eta}) - \log(\tilde{p}_r!). \tag{9.61}$$

It is possible to evaluate this criterion function for any given $\tilde{\mathbf{p}}_r$ and chosen $\boldsymbol{\eta}$: the exact ray integral $\bar{p}_r(\boldsymbol{\eta})$ can be evaluated using the discrete geometrical model, and consequently, as $\tilde{p}_r$ values are known from measurement, the complete sum can be calculated.

In the optimization, the last term in each summed expression may be omitted as independent on $\boldsymbol{\eta}$. Using a suitable optimization strategy, the optimum (Equation 9.58), representing the (unconstrained) ML estimate of the image, can be in principle found. Note that besides taking into account the probabilistic nature of the problem, the model of measuring projections is still present in calculating $\bar{p}_r(\boldsymbol{\eta})$.

However, the unconstrained optimization would not yield reasonable images as the problem is ill-conditioned— roughly speaking, the maximum of $L(\boldsymbol{\eta})$ is not distinctive enough. Thus, additional constraints are needed to better specify the choice of $\boldsymbol{\eta}$. There are many such *regularization* approaches, introducing the needed constraints, all basically aiming at suppressing the noise in the image. The main idea is to introduce a measure of noise $R(\boldsymbol{\eta})$ in the image and minimize it while simultaneously maximizing the likelihood function, or alternatively, set a limit to the noise measure. The latter task is thus formulated as a constrained restoration problem, similar to restoration problems that will be more systematically treated in Chapter 12.

While the second mentioned approach means constrained optimization, the former one leads to a compromising formulation of the optimization criterion, combining the likelihood function and the noise measure. Thus the problem becomes [8]

$$\hat{\boldsymbol{\eta}} = \arg\max_{\boldsymbol{\eta}} \Phi(\boldsymbol{\eta}), \qquad \eta_n \geq 0, \forall n, \qquad \Phi(\boldsymbol{\eta}) = L(\boldsymbol{\eta}) - \beta R(\boldsymbol{\eta}). \tag{9.62}$$

The parameter β enables the choosing of a degree of noise suppression at the cost of a certain loss in spatial resolution; a reasonable trade-off should be found. The noise measure $R(\boldsymbol{\eta})$ (also called *penalty function*) is usually based on the idea that the reconstructed image is a piece-wise smooth function so that, except for distinct edges, the areas in the image should provide only a low contribution to a selected difference operator, e.g.,

$$R(\boldsymbol{\eta}) = \sum_{i=0}^{N} \sum_{k=0}^{N} \alpha_{i,k} \psi(\eta_i - \eta_k) \tag{9.63}$$

where $\alpha_{i,k}$ are chosen weights and the function $\psi(\ldots)$ represents the cost of the differences (e.g., absolute value or square).

Different optimization strategies and corresponding reconstruction algorithms for both transmission and emission modalities are discussed in detail in [8] and references therein.

9.3 OTHER APPROACHES TO TOMOGRAPHIC RECONSTRUCTION

Two of the most frequently used tomographic modalities, namely, MRI and ultrasonography, are not commonly based on projection measurements, and consequently, the methods of image reconstruction from the raw data are different. In the following paragraphs, we shall briefly mention the methods of reconstructing the tomographic slices in these modalities.

9.3.1 Image Reconstruction in Magnetic Resonance Imaging

The image data reconstruction in MRI is based on the model of raw (base-band) MR signal (Equation 5.20),

$$S(t) \propto \iiint_V M_{xy}(x, y, z, t)\, e^{-j(k_x(t)x + k_y(t)y + k_z(t)z)} \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z,$$

which is obviously a three-dimensional Fourier transform of the instantaneous spatial distribution of the transversal magnetization M_{xy} to be imaged. As already mentioned in Chapter 5, the instantaneous values of the signal thus correspond to the spectral components of the instantaneous magnetization scene at the spatial frequencies $k_x(t)$, $k_y(t)$, and $k_z(t)$, i.e., the components of the **k**-term (Equation 5.14),

$$\mathbf{k}(t) = \gamma \int_0^t \mathbf{G}^T(\tau) \mathrm{d}\tau.$$

It has been described in Chapter 5 how the gradient field $\mathbf{G}(t)$ can be controlled in order to provide gradually a sufficiently dense description of the complete spectral (**k**-)space. This may be organized in two different ways.

9.3.1.1 Projection-Based Reconstruction

As shown in Section 5.4.4, when only a thin slice of the object has been excited, if time dependence of the magnetization $M_{xy}(\mathbf{r})$ during readout time is negligible, and when a perpendicular x-gradient is switched on for the period of measurement, then the above model

can be simplified to a two-dimensional integral (Equation 5.23) over the slice area A (i.e., field of view (FOV)),

$$S(t) \propto \Delta z \iint_A M_{xy}(x,y)\, e^{-jk_x(t)x}\,\mathrm{d}x\,\mathrm{d}y$$

and further to (Equation 2.25)

$$S(t) \propto \Delta z \int_{D_x} P_y(x) e^{-j\gamma G_x t\, x}\,\mathrm{d}x\,,$$

as the integral along y means the projection. The expression is obviously the one-dimensional Fourier transform of the projection, where the spatial frequency is $k_x(t) = \gamma\, G_x t$; in other words, the instantaneous value of $S(t)$ provides the value of the one-dimensional FT at the space frequency $\gamma\, G_x t$, with the range of spatial frequencies corresponding to the time of measurement according to Equation 5.26. Other imaging parameters (spatial resolution, image size), as well as the needed sampling frequency, have been derived and presented in Equations 5.27 to 5.30. By a single RF irradiation, the signal provides the one-dimensional spectrum of a single projection, perpendicular to the gradient. This image projection could easily be recovered by inverse Fourier transform. As the gradient field of any other direction in the (x,y)-plane can be formed as $\mathbf{G}_{xy} = G_x\mathbf{u}_x + G_y\mathbf{u}_y$, the spectra of arbitrarily oriented projections in the plane can be provided, with the previous expressions valid after simple rotation of coordinates.

Repeating the measurement sequence for different gradient directions may thus provide a number of projections sufficient for any method of reconstruction from projections. As one-dimensional spectra of projections are directly provided by the measurement, it is advantageous to use a method that would otherwise require computing of these spectra from projections as the first step. This is obviously the frequency-domain reconstruction based on the projection slice theorem (Section 9.1.3); the disadvantage of the need to interpolate the spectral values from the polar grid into the rectangular grid, as required for the inverse two-dimensional DFT, naturally remains. Alternatively, the filtered back-projection using the projection filtering via frequency domain (Section 9.1.4) may be used with a similar advantage.

9.3.1.2 Frequency-Domain (Fourier) Reconstruction

Modern MRI systems use mostly the direct reconstruction of image data from the spectral representation, without introducing the concept

of projections. However, the two-dimensional imaging remains based on a slice excitation, as in the previous paragraph. As shown in Section 5.4.5, the line of the spectrum, obtained as a single echo, may be shifted from the central position in the spectral domain in the direction of k_y by applying the time-invariable y-gradient during a period T_{ph} before echo readout. This way, the $k_y(t)$ value is adjusted to a constant $\gamma G_y T_{ph}$ valid during the complete signal acquisition. The signal model then becomes (Equation 5.32),

$$S(t) \propto \Delta z \iint_A M_{xy}(x,y) e^{-j\gamma(G_x tx + G_y T_{ph} y)} \mathrm{d}x\, \mathrm{d}y,$$

which is the two-dimensional Fourier transform of M_{xy}. Changing either the magnitude or the duration of the y-gradient in each measuring sequence shifts the obtained spectral-domain line, which enables filling in of the k-space gradually (sc., *phase encoding*).

When the k-space is filled with the spectral values, the image data are obtained by simple inverse two-dimensional DFT. In this case, no interpolation of spectral values is usually needed, as the sampling grid is naturally rectangular.

An example of experimental nuclear magnetic resonance (NMR) data from a laboratory MRI system and the reconstructed image are shown in Figure 9.17.

Three-dimensional reconstruction is a direct generalization of the two-dimensional approach. It differs primarily in the excitation that covers a substantial volume, i.e., even along the z-direction; the thin slice excitation is not applied. The spectral values are provided, as described in Section 5.4.6 via frequency and double phase encoding, which means that the signal model (Equation 5.36) remains three-dimensional,

$$S(t) \propto \iiint_V M_{xy}(x,y,z) e^{-j(G_x tx + G_y T_{ph} y + G_z T_{ph} z)} \mathrm{d}x\, \mathrm{d}y\, \mathrm{d}z\,.$$

Two of the spatial frequencies are fixed during each echo measurement by applying two perpendicular gradients before the readout. Each row in the k-space (now three-dimensional matrix) is again provided by a single echo, so that to fill in all the rows, gradually all the combinations of the discrete phase factors $G_y T_{ph}$ and $G_z T_{ph}$ must be formed before reading out of each individual echo (sc., *double phase encoding*).

When the three-dimensional discrete spectrum is provided, again on a rectangular grid, it suffices to apply the inverse three-dimensional DFT to obtain the three-dimensional image data.

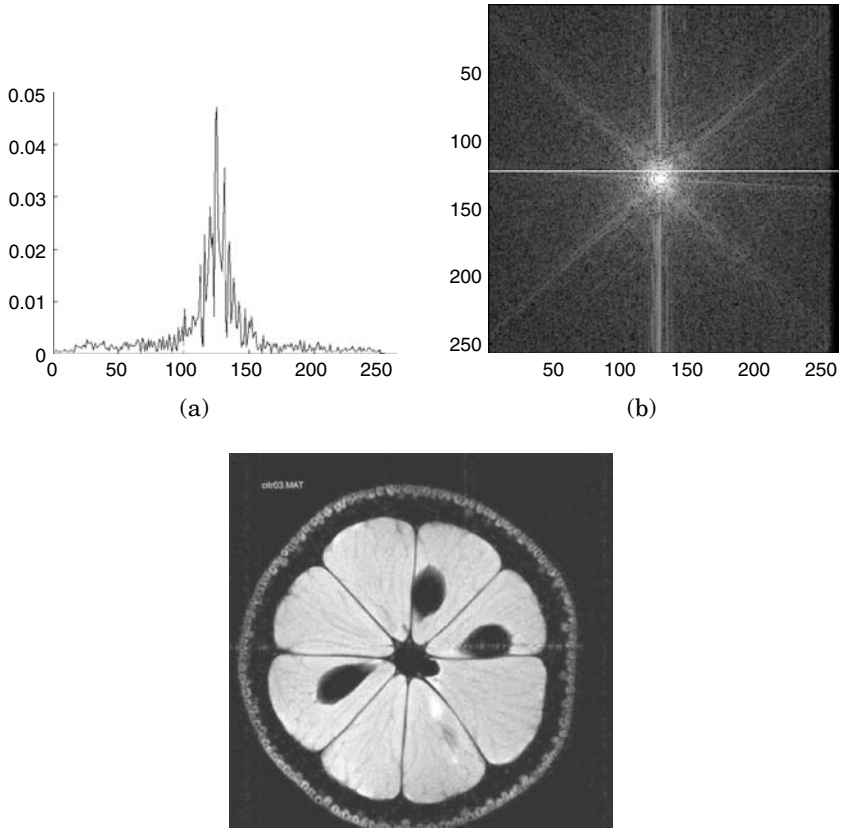


Figure 9.17 Example of experimental MRI data and reconstruction: (a) measured RF signal, i.e., a gradient-controlled echo forming the indicated line in the k-space (b); and (c) image reconstructed via two-dimensional DFT. (Courtesy of Institute of Scientific Instruments, Academy of Sciences of the Czech Republic Brno, Assoc. Prof. M. Kasal, Ph.D.)

The classical phase encoding, when each echo requires a new excitation, is not the only way of providing the two-dimensional or three-dimensional spectra. As shown in Section 5.4.7 on *fast MRI*, faster ways exist to provide the spectral-domain description during a few or even only a single excitation. However, the result — the two-dimensional or three-dimensional spectrum in the k-space — is basically the same. The only difference may be that in favor of a simpler gradient forming, the sequence of spectral values would not follow

the lines on the needed rectangular grid; consequently, interpolation in the k-domain is necessary before the inverse two-dimensional or three-dimensional DFT is executed. With computational hardware development, it becomes gradually advantageous to aim at maximizing the achievable speed of imaging by simplifying the measuring sequences, and to leave the burden of processing possibly more complex data on the computing equipment.

9.3.2 Image Reconstruction in Ultrasonography

9.3.2.1 Reflective (Echo) Ultrasonography

Standard ultrasonography utilizes the principle of receiving the responses to a short ultrasonic impulse sent along a narrow directional ray of known orientation, as described in Section 7.1. The responses are generated due to inhomogeneities of acoustic impedance in the imaged objects (tissues) that cause primarily diffraction and scattering of the ultrasonic waves, exceptionally also specular reflections. With respect to very complicated nonlinear phenomena causing the responses, it is impossible to define what exactly the measured and imaged parameters are. The responses are mostly evaluated only as to their amplitude, which is interpreted in terms of gray scale. The imaging principle is simple, and the reconstruction of the image is straightforward, as described in Section 7.1.2. The only necessary computational reconstruction concerns the format conversion, which is needed when the formats of the display device (usually rectangular) and ultrasonic scanner (often nonrectangular) differ. In the case of using a linear array probe, a simple conversion may be needed when the aspect of the displayed field on the monitor would differ from that of the FOV; then a simple one-dimensional interpolation suffices. When a fan probe is used, the data are provided in polar coordinates (r, ϑ) and the conversion is needed to prevent geometrical distortion of the image displayed on the Cartesian grid (x,y) of a standard monitor. The conversion formulae are (Equation 7.11)

$$r = \sqrt{x^2 + y^2}, \qquad \vartheta = \arctan\frac{x}{y}.$$

The corresponding elementary algorithms are described in Section 7.1.2.2. The only complication is due to the need to interpolate the real-time data, usually by bilinear backward interpolation (see Section 10.1.2).

The similar though more difficult problems of three-dimensional ultrasonographic data processing, needed to reconstruct the volume distribution of the response sources, are discussed in Section 7.3.2.

9.3.2.2 Transmission Ultrasonography

From the viewpoint of tomographic reconstruction, the transmission ultrasonography is more interesting (see, e.g., [37]). However, it is still rather in an experimental stage, as the application problems and practical complications are severe. The measuring arrangement is similar to the first-generation CT scanners (Figure 4.1) — the transmitting transducer on one side of the object and the receiving probe on the other side define the ray on which the ray integral of the chosen parameter is accumulated. This measuring couple is mounted on a frame that allows lateral shift of the ray, yielding a single projection via parallel rays per each frame position; the complete frame may be rotated in the range of 180 or 360° to provide projections at different angles. The main practical problem is in providing a reliable acoustic contact of both transducers with the imaged object, as the ultrasound does not propagate in air. Thus, the object is usually submerged in water, together with the measuring apparatus.

From the viewpoint of imaging properties, the difficulty is primarily in a rather fuzzy definition of the rays. Though this problem is met in the echo ultrasonography as well, it is more severe here, as the immediate correction (or rejection) of false data that is possible in B-mode scanning, thanks to the reaction of examining experienced staff, is missing. The ray may be repeatedly refracted, scattered, or reflected even to more directions, which may lead to a path substantially different from the expected ray. However, when only refraction is taken into consideration, it is in principle possible to compensate for the curved rays by inclusion of the estimated refraction values from the gradually reconstructed velocity image into the iterative reconstruction procedure.

There are basically two local acoustic parameters that can be imaged: the refractive index and the attenuation coefficient.

Ultrasonic *refractive index tomography* may be based on measuring the time of flight of an ultrasonic impulse along individual rays. The transit time is inversely proportional to the average ultrasound velocity and therefore also corresponds to the average of the refraction index. Measuring the transit time is relatively easy and precise.

Ultrasound *attenuation tomography* is more complicated, as there are difficulties with the unknown attenuation at the interfaces

to transducers. Another problem is due to frequency dependence of the attenuation, which is approximately linear in the range of frequencies used in medicine. When using narrowband signals, the attenuation would be defined approximately for the central frequency; with wideband signals, the frequency-independent attenuation parameter (in dB m^{-1} MHz^{-1}) is estimated. The direct attenuation measurement based on amplitudes of transmitted and received signals is not feasible due to the mentioned interface uncertainties. However, thanks to the known linear dependence, the attenuation may be estimated from the relation of the input and output spectra. (Roughly, the higher frequencies are more attenuated than the low ones, and the difference is greater, the higher the attenuation; thus, the absolute amplitudes are irrelevant as long as the interface problems are frequency independent.) The two methods described in [37] are based either on the comparison of two nonoverlapping frequency bands or on the shift of the mean frequency of a broadband signal. The latter method is also widely used in many other trials to determine the attenuation from the received ultrasonic signal, e.g., [33].

The published experimental results of ultrasonic transmission tomography are often impressive, but the practical obstacles prevent this modality from routine clinical use so far.

10

Image Fusion

This chapter deals with a relatively new direction in image processing that proves to be of high importance in many areas, not just in medical applications — obtaining qualitatively new information by combining the image data from several different images (or sequences of images). The term *image fusion* is used here in a very generic sense of synthesizing the new information based on combining and mutually complementing the information from more than a single image. Although we are dealing largely with two-dimensional images in this book, it should be understood that medical image fusion often concerns three-dimensional image data provided by tomographic modalities. However, the three-dimensional fusion is, disregarding rare exceptions, a conceptually simple generalization of the two-dimensional case.

The most common and obviously clinically needed is the *image registration* in its many variants: intramodality registration and intrapersonal (intrapatient) registration on one side, and intermodality and interpersonal registration on the other (the registration of measured image data to atlas images or to physical reality in the interventional radiology also belongs here). Registering images may

enable better comparison of the image information on the same area provided by different modalities, or by a single modality in the course of time, an assessment of anatomy differences inside a group of patients or volunteers, etc. In these cases, registration itself is the aim of processing, and further derivation of conclusions is left to the medical staff. However, even in this simplest case, fusing information from more images leads to a new quality.

Registration may also be the first step of a more complex processing leading to the automatic narrower-sense *image fusion* serving to derive new images. The simplest and perhaps most frequent example is the subtraction of two images (as in the subtractive angiography) in the hope that the difference image would reveal relevant dissimilarities; the preliminary registration is used to prevent enhancing of irrelevant differences due to field-of-view (FOV) shift, differing geometrical distortion, patient movement, etc. The registered images from different modalities may be fused into a single vector-valued image that may serve as a better ground for segmentation or another automatic analysis, or may be presented to the evaluating staff by means of, e.g., false colors or overlay techniques. On the other hand, fusion of single-modality multiview images covering contiguous parts of an object and overlapping at margins may enable the forming of images of substantially greater FOV; the fusion usually requires both geometrical and gray-scale (or color) transforms to enable reasonably seamless joining.

The registration in its simplest form — finding the best rigid transform, described in three dimensions by only six parameters of shift and rotation—is routinely used. However, in many instances, more sophisticated transforms are needed, allowing for compensation of complicated distortions, including multiview perspective, partial-area movement, area shrinking or expanding (possibly space variable), etc. The first step in such cases is to find an appropriate description of the geometrical relations between two (or possibly even more) images. The number of parameters describing a complicated registration transform may be very high—from tens or hundreds for complex flexible formulae to millions in the case of detailed disparity maps. It is obvious that such a parameter set carries information on the relations among the images to be registered (even though no factual registration needs to follow). Thus, the transform design itself may also be considered a fusing operation yielding output data, not available otherwise, which may be utilized in the following analysis. Depending on the geometrical transformation to be applied or on the way of further processing, the description needed as the

input to the transform design may concern only a few points selected manually or automatically, or, on the other hand, it may form a dense grid of vector values of the mutual local shifts—sc., *disparities*. Such a *disparity map* itself may therefore be considered another output of fusing the content of images. Its practical usage is manifold, from evaluation of a partial-area change of size and shift of the relevant borders (e.g., in growth or motion estimation) to automatic processing, yielding, e.g., a three-dimensional scene description based on the disparity map of a stereo image pair.

In this chapter, we shall follow the naturally sequenced steps appearing in image fusion:

- Preprocessing measures, leading in a sense to consistency of involved images, including geometrical transforms and image data interpolation as procedural components
- Distortion identification via disparity analysis, including description of similarity measures and possibly followed by some wider utilization of disparity maps
- Proper image registration as a preparatory step for consequent analysis or image fusion
- Proper image fusion (in the narrower sense) yielding a composite image of a new quality

The most important sources among those cited at the end of Part III are [16], [9], [52], [78], and also [59], [12], [66], [49], [18], [53], [10], [46], [27], [31], [45], [68].

10.1 WAYS TO CONSISTENCY

It is intuitively felt that the images, the information of which is to be combined, should be in a sense consistent. It depends on the character of the desired output information what the necessary consistency exactly means, but obviously, the mutually closer are the geometrical and other characteristics of the fused images, the more straightforward it is to combine them and analyze the similarities and differences, often carrying the desirable information.

The images should primarily be presented in similar geometrical format and coordinates; should the image sources provide the data in different coordinate systems, the first step would be the format conversion (e.g., for fan-scanned ultrasonographic data vs. any images provided in rectangular coordinates; see Section 9.3.2).

However, the images to be fused are usually geometrically distorted, due to different and often combined causes. It is obviously

the case when multimodal fusion is to be performed, as different modalities suffer from different geometrical distortions. Another source of geometrical distortion may be differing imaging geometry, namely in a multiview image set, or it may be due to motion of the object or its parts during acquisition of the image set, etc.

The crucial step that must precede the fusion is thus to align the data of the to-be-fused images so that the corresponding image features are identically positioned in the image matrices; this procedure is denoted as *image registration*. A high degree of accuracy and precision is required in registering medical images: imprecise registration would lead to a loss of resolution or to artifacts in the fused images, while unreliable and possibly false registration may cause misinterpretation of the fused image (or of the information obtained by fusion), with possibly even fatal consequences. When interpreting an image as mapping of reality, ideally, a point in the original physical space should, after registration, be represented by points of identical coordinates (i.e., identically indexed pixels or voxels in image matrices) in all images to be fused. It is usually not possible to achieve this ideal situation, so that tolerances as to the precision must be stated with respect to the fusion purpose and practical possibilities. We shall deal with the image registration in greater detail in Section 10.3; here, only the prerequisites needed for this goal will be treated.

The way to registration leads via *geometrical transforms* (Section 10.1.1) that provide for mapping the points of one (transformed) image to the coordinates of the other (base) image, thus registering the corresponding image elements. The used transforms may be classified into one of two groups. *Rigid transforms* do not change the internal geometrical characteristics of the transformed image and only shift and possibly rotate the complete image (generally three-dimensional image data set) to achieve the best possible alignment. *Flexible transforms*, on the other hand, allow a certain geometrical deformation to compensate for the image distortion with respect to reality if known, or at least to match the (unidentified) distortion of the other, s.c., base image. Usually, the latter is the case: with the lack of other information, the registration task is formulated as finding the transform that allows aligning the features (landmarks) of the transformed image onto the corresponding features in the base image. Sometimes, this may represent the distortion compensation, when the base image is calibrated, via, e.g., pretransformation based on landmarks with known positions.

Complexity of the transform with respect to the character of needed deformation determines the achievable degree of matching.

Naturally, transforms that are more complex generally allow for better matching, but their flexibility is paid for by a more complicated identification of their (often numerous) parameters. Sometimes, a superfluous flexibility may lead to undesirable local deformations in between the exactly registered landmarks.

The fused images are discrete in space; therefore, any continuous geometric transformation either requires or yields intensity values at points that are not on the sampling grid. It is then necessary to interpolate among the known values in order to provide the intensities at the grid nodes. The *interpolation methods*, as a substantial part of the transforms, will also be treated here (Section 10.1.2), though they have uses that are more generic.

The identification of transform parameters (Section 10.3) may be based on a *global measure of similarity* (Section 10.3.1), which serves as a goal or cost (penalty) function in optimization. Alternatively, the identification may require knowledge of a certain number of *correspondences* between concrete features (landmarks) in both registered images. A correspondence is defined as the couple of position vectors ($\mathbf{r}_A$, $\mathbf{r}_B$) of the corresponding features in the two images A, B to be registered. The correspondences are sought either interactively by the operator, utilizing his or her experience, or automatically. The search for similar areas, whether they are located at distinctive features (edges, corners, etc.) or at nodes of a regular grid imposed on the base image, necessarily needs a *similarity criterion*, enabling recognition of the corresponding areas. As the choice of similarity criterion crucially influences the reliability of the correspondences, we shall discuss them in detail (Section 10.1.3).

With respect to efficacy of intensity-based similarity criteria, but sometimes also accounting for the intended fused result, it is often useful to preprocess the to-be-fused images. By transforming the contrast (or color) scale of the fused images, either compensation for known or differing nonlinearities of the imaging modality or modalities, or approximately identical gray scales in the fused monomodality images, may be achieved. Restoring the images as to blur and noise or field nonuniformity concerns, in the preparatory step before fusion, may contribute significantly to the consistency as well. The restoration may be based on known deficiencies of the imaging modalities; however, even when unknown, blind restoration may be attempted. The contrast transforms will be mentioned in Chapter 11, while the restoration is the subject of Chapter 12, so these prerequisite operations will not be discussed in this chapter. On some occasions, it may be useful or even necessary to remove a part of the image from

registration or fusion, as it might confuse the registering procedure. This is usually done based on some preliminary segmentation, the basic image analysis step that will be discussed in Chapter 13.

Having settled the intensities (or colors) and possibly also corrected other deficiencies, chosen the proper part of the image, and defined the correspondences, we are prepared for the proper tasks of generic image fusion—applications of disparity analysis, image registration and image fusion (in the narrower sense), and derivation of new quality information based on the fused images.

10.1.1 Geometrical Image Transformations

The geometrical transform of an image A means a mapping T between its spatial coordinates $\mathbf{r} = (x, y)$ and the coordinates $\mathbf{r}' = (x', y')$ of the transformed image A',

$$\mathbf{r}' = T(\mathbf{r}). \tag{10.1}$$

This way, the transformed image values $f(x', y')$ are the same as the values of the image $f(x, y)$,

$$f(\mathbf{r}') = f(T(\mathbf{r})); \tag{10.2}$$

only shifted to a new position. Image A' may be thus geometrically deformed (warped) when the transform is nonrigid (see below). From the informational viewpoint, the information content of the transformed image remains obviously untouched, if the transform is reversible; i.e., if the inverse transform T^{-1} exists,

$$\mathbf{r} = T^{-1}(\mathbf{r}'), \tag{10.3}$$

because in this case the original image can be restored. Physically, the transformation is reversible when the mapping is of the one-to-one type; when the forward transform is of the many-to-one type, obviously the inversion is impossible.

The transforms are easier to explain in the two-dimensional case of planar images; however, they can be generalized to the three-dimensional case in a straightforward way—when interpreting the involved vectors as three-dimensional. As the registration may frequently concern three-dimensional image data, the vector equations, including Equations 10.1 to 10.3, should be understood as representing both dimensionalities. When detailed descriptions of vectors are used, we will present both two-dimensional and three-dimensional formulations.

The complete image may be transformed by a function, expressed by a single formula; then it is called *global transforms*. Alternatively, it is possible to use different functions for individual partial areas of the image, only guaranteeing continuity (and perhaps smoothness) at the borders of the areas. The mapping is then denoted as *piece-wise transformation*.

10.1.1.1 Rigid Transformations

In many instances, the image to be transformed is not geometrically distorted with respect to the desirable shape (e.g., compared with a base image of an image set to be registered), but only moved; the match is then achieved without any warping simply by shifting and rotating the images.

In the simplest case of a *plain shift* $\Delta\mathbf{r} = [\Delta x, \Delta y]^{\mathrm{T}}$, the needed transform is simply

$$\mathbf{r}' = \mathbf{r} + \Delta\mathbf{r}. \tag{10.4}$$

This transform has two parameters: the spatial shifts along individual coordinates (three in the three-dimensional case).

The *rotation transform*,

$$\mathbf{r}' = \mathbf{B}\,\mathbf{r}, \tag{10.5}$$

is defined by a matrix $\mathbf{B}$ of a special form. In the two-dimensional case, the rotation is defined by a single angle θ and the transform matrix is

$$\mathbf{B} = \begin{bmatrix} \cos\theta & -\sin\theta \\ \sin\theta & \cos\theta \end{bmatrix}. \tag{10.6}$$

The three-dimensional rotation is naturally described by a 3×3 matrix whose rather complex elements are determined by the three rotational angles θ_x, θ_y, θ_z about the respective coordinate axes. As the generic rotation can be interpreted as successive rotations about individual axes, the resulting matrix may be obtained as the product

$$\mathbf{B} = \begin{bmatrix} \cos\theta_z & -\sin\theta_z & 0 \\ \sin\theta_z & \cos\theta_z & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta_y & 0 & \sin\theta_y \\ 0 & 1 & 0 \\ -\sin\theta_y & 0 & \cos\theta_y \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\theta_x & -\sin\theta_x \\ 0 & \sin\theta_x & \cos\theta_x \end{bmatrix}. \tag{10.7}$$

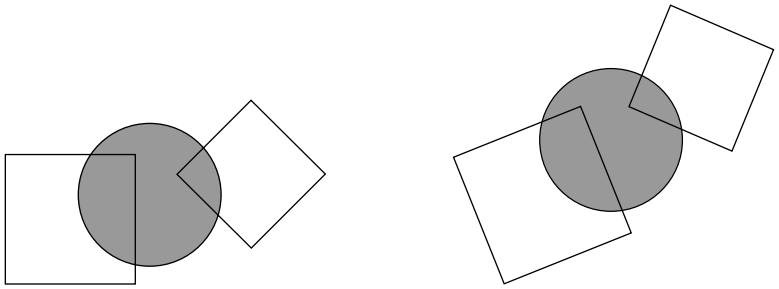


Figure 10.1 Rigid transform of a simple image: (left) original and (right) shifted and rotated image.

Notice that the first partial matrix is equivalent to the two-dimensional case (Equation 10.6). The three-dimensional rotation is thus described by three parameters; besides the mentioned triad of angles, other formulations are also used, e.g., by the unit vector defining the rotation axis and the single angle of rotation (see, e.g., [9]).

The *generic rigid transform* consists of rotation and shift (Figure 10.1),

$$\mathbf{r}' = \mathbf{B}\mathbf{r} + \Delta\,\mathbf{r}. \tag{10.8}$$

It has three and six parameters in the two-dimensional and three-dimensional cases, respectively. It is obviously a *linear* transform.

It would be advantageous to express the complete transform by a single matrix. This is possible when replacing the plain coordinate vector (x, y, z) by an extended vector expressing *homogeneous coordinates*, which carries identical information, $(cx, cy, cz, c)^{\mathrm{T}}$, where c is an arbitrary real constant. Equation 10.8 then becomes (preserving the previous symbols in the new meaning and selecting $c = 1$)

$$\mathbf{r}' = \mathbf{R}\mathbf{r}, \tag{10.9}$$

which has the following obvious two-dimensional interpretation,

$$\mathbf{r}' = \begin{bmatrix} x' \\ y' \\ 1 \end{bmatrix} = \begin{bmatrix} \cos\theta & -\sin\theta & \Delta x \\ \sin\theta & \cos\theta & \Delta y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}, \tag{10.10}$$

while in three dimensions it becomes (when denoting $\mathbf{B} = [b_{ik}]$)

$$\mathbf{r}' = \begin{bmatrix} x' \\ y' \\ z' \\ 1 \end{bmatrix} = \begin{bmatrix} b_{11} & b_{12} & b_{13} & \Delta x \\ b_{21} & b_{22} & b_{23} & \Delta y \\ b_{31} & b_{32} & b_{33} & \Delta z \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}. \tag{10.11}$$

This way, the linear transformation is expressed by a square matrix to which an inverse may exist; if so, the inverse rigid transform is thus

$$\mathbf{r} = \mathbf{R}^{-1}\mathbf{r}'. \tag{10.12}$$

Although 12 elements are to be set in the matrix $\mathbf{R}$ to define the three-dimensional transform, the coefficients b_{ik} are not mutually independent, and therefore the generic three-dimensional rigid transform is still defined by only six parameters.

10.1.1.2 Flexible Transformations

Transformations other than those, expressible in the form of Equation 10.10 or Equation 10.11 with the constrained matrix elements, are deforming the image in the sense that the distances among chosen image details (features) differ in the original and transformed images. It may be imagined that the image is printed on a flexible (e.g., rubber) sheet, parts of which may be stretched or constricted—this leads to the general name of *flexible transforms*.

The simplest transform of this kind is *plain scaling*,

$$\mathbf{r}' = \mathbf{S}\mathbf{r}, \tag{10.13}$$

where

$$\mathbf{S} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \text{ for 2D case and } \mathbf{S} = \begin{bmatrix} s_x & 0 & 0 & 0 \\ 0 & s_y & 0 & 0 \\ 0 & 0 & s_z & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \text{ in 3D.} \tag{10.14}$$

When all s_i values are identical, the scaling is isotropic—simple magnification (or reduction) of the two-dimensional image or three-dimensional scene. Otherwise, the image will be distorted

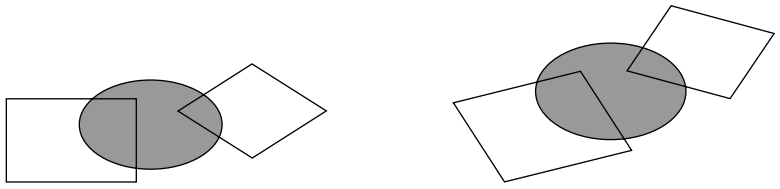


Figure 10.2 Plain scaling transformation of the previous two images (Figure 10.1) with differing s_x and s_y.

by changing proportions (Figure 10.2). Note that the right picture is now shifted, rotated, and scaled.

Combining the rigid transform and scaling leads to the (generic) *scaling transformation*

$$\mathbf{r}' = \mathbf{SR}\,\mathbf{r} \tag{10.15}$$

(notice that the sequence of both partial transforms generally matters, $\mathbf{SR} \neq \mathbf{RS}$). For two dimensions, its matrix becomes

$$\mathbf{SR} = \begin{bmatrix} s_x & 0 & 0 \\ 0 & s_y & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \cos\theta & -\sin\theta & \Delta x \\ \sin\theta & \cos\theta & \Delta y \\ 0 & 0 & 1 \end{bmatrix} = \begin{bmatrix} s_x\cos\theta & -s_x\sin\theta & s_x\Delta x \\ s_y\sin\theta & s_y\cos\theta & s_y\Delta y \\ 0 & 0 & 1 \end{bmatrix}. \tag{10.16}$$

Generic scaling transformation reduces to the *similarity transformation* when the scaling is isotropic.

The *image shearing* (Figure 10.3, gradually and uniformly shifting the rows or columns of the image matrix; in three-dimensional

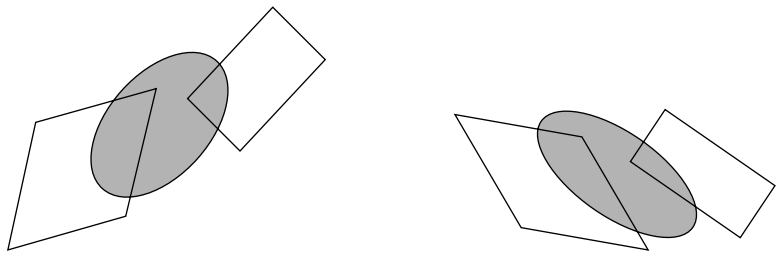


Figure 10.3 Image shearing with differing g_x, g_y.

data, possibly also layers along the z-axis) is defined in two dimensions as

$$\mathbf{G} = \mathbf{G}_x\mathbf{G}_y, \quad \mathbf{G}_x = \begin{bmatrix} 1 & g_{xy} & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad \mathbf{G}_y = \begin{bmatrix} 1 & 0 & 0 \\ g_{yx} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \tag{10.17}$$

while in three dimensions it becomes

$$\mathbf{G} = \mathbf{G}_x\mathbf{G}_y\mathbf{G}_z,$$

$$\mathbf{G}_x = \begin{bmatrix} 1 & g_{xy} & g_{xz} & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{G}_y = \begin{bmatrix} 1 & 0 & 0 & 0 \\ g_{yx} & 1 & g_{yz} & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \mathbf{G}_z = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ g_{zx} & g_{zy} & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{10.18}$$

The most generic linear image transform — the *affine transform* — adds shearing (skewing) to the scaling transformation, thus allowing for cascading of shearing, scaling, rotation, and shift,

$$\mathbf{r}' = \mathbf{GSR}\,\mathbf{r} = \mathbf{A}\mathbf{r}, \tag{10.19}$$

with the resulting matrices for two and three dimensions

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & t_1 \\ a_{21} & a_{22} & t_2 \\ 0 & 0 & 1 \end{bmatrix}, \quad \text{or} \quad \mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & t_1 \\ a_{21} & a_{22} & a_{23} & t_2 \\ a_{31} & a_{32} & a_{33} & t_3 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{10.20}$$

There are no restrictions imposed on the elements of the matrices. The affine transform therefore has 6 or 12 independent parameters for two dimensions and three dimensions, respectively. It preserves planarity of surfaces, straightness of lines, and parallelism among them, while angles between the lines or planes are generally changed. The affine transform is often used as the maximally flexible linear transformation; it preserves the advantage of easy identification thanks to linearity.

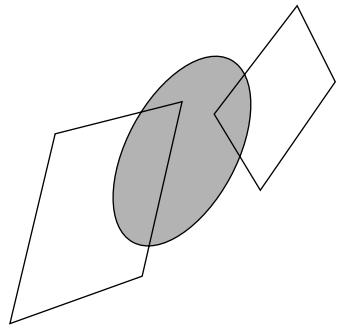


Figure 10.4 Projective transformation of the left part of Figure 10.1.

All other flexible transformations are *nonlinear*. From the viewpoint of geometrical distortion type, they may be divided into two groups. Projective transformations, forming the first group, still preserve planarity of surfaces and straightness of lines (Figure 10.4), while parallelism is not preserved. All the other transformations are *curved*; i.e., they generally bend the lines and surfaces.

The *projective transformation* is, depending on dimensionality, defined as

$$\mathbf{r}' = \frac{\mathbf{A}\mathbf{r}}{\mathbf{p}^T\mathbf{r} + \alpha}, \quad \text{with} \quad \mathbf{p}^T = [p_1 \quad p_2] \quad \text{or} \quad \mathbf{p}^T = [p_1 \quad p_2 \quad p_3], \quad (10.21)$$

where constants $\mathbf{A}$, $\mathbf{p}$, and α are the transformation parameters. The numerator is just the affine transform, and its results are divided by a value dependent on $\mathbf{r}$; this is the cause of nonlinearity. The projective transform has 9 and 16 scalar parameters (for two and three dimensions, respectively) and is therefore even more flexible than the affine transform. The projective transform may be formulated in two steps, the first being the linear transform

$$\mathbf{w} = \mathbf{P}\mathbf{r}, \quad \mathbf{P} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & t_1 \\ a_{21} & a_{22} & a_{23} & t_2 \\ a_{31} & a_{32} & a_{33} & t_3 \\ p_1 & p_2 & p_3 & \alpha \end{bmatrix}, \quad \mathbf{r} = \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}, \quad (10.22)$$

followed by the second step,

$$\mathbf{r}' = g(\mathbf{w}) = \begin{bmatrix} w_1/w_4 \\ w_2/w_4 \\ w_3/w_4 \\ 1 \end{bmatrix}. \tag{10.23}$$

A special case of the projection transform, for $\mathbf{A} = \mathbf{I}$, is the *perspective transform* projecting three-dimensional data onto a two-dimensional plane.

In a sense, the *polynomial transforms* may be considered a generalization of the linear transforms. The output vector $\mathbf{r}'$, in the three-dimensional version, can be calculated based on the components of the input vector $\mathbf{r} = (x, y, z)$ as

$$\mathbf{r}' = \sum_i \sum_k \sum_m \begin{bmatrix} {}_{ikm}c_x \\ {}_{ikm}c_y \\ {}_{ikm}c_z \end{bmatrix} x^i y^k z^m, \tag{10.24}$$

where the constants c are the coefficients of polynomials defining the transformation. For example, the second-order two-dimensional polynomial transformation is given by the expression controlled by 12 parameters:

$$\mathbf{r}' = \begin{bmatrix} x' \\ y' \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} & a_{15} & a_{16} \\ a_{21} & a_{22} & a_{23} & a_{24} & a_{25} & a_{26} \end{bmatrix} [1 \quad x \quad y \quad x^2 \quad y^2 \quad xy]^T. \tag{10.25}$$

A fundamental problem of higher-order polynomial transforms is their tendency to spurious oscillations that are difficult to control and that may cause undesirable local warping or even folding over of the transformed image. Therefore, polynomials of higher than the third order are rarely used.

Among many other types of functions used for geometrical transformations, the category of *radial basis functions* receives exceptional attention. These are functions based on known control points $\tilde{\mathbf{r}}$ for which the correspondence between $\tilde{\mathbf{r}}'$ and $\tilde{\mathbf{r}}$ is known (tildes distinguish the control points from general position vectors). The radial basis functions are rotationally symmetric functions dependent on the distance r from the respective control point,

$$g_k(\mathbf{r}) = g(|\mathbf{r} - \tilde{\mathbf{r}}_k|) = g(r_k). \tag{10.26}$$

The set of the functions for $k = 1, 2, \ldots, M$, when there are M correspondences available, serves then as a basis for construction of the transform functions. Different function types have been used, as, e.g., Gaussian function $\exp(-ar^2)$, but most attention in the image registration has been on *thin-plate splines*,

$$\begin{aligned} g_k(\mathbf{r}) &= r_k^2 \ln(r_k) \quad \text{in 2D case,} \\ g_k(\mathbf{r}) &= r_k \qquad\quad \text{in 3D case.} \end{aligned} \tag{10.27}$$

It is required that the transform function must fulfill

$$\tilde{\mathbf{r}}'_k = T(\tilde{\mathbf{r}}_k), \quad k = 1,2,\ldots,M \tag{10.28}$$

either exactly when the transform is interpolating or in the least mean square error sense, should the number of correspondences suffice. The transformation is defined as

$$\mathbf{r}' = \mathbf{A}\mathbf{r} + \sum_{k=1}^{M} \mathbf{b}_k g(|\mathbf{r} - \tilde{\mathbf{r}}_k|), \qquad \mathbf{b}_k = \begin{bmatrix} {}_k b_x \\ {}_k b_y \\ {}_k b_z \\ 0 \end{bmatrix} \tag{10.29}$$

where $\mathbf{A}$ is the matrix (Equation 10.20) characterizing the affine part of the transform (12 coefficients) and $\mathbf{B} = [\mathbf{b}_1 \quad \mathbf{b}_2 \quad \cdots \quad \mathbf{b}_M]$ is the matrix of $3M$ coefficients of the flexible part. To define a concrete interpolating transform, the matrices $\mathbf{A}$ and $\mathbf{B}$ must be derived based on the conditions in Equation 10.28, which represent $3M$ linear equations. The remaining 12 equations are formulated based on constraints imposed on the flexible part of the transform. Further details can be found, e.g., in [16].

The internal geometric converting algorithms of imaging systems belong to geometrical transformations of the discussed type as well. Examples may be the embedded conversion between the ultrasonographic acquisition fan format in polar coordinates and the rectangular display format in Cartesian coordinates or embedded algorithms to compensate for pincushion or barrel distortion in optical instruments via polynomial transforms.

This section is devoted to explicit functions defining the transforms. Alternatively, a rather different approach is possible that formulates the task as matching the original image (scene) to the base

image by deforming the material supporting the image (as if the image data were carried by an elastic, or fluid, sheet or volume). The disparities in coordinates of corresponding features act as forces deforming the volume, which—depending on the material properties like elasticity or viscosity—interpolates the coordinates in the space between the control points. This formulation describes the transformation implicitly by partial differential equations that are to be solved by the method of finite elements or finite differences. However, as there is no explicit formula for the resulting transform, we shall not elaborate on this matter here; a more detailed discussion and references can be found in [65].

10.1.1.3 Piece-Wise Transformations

So far, we have dealt only with global transforms that apply the same formula on the entire image plane or space. When the required transform becomes too complicated, it may be advantageous to split the image to adjoined but nonoverlapping areas that are transformed individually. The type of splitting depends on the character of the problem as well as the available information. The area borders may be given by the control points (see the following paragraph) or be defined by an *a priori* regular (most often rectangular) mesh, or they may depend on the image content and be determined by previous segmentation (e.g., to distinguish between different distortion of soft tissue area in contrast to skeleton elements).

The principle of the approach may be simply illustrated on the piece-wise linear two-dimensional transformation. Let us have a set of control points $\mathbf{r}_1, \mathbf{r}_2, \ldots \mathbf{r}_M$ with the corresponding proper positions $\mathbf{r}'_1, \mathbf{r}'_2, \ldots \mathbf{r}'_M$ (Figure 10.5). The image area may be partitioned to triangles, the vertices of which are the controlling points; inside of each of the triangles, we shall apply a separate transform. Let us

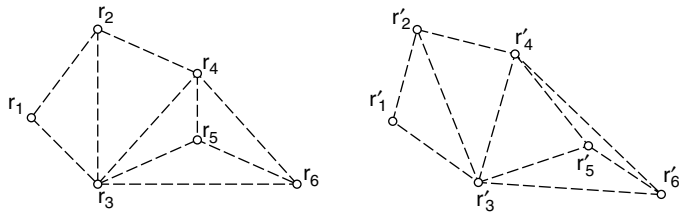


Figure 10.5 Corresponding sets of control points in a couple of images.

choose the most generic linear (i.e., affine) transform (Equation 10.19). In analogy to Equation 10.28, the transform should be precise at vertices; thus, for a particular k-th triangle and its i-th vertex,

$$\tilde{\mathbf{r}}'_{k,i} = T(\tilde{\mathbf{r}}_{k,i}) = \mathbf{A}_{k,i}\,\mathbf{r}_{k,i}, \quad i = 1,2,3. \tag{10.30}$$

By rewriting, we obtain a system of six linear equations,

$$\begin{aligned} x'_{k,i} &= {}_k a_{1,1} x_{k,i} + {}_k a_{1,2} y_{k,i} + {}_k t_1 \\ y'_{k,i} &= {}_k a_{2,1} x_{k,i} + {}_k a_{2,2} y_{k,i} + {}_k t_2, \quad i = 1,2,3, \end{aligned} \tag{10.31}$$

so that when denoting

$$\mathbf{a}_k = [{}_k a_{1,1} \quad {}_k a_{1,2} \quad {}_k t_1 \quad {}_k a_{2,1} \quad {}_k a_{2,2} \quad {}_k t_2]^T$$

$$\mathbf{x}'_k = [x'_{k,1} \quad y'_{k,1} \quad x'_{k,2} \quad y'_{k,2} \quad x'_{k,3} \quad y'_{k,3}]^T$$

$$\mathbf{X}_k = \begin{bmatrix} x_{k,1} & y_{k,1} & 1 & 0 & 0 & 0 \\ \vdots & & & & & \\ 0 & 0 & 0 & x_{k,3} & y_{k,3} & 1 \end{bmatrix},$$

the solution, yielding the transform parameters is

$$\mathbf{a}_k = \mathbf{X}_k^{-1}\mathbf{x}'_k. \tag{10.32}$$

This way, the proper transform is determined for every triangle. It is obvious that, thanks to linearity, the values interpolated in adjacent triangles at the common borders are identical and the transform is thus continuous; however, the continuity is not generally valid for derivatives, so that the piece-wise linear interpolation is not smooth at area borders and may only be used if smoothness is not needed. Naturally, the piece-wise linear interpolation may easily be generalized for the three-dimensional case.

To achieve smoothness at the borders, higher-order transforms must be used, and the requirement of equality of derivatives at the borders up to a chosen order must be included into the formulation. Let us consider the three-dimensional case of regular partitioning, where the image space is now divided *a priori* into rectangular cells by three sets of equidistant parallel planes, each set perpendicular to one of the coordinates x, y, z. Let the cells be cubic, $\Delta x = \Delta y = \Delta z = \Delta$ so that the cell corners $\tilde{\mathbf{x}}_{i,j,k}$, called *knots*, are uniformly distributed.

It is supposed that the proper correspondence $\tilde{\mathbf{x}}_{i,j,k} \rightarrow \tilde{\mathbf{x}}'_{i,j,k}$ is known for each knot. Every transform is valid just inside and at the borders of the cell; in other words, the support of each transformation function is finite and limited to its own cell.

Under such circumstances, the most commonly used functions are polynomials of a special type—*splines*, which are constructed so that their values and derivatives up to the $(p-1)$-th order at the borders continue to the corresponding functions of the neighboring cell (p being the polynomial order). The parameters of the transform functions (weights given to partial polynomials) are determined by the coordinates of corresponding knots; this simplifies the computations. Commonly, only separable splines $B_{3D}(x,y,z) = B(x)B(y)B(z)$ are used; mostly $p = 3$ is chosen and then the suitable partial one-dimensional splines may be shown to become

$$\begin{aligned} &B_n(s) = \mathbf{b}_n \mathbf{s}, \qquad \text{where} \qquad \mathbf{s} = [s^3 \quad s^2 \quad s \quad 1]^T \quad \text{and} \\ &\mathbf{b}_{-1} = [-1 \quad 3 \quad -3 \quad 1]/6, \qquad \mathbf{b}_0 = [3 \quad -6 \quad 0 \quad 4]/6, \\ &\mathbf{b}_1 = [-3 \quad 3 \quad 3 \quad 1]/6, \qquad \mathbf{b}_2 = [1 \quad 0 \quad 0 \quad 0]/6. \end{aligned} \tag{10.33}$$

The transform function in the (i, j, k)-th cell whose corner closest to the coordinate origin is $\tilde{\mathbf{x}}_{i,j,k}$ then becomes

$$\begin{aligned} &\mathbf{x}'_{i,j,k} = [x', y', z']^T = \sum_{l=-1}^{2} \sum_{m=-1}^{2} \sum_{n=-1}^{2} B_l(u)\, B_m(v)\, B_n(w)\, \mathbf{c}_{l,m,n}, \\ &u = (x' - \tilde{x})/\Delta, \quad v = (y' - \tilde{y})/\Delta, \quad w = (z' - \tilde{z})/\Delta, \quad \tilde{\mathbf{x}}_{i,j,k} = [\tilde{x}, \tilde{y}, \tilde{z}]^T, \\ &u, v, w \in \langle 0, 1 \rangle, \end{aligned} \tag{10.34}$$

where $\mathbf{c}_{l,m,n}$ are coefficient vectors determined by the given $\tilde{\mathbf{x}}'_{i+l,\, j+m,\, k+n}$. This interpolation by B-splines is smooth and well defined; its disadvantage may be only that it requires correspondences on a regular grid, as formulated above.

10.1.2 Image Interpolation

In the previous section on geometrical transformations, and on other occasions as well, the image is treated as continuous in space. However, all these transformations have to be realized in the discrete environment, where only samples of images are available, and the results of operations must be finally assigned to a discrete grid of pixels.

When not treated properly, the discrete approximation of operations, that were designed for and analyzed in continuous space, may lead to serious distortions and subsequent artifacts in the resulting images. In this section, we shall therefore present some of the approaches to *image value interpolation* that enable the obtaining, on a discrete grid, of results comparable with those obtained in continuous space.

The task of value interpolation is to recover, as precisely as possible, the image intensity at a spatial point different from sampling points, given the samples. This problem appears also when geometrical transforms are performed, due to the necessity to construct the values at nodes of a given (fixed) grid of the output image, while the transformation function itself would provide the values in between them. Other occasions are image resampling, or superresolution operations.

Let us have the discrete image data $f(k\Delta x, i\Delta y)$ on an equidistant rectangular grid (in two-dimensional or three-dimensional space), the nodes of which are the sampling points. If we are dealing with image samples obtained in accordance with the sampling theorem, the *exact* reconstruction formula exists. As explained in Section 2.1.2 for the two-dimensional case (Equations 2.17 and 2.18), the interpolated value is

$$f_r(x,y) = \sum_i \sum_k f(k\Delta x, i\Delta y)\, r(x - k\Delta x, y - i\Delta y), \qquad (10.35)$$

where

$$r(x,y) = \frac{u_m v_m}{\pi^2} \frac{\sin(u_m x)}{u_m x} \frac{\sin(v_m y)}{v_m y}, \qquad u_m = \pi/\Delta x,\ v_m = \pi/\Delta y.$$

Obviously, the reconstruction formula means interpolation between the samples using the interpolation kernel, which is separable into functions $\mathrm{sinc}(c_x\, x)$ and $\mathrm{sinc}(c_y\, y)$. Because these functions acquire values of 1 only at the position of the corresponding sample, while they are zero at locations of all other samples, the formula fulfills the interpolation condition

$$f_r(k\Delta x, i\Delta y) = f(k\Delta x, i\Delta y).$$

The calculated values between the samples are, under the above conditions, exact values of the original continuous image before sampling; this interpolation method thus represents the golden standard.

The result (Equation 10.35) may also be interpreted as the convolution, i.e., as the output of the linear two-dimensional filter having the impulse response $r(x, y)$, which corresponds to the ideal

low-pass frequency response (Figure 10.8a) up to the frequencies u_m, v_m (Equation 2.15).

10.1.2.1 Interpolation in the Spatial Domain

Though theoretically that simple, practically it is hardly feasible to obtain the interpolated values this way, due to complexity of the interpolating sinc(…) function. Primarily, it has infinite support with rather slow decay from the central point of the respective sample (envelope $\sim 1/|x|$), so that any reasonable approximation would have to include many even distant samples in opposite directions in every dimension; thus, every sample would influence a large area (theoretically total) of the image. Very many terms (theoretically infinite number) would have to be summed in the interpolation formula. Also, the calculation of the sinc(…) function is demanding, so that the interpolation would represent a high computational burden and consequently be time-consuming.

Hence, simpler interpolation formulae are to be found, still yielding reasonably precise values. As the exact formula is separable*, it may be expected that any good approximation may be separable as well. This simplifies the situation, as only one-dimensional interpolation kernels need to be considered. The requirements to the interpolation formula are the following:

- Good precision, i.e., acceptable deviations from Equation 10.35.
- Small support size—only a few samples around the calculated point should enter the calculation, so that there is only a low number of terms in the interpolation formula.
- Simple kernels (basis functions).
- Continuous and possibly even smooth behavior on the whole image area.
- Preserving approximately evenly the frequency content up to the Nyquist frequencies u_m, v_m, and maximally suppressing frequencies above them.

No approximation can optimally fulfill all the requirements. It depends on the character of the task which of the aspects should be emphasized. Obviously, the approximation may be formulated either as interpolation or as filtering—i.e., convolution with a chosen impulse response.

*The interpolation thus may be performed separately along rows and then along columns, or vice versa.

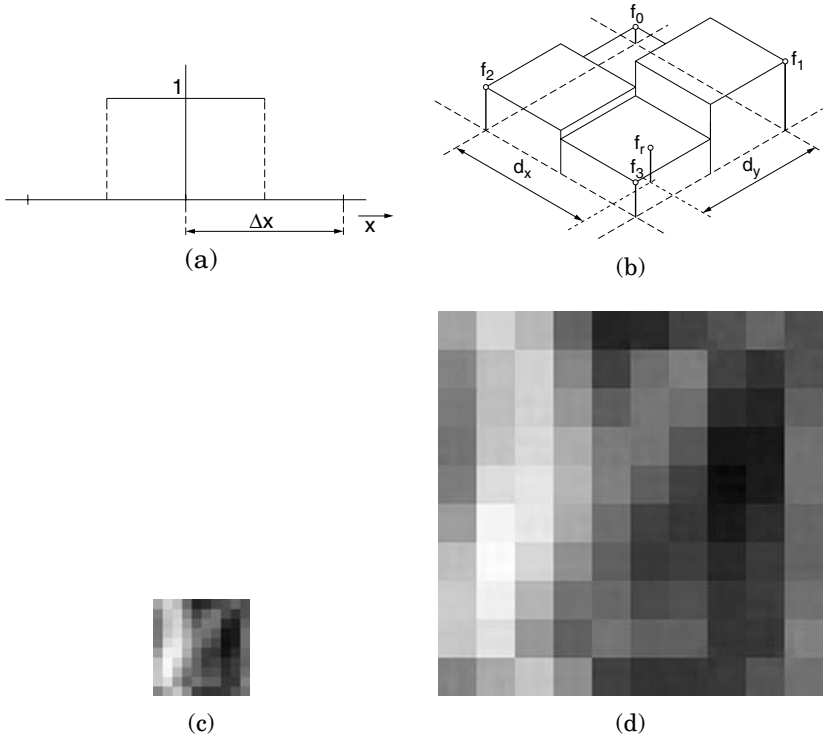


Figure 10.6 One-dimensional and corresponding two-dimensional nearest-neighbor interpolation: (a) one-dimensional interpolation kernel, (b) schematic two-dimensional interpolation, and (d) example of NN-interpolated image (four times magnified from the original (c)).

The simplest of the approximation methods is the *nearest-neighbor (NN) interpolation* (Figure 10.6a, b), in two dimensions covering the complete square area $\Delta x \Delta y$ around the sample with the sample value (up to the half distance to neighboring pixels). Otherwise expressed, and when considering the uniformly sampled three-dimensional case, a cube between eight neighboring samples forms the field of interpolation; the interpolated value in this cube may be expressed as

$$
\begin{aligned}
{}_{NN}f_r(x,y,z) = f_r(\mathbf{r}) = f((k+l)\Delta,(i+m)\Delta,(j+n)\Delta)\,, \quad l,m,n \in \{0,1\} \\
[l,m,n] = (\mathbf{r}_n - \mathbf{r}_0)/\Delta,\ \mathbf{r}_0 = [k,i,j]\Delta,\ \mathbf{r}_n = \arg\min_n |\,\mathbf{r} - \mathbf{r}_n\,|\,,\ n = 0,1,\ldots,7,
\end{aligned}
\tag{10.36}
$$

where $\mathbf{r}_0$ is the sampling position closest to the origin of coordinates and $\mathbf{r}_n$ is the nearest sampling node, both of the eight sampling positions surrounding the required position $\mathbf{r}$.

The resulting "staircase" function (Figure 10.6b) is not continuous, and its frequency content is heavily modified in comparison with the original image spectrum. The frequency response of the one-dimensional nearest-neighbor interpolation, as corresponding to the rectangular impulse response, is (see Figure 10.8b)

$$R_{NN}(u) = \frac{2}{\Delta x}\frac{\sin u\Delta x/2}{u\Delta x/2} = \frac{2}{\Delta x}\frac{\sin \pi u/u_m}{\pi u/u_m}, \tag{10.37}$$

where u_m is the Nyquist frequency in the direction x. Thus, the main lobe of the frequency response reaches zero only at the sampling frequency, allowing for aliasing interference; substantial side lobes in the stop-band further worsen this by allowing components of higher spectral replicas to pass. The main lobe response is uneven in the required passband, so that the NN interpolation partly limits the useful high-frequency components below u_m, thus reducing contrast of details. The discontinuities of the interpolated function may lead to severe distortion, namely when the interpolated image is further processed. The NN interpolation therefore should only be used when the image is sufficiently oversampled, in the final steps of processing immediately preceding display, or in initial iterative steps of processing, as a simple estimate that would be later improved by a more sophisticated method.

Although more demanding, substantially better is the two-dimensional *bilinear* (*BL*) *interpolation* (or *trilinear* (*TL*) in three dimensions), consisting of two (or three) perpendicular one-dimensional linear interpolations with the linear kernel (Figure 10.7a). As the one-dimensional triangular impulse response may be considered self-convolution of the rectangular impulse response of NN interpolation, $r_L(x) = r_{NN}(x) * r_{NN}(x)$, the one-dimensional frequency response of the linear interpolation is the squared NN response (see Figure 10.8c),

$$R_L(u) = \left(\frac{2}{\Delta x}\frac{\sin \pi u/u_N}{\pi u/u_N}\right)^2. \tag{10.38}$$

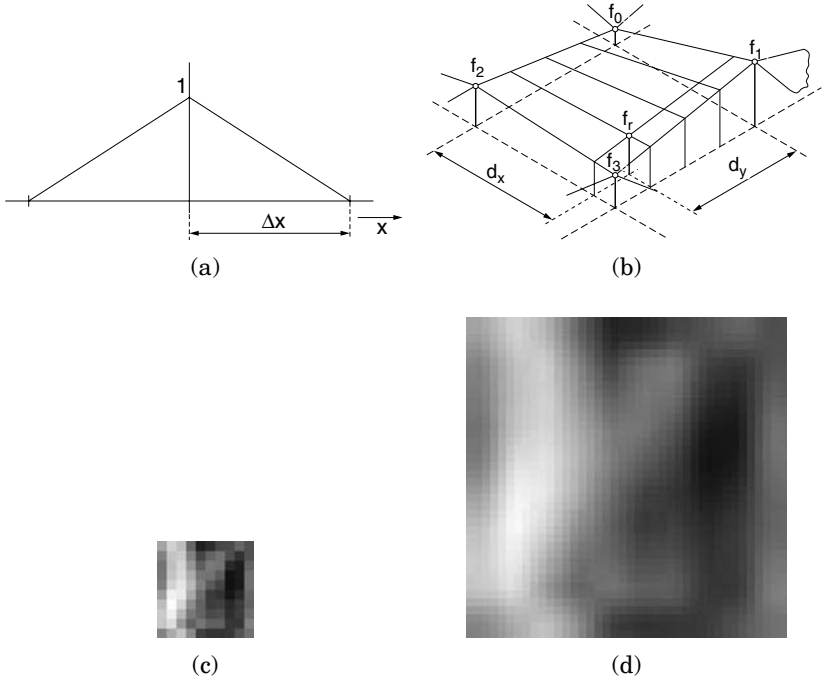


Figure 10.7 Bilinear interpolation: (a) one-dimensional interpolation kernel, (b) schematic two-dimensional interpolation, and (d) example of BL-interpolated image (four times magnified from the original (c)).

While obviously losing even more on proper high frequencies than NN, the function is continuous and the aliasing is substantially better suppressed. The two-dimensional bilinear interpolation may be formulated as

$$
\begin{aligned}
&{}_{BL}f_r(x,y) = f_r(\mathbf{r}) = \sum_{l=0}^{1}\sum_{m=0}^{1} f((k+l)\Delta,(i+m)\Delta)\,c_{l,m}, \quad \mathbf{r}_0 = [k,m]\Delta, \\
&[d_x,d_y] = (\mathbf{r}-\mathbf{r}_0)/\Delta, \qquad \mathbf{c} = \begin{bmatrix} (1-d_x)(1-d_y) & (1-d_x)d_y \\ d_x(1-d_y) & d_x d_y \end{bmatrix},
\end{aligned}
\tag{10.39}
$$

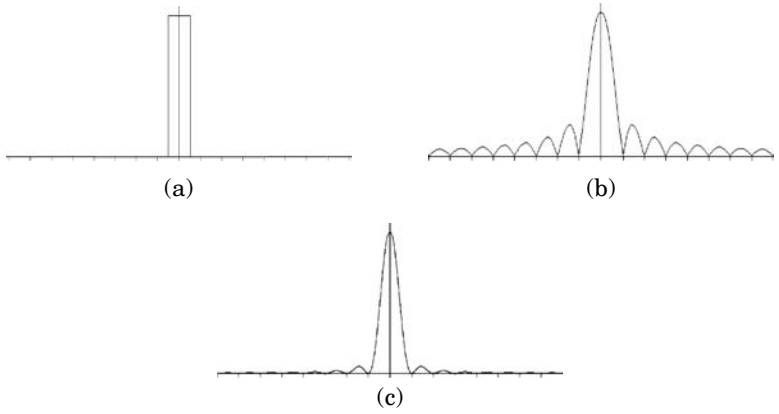


Figure 10.8 Normalized one-dimensional frequency responses of basic interpolation methods: (a) ideal (sinc) interpolation, (b) nearest-neighbor interpolation, and (c) linear interpolation (a unit on the frequency axis corresponds to the sampling frequency).

while for three dimensions, the trilinear interpolation becomes

$$_{TL}f_r(x,y,z) = f_r(\mathbf{r}) = \sum_{l=0}^{1}\sum_{m=0}^{1}\sum_{n=0}^{1} f((k+l)\Delta,(i+m)\Delta,(j+n)\Delta)\,c_{l,m,n},$$

$$\mathbf{r}_0 = [k,m,n]\Delta, \quad [d_x,d_y,d_z] = (\mathbf{r}-\mathbf{r}_0)/\Delta, \quad \mathbf{c}_{n=0} = (1-d_z)\mathbf{c}, \quad \mathbf{c}_{n=1} = d_z\mathbf{c}. \tag{10.40}$$

From the original-domain viewpoint, the bilinear (trilinear) interpolation is continuous at the borders of the individual interpolation areas (Figure 10.7b); however, the derivatives are generally discontinuous. Note that the two-dimensional BL interpolation function does not form a plane in the extent of the square, but rather a curved surface. The derivatives are discontinuous at the border lines or planes in the two-dimensional or three-dimensional case, because only the nearest sample values are taken into account (four in a two- or eight in a three-dimensional case). For display, the bi- or trilinear interpolation may be considered well acceptable (Figure 10.7c). However, the lack of smoothness may be prohibitive when further complex processing is intended.

When the continuity of up to the second derivative is required, *bicubic (or tricubic) interpolation* is needed, which provides a smooth surface. Commonly, cubic B-spline interpolation is used, guaranteeing

the continuity of up to the second derivative at the borders of individual squares (or cubes). The two-dimensional bicubic B-spline interpolation can be expressed as (see [59])

$$_{TC}f_r(x,y) = f_r(\mathbf{r}) = \sum_{l=-1}^{2}\sum_{m=-1}^{2} f((k+l)\Delta,(i+m)\Delta)\, r_C(l-d_x)\, r_C(d_y-m),$$

$$\mathbf{r}_0 = [k,m,n]\Delta, \quad [d_x,d_y,d_z] = (\mathbf{r}-\mathbf{r}_0)/\Delta,$$

$$r_C(x) = \begin{cases} 0.5\,|x^3| - x^2 + 2/3 & \text{for } |x| \in \langle 0,1\rangle. \\ -|x^3|/6 + 3 & \text{for } |x| \in \langle 1,2\rangle \end{cases} \tag{10.41}$$

Obviously, a greater set of 16 samples is involved, extended by one to every direction in comparison with BL; in the three-dimensional case, this means 64 samples surrounding the cube in which the interpolation is executed. Other relatively complicated cubic interpolation formulae with a possibility to optimize the frequency properties, and the sets of respective linear equations yielding the weighting coefficients, can be found, e.g., in [22], [59], [78]. The approximation error is low (of the order of d_x^4 for small d_x), and better spectral properties may be achieved than for any of the above methods, however, at the cost of more complicated computation.

Besides polynomials, many other interpolation kernels may be used. Among them, one deserves mentioning as a theoretically interesting concept: the shortened sinc(…) function. This approach is based on the well-known technique of designing finite impulse response (FIR) filters*—the finite support function is obtained from the sinc(…) by weighting with a finite-length window. As this means a multiplication in original domain, the ideal frequency response of the sinc(…) is modified by convolution with the window spectrum. Obviously, the oscillating spectrum of a rectangular window (corresponding to simple truncation) would lead to undesirable long tails of the desired low-pass characteristic; therefore, optimized windows (like Hamming) should be used. It can be shown that such truncated sinc functions are well approximated by the above-mentioned cubic polynomials.

The original-domain calculations, as above, are used when individual interpolated values are needed for different arguments d_x, d_y, d_z,

*An FIR filter is a one-dimensional discrete system designed with an *a priori* finite length of its impulse response (see, e.g., [30]).

as, e.g., in geometric transforms. On the other hand, when the interpolation aims at obtaining a resampled image on an n times denser grid, another approach is advantageous. The first phase consists in interleaving $n-1$ rows of zeros between every two original rows and below the lowest one; similarly, $n-1$ columns of zeros are interleaved between every two original columns and to the right of them. This way, a sparse image is obtained that obviously has a modified spectrum in comparison with the original. The idea of the interpolation is that a suitable filter realized as convolution with a two-dimensional kernel should restore the original spectrum. This is not exactly possible with small kernels; however, simple smoothing masks, sized 2×2, 3×3, up to about 5×5, have been designed and are used to this end, like the simplest,

$$\begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}, \quad \text{or} \quad \begin{bmatrix} 0.25 & 0.5 & 0.25 \\ 0.5 & 1 & 0.5 \\ 0.25 & 0.5 & 0.25 \end{bmatrix}. \tag{10.42}$$

When applying these two kernels to the image interleaved by zeros with $n = 2$, the first filter obviously provides the nearest-neighbor interpolation (slightly shifted due to lack of central element in the mask), while the second one realizes the bilinear interpolation (naturally only for the interleaved points). When a better approximation is needed, which would require large masks, frequency-domain-based interpolation may be a better alternative.

10.1.2.2 Spatial Interpolation via Frequency Domain

Taking into account the properties of the two-dimensional discrete Fourier transform (DFT), it is theoretically possible to obtain the resampled image exactly (with a reservation; see below) on an arbitrarily denser uniform sampling grid via frequency domain. The procedure starts with padding the standard spectral matrix of an image sized, say, $M \times N$ (with zero frequency at the left upper position) by zeros on the side of high frequencies, so that a new, bigger matrix of the size, say, $mM \times nN$, $m, n > 1$, mM, nN integers, is formed, in which the original spectrum is situated at the left upper corner,

$$\mathbf{A} = [f_{i,k}], \quad \mathbf{F} = \mathrm{DFT}_{2\mathrm{D}}\{\mathbf{A}\}, \quad \mathbf{F}_{ex} = \begin{bmatrix} \mathbf{F} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}. \tag{10.43}$$

This matrix may be considered a two-dimensional DFT of the same original image, but has a greater frequency extent up to mU, nV.

The spectral *content* remained naturally the same (the highest nonzero spectral component still at the original two-dimensional sampling frequency (U,V)), so that the components with frequencies higher than that are naturally zero. However, the highest frequency (though with zero intensity) in the spectral matrix is now (mU, nV), as if the image had been sampled m, n times more densely. Obviously, by inverse transforming this extended spectral matrix via two-dimensional DFT^{-1}, we obtain the discrete image $\mathbf{A}_{\mathrm{ex}}$ of the identical size as the spectrum, i.e., $mN \times nN$, representing the original unchanged spatial extent,

$$\mathbf{A}_{ex} = \mathrm{DFT}_{2\mathrm{D}}^{-1}\left\{\begin{bmatrix}\mathbf{F} & \mathbf{0}\\ \mathbf{0} & \mathbf{0}\end{bmatrix}\right\}, \tag{10.44}$$

with values *exactly* interpolated from the original samples. If m, n are integers, the original samples are preserved, and between them $m-1, n-1$ new values in the direction x, y, respectively, are exactly interpolated (Figure 10.9 for $m, n = 2$). Should m, n be a noninteger, generally all image samples acquire new values. This procedure works for arbitrary integers M, N, mM, and nN, although it is computationally advantageous when they can be factored to products of low integers, optimally to integer powers of 2.

Another possibility can be visualized as obtaining the interpolated values in between existing samples by shifting the original image with respect to the sampling grid by (d_x, d_y) and then resampling with the same resolution; both the original and the new sample sets may be merged afterwards. This can be done without really

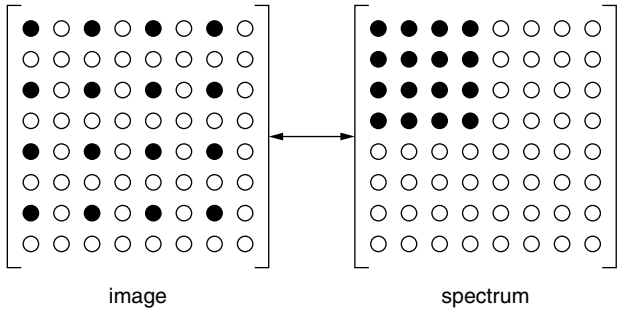


Figure 10.9 Doubling sampling density by interpolation via frequency domain: full circles, original samples; empty circles, padded spectral zeros on right and interpolated image samples on left.

resampling the image. Instead, the spectrum of the shifted image is obtained according to Equation 1.35 from the original spectrum by multiplying all its spectral components by the respective phase factor

$$\mathbf{A} = [f_{i,k}] = [f(i\Delta x, k\Delta y)], \quad \mathbf{F} = \mathrm{DFT}_{2\mathrm{D}}\{\mathbf{A}\}, \quad \mathbf{F}' = \left[F_{m,n} e^{-j(u_m d_x + v_n d_y)} \right], \tag{10.45}$$

where u_m, v_n are the frequencies corresponding to the (m, n)-th spectral element. The following inverse transform then provides the values of the samples on the grid shifted with respect to the image by (d_x, d_y),

$$\mathbf{A} = \mathrm{DFT}_{2\mathrm{D}}^{-1}\{\mathbf{F}'\} = [f(i\Delta x - d_x, k\Delta y - d_y)]. \tag{10.46}$$

Naturally, for interpolation purposes, $d_x < \Delta x$, $d_y < \Delta y$.

At first glance, both described approaches seem to be the ideal situation; however, they are not, because we are dealing with finite-size images. One of the two-dimensional DFT properties is that, due to discretization of spectrum, the original image is supposed periodical in both directions x, y, with the periods equal to the respective image dimensions. As the ideal interpolation is based on the sinc(…) kernel with infinite support, even distant samples influence the interpolated values. Therefore, the interpolated values near to image margins are substantially influenced by the samples from the opposite side of the image, which is obviously improper. This phenomenon can be reduced by *a priori* extending the image with extrapolated (e.g., mirrored) margins that will be abandoned after interpolation, but obviously, the result remains only an approximation anyway. Generalization of both methods to the three-dimensional case is straightforward.

10.1.3 Local Similarity Criteria

Similarity criteria serve to evaluate resemblance of two (and possibly more) images or their areas. Similarity on the global scale must be evaluated when matching two or more images via geometrical transforms (Section 10.3.1); in complex pattern recognition, the criterion may be needed semiglobally—the comparison then concerns significant areas of images. On the other hand, in disparity analysis, similarity must be evaluated locally, considering as small as possible areas, to obtain the local disparity vectors between positions of corresponding features in compared images. When disparities are determined for many pairs of landmarks, or on a dense grid in image space, these vectors form the disparity map or field (Section 10.2.1).

Except for a few differently based approaches, even the global similarity is based on local similarities. It may be interpreted in the sense that the resulting global similarity is simply a total of partial similarities (integral, sum, average). Alternatively, the matching errors of locally determined point correspondences are evaluated as a whole, which leads to a vector criterion, or possibly its length or another norm. The intensity-based local similarity criteria are necessary also for finding individual correspondences between pairs in a group of landmarks (or in perfecting manually assigned rough correspondences), as needed in point-based registration approaches.

Therefore, the local similarity evaluation is an important concept in image fusion. This section will thus be devoted to the local similarity criteria, based on intensity data. These relations are often called *intensity similarities*, although color may be involved as well. There is no definite limit on the size of the compared image areas for local or semiglobal comparisons; in every case, certain finite areas are compared, usually completely overlapping. The global comparison (see Section 10.3.1) differs in usually large compared areas and may be complicated by a mere partial overlap, the size and shape of which may change when the registering parameters are changed; however, the concept of intensity-based similarity criteria may remain basically unchanged even for global comparisons.

10.1.3.1 Direct Intensity-Based Criteria

The similarity criteria compare, in a sense, the intensities in two image areas, A and B, usually from two images to be fused. A discrete image area contains a certain finite number of pixels; thus, area A may be characterized by a vector **a** formed of N pixel values ordered systematically according to a chosen rule (e.g., by column scanning). Most local criteria require that both compared areas are of the same size so that the dimension of the vector **b**, similarly characterizing the area B, is identical to the dimension of **a**; all possible vectors of the same dimension thus form a vector space where the N coordinates describe intensities of individual pixels. This idea can be easily extended from intensity images to vector-valued (e.g., color or multimodal) images: each pixel value itself is then a vector (e.g., three-dimensional for color), so that the only difference is that the dimension of the vector space is several times higher (three times for color). When the comparison concerns areas after some geometrical transforms of the original images, it is the task of the transforms to ensure the identical area sizes, i.e., the compatible vector dimensions.

This concept enables us to explain an important class of similarity criteria in the unifying frame of *interpretation in the intensity vector space*, as in [31], [27]. It is obvious that identical vectors represent identical areas, while differing vectors correspond to differing though possibly similar compared areas. It is a matter of choice, which quantity characterizing the inequality will be considered relevant as a criterion of (dis)similarity; the efficacy of a particular criterion with respect to a certain problem should be analyzed and its suitability verified (at least experimentally).

The conceptually simplest criterion of dissimilarity is the *Euclidean distance*, the length of the difference vector,

$$C_E(\mathbf{a},\mathbf{b}) = |\mathbf{a}-\mathbf{b}| = \sqrt{\sum_{i=1}^{N}(a_i - b_i)^2}. \tag{10.47}$$

The square of it, the *sum of squared differences*, is an often used criterion with similar properties, only emphasizing greater dissimilarities. Both quantities may be normalized by N, thus obtaining comparable values, should the areas be of different size N in individual comparisons*. The geometrical interpretation of the Euclidean distance (Equation 10.47) and criteria similar to it show clearly that such criteria are sensitive to the difference in absolute lengths of the vectors; thus, a mere linear change of contrast in one of otherwise identical compared images may spoil the similarity in the sense of this criterion. Therefore, a criterion less sensitive to the contrast differences among images may be preferred.

A criterion often recommended in literature is the s.c. *cross-correlation* given by the dot product **ab**. It can readily be shown that it is a very unreliable similarity measure, and its use should therefore be avoided. Really, expanding the argument of the square root in Equation 10.47 gives

$$\sum_{i=1}^{N}(a_i - b_i)^2 = \sum_{i=1}^{N}\left(a_i^2\right) - \sum_{i=1}^{N}(a_i b_i) + \sum_{i=1}^{N}\left(b_i^2\right), \tag{10.48}$$

where the central term is obviously the above-mentioned cross-correlation. If both sums of squares were constant, as it is often

*Naturally, both compared areas in each comparison must be equally sized.

implicitly believed, it would suffice to maximize the cross-correlation

$$C_C(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{N}(a_i b_i) \propto \frac{1}{N}\sum_{i=1}^{N}(a_i b_i) \tag{10.49}$$

as the similarity measure to achieve the best match in the same sense as with Equation 10.47. Nevertheless, the assumption is false: only vector **b** of the base image area is constant (and thus even its square length given by the last term). The content of **a** is varied when the optimal similarity is sought, and therefore omission of the first term may substantially influence the similarity measure. Any attempt to find the best correspondence for **a** taken from an image with variable local mean (as is naturally common), would be unreliable and usually fails, leading to false correspondences.

The interpretation of the expression as correlation deserves explanation. As explained in Section 1.4.2, the correlation is the mean of products of two stochastic variables, while the criterion (Equation 10.49) is applied to deterministic images. However, when the vectors **a**, **b** may be considered realizations of homogeneous fields A, B, where all pixels in A (or in B) have the same statistics, then all pixel values of an image may be considered realizations of the same stochastic variable. Then the normalized sum on the right of Equation 10.49 is an average, approximating the mean of products of corresponding realizations of those two variables. Thus, the last expression really is an estimate of the correlation between the areas in this sense. It can be shown (below) that it is a measure of similarity providing that the image components have zero mean in the frame of the compared areas, which typically is not fulfilled in natural images.

Equation 10.49 may be rewritten as

$$\sum_{i=1}^{N}(a_i b_i) = \sum_{i=1}^{N}(a_i - \bar{a})(b_i - \bar{b}) + \sum_{i=1}^{N}\bar{a}\bar{b}, \quad \bar{a} = \frac{1}{N}\sum_i a_i, \quad \bar{b} = \frac{1}{N}\sum_i b_i. \tag{10.50}$$

The local means clearly do not carry any information on similarity of the areas, but they may substantially influence the criterion value. When the second term, spoiling the similarity criterion due to variability of $\bar{a}$, is removed, the *covariance criterion* is obtained,

$$C_{CV}(\mathbf{a},\mathbf{b}) = \sum_{i=1}^{N}(a_i - \bar{a})(b_i - \bar{b}), \tag{10.51}$$

which naturally gives better results than C_C. Obviously, it is sufficient to modify only one of the areas by subtracting the mean; it is easier for the invariant area **b**.

Returning to the geometrical interpretation in the N-dimensional space, we may suggest another (dis)similarity criterion: the difference angle δ between the image vectors **a** and **b**. Obviously, not only identical vectors have identical orientation; the same applies also to vectors describing images of which the intensities differ by a multiplicative factor. Thus, the sensitivity of the criterion to linear contrast changes is zero, and it has been shown experimentally that this convenient property is basically preserved even for nonlinear monotonous contrast transforms. Instead of minimizing the angle, its cosine may be maximized, as only the absolute value of δ matters. Taking into account that $\mathbf{ab} = |\mathbf{a}|\,|\mathbf{b}|\,\cos(\delta)$, the *angle criterion* (C_A, also called *cosine criterion*) may be formulated as

$$C_A(\mathbf{a},\mathbf{b}) = \frac{\mathbf{ab}}{|\mathbf{a}|\,|\mathbf{b}|} = \frac{\sum_i a_i b_i}{\sqrt{\sum_i a_i^2}\sqrt{\sum_i b_i^2}}. \tag{10.52}$$

This criterion [31] is naturally normalized to the range <–1, 1> and can therefore be used for estimating the absolute degree of similarity. The negative range is excluded because the intensities are nonnegative. However, when only relative evaluation of similarity is needed, as in the search of optimal match, the absolute values are unnecessary and then the criterion may be simplified to

$$C'_A(\mathbf{a},\mathbf{b}) = \frac{\mathbf{ab}}{|\mathbf{a}|} \tag{10.53}$$

because the length of the constant vector **b** does not influence the maximum position. This similarity criterion proves very reliable, though the differences in its values are small—the optimum must be found by careful evaluation of the criterion.

When each of the vectors is modified by subtracting the mean value from all its elements, we obtain the *norm-cosine criterion*, the vector space interpretation of another widely used *correlation coefficient* (CC) criterion:

$$C_{CC}(\mathbf{a},\mathbf{b}) = \frac{\sum_i (a_i - \bar{a})(b_i - \bar{b})}{\sqrt{\sum_i (a_i - \bar{a})^2 \sum_i (b_i - \bar{b})^2}}, \quad \bar{a} = \frac{1}{N}\sum_i a_i \quad \bar{b} = \frac{1}{N}\sum_i b_i. \tag{10.54}$$

The resulting angle differences are increased in comparison with Equation 10.52 as the non-negativity constraint is removed. This way, the differences in the criterion values are enhanced compared to the plain angle criterion (Equation 10.52); under certain circumstances, it should then be possible to detect the maximum of the criterion simply by crossing a certain *a priori* threshold instead of proper optimization. The CC criterion is similar to the angle criterion (Equation 10.52), but more complicated as the mean values of areas must be determined and subtracted (repeatedly for every **a** in a single sequence of estimates). The reasoning for the correlation-related name of the criterion is similar to that following Equation 10.49. Similarly as for the previous criterion, the simplified version, not preserving normalization to the range <–1, 1>, but with an equally positioned maximum, can be used in optimization.

In spite of higher complexity, extensive experiments on natural images [31] have shown that the CC criterion was less reliable (had a higher rate of falsely determined correspondences) than the simpler angle criterion C_A. According to the mentioned study on comparison of different similarity criteria used in optimization, the most reliable by far was the angle criterion under all varied conditions, followed with a considerable gap by Euclidian distance and correlation coefficient criteria, each of them suitable under certain circumstances. The covariance criterion was the worst, though still usable, on the difference to the classical correlation criterion, which naturally turned out useless.

It has been mentioned that the compared images may even be vector valued (multimodality or multispectral images), i.e.,

$$\mathbf{A} = [\mathbf{a}_{i,k}] = [(a_1, a_2, \ldots, a_n)_{i,k}], \quad \mathbf{B} = [\mathbf{b}_{i,k}] = [(b_1, b_2, \ldots, b_n)_{i,k}]. \tag{10.55}$$

Realizing that the order of components in the compared vectors is irrelevant as long as the same order is kept in all compared vectors, the new vectors to be compared can be formed simply by concatenating the partial vectors provided for the areas by individual modalities, as

$$\mathbf{a} = \left[\mathbf{a}_1^T, \mathbf{a}_2^T, \ldots, \mathbf{a}_n^T\right]^T, \quad \mathbf{b} = \left[\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_n^T\right]^T. \tag{10.56}$$

It can easily be shown that

$$\mathbf{ab} = \left[\mathbf{a}_1^T, \mathbf{a}_2^T, \ldots, \mathbf{a}_n^T\right]\left[\mathbf{b}_1^T, \mathbf{b}_2^T, \ldots, \mathbf{b}_n^T\right]^T = \mathbf{a}_1\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2 + \cdots + \mathbf{a}_n\mathbf{b}_n \tag{10.57}$$

and

$$|\mathbf{a}| = \left| \left[\mathbf{a}_1^T, \mathbf{a}_2^T, \ldots, \mathbf{a}_n^T \right]^T \right| = \sqrt{|\mathbf{a}_1|^2 + |\mathbf{a}_2|^2 + \cdots + |\mathbf{a}_n|^2}. \quad (10.58)$$

Thus, all the above criteria, based on the dot product and the vector length, can easily be generalized to vector-valued images. They can even be efficiently computed using the partial criteria calculated individually for the (single-color or single-modality) image components. Particularly, the angle criterion in Equation 10.53 may be calculated as

$$C'_A(\mathbf{a}, \mathbf{b}) = \frac{\mathbf{ab}}{|\mathbf{a}|} = \frac{\mathbf{a}_1\mathbf{b}_1 + \mathbf{a}_2\mathbf{b}_2 + \cdots + \mathbf{a}_n\mathbf{b}_n}{\sqrt{|\mathbf{a}_1|^2 + |\mathbf{a}_2|^2 + \cdots + |\mathbf{a}_n|^2}}. \quad (10.59)$$

As it can be seen, the individual terms in the fraction allow calculating the individual angle criteria for each modality (or color) separately, besides obtaining the joint criterion value, at only a small overhead computational effort. This property can be utilized for a kind of check of reliability: the agreement of all or the majority of partial results with the result of the joint criterion may be considered a measure of likelihood of the determined correspondence.

All the above criteria may be realized by means of generalized nonlinear matched filters [31] (Section 10.2.1). It is obvious that Euclidean distance criterion and criteria derived from it work well only when there is no contrast transform among the compared images. On the contrary, angle criterion and correlation coefficient are insensitive to linear contrast transform and have proved robust even to nonlinear monotonous nondecreasing transforms in experiments. From the medical image analysis point of view, they are suitable for intramodality comparisons, where these conditions are practically fulfilled.

In the intermodality comparisons, complex and ambiguous contrast transforms among different modality images must be expected, even when their geometrical similarity is high. Differently formulated criteria are then needed. One group of similarity criteria that forms a certain transition to intermodal similarity criteria (Woods, 1992, 1993 as mentioned in [72]) cannot easily be presented in the vector space frame: the *ratio image uniformity* (RIU), also called *variance of intensity ratio* (VIR), which is based on the vector of ratios of pixel values,

$$\mathbf{r} = \left[\frac{a_1}{b_1} \quad \frac{a_2}{b_2} \cdots \frac{a_N}{b_N} \right]^T. \quad (10.60)$$

The dissimilarity criterion evaluates the relative variance of the vector elements,

$$C_{RIU}(\mathbf{a},\mathbf{b}) = \frac{\sqrt{\sum_i (r_i - \bar{r})^2}}{\bar{r}}, \quad \bar{r} = \frac{1}{N}\sum_i r_i. \tag{10.61}$$

Obviously, this criterion considers, as ideally similar, the images that differ only by a linear contrast transform; it is therefore insensitive to such transforms, like the angle criterion. To be applicable in the intermodality registration, the algorithm has been modified to accept differently modified contrast for different types of tissue recognized by their intensities in one of the modalities (usually computed tomography (CT)); the idea is then related to the concept of joint histogram–based criteria (see below).

Still another group of s.c. *local feature criteria* is based on derived parametric images. When the intensity images ${}_0\mathbf{a} = [{}_0a_i]$, ${}_0\mathbf{b} = [{}_0b_i]$ are to be compared, new parametric images ${}_k\mathbf{a} = P_k\{\mathbf{a}\}$, ${}_k\mathbf{b} = P_k\{\mathbf{b}\}$ are derived by means of some selected local operators P_k, like local (e.g., 3×3 or 5×5) average, local variance, different band-pass filters, various edge detectors, etc. The operators may even be anisoplanar; however, the same operators must treat both corresponding areas. Then, the obtained vector-valued (parametric) images

$$\vec{\mathbf{a}} = [\vec{\mathbf{a}}_1, \vec{\mathbf{a}}_2, \ldots, \vec{\mathbf{a}}_N]^T = \left[\begin{bmatrix} {}_0a_1 \\ {}_1a_1 \\ \vdots \\ {}_Ka_1 \end{bmatrix} \begin{bmatrix} {}_0a_2 \\ {}_1a_2 \\ \vdots \\ {}_Ka_2 \end{bmatrix} \cdots \begin{bmatrix} {}_0a_N \\ {}_1a_N \\ \vdots \\ {}_Ka_N \end{bmatrix}\right]^T,$$

$$\vec{\mathbf{b}} = [\vec{\mathbf{b}}_1, \vec{\mathbf{b}}_2, \ldots, \vec{\mathbf{b}}_N]^T \tag{10.62}$$

provide rather rich information on local properties of the original images in the individual pixel vectors of the parametric images. It may then be sufficient to compare just the vectors of individual pixels (positioned at $\mathbf{r}$) instead of (small) areas; the similarity measure may be defined as, e.g., the Euclidean distance of the K-dimensional pixel vectors,

$$C_{LF}(\vec{\mathbf{a}}_{\mathbf{r}_A}, \vec{\mathbf{b}}_{\mathbf{r}_B}) = |\vec{\mathbf{a}}_{\mathbf{r}_A} - \vec{\mathbf{b}}_{\mathbf{r}_B}|. \tag{10.63}$$

As the derived parametric images may carry information that is not directly dependent on the contrast scale of the original images (e.g.,

edge representation), the local-feature criteria may work even in intermodality comparisons. However, the performance of this type of criteria depends on the dimension K of pixel vectors and crucially on the choice of operators with regard to the particular type of images.

The principle of using the parametric images may be extended: the local operators can prepare image data for area-based criteria or even for global similarity evaluation. The area or global criteria may then be applied to individual parametric images instead of original images, in the hope that the preprocessing would have enhanced the features (e.g., edges) enabling a better comparison. This idea may be further generalized to submitting the above-mentioned combined vector-valued images to the area-based or global criteria; it is feasible when these criteria can accept images with vector values.

The criteria so far mentioned were primarily intended to similarity evaluation of images with similar though differing contrast, as is the case in intramodality comparison. Then it may be expected that the corresponding points in both images have approximately the same or linearly dependent values, or that there is a simple monotonous function enabling estimation of the value of the corresponding point, knowing the appropriately positioned value in the other image.

10.1.3.2 Information-Based Criteria

However, there are many instances in practice when the contrast in both compared images differs substantially; in medical applications, it concerns namely intermodality fusion. Thus, to a certain intensity in one of the compared images, several or many different intensities correspond in the other image, often very different from the first image value. This situation can be described by a two-dimensional joint histogram, which is a generalization of the one-dimensional histogram concept mentioned in Section 2.1.1.

Let us first briefly repeat that the one-dimensional histogram of a discrete intensity image A with q levels of gray is the vector

$$\mathbf{h}^{\mathrm{A}}:\ h_l^{\mathrm{A}} = \sum_{A_l} 1, \quad l = 0,1,\ldots,q-1, \quad \mathrm{A}_l = \{[i,k]:\ ((a_{i,k} \in \mathrm{A}) = l)\}, \tag{10.64}$$

i.e., the l-th component of the histogram carries the information on how many pixels at the l-th intensity (gray) level are present in the image. A_l is thus the set of positions $\mathbf{x} = [i, k]$ of pixels a_{ik} acquiring the intensity l. Should the image be a realization of a homogeneous and ergodic stochastic field, the components of normalized histogram

approximate the probability of a generic pixel in image A acquiring a certain gray value,

$$p_A\{l\} \approx \frac{1}{N} h_l^{\mathrm{A}}, \tag{10.65}$$

where N is the total number of pixels in A. Let us denote the variable acquiring this distribution as A. Similarly, we obtain estimates of probabilities for the analogous variable B in image B,

$$p_B\{l\} \approx \frac{1}{N} h_l^{\mathrm{B}}. \tag{10.66}$$

These are *marginal histograms* and *marginal probabilities* with respect to the system of the two stochastic variables A and B.

The two-dimensional *joint histogram* (Figure 10.10) for two equally sized gray-scale images A and B is a matrix $\mathbf{h}^{\mathrm{AB}}$ of the size $q \times r$, where q is the number of gray shades in A and r is a similar quantity in B. The elements of the matrix are counts of pairs of equally positioned pixels in A and B that have a particular combination of intensities. More precisely, the (l, m)-th element of the histogram carries the information on the total number of different positions

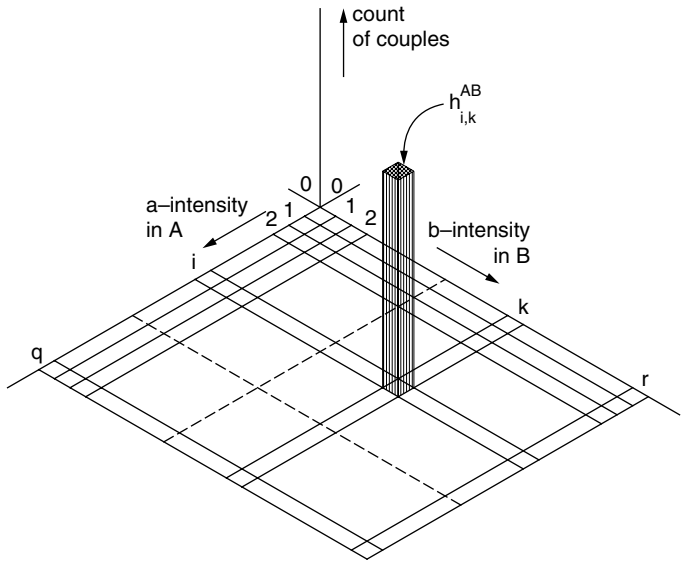


Figure 10.10 Joint histogram of two images schematically (only a single histogram value out of all qr values is depicted).

$\mathbf{x} = (i, k)$, such that the two pixels positioned at $\mathbf{x}$ in A and in B have the intensities l and m, respectively; thus

$$\mathbf{h}^{\mathrm{AB}}:\ h_{l,m}^{\mathrm{AB}} = \sum_{D_{l,m}} 1, \quad l = 0, 1, \ldots, q-1, \quad m = 0, 1, \ldots, r-1,$$
$$D_{l,m} = \{[i,k]:\ ((a_{i,k} \in \mathrm{A}) = l) \wedge ((b_{i,k} \in \mathrm{B}) = m)\}. \tag{10.67}$$

Like with one-dimensional histograms, a two-dimensional histogram is created by fully searching all image positions, but the information on concrete positions $\mathbf{x}$ is abandoned when incrementing the respective counts. The joint histogram therefore provides global information on relations between intensities in one and the other image at corresponding positions without concretely specifying the positional information.

A few simple examples (Figure 10.11) should help to interpret the joint histogram forms. When A ≡ B, obviously all the nonzero counts are located on the main diagonal of the histogram matrix, acquiring the values of the respective one-dimensional histogram (Figure 10.11a). When B is completely (deterministically) derived from A by a point operation (simple contrast transform $b = f(a)$), every l-th gray shade in A will generate only a single value in B and the nonzero counts will be situated on a (discretized) curve $b = f(a)$ (Figure 10.11b). Notice that in both cases, image A carries complete information on image B (when neglecting the quantizing errors). The reverse is obviously valid only if the function is monotonous.

The joint histogram clearly indicates the case when a single intensity value in A corresponds to a number of different intensities in B (and vice versa). Such a situation appears when image B is geometrically transformed (e.g., shifted or rotated) with respect to A, even when both images were originally identical (Figure 10.11c, d).

More importantly, such diversification in intensity correspondences arises also when different modality images of the same object are analyzed (e.g., CT vs. MRI). Then a single gray value in image A may describe different tissue types, which, however, may be discriminated in another modality (image B) by different gray shades, and vice versa. Then, even if the imaging geometry in both images is identical, the nonzero counts in the joint histogram will be scattered over the matrix. However, if there is a likeness between the images detectable by a human observer, it may be expected that there are also strong (although ambiguous) relations between the intensity values. Such relations are expressed by nonzero counts

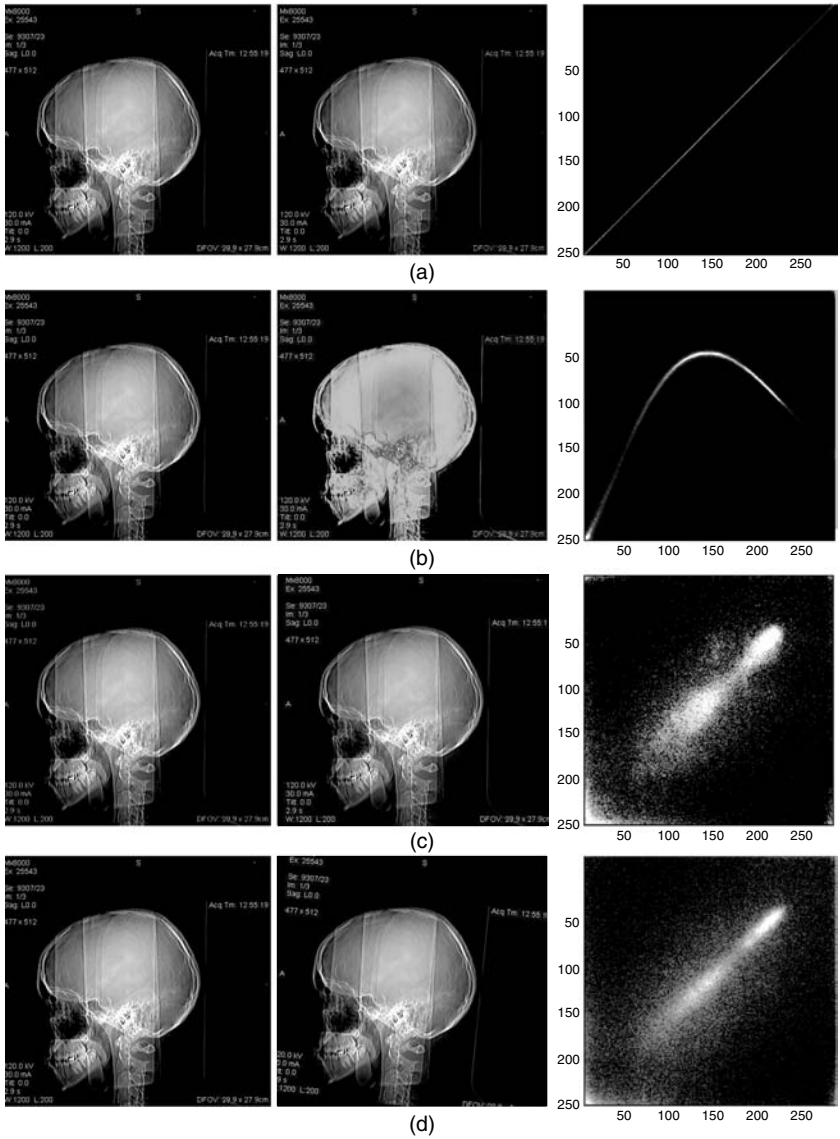


Figure 10.11 Examples of joint histograms. Left and middle columns: images A and B. Right column: joint histograms. From above: (a) identical images, (b) contrast-transformed images, (c) mutually horizontally shifted identical images, and (d) mutually rotated identical images.

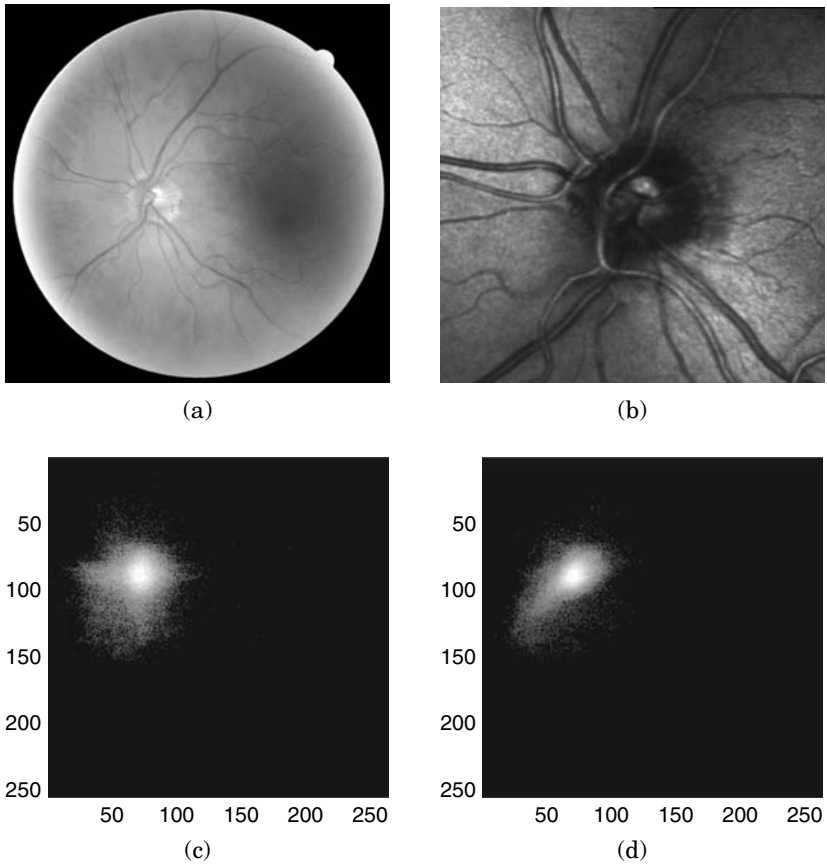


Figure 10.12 Retina images in two different modalities: compounded image from Heidelberg retina tomograph (b), image from optical fundus camera (a), and the corresponding joint histograms in (c) grossly misregistered and (d) registered states. Image (a) is resized and cropped before registration [45].

cumulated in relatively restricted areas (Figure 10.12d). This is the basic property of the joint histogram, enabling detection and evaluation of the similarity between images even if their gray scales are substantially different, so that diverse features are emphasized or suppressed differently in each of the modalities. Of course, when the images are misregistered, this cumulation is less pronounced; the less pronounced the more severe is the misregistration of the images (Figure 10.12c). Let us observe that the size of the compared areas of images is arbitrary (though identical in both images); thus, the

joint histogram may be constructed for complete images as well as for small "feature" areas.

Though interesting, the visual properties of joint histograms are too complex to allow simple evaluation of similarity of two compared images. A similarity criterion should be found that expresses in a simple way what can be deduced from the joint histogram. The questions involved are: Does a particular intensity of a pixel in A predicate anything about the intensity of an identically situated pixel in B? If so, how should the predictive power be quantified? Can this predictive power be evaluated by a single quantity for the complete image? Is there a relation between visual similarity of images and the prediction possibilities? Ultimately, a simple scalar measure based on the joint histogram is needed that might be used for similarity quantification, as essential, e.g., in image registration via optimization. The theory of information offers several such measures, should we interpret the histograms in terms of probabilities, i.e., when taking the images as realizations of stochastic fields.

Besides the individual variables A, B with distributions shown in Equations 10.65 and 10.66, we shall now define a new joint variable (A, B) acquiring the vector value (l, m), consisting of intensity in A and intensity in B for identically positioned pixels in both images. Obviously, this variable takes on $q \times r$ different vector values with the probabilities

$$p_{AB}\{l,m\} \approx \frac{1}{N} h_{l,m}^{\mathrm{AB}}. \tag{10.68}$$

The information content of a chain of symbols chosen from a finite alphabet (as, e.g., a chain of discrete and quantized pixels in the scanned representation of an image) is given by the probability distribution among pixel values. Intuitively, if only a single pixel intensity was possible while others were excluded, no information would be conveyed by the image, as its content would be known ahead of time. On the other hand, the maximum information is conveyed when no symbol is preferred, i.e., when the probability is uniformly distributed and the uncertainty in the estimate of the next symbol is maximum; the uncertainty, and consequently the information, increases with the number of symbols in the alphabet. All other cases of probability distribution are in between, and the information measure should obviously be continuous in probabilities. During the 1940s Shannon showed that a function evaluating the average amount of information conveyed by a symbol

in a set of symbols $\{a_l\}$ with given probabilities $\{p(a_l)\}$ is the *entropy*,

$$H_A = -\sum_{l=1}^{q} p(a_l) \log p(a_l). \tag{10.69}$$

It can easily be checked that the previous extreme cases and intuitive requirements correspond to the properties of this function.

So far, we have three entropies as information measures, H_A, H_B, and H_{AB}, the last being

$$H_{AB} = -\sum_{l=1}^{q} \sum_{m=1}^{r} p(a_l, b_m) \log p(a_l, b_m). \tag{10.70}$$

While the first two may be called individual entropies, the latter is the *joint entropy*. They are proportional to the information carried by the individual image A, the individual image B, and the image union (A,B), respectively. A question now arises: Do these information measures express anything about the relation between both images?

Intuitively, the information carried by the image union H_{AB} should be maximally equal to the sum of information of both images individually, $H_A + H_B$; the equality arises in case of complete independence of both images. The more the images are dependent, the higher is the predictability of B given A (and vice versa), and the union then carries obviously less information. It may be demonstrated on the two following extremes: In the case A ≡ B, obviously $H_A = H_B$, and $H_A = H_{AB}$, as the joint histogram has only q nonzero elements on the main diagonal. The other extreme is A ≠ B with uniform joint histogram, i.e., with q^2 nonzero entries (providing that $q = r$), which implies equally probable q^2 intensity couples (l, m); the joint entropy is then

$$H_{AB} = -\sum_{l} \sum_{m} \frac{1}{q^2} \log\left(\frac{1}{q^2}\right) = 2\log(q). \tag{10.71}$$

It can easily be shown that the differing images A, B must both have (almost) uniform one-dimensional histograms approximating uniform probability distribution when the joint histogram is to be uniform. Then

$$H_A = H_B = -\sum_{l} \frac{1}{q} \log\left(\frac{1}{q}\right) = \log(q). \tag{10.72}$$

Obviously, in the first case, all the information is contained in image A, and as the joint entropy shows, it is equal to the information contained in the union (A, B). In the second case, no information on B is contained in A (and vice versa), and consequently, the image union carries doubled information (Equation 10.71) compared to each of its components.

This leads to the definition of a new quantity, called *mutual information* (MI), as the difference between the sum of information in individual images and the joint information in the union (A,B),

$$I_{AB} = H_A + H_B - H_{AB}. \tag{10.73}$$

This quantity is obviously a good measure of the above-mentioned predictability of values of B knowing A (and vice versa). In the above two examples, when A ≡ B, the predictability is maximum and $I_{AB} = H_A = H_B$; on the other hand, when A and B are random and completely independent, $I_{AB} = 0$.

The predictability can now be explained more rigorously in terms of conditional probabilities. When images A and B, and thus the variables A and B, are dependent, knowing the value of variable A changes the probability distribution of B from the *a priori* probability $p_B\{m\}$ to the conditional probability distribution $p_{B|A}\{m \,|\, (A = l)\}$, or briefly $p_{B|A}\{m \,|\, l\}$, which is approximated by the normalized l-th row in the joint histogram,

$$p_{B|A}\{m \,|\, l\} \approx \frac{1}{N_l} h^{\mathrm{AB}}_{l,m}, \quad N_l = \sum_{m=1}^{r} h^{\mathrm{AB}}_{l,m} = h^{\mathrm{A}}_{l}. \tag{10.74}$$

Conversely, $p_{A|B}\{l \,|\, m\}$ is approximated by the normalized m-th column of $\mathbf{h}^{\mathrm{AB}}$,

$$p_{A|B}\{l \,|\, m\} \approx \frac{1}{N_m} h^{\mathrm{AB}}_{l,m}, \quad N_m = \sum_{l=1}^{q} h^{\mathrm{AB}}_{l,m} = h^{\mathrm{B}}_{m}. \tag{10.75}$$

Here, we used simplified notation, supposing $a_l = l$, $b_m = m$, as in Figure 10.10. Each of these conditional distributions has its own entropy,

$$H_{A|m} = -\sum_{l=1}^{q} p_A(l \,|\, m) \log p_A(l \,|\, m), \quad H_{B|l} = -\sum_{m=1}^{r} p_B(m \,|\, l) \log p_B(m \,|\, l), \tag{10.76}$$

which may be denoted as individual conditional entropy (for the condition m or l, respectively). The mean value of $H_{A|m}$ with respect to *a priori* probabilities of B becoming m is

$$H_{A|B} = -\sum_{m=1}^{r} p_B(m) H_{A|m} = -\sum_{m=1}^{r}\sum_{l=1}^{q} p_{AB}(l,m) \log p_{A|B}(l|m); \qquad (10.77)$$

this is the mean conditional entropy, or simply the *conditional entropy* $H_{A|B}$. Similarly, the conditional entropy $H_{B|A}$ is

$$H_{B|A} = -\sum_{l=1}^{q} p_A(l) H_{B|l} = -\sum_{l=1}^{q}\sum_{m=1}^{r} p_{AB}(m,l) \log p_{B|A}(m|l). \qquad (10.78)$$

The last conditional entropy can be interpreted as the measure of the mean information remaining in image B when knowing A. Thus, if B can be derived deterministically from A, the conditional entropy is zero; on the other hand, it is maximized when both images are independent. It can be shown that

$$I_{AB} = H_A - H_{A|B} = H_B - H_{B|A}. \qquad (10.79)$$

The mutual information thus shows how much is the *a priori* information content of one image changed by obtaining the knowledge of the other image.

In order to maintain a straightforward explanation, the considerations were limited to two-dimensional images. It is nevertheless easy to generalize the conclusions to three-dimensional data: when two three-dimensional data blocks are compared, the joint histogram remains obviously two-dimensional, although the voxel positions are given by $\mathbf{x} = (i, k, n)$; thus the derived entropies as well as mutual information are given by identical formulae, as before. As already mentioned, the use of MI criterion is not limited only to original (intensity) image data: it can evidently also be applied to derived (parametric) images, as, e.g., edge representation, with the advantage that differing feature representations with diverse intensity profiles in different modalities need not preclude recognition of similarity.

The mutual information has proved to be a rather good criterion of similarity; so far, it seems to be the best choice for multimodality fusion. However, mutual information is only a special case from a more general class of *information-theoretic measures*; other criteria derived from the entropies have been

suggested and tested [16], [52], but on the difference to MI they have not reached a widespread acceptance yet. Nevertheless, the background remains the same, only the resulting entropies are combined differently, as, e.g., in

$$\tilde{I}_{AB} = \frac{H_A + H_B}{H_{A,B}}. \tag{10.80}$$

This criterion, obviously closely related to MI, can be considered a quantity normalized with respect to the size of the compared regions; it has been shown to be more robust with respect to changes in the overlap area of A and B in optimization-based registration (see Section 10.3.1).

10.2 DISPARITY ANALYSIS

Disparity analysis aims at determining shifts among corresponding points in two (or more) images. Its use is manifold: reconstruction of spatial objects or scenes based on stereo image couples (or sequences), detection and evaluation of temporal developments of spatial relations, e.g., of growth and motion in the FOV, etc. Disparity analysis in the narrower sense usually concerns images that are preliminarily registered so that the gross misregistration need not be considered. Thus, the mean of disparities should be approximately zero, or there should be a reference frame with zero disparities; otherwise, the precision of disparity analysis may be influenced. Disparities are usually determined with pixel accuracy, but subpixel resolution is possible in principle, providing that the input images are of high quality.

The applications of disparity analysis thus have three sometimes partly merging or overlapping steps:

- *Preprocessing*: Image enhancement (e.g., noise suppression, sharpening, band-pass filtering, conversion into a parametric image, e.g., edge representation, etc.) and preregistration.
- Actual *disparity analysis* yielding the disparity map based on a pair (or series) of preprocessed images.
- *Exploitation* of the disparity information to derive the final results: three-dimensional scene, motion, or time development, detection of defects in a spatial structure, etc.

We shall discuss the second step here; preprocessing is the topic of Chapter 11, and preregistration uses the techniques described in

Section 10.3. The applications in stereo-based surface determination and in motion estimation are mentioned in Sections 10.4.5 and 10.4.6.

10.2.1 Disparity Evaluation

10.2.1.1 Disparity Definition and Evaluation Approaches

Let us have two images A and B to be compared with respect to differences in position of corresponding points. Image B will be used as the base image, to which the differences are related. When $\mathbf{r}_A$, $\mathbf{r}_B$ are position vectors of a feature present in both images, the two-dimensional difference vector $\Delta\mathbf{r} = \mathbf{r}_B - \mathbf{r}_A$ is usually denoted as *disparity* at $\mathbf{r}_B$. It is a vital step for every disparity-based application to find *reliably* a necessary number of different correspondences with a needed degree of accuracy. The required number of correspondences starts from a few for simple orientation tasks and rigid registration and increases to several tens for global (whole-image) flexible transforms; on the other hand, it may reach up to a fraction, or even the total, of the number of image pixels in the case of detailed disparity analysis. In the latter case of a dense set of disparities among which it may be further interpolated, the resulting spatial function $\Delta\mathbf{r}(x, y)$ or, in discrete representation $\Delta\mathbf{r}_{i,k}$ is called the *disparity map*.

Obviously, the disparity map may be represented as a vector-valued image (or a pair of gray-scale images, representing x- and y-disparities, or absolute disparity and its direction). Hence, the disparity map itself may be considered an output of image fusion, independently of what it would be subsequently used for. Nevertheless, it is usually utilized as the input for a higher-level analysis, e.g., estimation of motion or shape and size changes of image subareas (Section 10.4.6), or for three-dimensional surface determination or scene reconstruction in stereo analysis (Section 10.4.5).

The set of points in the base image, for which the correspondences are to be determined, is given either by preliminarily determined prominent features (edges, corners, crossings, branching, singular point objects, etc.) or by a regular grid of an *a priori* chosen density, regardless of the image content.

When the set of correspondences is to be based on recognized features, the recognition may be either done interactively by a human operator or automatically based on image analysis. The manual approach—when the features are selected not only in the base image

B but also pointed to approximately in A—is usually much more reliable; the disparity analysis then consists of elaborating the suggested correspondences to improve the disparity precision. Automatic finding of features is still not fully dependable. The feature detection itself, combining different analytic techniques (Chapter 13), is not entirely reliable and depends on the image type, so that some important features often pass undetected in one or both images or false detection adds nonexistent features. Moreover, occlusion or glitter may preclude seeing a feature in one of the images. All this makes the two detected feature sets in A and B usually partly incompatible with respect to finding pairs of corresponding points. The second step of feature-based disparity analysis then consists of approaching a difficult problem of finding the valid correspondences among all possible combinations between both sets of landmarks. Finally, the feature-based disparity maps, whether detected automatically or manually, usually suffer with a rather sparse network of correspondences, so that the disparities for other points can only be approximated by interpolation.

The other approach—determining correspondences on a regular grid—eliminates the difficult and unreliable phase of feature detection as well as the following complex determination of valid correspondences. The patterns to be found in A are now simply small areas around the node points in B; however, nothing is *a priori* known on the corresponding areas in A except that these areas may be expected in a vicinity of the respective node points in A. This approach relies on local patterns and textures in both images being adequately similar in corresponding areas, while at the same time being sufficiently diverse in other image areas not to allow confusion of the similarity criterion.

In every case, the fundamental problem of disparity analysis reduces to precise determination of the vector $\mathbf{r}_A$ in A corresponding to a known position $\mathbf{r}_B$ in B. A defined (small) area (e.g., a square), surrounding the node point $\mathbf{r}_B$ in B, carries the pattern to be found in image A. By shifting the pattern as a mask on image A in the vicinity of $\mathbf{x}_A = \mathbf{r}_B$, the best match of local content of image A with the mask pattern is found at a point $\mathbf{r}_A$. The optimum disparity $\Delta\mathbf{r} = \mathbf{r}_B - \mathbf{r}_A = (\Delta x, \Delta y)$ is thus determined for the point $\mathbf{r}_B$. The match is optimal in the sense of the chosen similarity criterion; therefore, finding a single disparity is an optimization problem—to minimize or maximize the criterion dependent on $(\Delta x, \Delta y)$ or, in the discrete version, on index differences $(\Delta i, \Delta k)$. Besides the choice of similarity criterion, two parameters must be defined: the size of the areas taken as patterns and the extent around $\mathbf{x}_A$, in which the pattern is searched. As for the pattern size, it is obvious that the greater the area, the less

is the probability of finding an improper correspondence due to a seeming similarity; on the other hand, a large area may cover a region with different disparities. This may preclude similarity recognition and, when it is at all recognized, obviously worsens the disparity resolution. The second parameter is no less critical: too large a search extent may lead to finding a distant false correspondence due to an accidentally higher seeming similarity than the proper one; too small a search extent may have paradoxically the same result, as the pattern may be included in the searched area only incompletely. Both parameters thus influence the reliability of found correspondence; moreover, they obviously influence the computational requirements. Further discussion may be found in the next section on disparity map computation.

Usually, the images subject to disparity analysis are provided by a single modality, or both may be preliminarily formed as vector-valued images obtained by fusion from more modalities (e.g., both A and B are fused images from electronic microscopy containing backscattered electron imaging (BEI) and secondary electron imaging (SEI) components). Therefore, usually the intensity-based similarity criteria are used, namely, the cross-correlation coefficient (Equation 10.54) or the simpler and often better angle criterion (Equation 10.53 or 10.52). Using the mutual information criterion in disparity analysis is worth the much more demanding computation only when there are substantial contrast differences between A and B, e.g., due to different illumination.

The above point-wise (locally optimizing) approach may suffer with falsely determined correspondences that may accumulate and finally provide a partly (or even completely) false disparity map. This is the reason for the high requirements concerning the reliability of the similarity criterion used. The prevention of the possible failures is in a running check of topographic consistency of the disparities during the complete analysis. An alternative global approach providing the complete disparity map as a set of parameters may include all the physical constraints into the global optimization procedure. The methodology stating the stereo reconstruction as a global optimization problem solved by dynamic programming is treated in detail in [10].

In exceptional cases, the local disparity as a mere shift may be substituted by a more generic geometrical transform, e.g., by adding rotation to the shift; in such a case, a multidimensional vector of transform parameters expresses each local disparity. We will limit the discussion here to the case of mere space-variant two-dimensional shift, as mostly sufficient when the images are well preregistered. The reader interested in more general disparity analysis will find some related ideas in Section 10.3 on registration.

The generalization of the above considerations to three-dimensional disparities in spatial image data is straightforward and does not need special treatment.

10.2.1.2 Nonlinear Matched Filters as Sources of Similarity Maps

The local optimization mentioned above requires that the similarity measure of the couple $\mathbf{r}_A$, $\mathbf{r}_B$, belonging to a fixed $\mathbf{r}_B$, can be determined for any point $\mathbf{r}_A$ of the search area in image A. As far as the full-search optimization is implemented for every $\mathbf{r}_B$ (which turns out reasonable, taking into account the complex shape of the similarity function and its relatively small extent), the similarity criterion must be known for each sample point of the search area in A. This way, a two-dimensional *similarity map* is defined for each particular $\mathbf{r}_B$, as in Figure 10.13. The local optimization then searches the optimum of the map; this has to be done repeatedly for each $\mathbf{r}_B$ of the chosen grid. An efficient algorithm for providing the similarity maps is thus needed. The intensity-based criteria in Equations 10.47, 10.49, and 10.51 to 10.54 all can be implemented via generalized two-dimensional *nonlinear matched filters* [31], [27]. The matched filter concept comes from communications; the purpose of the filter is to detect (localize in time) an occurrence of a known finite-length signal in the received noisy signal. We have the same problem in two dimensions: to localize an occurrence (or rather the best fit) of a sought pattern from B on image A; the pattern may naturally be distorted here. The shape of the sought pattern may be arbitrary, but for practical reasons usually only rectangular segments are searched for, characterized by intensity distribution, as will be the case below. In the communication application, the situation is simpler, as even the short-time mean value of the received signal is zero and the noise is usually broadband, if not white. Thus, the linear correlation filter (one-dimensional counterpart of Equation 10.49) is exclusively used. As explained above, such a filter is usually useless in disparity analysis and its nonlinear modifications must be used. Application of such filters provides automatically the complete similarity maps of the searched areas.

The description of matched filters in matrix notation (instead of complicated double-sum equations) paraphrases that in [27]. Let us have a matrix $\mathbf{A}$ sized $M \times N$ (the search area in A) and a matrix $\mathbf{B}$ sized $m \times n$ (the pattern from image B), $M, N >> m, n$. The goal is to find the index position vector $\mathbf{p} = (i, j)$, so that the submatrix $\mathbf{C}$

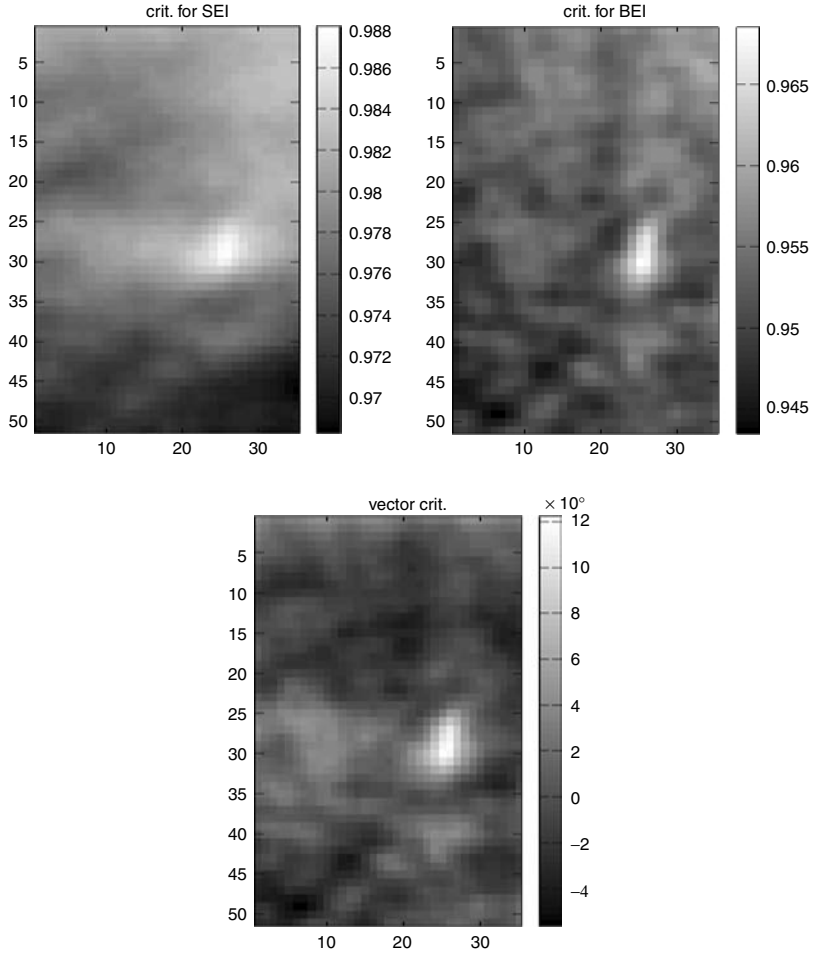


Figure 10.13 Example of similarity maps (angle criterion) of mutually inclined images in scanning electron microscopy, based on: (a) SEI images, (b) BEI images, and (c) joint bimodal vector images, for a particular position $\mathbf{r}_B$. (From Jan, J., Janova, D., *Mach. Graphics Vision*, 10, 261–288, 2001. With permission.)

of $\mathbf{A}$, $\mathbf{C} = \mathbf{A}_{i..\,i+m-1,\,j..\,j+n-1}$ is maximally similar to $\mathbf{B}$ in the sense of the chosen similarity criterion F,

$$\mathbf{p} = \arg\max_{\mathbf{p}} F(\mathbf{C}(\mathbf{p}), \mathbf{B}). \tag{10.81}$$

A mild geometric and intensity distortion of the pattern in $\mathbf{C}$ with respect to its counterpart in $\mathbf{B}$ should be allowed, though it is limited by the sensitivity of the used criterion to geometric transforms (e.g., rotation) and to contrast differences. When the distortion is acceptable, the sought parameters are just the shift components i, j of the local disparity.

The linear versions of the matched filter can be described by two-dimensional discrete convolution (denoted by *). For the *plain matched filter* (correlation as in Equation 10.49), the similarity map $\mathbf{Y}$ is

$$\mathbf{Y} = \mathbf{H} * \mathbf{A}, \quad \mathbf{H} = \mathbf{B}'', \tag{10.82}$$

where the point-spread function (PSF) of the filter is the matrix $\mathbf{B}''$, meaning $\mathbf{B}$ rotated by 180°. The *covariance filter* (simplified Equation 10.54) may be implemented similarly, with the filter's PSF modified by subtracting the average of elements of $\mathbf{H}$ from all its elements,

$$\mathbf{Y} = (\mathbf{H} - \bar{h}) * \mathbf{A}, \quad \bar{h} = \frac{1}{mn} \sum_{i=1}^{m} \sum_{k=1}^{n} h_{i,k}. \tag{10.83}$$

The nonlinear two-dimensional filters are all based on the plain matched filter (Equation 10.82), but its output is modified pixel by pixel by either subtraction of or division by some auxiliary quantities derived from (suitably modified) matrix $\mathbf{A}$. The below given formulae, complemented with algorithmic descriptions, can easily be derived from the above-mentioned expressions. The *Euclidean distance filter* (Equation 10.47) yields the similarity map

$$\mathbf{Y} = \mathbf{H} * \mathbf{A} - \mathbf{P}/2, \tag{10.84}$$

where $\mathbf{P}$ is the resulting matrix of the sequence: squaring elements of $\mathbf{A}$ and convolving the result with the matrix of ones, sized as $\mathbf{H}$.

The *cosine filter* according to Equation 10.53 responds with

$$\mathbf{Y} = \left[\frac{q_{i,j}}{\sqrt{p_{i,j}}} \right], \quad \mathbf{Q} = [q_{i,j}] = \mathbf{H} * \mathbf{A}, \quad \mathbf{P} = [p_{i,j}], \tag{10.85}$$

where **P** is defined as above. This criterion proved superior to others in disparity analysis, as already mentioned.

For the *norm-cosine filter* (simplified cross-correlation coefficient criterion, Equation 10.54) we obtain the similarity map as

$$\mathbf{Y} = \left[\frac{s_{i,j}}{\sqrt{r_{i,j}}} \right], \qquad \mathbf{S} = [s_{i,j}] = (\mathbf{H} - \bar{h}) * \mathbf{A}, \qquad \mathbf{R} = [r_{i,j}], \quad (10.86)$$

where **R** is the matrix resulting from the sequence: subtracting the average $\bar{a}$ from all elements of **A**, squaring the elements of the result, and convolving the resulting matrix with a matrix of ones, sized as **H**.

The interpretation of similarity criteria computation as nonlinear matched filtering and in turn the expression of the nonlinear filters as modifications of linear ones allows a substantial speedup of computations by calculating all the convolutions (or correlations) via frequency domain, exploiting the convolutional property of two-dimensional DFT. An additional possibility enabling much faster computation of the responses of nonlinear filters is to provide the auxiliary quantities beforehand, by precalculating the auxiliary matrices **P** (or **R**) of partial averages as other two-dimensional convolutions. This way, the modifications of the linear matched filter output consist only of element-by-element operations between two matrices; the same matrix **P** (or **R**) can be used repeatedly for different patterns **B**.

When the compared images are vector-valued, the joint similarity criterion may be calculated using individual values of scalar-based criteria for individual component images, utilizing the principle described by Equations 10.55 to 10.59. An example of the individual and joint similarity maps for the vector-valued images from scanning electron microscopy (SEM) formed of BEI and SEI images is in Figure 10.13.

10.2.2 Computation and Representation of Disparity Maps

10.2.2.1 Organization of the Disparity Map Computation

Local optimization based on repeated evaluation of the above similarity criteria is computationally rather demanding; it is thus desirable to limit the count of these evaluations. As already discussed, the size of the pattern **B** ($m \times n$) and of the searched area **A** in A both

influence these requirements as well as the reliability of the disparity computation, and the **B** size also the resolution, i.e., the precision of the determined disparity. The general policy is thus to use initially large both **B** and **A** sizes in order to determine reliable though not very precise disparities, then successively decrease both sizes to improve gradually the precision. However, the computation would be very demanding when the full resolution disparity map was calculated for all gradually decreasing sizes of **B** and **A**. The overall computation load is then linearly dependent on the number of disparities in the map that should be dense enough to describe the differences in detail—up to the image resolution (or not much coarser).

It is tempting to use the pyramidal multiresolution approach—to start with a coarse approximation of the images, determining a sparse grid of rough disparities, which would be gradually improved on successive more detailed levels of the image representation. However, this approach only works satisfactorily when the images preserve sufficiently distinctive textures even in coarser representations at higher pyramidal levels. This fractal character is often not present (e.g., in microscopic images); the areas usually become flat on coarser resolution.

Another approach proved useful: the *leveled computation*, working with the full image resolution on all levels while gradually increasing the density of the disparity map. Initially, a sparse map of only rough but reliable disparities is calculated with large **B** and **A**. On further levels, the density of the map is gradually increasing together with decreasing the sizes of **B** and **A**; the disparity map from the previous level provides, via interpolation, initial estimates for the less reliable but more precise calculation of more dense disparities. This way, the overall low error rate may be preserved (at the level of 10^{-3}), which is necessary for most applications. An example of initial and final estimates of a disparity map, based on local similarity maps, the example of which was depicted in Figure 10.13, may be seen in Figure 10.14.

10.2.2.2 Display and Interpretation of Disparity Maps

The disparity maps are generally two-dimensional or three-dimensional vector-valued fields (the field dimensionality corresponding to the compared images), discretized on a regular grid or irregularly, dependent on the used method of disparity analysis. To display such

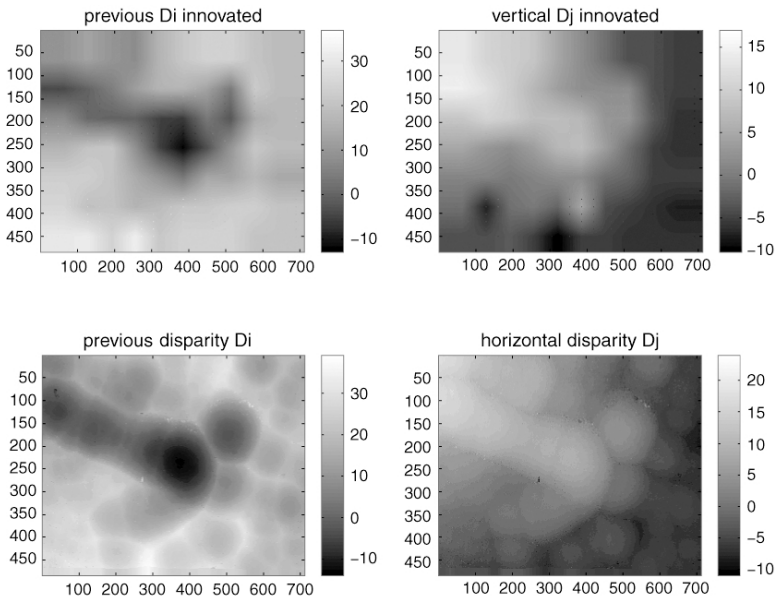


Figure 10.14 Example of disparity map estimates. Left column: initial sparse-grid estimates (interpolated to the dense grid). Right column: final maps. (From Jan, J., Janova, D., *Mach. Graphics Vision*, 10, 261–288, 2001. With permission.)

fields on the available two-dimensional screens designed to display intensity and colors, we have to code the two (three) components of the disparity vectors appropriately. The easiest way is to display each component separately in split images (Figure 10.14); however, interpreting such representation is not very graphic and requires experience.

Another possibility is displaying small arrows representing both the magnitude and direction of disparities. Such a directional field is well understandable unless the arrows are too dense; a reasonable reduction of information in case of dense disparity maps is necessary. The arrows are displayed uniquely for two-dimensional maps; for three-dimensional maps, a suitable representation of three-dimensional space must be chosen so that the arrow directions can be well identified (a comprehensive possibility is a rotating axonometric or perspective projection).

As the disparity map is a vector-valued image (regularly or even irregularly sampled), various techniques to display such images (Section 10.4.2) also represent options.

As for the interpretation of the disparity maps, it depends on the problem solved and the corresponding character of the compared images. The most frequent applications are the recovery of spatial information from two-dimensional images, taken with diversified views, via stereo techniques (Section 10.4.5), and the tracking of temporal changes in a couple (or series) of consecutive images for estimation of motion or growth (Section 10.4.6).

10.3 IMAGE REGISTRATION

The purpose of *image registration* is to provide a pair of (or several) images, possibly from different modalities, that are spatially consistent; i.e., a particular pixel in each of the images corresponds to the same unique spatial position in the imaged object (e.g., a patient). It must be admitted that the imaging geometry for each of the images is different due to possibly different physical properties and distortions inherent to different modalities; also, the imaged scene itself may change because of patient movements, physiological or pathological deformations of soft tissues between taking individual images, etc. It is therefore necessary to transform the images geometrically, in order to compensate for the distortions, so that the above consistency condition becomes fulfilled.

Usually, one of the images may be considered undistorted; it is then taken as the reference (base) image, or a reference frame is somehow defined, e.g., based on calibrated markers. Alternatively, a standard that would define the proper geometry (i.e., with a defined relation to the scene geometry) may be unavailable; the supposedly least distorted (or best-quality) image serves then as the reference. In all cases, the images (possibly excluding the reference image) are to be transformed to match the reference geometry. Because the generalization to multi-image registration, as successive matching of individual images to the reference image, is straightforward, we shall discuss only the case of two images.

Let image B be the base and image A the to-be-matched image. Let $\mathrm{X_A}$ be the field of view (the complete set of position vectors $\mathbf{x}_\mathrm{A}$) of image A; similarly, $\mathrm{X_B}$ is the spatial support of image B; the position vectors are mapped to the respective intensity values $A(\mathbf{x}_\mathrm{A})$, $B(\mathbf{x}_\mathrm{B})$,

$$\mathrm{X_A} = \{\mathbf{x}_\mathrm{A}\}, \quad \mathrm{X_B} = \{\mathbf{x}_\mathrm{B}\}; \qquad \mathrm{X_A} \mapsto \{A(\mathbf{x}_\mathrm{A})\}, \quad \mathrm{X_B} \mapsto \{B(\mathbf{x}_\mathrm{B})\}. \quad (10.87)$$

The process of registration uses a chosen geometrical transform $T_{\boldsymbol{\alpha}}$, controlled by the parameter vector $\boldsymbol{\alpha}$ that transforms image A into the transformed one, A′, which is then laid on (in other words, spatially identified with) image B,

$$T_{\boldsymbol{\alpha}} : \mathbf{x}_{\mathrm{A}'} = T_{\boldsymbol{\alpha}}(\mathbf{x}_{\mathrm{A}}), \qquad A'(\mathbf{x}_{\mathrm{A}'}) = A(\mathbf{x}_{\mathrm{A}}), \quad \mathbf{x}_{\mathrm{A}'} = \mathbf{x}_{\mathrm{B}}, \tag{10.88}$$

so that both images can be compared. Generally, the spatial range $\mathrm{X}_{\mathrm{A}'}$ differs from X_{B} so that the image content comparison is only possible on the common area; this overlap is the intersection $\Omega_{\boldsymbol{\alpha}}$ of X_{B} with the transformed image area $\mathrm{X}_{\mathrm{A}'}$,

$$\Omega_{\boldsymbol{\alpha}} = \mathrm{X}_{\mathrm{A}'} \cap \mathrm{X}_{\mathrm{B}}, \quad \mathrm{X}_{\mathrm{A}'} = \{T_{\boldsymbol{\alpha}}(\mathbf{x}_{\mathrm{A}})\}, \quad \mathrm{X}_{\mathrm{B}} = \{\mathbf{x}_{\mathrm{B}}\}. \tag{10.89}$$

Obviously, the overlap $\Omega_{\boldsymbol{\alpha}}$ is dependent on the concrete transform determined by $\boldsymbol{\alpha}$.

The registration aims at finding the parameter vector $\boldsymbol{\alpha}$ such that the registration of B with the transformed image A′ is optimal with respect to the chosen criterion of image similarity, $c(B(\mathbf{x}_{\mathrm{B}}), A'(T_{\mathbf{x}_{\boldsymbol{\alpha}}}(\mathbf{x}_{\mathrm{A}})))$. In simple cases, the optimum $\boldsymbol{\alpha}_0$ may be found by direct computation; otherwise (and more commonly), it is determined by optimization,

$$\boldsymbol{\alpha}_0 = \arg\max_{\boldsymbol{\alpha}} c\,(B(\mathbf{x}_{\mathrm{B}}), A'(T_{\boldsymbol{\alpha}}(\mathbf{x}_{\mathrm{A}}))), \quad \mathbf{x}_{\mathrm{B}}, T_{\boldsymbol{\alpha}}(\mathbf{x}_{\mathrm{A}}) \in \Omega_{\boldsymbol{\alpha}}. \tag{10.90}$$

The above-mentioned dependence of $\Omega_{\boldsymbol{\alpha}}$ on the transform parameters implies that the common area is variable during optimization, which requires that $c(\ldots)$ is normalized with respect to the size of $\Omega_{\boldsymbol{\alpha}}$. Dependent on the character of the similarity measure, the optimal value may be minimum instead of maximum.

10.3.1 Global Similarity

In Section 10.1.3, we discussed similarity measures primarily as local, evaluating similarity of small image areas, as needed, e.g., in defining pairs of landmarks for landmark-based registration, in disparity map calculation, etc. However, for registering images, a global measure of similarity is needed, which serves as the goal (or cost) function in the optimization procedure searching for the optimum parameter vector $\boldsymbol{\alpha}_0$ of the registering geometrical transform, or is used to derive an equation system allowing determination of $\boldsymbol{\alpha}_0$ directly.

As the size of compared regions in Section 10.1.3 was generally not limited, the local intensity-based criteria (all direct vector space

comparisons, including correlation, or information-based criteria) can be in principle used as the global similarity measures of complete images as well. Such a criterion evaluates the overall registration in the complete overlapping area of both images; i.e., roughly said, every detail counts. The development in the registration methods seems to lead to preference of methods based on this approach, as they are conceptually simpler—they do not need any landmark determination and have a rather generic character being more or less image content independent. We shall have primarily this approach in mind.

Nevertheless, other established global similarity criteria are also in common use, namely, point-based and surface-based criteria; these will be mentioned as well.

10.3.1.1 Intensity-Based Global Criteria

All the intensity-based criteria defined in Section 10.1.3 can serve as global optimization criteria. However, there are two important differences compared to local similarity evaluation. In the local use, we supposed that the size and shape of compared image areas A, B were fixed and that both areas were entirely inside of respective images during the complete optimizing procedure searching for the best local fit; also, we usually assumed that the intensities of both complete images were fixed during the search (no interpolation involved).

In global comparison of discrete images, the situation during optimization (see Section 10.3.2) is more complex. The overlap $\Omega_{\boldsymbol{\alpha}}$ is subject to variation in both size and shape due to different values of $\boldsymbol{\alpha}$ in individual optimization steps; not only the area of A′, but naturally also the used part of B is variable during the registration process. Further, the intensities of A′ are slightly modified in each step due to differing interpolation; the intensities of A′ on the B-grid must be interpolated from the samples of the original image A as its grid is differently distorted in each step by the varying transform. However, the definition of any chosen similarity criterion remains the same—it is still a measure of similarity between two given areas of identical size and shape (in an individual optimization step).

Thus, when the vectors $\mathbf{a}'$, $\mathbf{b}$ of intensities at all the grid node positions $\mathbf{x}_{\mathrm{A}'} = \mathbf{x}_{\mathrm{B}}$ represent the discrete compared areas, the *direct intensity-based criteria* may be generally written as

$$c_d(\mathbf{a}'(\boldsymbol{\alpha}), \mathbf{b}(\boldsymbol{\alpha})), \quad \mathbf{x}_{\mathrm{A}'}, \mathbf{x}_{\mathrm{B}} \in \Omega_{\boldsymbol{\alpha}} \tag{10.91}$$

where now both $\mathbf{a}'$ and $\mathbf{b}$ depend on the present transform parameter vector $\boldsymbol{\alpha}$ and also on the used interpolation method. Of the direct intensity-based criteria, the correlation coefficient criterion (Equation 10.54) is usually preferred. Though more complex than the angle criterion Equation 10.52 or 10.53, the more diversified shape of the function Equation 10.54 may be more appropriate for optimization than the angle criterion that is rather flat (though with usually more reliable absolute maximum). Other direct-comparison intensity-based criteria are occasionally used as well, the choice depending on the particularities of the problem. All the direct criteria require that the contrast scales of both compared images are compatible (about equal or at least approximately linearly dependent). This restricts application of these criteria to only intramodality images.

The ever-increasing interest in *information-based criteria*, primarily mutual information (Equations 10.73 and 10.79 or its modification, Equation 10.80), is obviously due to their capability to compare images with differing directly incompatible contrasts—the comparison being implicitly based on common geometrical features rather than on intensity similarity. These criteria are therefore suitable even for intermodality comparisons as needed in multimodal registration. The information-based criteria are all calculated based on the joint histogram, which must be derived, besides from original B values, from the interpolated values in A′. As the compared areas contain now, in contrast to the local use, substantial parts of images, the calculation of the joint histogram is rather demanding. The behavior of the MI criterion in the sense of smoothness can be improved by grouping the gray-shade values so that the intensity resolution is lowered, e.g., to 8 to 16 grades from the original 256 levels. This way, the histogram becomes much more compact, which not only decreases the memory requirements, but also primarily improves the histogram's informational value as the individual bins are better filled, thus decreasing the variance of the probability estimates.

Because the used interpolation methods are only approximate, the differing interpolation in the individual optimization steps may cause certain local variations in the criterion values, complicating the optimization procedure. There is another (and probably better) possibility, sc., *partial-volume interpolation* (Colignon, 1995, see [16]): instead of interpolating in the to-be-transformed image A among four (in two-dimensions) or eight (in three-dimensions) neighbors, a different approach is used. All these neighbors directly influence the histogram, though with weighted contributions, using linear interpolation weights. Thus, generally, up to four or eight bins

of the histogram are incremented, though only by fractions corresponding to the interpolation coefficients; the total histogram increment therefore always remains at 1. However, even with this approach, some problems with local variations seem to persist [68].

10.3.1.2 Point-Based Global Criteria

Another approach, in a sense simplified and historically older, considers a limited set of n pairs of corresponding points—*homologous landmarks* (also called *fiducial markers*) $\{({}_i\mathbf{x}_{A'} = T({}_i\mathbf{x}_A), {}_i\mathbf{x}_B), i = 1 \dots n\}$ in two-dimensional or three-dimensional images (${}_i\mathbf{x}_{A'}$ being a positional vector in A′, etc.). The main problem of this approach is to define such a landmark set, based on local similarity criteria and perhaps on consequent selection of valid pairs out of possible couples (see Sections 10.1.3 and 10.2.1); the task is simplified when easily identifiable artificial markers are attached to the patient.

When the point set is defined, the geometrical similarity criterion is simply based on individual error vectors

$$\mathbf{e}_i = {}_i\mathbf{x}'_A - {}_i\mathbf{x}_B = T({}_i\mathbf{x}_A) - {}_i\mathbf{x}_B, \tag{10.92}$$

and the global criterion may be based on either the lengths or square lengths of the vectors,

$$E_{abs} = \frac{1}{n}\sum_i w_i |\mathbf{e}_i| \quad \text{or} \quad E_{sq} = \frac{1}{n}\sum_i w_i |\mathbf{e}_i|^2, \quad \sum_i w_i = n \tag{10.93}$$

where the coefficients w_i reflect the reliability or importance of the point pairs.

Often, the number of conditions given by the point pairs is equal to (or greater but comparable with) the number of the registration transform parameters. Then it is possible, instead of using the criterion explicitly in optimization, to derive a system of equations for the parameters, which may be directly solved by inversion (or pseudoinversion, leading to the least square error solution, when there are more conditions than necessary).

10.3.1.3 Surface-Based Global Criteria

The surface-based registration can be interpreted as fitting two similar surfaces in the compared images. It is usually used in three-dimensional images; however, the simplified two-dimensional version

is possible as well. The surfaces can be represented by sets of points, by triangle sets or other faceted surfaces, or by parametric piecewise approximated functions, e.g., by splines. The first phase of the surface-based registration thus consists in providing the surfaces via image segmentation (see Chapter 13) that must be reliable, which imposes restrictions on the types of imaging modalities for which this approach is feasible.

In the medical imaging field, surface-based registration is primarily used in head (brain) imaging so that the surfaces (e.g., skull–skin interfaces) may be considered convex, which simplifies the problem. One of the surfaces is usually in a higher-resolution modality (e.g., CT), while the other may have a lower resolution (e.g., PET). The higher-resolution image is then expressed in (or converted into) the form of a continuous (perhaps faceted) surface Q; the other surface is expressed by a set P of points $\mathbf{p}_i$. The surface-based criteria are used mostly only in rigid transform registration; there is generally a nontrivial relation between a flexible transform and the corresponding deformation of the surface, which would complicate the criterion.

All surface-based methods use basically the same approach to defining a similarity criterion. For each point $\mathbf{p}_i$ of P, the nearest point $\mathbf{q}_i$ on the surface Q is found, which is implicitly considered the corresponding point of the pair $\{\mathbf{p}_i, \mathbf{q}_i\}$. Then, the criterion is a metric function, defined similarly as in the *a priori* point-based case,

$$E_{abs} = \frac{1}{n}\sum_i w_i |\,\mathbf{p}_i - \mathbf{q}_i\,| \quad \text{or} \quad E_{sq} = \frac{1}{n}\sum_i w_i \,|\,\mathbf{p}_i - \mathbf{q}_i\,|^2, \quad \sum_i w_i = n. \tag{10.94}$$

Determining the set $\{\mathbf{q}_i\}$ exactly is computationally demanding; therefore, approximations are used. A simple possibility is that the intersection point of the surface Q with the line between $\mathbf{p}_i$ and the centroid of the Q surface may be taken as the nearest point $\mathbf{q}_i$. When the Q surface is triangulated, $\mathbf{q}_i$ may be the projection to the triangle with the shortest distance to $\mathbf{p}_i$ — exactly the nearest point given the triangular approximation of Q.

10.3.2 Transform Identification and Registration Procedure

The first task in image registration is to determine which registering transform should be used. It may follow from the nature of the problem: if the imaged object may be considered rigid, and the

same imaging geometry is used in both images, a *rigid transform* may be sufficient, possibly even simplified to the plain shift or rotation. When such a presumption is not justifiable, *flexible transforms* must be considered; usually only experiments with real data allow a decision on the necessary complexity of the transform. In most cases, the affine transform provides sufficient flexibility even in complicated cases; otherwise, perspective transform or an even more flexible deforming transform may be necessary. The physical nature of the problem may suggest that *modeling the image deformation* as corresponding to a certain type of mechanical deformation (of, e.g., elastic but incompressible matter [65]) may be appropriate.

When the transform type is chosen, it remains to determine the concrete parameters of the particular transform leading to the best fit of the registered images, either directly or via optimization, as is described in the next two sections. The final goal is to transform image A so that the transformed image A′ matches B in the overlapping area as well as possible. This may be a single-step procedure when direct computation is feasible, or a result of an iterative optimization in the frame of which many attempts have to be done in order to improve the match gradually.

10.3.2.1 Direct Computation

As mentioned in the previous section, point-based criteria enable the formulation of systems of equations, the solving of which may directly determine the transform parameters. The exact solution is possible when the number of conditions given by the point correspondences is equal to the number of transform parameters (providing that the correspondences are correct and exact). More often, the number of correspondences is higher than necessary, though possibly they may not be quite precise. In this case, the solution in the least mean square error (LMSE) sense is at hand, which would hopefully lead to mutual compensation of the errors of individual correspondences.

The following step is then the transformation of A using the obtained parameter vector, which leads to either a complete match of the corresponding markers (landmarks) or the best match in the LMSE sense. The areas in between the landmarks are better matched the more appropriate is the used transform for the particular image distortion. Obviously, increasing the number of corresponding landmarks generally leads to a more reliable and possibly more precise overall correspondence. Nevertheless, increasing the complexity (number of degrees of freedom) of the transformation

need not lead to an improvement; as with any approximation, it may even cause increased errors due to local transform oscillations when the flexibility is excessive.

10.3.2.2 Optimization Approaches

More complex similarity criteria either would lead to formulation of nonlinear systems of equations that are difficult to solve, or may even be unsuitable for equation system derivation, when the criteria are defined algorithmically rather than by closed formulae. Then, the optimization approach that searches for $T_{\boldsymbol{\alpha}}$ yielding the best fit is used, which maximizes (or minimizes) a goal (cost) function based on the chosen global similarity measure. When the transform type has been selected, the optimization aims to determining the optimal vector $\boldsymbol{\alpha}$ of transform parameters.

Generally, the registering procedure is rather complex due to two above-mentioned phenomena: only partial overlap $\Omega_{\boldsymbol{\alpha}}$ of the registered images A′ and B, and the necessity to interpolate among the known samples of image A (see Section 10.1.2). In every step of optimization, the following partial actions must be done:

- Setting of $\boldsymbol{\alpha}$ according to the chosen optimization strategy, utilizing the information from the previous course of optimization
- Calculation of positions $\mathbf{x}_{\mathrm{A}}$ corresponding to grid node positions in $\mathrm{A}' = T_{\boldsymbol{\alpha}}(\mathrm{A})$
- Determination of the overlap area $\Omega_{\boldsymbol{\alpha}}$
- Calculation of the interpolated values $\mathrm{A}'(T_{\boldsymbol{\alpha}}(\mathbf{x}_{\mathrm{A}}))$ based on samples of A
- Goal function calculation based on the chosen similarity criterion
- Test on having reached the optimum: When positive, termination of the algorithm with the present $\boldsymbol{\alpha}$ considered $\boldsymbol{\alpha}_0$; otherwise, continuing from the first item of the cycle

The individual items of the procedure may overlap or be simultaneous. The definition of the similarity criterion should either be independent of the size of $\Omega_{\boldsymbol{\alpha}}$ or be normalized with respect to the area size; otherwise, the dependence on the size may contribute to irregular behavior of the criterion and, consequently, to unreliable or slow convergence. The first and last items of the cycle depend on the chosen optimization strategy.

It has been repeatedly shown that the goal (cost) functions, using any of the mentioned criteria, cannot be approximated by a

smooth radially monotonous function without side extremes; particularly, it can be considered a second-order polynomial only in a very narrow neighborhood of the optimum, and only roughly, if at all. The complex character of the criterion behavior, with many side extremes and possibly small local oscillations in the complete range, thus complicates use of any methods that rely on the simple character of the goal (or cost) function and on availability of its gradient, as do the steepest ascent/descent methods, simplex methods, Newtonian approaches, etc. The optimization by components relying on sequential iterative one-dimensional optimization along individual parameters or conjugate gradients is also hardly applicable, though a good experience with the Powell method for rigid and affine registrations is reported in [68]; however, in other works it has been found ineffective. The rating of individual standard optimization methods as useful or unreliable depends obviously on the average size of the searched area, i.e., whether the initial guess is usually chosen close enough to the optimum. The full search, as the absolutely reliable but extremely computationally demanding golden standard, is too heavy going for practical applications with the available computing power. The published registration results show that the optimization methods with a stochastic element in the iteration strategy, namely, the *simulated annealing* or the *controlled random search* (CRS), represent a good compromise between reliability and computational burden; nevertheless, tuning the annealing parameters for a particular type of problems is an uneasy task [57]. The optimization methods, the use of which is quite general, are described in the rich specialized literature and will not be further treated here.

The *pyramidal multiresolution approach* (coarse-to-fine estimation)—estimating $\boldsymbol{\alpha}$ initially in the downsampled image and then increasing the resolution gradually—has the advantage that the initial search in a large part of the parameter space uses less (though averaged) data, thus saving a substantial part of calculations. More demanding optimization using higher-resolution images on lower pyramidal levels is then restricted to a smaller subspace of $\boldsymbol{\alpha}$ around the result provided by the previous pyramidal level. However, this promising idea turns out to have a limited use; many images lack the necessary fractal character preserving sufficient detail even under low resolution on initial pyramidal levels. Therefore, only a few (one to three) levels of undersampling are mostly useful.

The indicated optimization approach is very general and can be used for image data of any dimensionality. The formalism covers, besides two-dimensional registration, demonstrated on figures,

common three-dimensional spatial registration; four-dimensional data can be basically registered with the same approach as well, though the transforms become more complex. The multidimensional optimization may then naturally become rather difficult and time-consuming. Interesting and medically important is the case of registering two-dimensional data (either slice or projection images) with the three-dimensional volume image data of the same object. It can be accomplished based on similar principles as registering the data of identical dimension; however, the details of additional steps are beyond the scope of the basic explanation.

10.3.3 Registration Evaluation and Approval

Using the optimal transform T_{α_0} derived either directly or via optimization, the spatially registered image pair (on the overlap area (Ω_{α_0}) is obtained. It is then necessary to evaluate the validity and quality of the registration, primarily in order to prevent a gross error due to the optimization being trapped in a local extreme. In medical applications, a medical expert should do this by visual inspection of the registered images. There are several possibilities of suitable presentation enabling assessment of the registration of two-dimensional images: the mosaic checkerboard presentation with alternated images in the neighboring fields (Figure 10.15), temporally alternating display of the complete registered images in the same spatial area (possibly combined with the previous mosaic presentation), or color representation as the fused vector-valued image (Section 10.4.2). Assessing the three-dimensional data registration quality requires either comparison of several carefully chosen pairs of identically positioned two-dimensional slices or comparison of three-dimensional reconstructions, possibly also alternating in the same space, and rotated by the operator. Human observers have been found to be well capable of discovering gross misregistrations, while the fine-tuning of the registration parameters is usually better done by computer.

The basic problem in evaluating registration is usually a lack of real golden standards, namely in intermodality registration. The exceptions are cases when reliable matching points are available, as in imaging with stereotactic frame or other artificial landmarks fixed invariably to a rigid structure of the object (e.g., a bone). Often, the registration abilities of human visual systems are considered a golden standard; however, this is obviously rather vaguely defined.

To evaluate the registration quantitatively, a quality criterion is needed, which is not easy to define when it should be relevant in

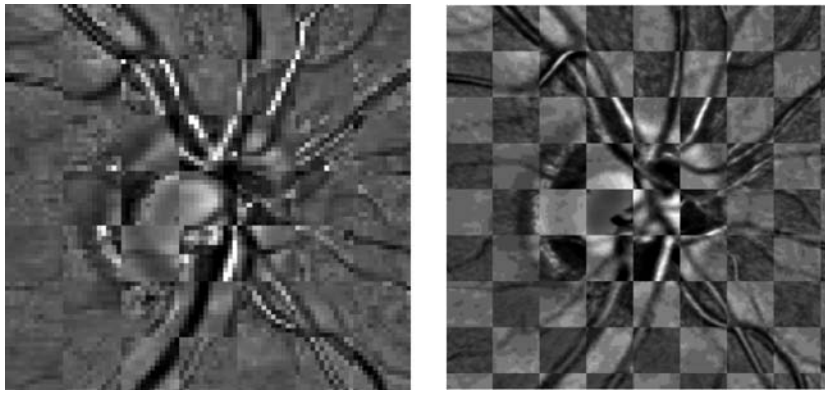


Figure 10.15 Mosaic presentation of the registered retina images of Figure 10.14. Left: Only roughly registered using four times subsampled images. Right: Result of fine-tuned registration. (Courtesy of M. Skokan and L. Kubecka, Brno University of Technology.)

practical medical applications. It is always derived from the global similarity criteria, perhaps partly modified. The commonly used group of criteria is based on comparison of corresponding sets of isolated points. The *fiducial registration error* (FRE) refers to correspondences between both images established on the corresponding point (landmark) sets, and is basically identical to the criterion in Equation 10.93. Obviously, this criterion only evaluates the suitability of the chosen transform and correctness of the computations with respect to the given points; it does not say anything about the registration in the medically important (target) areas of the image, off the fiducial points. Therefore, the similarly defined *target registration error* (TRE) concerns correspondences of clinically relevant features in the target areas; however, its evaluation requires that such correspondences be somehow defined.

When such sets of corresponding points are not available, the overall quality of registration could only be estimated using the global criteria based on a complete set of image intensities: either the direct intensity-based criteria or the criteria evaluating the informational similarity, as described in Section 10.3.1. However, in the lack of golden standards and reliable normalization, the criteria values should be understood as only relative measures evaluating comparatively the success of the applied optimization procedures and global suitability of the used transforms.

10.4 IMAGE FUSION

Image fusion can generally be defined as providing new quality information based on two or more images of the same scene, in the frame of a single modality or using multimodal image sets. In the simplest case, the image fusion may take on the elementary form of providing a set of well-comparable images, which allows, with proper knowledge, training, and experience, the obtaining of new information by mental fusion, performed by the user of the images. This is a characteristic work mode of a radiologist, forming the mental three-dimensional image of an anatomic scene based on a set of two-dimensional slices or projections, possibly in different modalities. In the following sections, several typical cases of more formalized fusion methods will be shown.

10.4.1 Image Subtraction and Addition

Subtraction of an image from another one of the same scene (usually provided with the same modality) serves to enhance the differences between them so that even tiny differences become detectable and can be well evaluated. A typical example of medical use is *subtractive angiography*, where the base image taken before applying a contrast agent is subtracted from the image of the same scene after the contrast agent application*. This way, blood vessels are visualized, and vascularization as well as perfusion of organs may be evaluated. In other fields, like astronomy and ecology, the subtraction is used to detect temporal changes that took place between the first and second expositions; the same approach may find use in medical applications as well.

Obviously, perfect registration of both images is the crucial condition for successful subtraction. Every imperfection, either due to different imaging geometry of the images or because of scene deformation (e.g., patient movement), manifests itself as local "plasticities"—the unmatched edges become doubled with high-contrast rims (partly also in Figure 10.16a). Thus, the registration phase, mostly with flexible transforms, constitutes the basic (and usually more difficult) part of image subtraction.

Similarly, the images may be summed together (or averaged, with normalization to the available brightness extent) in order to

*On the necessary image contrast transforms preceding the subtraction, see Section 3.2.

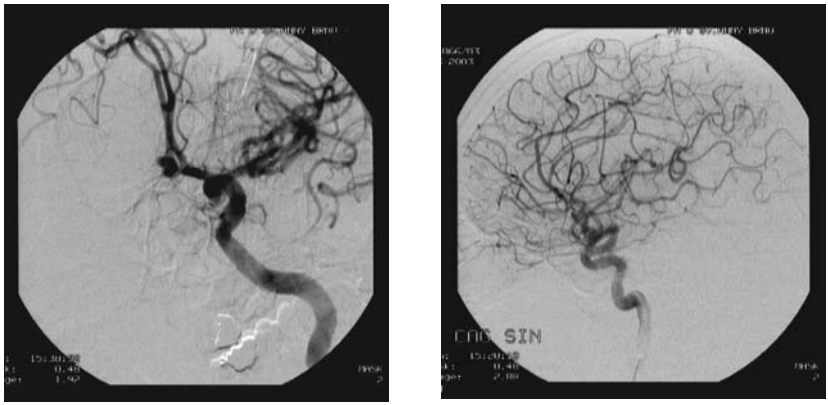


Figure 10.16 Angiographic images obtained by subtraction. Left: Locally imperfect registration of input images. Right: A globally good registration of input images. (Courtesy of the Faculty Hospital of St. Anne Brno, Clinic of Radiology, Assoc. Prof. P. Krupa, M.D., Ph.D.)

increase contrast of interesting structures and suppress random zero-mean noise. Each of the partial images ${}_j g$ then consists of the invariable interesting structure f forming the informational content of the image, and of noise that may be considered a realization ${}_j n$ of a (usually homogeneous) stochastic field,

$$ {}_j g_{i,k} = f_{i,k} + {}_j n_{i,k}. \tag{10.95} $$

The image averaged from N realizations is

$$ \hat{f}_{i,k} = \frac{1}{N} \sum_{j=1}^{N} {}_j g_{i,k} = f_{i,k} + \frac{1}{N} \sum_{j=1}^{N} {}_j n_{i,k}; \tag{10.96} $$

while the image content remains untouched, the noise is averaged. If the stochastic variable $n_{i,k}$ has zero mean, it may be estimated intuitively that the average would tend to zero for large N values. More precisely, the variance ${}_N \sigma_n^2$ of the average that determines the resulting noise power is reduced N times in comparison with the original variance σ_n^2; thus, the standard deviation, which corresponds to the mean amplitude of noise, reduces $\sqrt{N}$ times. Therefore, the amplitude signal-to-noise ratio (SNR) is improved by the factor $\sqrt{N}$ using such simple averaging. Naturally, the invariable useful content must be exactly registered among the individual images, and the noise should be stationary during the acquisition

period. This may be a useful technique in gamma imaging, where the SNR of individual images is usually low. Weighing of the images in the average need not be only uniform, as in Equation 10.96; other weight sequences are used as well. Namely, exponentially decreasing weighing from present to past is often used in *exponential averaging* of image sequences.

Improving SNR belongs generally to image restoration (see also, Chapter 12); averaging thus represents one of the fusion-based restoration techniques.

Addition of partial images may also serve to complement information from different modalities, by combining, e.g., images provided in different parts of visible light, infrared and ultraviolet parts of the spectrum, and the like.

Both sum and difference images usually need normalization of the resulting intensity values to the standard gray-scale range used by the concrete image processing system.

10.4.2 Vector-Valued Images

The natural aim of image fusion may be to consider the intensity values of corresponding pixels in the fused images A, B, C, etc., as elements of vectors forming the pixel values of the resulting fused image F, which thus becomes vector-valued,

$$\mathrm{F} = \begin{bmatrix} \mathbf{f}_{1,1} & \cdots & \mathbf{f}_{1,\mathrm{N}} \\ \vdots & \ddots & \vdots \\ \mathbf{f}_{\mathrm{M},1} & \cdots & \mathbf{f}_{\mathrm{M},\mathrm{N}} \end{bmatrix}, \quad \mathbf{f}_{i,k} = [a_{i,k}, b_{i,k}, c_{i,k}, \ldots]^T. \tag{10.97}$$

Such *vector-valued images* may serve different purposes: they may be visually evaluated when appropriately presented (see below), but mostly they are intended to serve as input images to analytic procedures where the integrated information contained in all partial images may be utilized in a formalized way, should the algorithms accept the complex-valued images. The joint information may have a synergetic effect: the vector image analysis may be more successful than analyzing each component image separately and combining the results subsequently. A typical example is the segmentation based on vector-valued images that may provide substantially better results than the ordinary sequence of segmenting each partial image separately and fusing the results; an example of such an approach application is segmentation of bimodal retina images [21].

10.4.2.1 Presentation of Vector-Valued Images

The vector-valued images (including any registered images fused in the sense of Equation 10.97, and also disparity maps) may be represented visually on two-dimensional displays in a number of ways. The static display can effectively represent only images with a low number of vector elements (mostly two or three). The possibilities of static presentation are manifold:

- Representing each component image separately *in parallel*, in identically large display fields, and possibly with a multiple pointer pointing to corresponding positions in all component images, thus enabling easy comparison of local features.
- Sharing of the same display area by both (all) component images in a *spatially interleaved manner*, usually in a checkerboard mosaic (as in Figure 10.15). The size of the mosaic squares depends on the type of image and purpose of displaying it; it should enable, e.g., good assessment of the degree of registration. In the extreme case, the square size may degenerate to individual pixels, possibly with differently colored component images; then the result is similar to the display based on color layers (see below). The other extreme of the same principle is the *split view* — dividing the image area into two (or more) parts, each belonging to an individual component image, possibly with borders that may be moved by the user.
- Sharing the same display area in time — dynamically *alternating* two (or more) images upon the user's request or automatically with user-adjustable fast/slow pace. A fast version of this approach is also used for stereo image couple presentations, in connection with special glasses allowing the light to pass alternately to the left and to the right eye of the observer in synchronism with the image switching.
- Color layers-based display: Each component image is presented by a different color, or the base image in gray scale and further component images (possibly segmented, thus showing only selected details) in different colors. It should be noted that a unique color display is only possible for three component images because of three-component representation of any color.
- In some special cases, the representation of the vector fields by little arrows showing the amplitude and spatial

direction of local vectors is useful, namely when disparity maps are presented. Alternatively, the disparity maps may be represented by the deformed grid on one of the images, describing the positions corresponding in the other image of the compared pair to a uniform grid of parallel lines. This type of display enables easy recognition of the erroneous nonmonotonity of disparities — it manifests itself as folding of the grid.

Image series describing time development can obviously also be represented by a vector-valued image; however, when the component images form a longer sequence (though perhaps modified by analysis, registration, etc.), this can be most effectively presented as a video sequence visualizing a motion or growth, possibly in slow or accelerated motion.

10.4.3 Three-Dimensional Data from Two-Dimensional Slices

The already mentioned mental process of forming a three-dimensional vision from two-dimensional slices (typical, e.g., in ultrasonography) may be formalized and automated. The three-dimensional imaging is one of the fast-developing directions of ultrasonography, nowadays already established in gynecology and obstetrics and aiming even at imaging of fast-moving structures like the heart. The major promise of three-dimensional ultrasonography lies in elucidating specific features and abnormalities that are not easily resolved in two-dimensional images; also, the higher reproducibility and observer independence of three-dimensional reconstructions is important, namely in shape and volume estimation. There are several problem areas connected with the three-dimensional (or four-dimensional) data reconstruction from two-dimensional slices:

- Spatially irregular two-dimensional data acquisition, so that together with the image data, localizing information for each slice either must be provided by a positioning subsystem or might be possibly derived from the images themselves based on inherent spatial correlations among frames. In case of echocardiography, timing information related to the phase of cardiac cycle must also be provided.
- Compilation of consistent three-dimensional data should take into account the nonideal character of slices (nonzero and variable thickness, possibly spatial distortion so that

a nonplanar slice is acquired, anisoplanar defocus, space-variable geometric distortion in the frame of a single slice, etc.). The compilation procedure should enable partial fusion of overlapping data and interpolation needed for resampling to a three-dimensional regular grid; it may even involve some restoration procedures, taking into account the varying quality of acquired voxel information due to inhomogeneous conditions for ultrasound propagation, reflection, and scattering in tissues.

- Visualization of the usually low-quality three-dimensional data.

10.4.4 Panorama Fusion

Another common task is fusion of partly overlapping images into an image of a greater field of view. It requires registration of overlapping parts, including identification of the optimum transform of one (or both) neighboring images so that the differences in geometrical distortion due to imaging properties may be respected and equalized. The transform determination should also take into account the imaging geometry—the differing position and focus of the camera with respect to the scene in individual images. When used for purposes where geometrical fidelity is required, it should be noticed that the transform is determined from a relatively small part of the images, but influences the image geometry even in distant parts; this may lead to gross geometrical distortions in some cases.

10.4.5 Stereo Surface Reconstruction

Stereo vision is based on mental fusion of images seen separately by each eye of an individual. More generally, the stereo analysis deals with sets of several images of the same scene provided with different imaging geometry so that there are parallax differences carrying the information on the distance of imaged objects from the observer, i.e., on the third dimension perpendicular to the image plane. In the general case, special stereo (Hering) geometry has to be used (see, e.g., [53]). As an example of a geometrically simpler case, we shall limit ourselves to the case of stereo-based surface reconstruction of samples in SEM [32], [27].

The imaging geometry in this special case is given by mounting the sample on an inclinable support, so that at least two images of

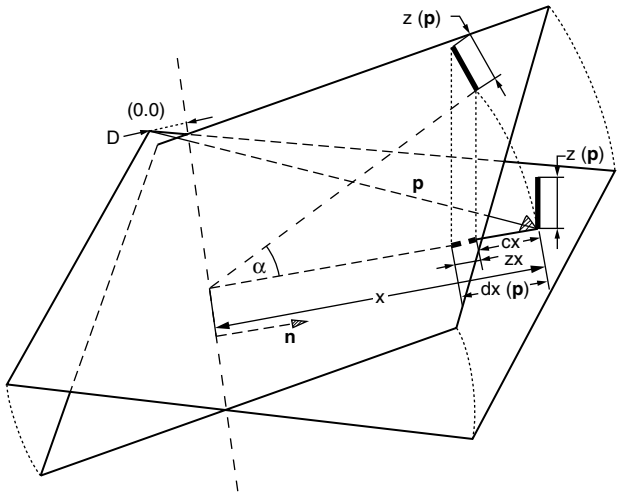


Figure 10.17 Geometry of stereo measurement of SEM samples. (From Jan, J., Janova, D., *Mach. Graphics Vision*, 10, 261–288, 2001. With permission.)

the same area differing slightly by the angle of view can be provided, one in the basic position (perpendicular to the microscope axis) and the other in a slightly tilted position (Figure 10.17).

Quantitative data on the surface $z(\mathbf{p})$ can be obtained algorithmically based on this geometrical model, when we utilize the information contained in the disparity map derived from the stereo couple of images. The disparity $\mathbf{d}(\mathbf{p})$ for a particular point $\mathbf{p} = (x, y)$ is a vector quantity primarily because the tilt axis direction is in general not parallel with a side of the image. There is also a certain variance in the disparity direction due to several parasitic phenomena, but only the disparity components perpendicular to the tilt axis reflect the surface height. When introducing the space coordinates so that the reference plane (x, y) is horizontal with the y-axis parallel to the tilt axis and the z-coordinate is vertical, the scalar field of the disparities is given by the dot product

$$dx(\mathbf{p}) = \mathbf{n}^{\mathrm{T}}\, \mathbf{d}(\mathbf{p}), \tag{10.98}$$

where $\mathbf{n}$ is the normal unit vector to the tilt axis. According to Figure 10.17, the shift $dx(\mathbf{p})$, formed by vertical projection of the inclined surface to the horizontal plane, consists of two components: $cx(\mathbf{p})$ is only due to projection shortening of distances, while $zx(\mathbf{p})$ reflects the sought vertical position of the measured point in relation

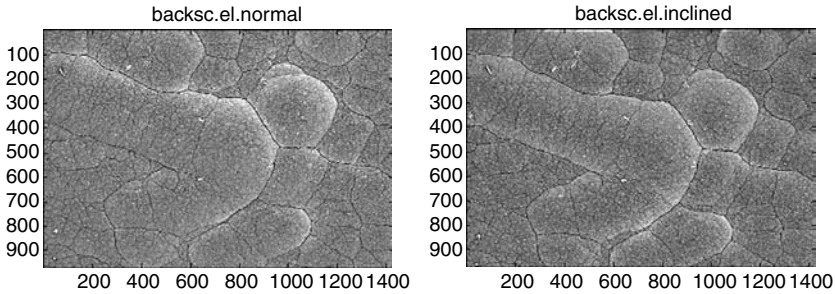


Figure 10.18 Example of a stereo pair of SEM images (a biosubstrate surface). (From Jan, J., Janova, D., *Mach. Graphics Vision*, 10, 261–288, 2001. With permission.)

to the reference plane of the sample. As derived in [32], the z-component (depth or height) at the point $\mathbf{p}$ is

$$z(\mathbf{p}) = \frac{-dx(\mathbf{p}) - cx(\mathbf{p})}{\sin\alpha}, \tag{10.99}$$

where α is the tilt angle and $cx(\mathbf{p})$ is the component of $dx(\mathbf{p})$ dependent only on the distance $x(\mathbf{p})$ of the point $\mathbf{p}$ from the axis of tilt,

$$cx(\mathbf{p}) = x(\mathbf{p})(1-\cos\alpha) = (\mathbf{n}^T\mathbf{p} - D)(1-\cos\alpha), \tag{10.100}$$

D being the distance of the tilt axis from the origin of coordinates.

An example of a stereo image pair is in Figure 10.18. Using the angle criterion of similarity (Equation 10.53) realized by the nonlinear matched filter (Equation 10.85), the disparity matrix on the grid, only two times less dense than pixel density, can be provided (Figure 10.19). After the disparities have been calculated, Equations 10.99 and 10.100 enable determination of the z-coordinates on this dense grid. The corresponding result can be seen in Figure 10.20.

10.4.6 Time Development Analysis

Time development analysis concerns images of basically the same scene taken sequentially with regular or irregular time intervals. Such a situation can be understood as acquisition of spatiotemporal data: in the case of planar images, the obtained data set is three-dimensional, while four-dimensional data are obtained when three-dimensional image data are acquired. This data set can be regarded

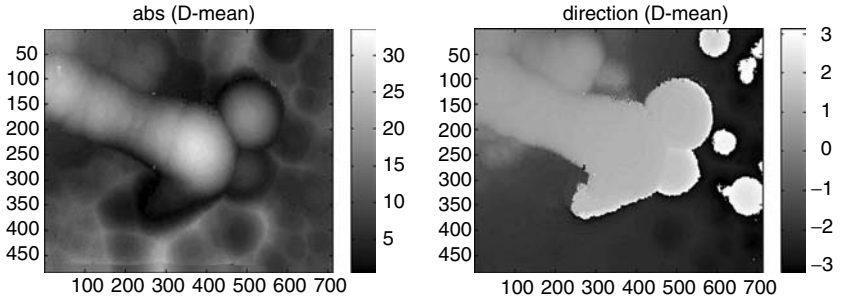


Figure 10.19 The disparity map (left, absolute disparity; right, direction) belonging to the image pair in Figure 10.18. (From Jan, J., Janova, D., *Mach. Graphics Vision*, 10, 261–288, 2001. With permission.)

as a sampled version of a three-dimensional function (or four-dimensional function, respectively),

$$f(x,y,t) \qquad \text{or} \qquad f(x,y,z,t). \tag{10.101}$$

Let us suppose that the time intervals are regular; i.e., the sampling period T is a constant—its absolute value may range from small

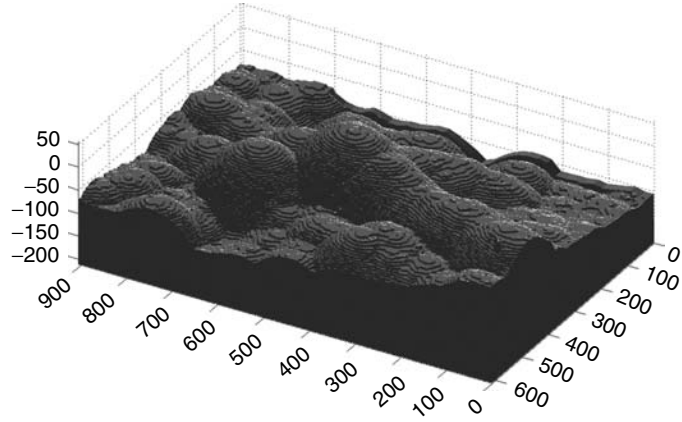


Figure 10.20 The surface, determined using the disparity map in Figure 10.19. (From Jan, J., Janova, D., *Mach. Graphics Vision*, 10, 261–288, 2001. With permission.)

fractions of a second, as, e.g., in a video sequence, to years, when tracking spatial development of organs or lesions during anatomic growth or a disease treatment. From the conceptual image analysis view, the sampling period is irrelevant as far as the images are available; the value of T only influences the interpretation of the analytic results concerning motion or growth velocities. Obviously, the long sampling periods require careful registration of the images that should also be able to respect the spatial changes that are not the subject of analyzed motion or growth.

There are in principle two approaches to the motion or growth analysis: the conceptually simple pair-wise disparity analysis, and the more involved analysis based on the interesting concept of optical flow, which we will discuss below.

10.4.6.1 Time Development via Disparity Analysis

In many medical pairs or series of images provided in one of the above ways, the development between images, neighboring in time, is in multipixel range and is thus grossly discontinuous. Then the below described differential concept of optical flow cannot be applied, and the only choice is to describe the temporal changes step by step by *disparity maps* each derived from a pair of immediately consecutive images (see Section 10.2). The vector-valued disparities related to the time difference between the consecutive images yield the local velocities of motion or the rate of growth (both magnitudes and directions).

10.4.6.2 Time Development via Optical Flow

The concept of *optical flow*[*] will be explained in the three-dimensional case of planar two-dimensional time-variable images, as follows. The generalization to the four-dimensional case (three-dimensional data varying in time) is again straightforward.

This local concept assumes that the image content of a moving object (a certain area of the image) is constant in time; the changes of the intensity $f(x, y, t)$ are therefore only due to the movement.

[*]The term *optical flow* is a purely image processing concept (for further references, see, e.g., [18]; note that there is no connection with the energy flow in optical fields or with optics as such. Perhaps, the *local velocity* forming the *velocity field* on the total image area would be more descriptive.

When starting from a certain point $(x, y, t) = (\mathbf{x}, t)$ in the spatiotemporal space, the intensity values in its close neighborhood may be estimated using the first-order terms of the Taylor expansion,

$$f(x+\Delta x, y+\Delta y, t+\Delta t) \cong f(\mathbf{x},t) + \frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \frac{\partial f}{\partial t}\Delta t. \quad (10.102)$$

When the above constraint concerning the intensity applies, the intensities in the frame of the moving object remain equal, only spatially shifted in the subsequent image taken at $t + \Delta t$,

$$f(x+\Delta x, y+\Delta y, t+\Delta t) = f(\mathbf{x},t). \quad (10.103)$$

By comparison, we obtain

$$\frac{\partial f}{\partial x}\Delta x + \frac{\partial f}{\partial y}\Delta y + \frac{\partial f}{\partial t}\Delta t = 0, \quad \text{or} \quad \frac{\partial f}{\partial x}\frac{\Delta x}{\Delta t} + \frac{\partial f}{\partial y}\frac{\Delta y}{\Delta t} + \frac{\partial f}{\partial t} = 0 \quad (10.104)$$

and, in the limit for $\Delta t \to 0$,

$$\frac{\partial f}{\partial x}\frac{\mathrm{d}x}{\mathrm{d}t} + \frac{\partial f}{\partial y}\frac{\mathrm{d}y}{\mathrm{d}t} + \frac{\partial f}{\partial t} = 0, \quad \text{i.e.,} \quad (\nabla f(\mathbf{x}))^T \mathbf{v}(\mathbf{x}) = -\frac{\partial f}{\partial t} = -f_t'(\mathbf{x}), \quad (10.105)$$

where $\mathbf{v}(\mathbf{x})$ is the local velocity vector (*flow vector*)

$$\mathbf{v}(\mathbf{x}) = \begin{bmatrix} v_x \\ v_y \end{bmatrix} = \begin{bmatrix} \mathrm{d}x/\mathrm{d}t \\ \mathrm{d}y/\mathrm{d}t \end{bmatrix}. \quad (10.106)$$

In the linear formula in Equation 10.105, the spatial gradient $\nabla f(\mathbf{x})$ as well as the time derivative $f'_t(\mathbf{x})$ can be approximated by the (local) normalized finite differences of sample values. Consequently, compilation of the equation is feasible, providing that the consequent images are separated by such a time interval that the spatial shifts are sufficiently small (i.e., roughly in a few-pixel range), but not negligible so that the derivatives can reasonably be approximated.

However, there is only a single equation for two components of the flow; thus, every combination of v_x, v_y lying on the line described by this equation in (v_x, v_y)-coordinates represents a possible solution. Without further information, only the *normal flow*, corresponding to the smallest $\mathbf{v}$ (the point of the mentioned line nearest to the origin), can be determined unambiguously, which naturally does not need to be the right solution. This ambiguity (sc., *aperture problem*) has to be resolved by additional information from neighboring points in a small vicinity of $\mathbf{x}$ that may be considered to belong to a rigidly moving image structure. If the gradient ∇f is diversified in the vicinity range

(i.e., the range is filled with a certain structure or texture), Equation 10.105 for differing points of the neighborhood describes different lines and the proper solution is given by their intersection.

Due to noise and possible variation of **v** (nonrigidity) in the frame of the chosen vicinity, the lines would not intersect at exactly a single point. It is then desirable to find an approximate solution in the sense of least square errors,

$$\mathbf{v}(\mathbf{x},\mathbf{v}) = \arg\min_{\mathbf{v}}\{\varepsilon^2(\mathbf{x},\mathbf{v})\}, \qquad \varepsilon^2 = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(\mathbf{x}-\mathbf{x}')((\nabla f)^T\mathbf{v} + f_t')^2 \mathrm{d}x'\mathrm{d}y', \quad \mathbf{x}' = [x', y'], \tag{10.107}$$

where $w(\mathbf{x})$ is a two-dimensional window determining the vicinity area, usually of the Gaussian shape. This way, the result is averaged over the area and the velocity map therefore partly smoothed. This leads to a constraint as to the area size: though the size must be chosen sufficiently large to allow for texture diversity, too large an area may not fulfill the requirement of rigid motion and might lead to undesirable smoothing of the optical flow field.

The least square optimization leads to the requirement

$$\delta_{\mathbf{v}}\varepsilon^2(\mathbf{x},\mathbf{v}) = \mathbf{0}\text{, i.e., } \frac{\mathrm{d}\varepsilon^2}{\mathrm{d}v_x} = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(\mathbf{x}-\mathbf{x}')((\nabla f)^T\mathbf{v} + f_t')\frac{\partial f}{\partial x}\mathrm{d}x'\mathrm{d}y' \overset{!}{=} 0,$$

$$\frac{\mathrm{d}\varepsilon^2}{\mathrm{d}v_y} = 2\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} w(\mathbf{x}-\mathbf{x}')((\nabla f)^T\mathbf{v} + f_t')\frac{\partial f}{\partial y}\mathrm{d}x'\mathrm{d}y' \overset{!}{=} 0, \tag{10.108}$$

which obviously yields a linear equation system for **v**, $\mathbf{v} = \mathbf{A}^{-1}\mathbf{b}$, in which both the matrix **A** and the right-side vector **b** are determined by the integrals of window-weighted products of the partial derivatives. As the (position-dependent) derivatives can be approximated by finite differences as mentioned above, the computation of the optical flow based on original-space analysis is feasible. Optimization approaches based on the same background have also been published; see [18], [72].

An alternative way to determine the optical flow **v** leads via frequency domain. Let us suppose momentarily that the complete

image in the unlimited (x, y)-plane is moving as a rigid structure with the velocity $\mathbf{v}$. Obviously, then, the three-dimensional function $f(x,y,t)$ becomes

$$f(\mathbf{x},t) = f(\mathbf{x} - \mathbf{v}t, 0), \tag{10.109}$$

the spectrum of which is

$$F(\mathbf{u},\omega) = F(\mathbf{u},0)\,\delta(\mathbf{u}^T\mathbf{v} - \omega). \tag{10.110}$$

The argument of the Dirac function, when set to zero, represents the equation of a plane in the frequency space $(\mathbf{u}, \omega)$, which is perpendicular to the direction of $\mathbf{v}$, and to which the values of the spectrum $F(\mathbf{u}, 0)$ of the (stationary) image $f(\mathbf{x}, 0)$ are projected in parallel with the axis ω. Therefore, when Fourier transforming three-dimensional data, representing a rigid moving image, the nonzero values would fill only such a plane; otherwise, the spectrum is empty. Identifying the normal vector of such a plane, theoretically from only three $\mathbf{u}$-points, reveals the velocity vector. More practically, $\mathbf{v}$ can be determined via $\nabla_{\mathbf{u}}\omega$ for values $\omega(\mathbf{u}) = \mathbf{u}^{\mathrm{T}}\mathbf{v}$ on the plane,

$$\mathbf{v} = \nabla_{\mathbf{u}}\omega(\mathbf{u}). \tag{10.111}$$

It is therefore only necessary to approximate the plane from the nonzero spectral points by linear regression; this suppresses the influence of spectral errors due to noise.

In an image of a natural scene, different velocities appear in separated areas of the image. The above methodology must then be applied to small image areas where the rigidity may be expected. This implies using original-space windowing, which means that the resulting spectrum is blurred, becoming the convolution $F(u, \omega)^*W(u, \omega)$, with the spectrum $W(u, \omega)$ of the window (e.g., three-dimensional sinc function for a rectangular window). This causes a smeared point set to be obtained, with points around a plane, instead of a distinct plane. The accuracy of velocity determination is thus obviously worse the smaller the chosen windows are; i.e., it deteriorates with increasing spatial resolution, as a plane, in the original-domain difference techniques.

The choice between the two above-mentioned motion (or growth) estimation approaches (based on disparity or optical flow) depends on the extent and smoothness of the changes among the images. When the motion between subsequent images manifests as large shifts, the disparity approach is only possible, as the

estimation of partial derivatives is then infeasible. On the other hand, when the motion is relatively slow with respect to the pace of image acquisition, so that the motion changes are small and continuous, the optical flow approach may be preferable, as it can more easily detect and evaluate even subpixel shifts. In both cases, the fusion of images from the measuring set enables the construction of an image with a new quality — a single motion or growth map, possibly even with the development in time, as a sequence of such maps.

10.4.7 Fusion-Based Image Restoration

Though this area is still mostly beyond routine use in medicine, it should be mentioned that using sets of multichannel, multitemporal, or multiaspect images may assist the difficult task of image restoration. The additional information contained in the image sets, in comparison with the simple sum of information carried by the involved individual images, may substantially improve the possibilities of the distortion and noise identification.

An example may be the compensation for aberrations of lens systems in color optical imaging. The color image may be considered a vector-valued fused image, the components of which are generally misaligned and differently distorted images. When the individual partial images (e.g., R, G, and B color components) are only individually restored with respect to blur and noise, the resulting color image may still suffer with heavy aberrations. When, on the other hand, the task is considered a multiparameter optimization, including, besides color-dependent deconvolution, flexible registration of the components, the result may be substantially better. Another simple example of fusion-based image restoration is the already mentioned averaging of a group of images aiming at noise suppression (Section 10.4.1).

A generalized method of restoration of defocus or atmospheric blur utilizing the extra information contained in series of generally misaligned images that are differently blurred is presented in [73]. Here, the mutual spatial shift is considered a parameter of the distortion and compensated for in the frame of the restoration procedure, without a need for preliminary registration. The convolutional distortion is identified at the same time, based on the diversified blur of the component images, thus enabling the blind deconvolution without a preliminary identification of the PSFs of blur.

11

Image Enhancement

Image enhancement is defined as processing that takes an image as its input and provides, as the output, another image, hopefully better in a sense. What the improvement exactly means depends on the intended use of the resulting image; the result is mostly evaluated based on the subjective visual impression, taking into account the purpose of the image, or the evaluation may be derived from the results of subsequent analysis applied to the enhanced image, e.g., in diagnostics.

The enhancement may also be interpreted as *ad hoc* compensation of imperfections in the imaging process providing the image; however, on the difference from image restoration, the imaging properties, distortion, or noise parameters usually need not be identified for enhancement, or only roughly. From this viewpoint, some operations, like contrast transforms or sharpening, found generally useful for a particular modality or a certain type of imaged objects, may be done in batch mode automatically for all images of a group.

Nevertheless, the general approach to image enhancement relies on individual trial-and-error attempts, the results of which are mostly evaluated immediately by visual inspection; when the output image is not satisfactory, a modified operation will be tried. This is particularly useful in diagnostic evaluation because the visual properties of the

inspected image, such as contrast, sharpness, noise smoothing, or geometrical magnification, directly influencing the detectability of important features, may be easily adapted to a concrete image. It is usually done interactively by the inspecting medical staff (e.g., a radiologist) on a local workstation, even when the original image is fetched from a hospital information system.

The main purpose of the enhancement concerns:

- Contrast (and possibly color) transforms
- Sharpening of images, thus enhancing detail resolvability
- Suppression of noise and other disturbing image components
- Geometrical transforms, namely, zooming of partial image areas, and removal of distortion due to imperfection of the imaging systems

The contrast (and color) transforms are typically simple nonlinear point operations, though possibly based on nontrivial algorithms of calculating the brightness transfer functions. Sharpening and noise smoothing are conflicting operations and should therefore be considered simultaneously. They are, as far as enhancement concerns, realized mostly by local (mask) operators, linear and nonlinear. Both point and local operations may be space invariant, or they may depend on regional or local properties of the image in an area surrounding the processed pixel; in this case, it is called adaptive enhancement. Geometrical modifications rely on the transform and interpolation methodology (Sections 10.1.1 and 10.1.2); usually only a limited repertoire of zooming and common distortion elimination is available. The enhancement of images is the most frequently treated topic in literature on image processing and computer vision. For a good overview, as well as for further details, see [59], [26], [64], [12], [72].

11.1 CONTRAST ENHANCEMENT

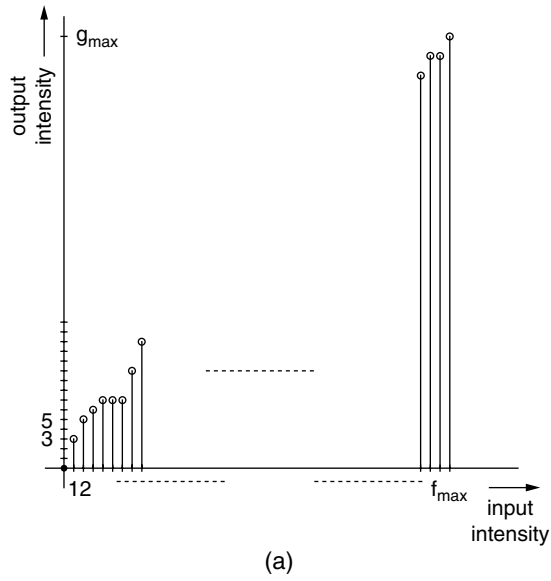
The appearance of an image, as well as possibilities of recognizing the diagnostic features, depends crucially on the image contrast. Though primarily determined by the properties of the imaging modality and adjustment of the imaging system, the displayed contrast can be substantially influenced by technically simple contrast transforms and the image thus adapted to the needs of the human observer or the consequent automatic analysis.

Contrast transform is typically a point operation (see Section 1.3.4). It relates the input pixel value $f_{i,k}$ and the output value $g_{i,k} = \mathcal{N}(f_{i,k})$ controlling the brightness of the respective displayed pixel. Without a loss in generality, we shall suppose that the gray scales in both images start at zero for black, while the brightest white is represented by the maximum values $f_{\max}$, $g_{\max}$, respectively. Most commonly, $f_{\max} = g_{\max} = 2^8 = 256$ in contemporary standard imaging systems with pixels represented by single bytes; less frequently, the upper bound is $2^{16} = 65{,}536$ for two-byte representation. Sometimes $f_{\max} > g_{\max}$; i.e., the internal image representation is then finer than the display brightness resolution. The reason for this arrangement is in the limited dynamic resolution of the human sight, which can only resolve about 50 levels under favorable conditions, so that the single-byte representation is more than sufficient for display*. On the other hand, substantially higher resolution may be required internally, when complicated procedures are applied to the image and the rounding errors (including input quantizing noise) should remain negligible.

Any contrast transform $g_{i,k} = \mathcal{N}(f_{i,k})$ between quantized intensities $f_{i,k}, g_{i,k}$ (Figure 11.1) can be expressed by a lookup table (LUT), in which the input column contains the input intensities, while the values in the second column are the corresponding output values. It can be realized easily in real time via random-access memories, where the input value $f_{i,k}$ addresses the memory location containing the corresponding output value $g_{i,k}$. Contrast transforms can thus be realized even in real time, if the image values are repeatedly read from the display memory and each intensity value, before reaching the monitor, is modified via such a LUT memory. Obviously, changing the memory content changes immediately the contrast transform, and consequently the appearance of the image.

The contrast transforms are most frequently isoplanar; i.e., the same transform $\mathcal{N}(\ldots)$ is applied to all pixels independently of their position in the image. This is also the case when the common LUT approach, as above, is used. The anisoplanar contrast transforms $\mathcal{N}_{i,k}(\ldots)$, on the other hand, depend on the position of the processed pixel. The dependence may either be *a priori* given by the character of a type of images or, more flexibly, be derived from

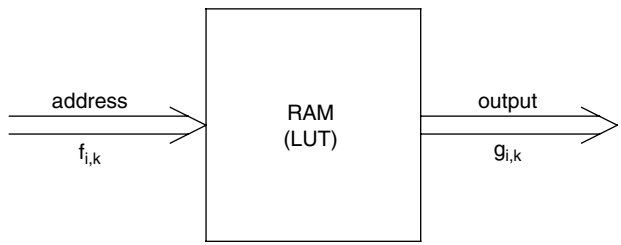
*For color display, commonly 3-byte representation for RGB components is used today. This provides $(2^8)^3 = 16{,}777{,}216$ possible colors, more than required by the human sight color resolution.



LUT

f	g
0	0
1	3
2	5
3	6
4	7
5	7
⋮	⋮
252	251
253	253
254	253
255	255

(b)



(c)

Figure 11.1 (a) Quantized contrast transform function, (b) its lookup table representation, and (c) the hardware realization via RAM.

local image properties, like local variance or a histogram in a small neighborhood. We then speak about *adaptive contrast transforms*. Naturally, the LUT content is then space-variable and the hardware realization becomes more complicated than that mentioned above.

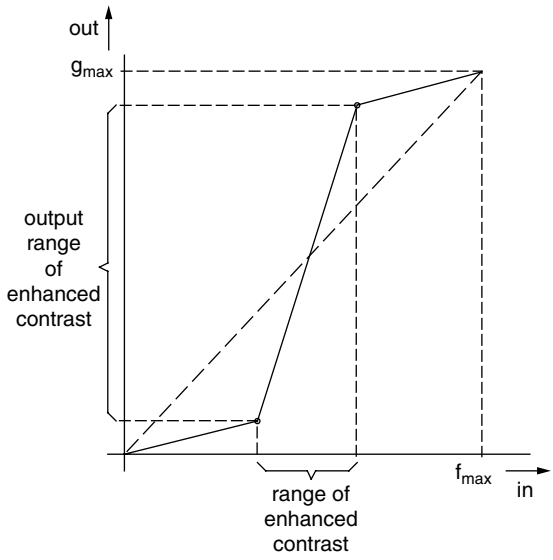


Figure 11.2 Linear and piece-wise linear contrast transfers.

11.1.1 Piece-Wise Linear Contrast Adjustments

The simplest linear contrast transforming function $g_{i,k} = cf_{i,k}$ is depicted in Figure 11.2*; in order to utilize fully the dynamics of the display, the constant factor of the linear function must obviously be $c = g_{\max}/f_{\max}$. A lower slope of the linear function is rarely required, as the full contrast of the display would not be exploited. The opposite situation, when the image contrast is low and utilizes only a part of the input range is quite frequent. It is then advisable to expand the narrower range of the input values $<f_1, f_2>$ so that the displayed range would be fully utilized, i.e.,

$$g_{i,k} = \frac{g_{\max}}{f_2 - f_1}(f_{i,k} - f_1); \tag{11.1}$$

any image values out of the range $<f_1, f_2>$ would be clipped off and displayed as pure black or brightest white, respectively. The information on the suitable limits f_1, f_2 can best be obtained from the

*For simplicity, the schematic figures of contrast transforms are plotted as continuous, though they are discrete functions.

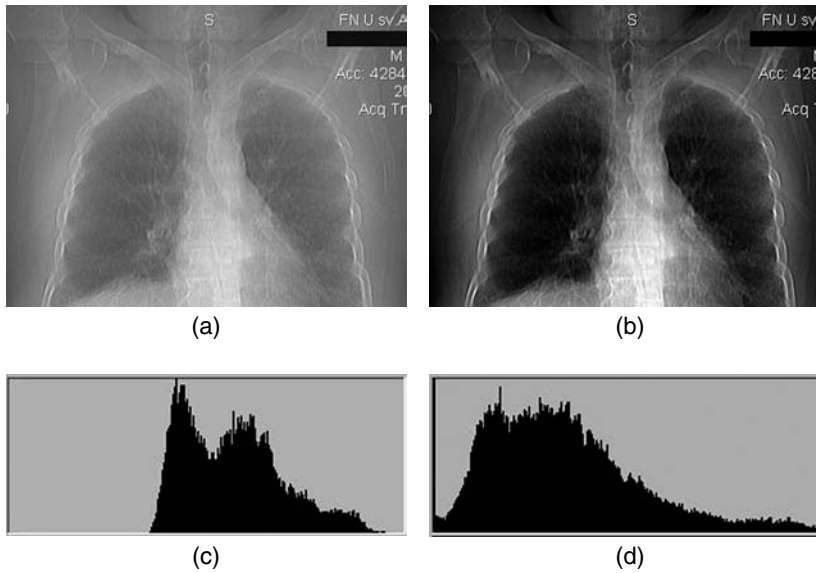


Figure 11.3 Examples of a contrast transform of an image not utilizing fully the range of pixel values. (a) Original image and (b) image after linear contrast enhancement. (c, d) Corresponding histograms.

histogram of the input image. An example of contrast enhancement by a narrowed input gray-scale range is in Figure 11.3.

A more generic case is the piece-wise linear transform in Figure 11.2, which enhances the central part of the gray-scale while certain contrast is still preserved in the outer parts, though naturally lower. Where to place the limits of the central steeper part, and what should be the corresponding g-values, may again be best estimated from the histogram; if not available, experiments are needed. Modern image processing systems often allow determination of the mentioned limits by a suitable interface device such as a joystick or tablet pen; as the global contrast transforms can be realized in real time, changes in the image's appearance may be evaluated immediately.

Usually, it is required that the transform is monotonous in order for the mapping to be unique and the gray scale not confusing; however, there are occasions when the nonuniqueness is accepted. An example—the sawtooth transform—is depicted in Figure 11.4a. This transform provides a high contrast in the complete input range at the cost of discontinuities in the output values; the image takes on a zebra appearance in areas of originally continuously increasing intensity. This is also perhaps the simplest way to estimate isolines

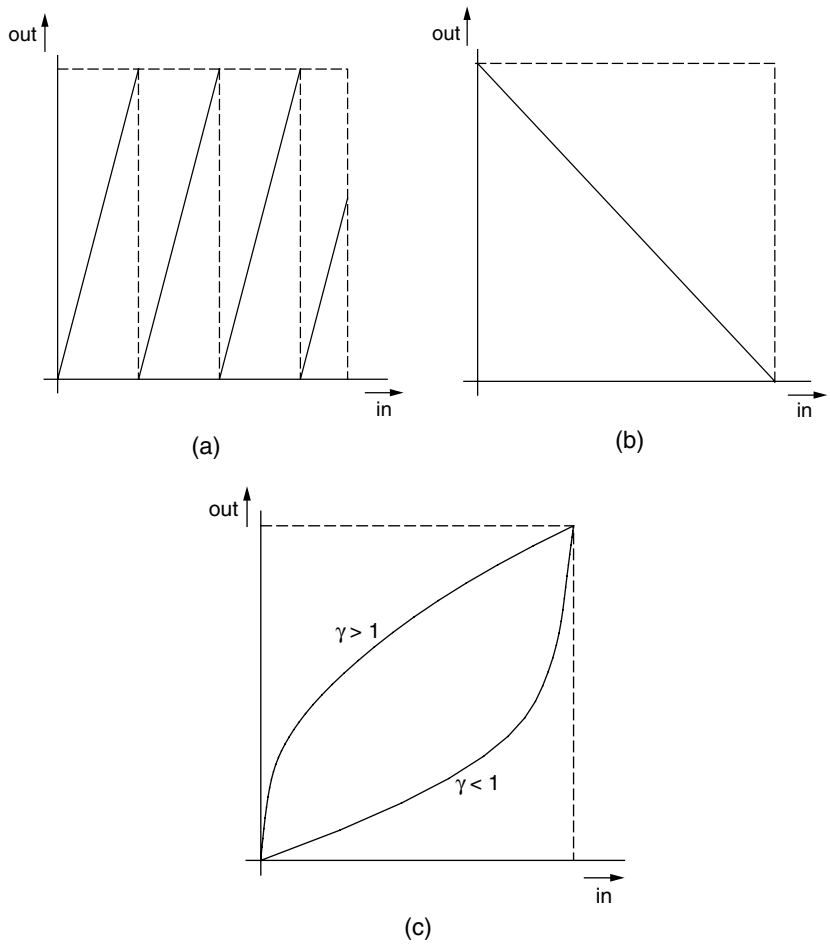


Figure 11.4 Some other contrast transforms: (a) zebra transform, (b) negative transform, and (c) gamma correction.

in a noisy image where the dedicated transform (isolated periodical white points among black ones) would not provide continuous curves. The linear transform with a negative slope, as in Figure 11.4b, provides negative images with a contrast corresponding to the slope.

11.1.2 Nonlinear Contrast Transforms

The brightness mapping may be more generally described by nonlinear functions. It is then possible — usually experimentally — to

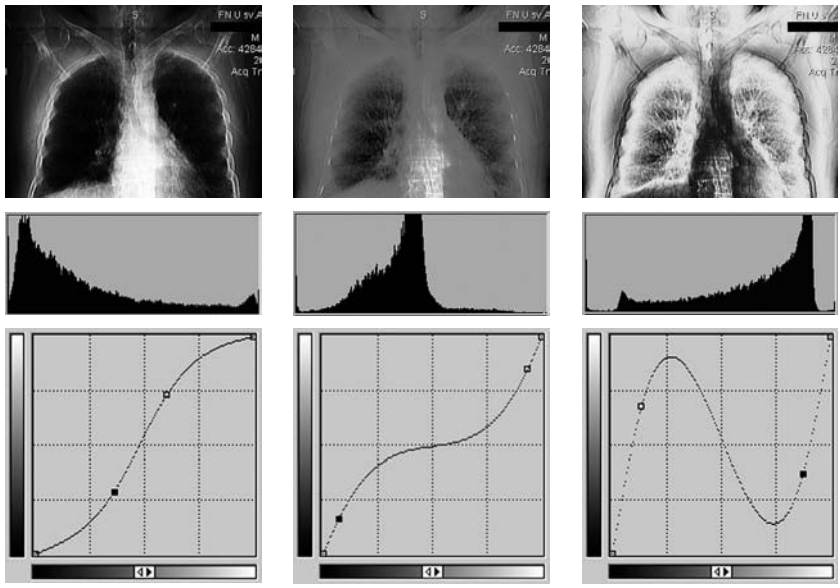


Figure 11.5 Nonlinear contrast transforms: (a) enhancing middle range contrast, (b) enhancing contrast in dark areas and highlights, and (c) a nonmonotonous transform.

adjust the shape of the curves as needed for particular image content, possibly even individually for RGB color components, thus influencing not only gray scale, but also the color shades. Typical examples are in Figure 11.5, where the first curve obviously increases contrast in the middle range of brightness, while the second one does so in the marginal shadow and highlight ranges — in both cases at the cost of low contrast in the remaining areas. The third curve is nonmonotonous, thus inverting the middle brightness range to a negative and therefore changing the appearance of the image completely.

A special type of nonlinear contrast transform deserves mention — the *gamma correction*. Originally, as the name suggests, it was intended to compensate for exponential characteristics of image sensors (or films) and displays. During image enhancement, we usually do not have the information on the degree of these nonlinearities, so that the exact compensation is not possible (and

also usually not required). However, changing the gamma provides a single-parameter nonlinear contrast transform (Figure 11.4c). This transform keeps the extreme values of the dependence $g = \mathcal{N}(f)$ at given fixed positions $(f_{\min}, g_{\min})$ and $(f_{\max}, g_{\max})$, usually black and white in both input and output scales, while the value of γ determines whether the curve between these two points is straight (for $\gamma = 1$) or bent, dependent on how much γ differs from 1. The usual range of γ is about <0.6, 2.5>, but it may occasionally acquire values outside of this range. Values less than 1 provide concave curves so that the contrast is increased in the highlight end of the scale; on the other hand, $\gamma > 1$ bends the line to a convex shape, thus lightening the shadows and increasing the contrast at the dark end of the scale at the cost of the highlight-end contrast. It turns out that the practical utility of this simply controlled transform is considerable for many kinds of images (see Figure 11.6).

Another type of nonlinear contrast transforms are thresholding and clipping functions. The purpose of such functions, partly already mentioned, is to emphasize certain areas of the image marked by a defined extent of intensities, by either completely suppressing the other parts or decreasing their contrast. Conversely, the pixels in the

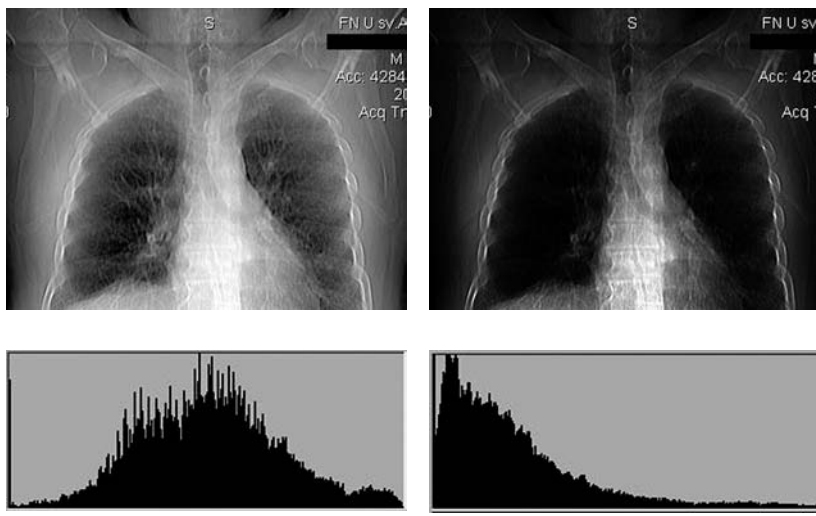


Figure 11.6 Application of gamma correction enhancing either shadows (left, $\gamma = 1.75$) or highlights (right, $\gamma = 0.5$); corresponding histograms below.

spatial extent of an interesting object, the intensities of which are in a certain range, may be set to either zero or the maximum, so that the object appears either black or white against the background of the opposite color or natural gray-scale background. The plain thresholding—converting the input image into a binary representation—is the limit case of contrast enhancement when there is a sudden jump from zero to maximum output level; the threshold is determined by the discontinuity position on the input axis.

11.1.3 Histogram Equalization

A good gray-scale image should reach both brightness extremes (black and white); besides that, the visual impression is usually good when all the intermediate shades are distributed evenly, i.e., when the histogram of the image is approximately flat—all the counts are about equal. The nondecreasing contrast transform, which changes the pixel intensities so that the histogram is optimized with respect to such equality, is called *histogram equalization*.

If the image intensities $f_{i,k}$ were from a continuous range (i.e., nonquantized), and the image might be considered a realization of a homogeneous field, the histogram would approximate the probability density $p_f(f)$ of the stochastic variable controlling all pixel values. We may construct the needed (continuous) contrast transform $\mathcal{N}(\ldots)$, utilizing the method of transforming the probability distribution, as derived in the probability theory. When f is a stochastic variable with the distribution function $P_f(f)$, and if $P_g(x)$ is a nondecreasing function, then the transformed stochastic variable $g(f) = P_g^{-1}(P_f(f))$ has the distribution function $P_g(\ldots)$. In our case, the distribution function corresponding to the desired constant probability density is

$$P_g(g) = \int_0^g \frac{1}{g_{\max}} \mathrm{d}g = \frac{g}{g_{\max}}; \tag{11.2}$$

the inverse function is $P_g^{-1}(x) = g_{\max} x$, and consequently the transform function is

$$g(f) = g_{\max} P_f(f). \tag{11.3}$$

Thus, it would seemingly only be needed to provide an approximation of $P_f(f)$ by summing the normalized histogram counts, and the output values could be calculated. Unfortunately, while this may

serve for explanation, the discrete character of digital images does not allow the deriving of a satisfactory contrast transform this way.

We shall therefore present a different approach that is entirely based on discrete representation. The contrast transform may be considered a mapping $\mathcal{N}$,

$$\{0,1,\ldots,r-1\} \rightarrow \{0,1,\ldots,q-1\} \tag{11.4}$$

between the input discrete scale of r shades of gray and the output scale of q shades. We thus transform the values of the input image $\mathrm{P} = \{p_{i,k}\}$ sized $m \times n$ pixels to the equally sized output image $\mathrm{D} = \{q_{i,k} = \mathcal{N}(p_{i,k})\}$; let $q \leq r$. The histogram of the output image is the vector

$$\mathbf{h}^{\mathrm{D}}:\ h_l^{\mathrm{D}} = \sum_{D_l} 1, \quad l = 0,1,\ldots,q-1, \quad D_l = \{[i,k]:\ ((d_{i,k} \in \mathrm{D}) = l)\}, \tag{11.5}$$

which should ideally be flat; i.e., all its components should approach the count mn/q as the sets D_l are mutually disjunct, $D_l \cap D_m = 0$, $l \neq m$. In the similarly defined histogram of the input image $\mathbf{h}^{\mathrm{P}}$, the sets P_k are disjunct as well. Therefore,

$$\sum_{k=0}^{r-1} h_k^{\mathrm{P}} = \sum_{l=0}^{q-1} h_l^{\mathrm{D}} = mn. \tag{11.6}$$

The desired mapping $\mathcal{N}$ is a discrete (staircase) function that can be expressed as a lookup table and must be nondecreasing. From this requirement, it follows that any set D_l can only be a union of neighboring sets P_k,

$$D_l = \bigcup_{k=k_l}^{k_l+\Delta k_l} P_k, \quad \Delta k_l \geq 0. \tag{11.7}$$

Therefore, the components of the output histogram $\mathbf{h}^{\mathrm{D}}$ are formed by sums of the neighboring components of the original histogram,

$$h_l^{\mathrm{D}} = \sum_{k=k_l}^{k_l+\Delta k_l} h_k^{\mathrm{P}}. \tag{11.8}$$

Based on these properties, we may formulate the following algorithm: Starting from the 0th class D_0, we shall construct

gradually all the classes D_l as follows. D_0 will be formed by the union $D_0 = \bigcup_{k=0}^{k_1-1} P_k$ fulfilling

$$\sum_{i=0}^{k_1-1} h_i^{\mathrm{P}} \leq \frac{mn}{q} < \sum_{i=0}^{k_1} h_i^{\mathrm{P}}. \tag{11.9}$$

Similarly, D_1 would contain the classes $P_{k_1} \ldots P_{k_2-1}$ with

$$\sum_{i=0}^{k_2-1} h_i^{\mathrm{P}} \leq \frac{2mn}{q} < \sum_{i=0}^{k_2} h_i^{\mathrm{P}}, \tag{11.10}$$

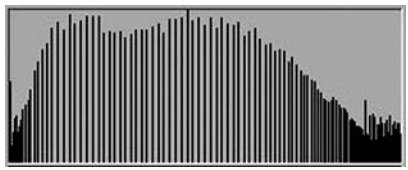
etc.; generally, for the class D_{l-1},

$$\sum_{i=0}^{k_l-1} h_i^{\mathrm{P}} \leq \frac{lmn}{q} < \sum_{i=0}^{k_l} h_i^{\mathrm{P}} \tag{11.11}$$

The summing character of the criterion in Equation 11.11 prevents cumulating of errors that necessarily appear due to the discrete character of the construction. Simple inspection shows that the algorithm provides an almost uniform output histogram when $q << r$. When, as more common, $q = r$, it is obviously not possible to obtain even distribution, as cumulating of neighboring original classes necessarily leads to low or zero counts in some of the D-classes. Nevertheless, the averages over neighboring values of h_l^{D} are well equalized, and consequently, the appearance of the output image is close to that of the ideally transformed image in the case of continuous intensities according to Equation 11.3. An example of the application of histogram equalization can be seen in Figure 11.7; the upper histogram corresponds to the plain equalization, when some output classes resulted in high counts while the neighboring ones remained empty. The lower histogram corresponds to the same image after a mild smoothing that provides the intermediate gray-scale values, thus leveling the histogram (like when averaging neighboring classes in the first histogram).

To overcome the difficulty with the input intensities at the borders of output classes, causing a single output level to be too numerous and the count of the following level to be too low, it is possible to redistribute the pixels at this input level between the two neighboring output levels. It may be done randomly, based on

(a)



(b)

(c)

Figure 11.7 (a) The originally low-contrast image Figure 11.3a after histogram equalization and moderate smoothing. (b) The histogram after mere equalization. (c) The resulting histogram after image smoothing.

a simulated stochastic decision, or it may depend on the context in the image, i.e., basically on the pixels in a neighborhood. It naturally complicates the procedure, but the resulting image may be smoother — the artificial edges caused by the discontinuities in the contrast transform may be significantly suppressed.

Though equalizing the histogram is a formalized operation, it aims primarily at a good appearance of the output image and cannot be considered a restoration method. It is often used in general imaging to achieve full-contrast good-looking "brilliant" images in a single operation. However, in the field of diagnostic imaging, the plain equalization hardly ever fulfills the requirement of optimally displaying the details in all parts of the image. The local equalization, based on the local histogram of a small image area and applied only in a corresponding neighborhood, may provide results that are more useful (Figure 11.8). When this is done gradually on the whole image plane with a suitable interpolation among the partial transform functions, we obtain the *adaptive histogram equalization* with a position-dependent transform function $\mathcal{N}_{i,k}(\ldots)$.

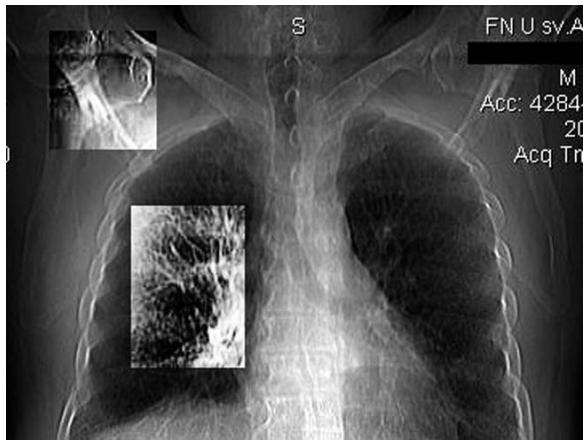


Figure 11.8 Local histogram equalization based on corresponding local histograms — the basic idea leading to adaptive histogram equalization.

The histogram equalization is sometimes falsely interpreted as increasing the informational content of the image. This paradoxical statement seems reasonable at first glance, as the equalized probabilities would seemingly lead to higher entropy. Of course, a more detailed inspection reveals that the transform may at most preserve the information and usually leads to factual informational deterioration. However, this does not mean that the transformed image is not more suitable for visual inspection, and thus the share of the *exploited* information is higher than that for the original image.

11.1.4 Pseudocoloring

Color finds its way into predominantly gray-scale medical imaging primarily in the form of false colors, used to enhance visibility of details in areas where the differences in gray shades are minor. The pseudocoloring — expressing the shades of gray by means of colors — utilizes the much higher resolution of human sight to color shades than to degrees of gray.

The principle is quite simple: each gray-scale intensity determines a displayed color via independent transforms for each color component,

$$\begin{aligned} g^{\mathrm{R}}_{i,k} = \mathcal{N}_{\mathrm{R}}(f_{i,k}), \quad g^{\mathrm{G}}_{i,k} = \mathcal{N}_{\mathrm{G}}(f_{i,k}), \quad g^{\mathrm{B}}_{i,k} \\ = \mathcal{N}_{\mathrm{B}}(f_{i,k}), \quad \text{i.e.,} \quad \mathbf{g}_{i,k} = \boldsymbol{\mathcal{N}}_{\mathrm{RGB}}(f_{i,k}). \end{aligned} \tag{11.12}$$

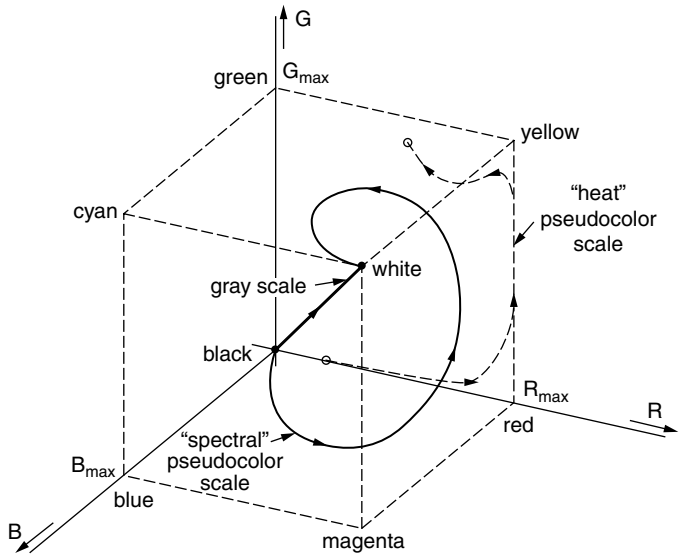


Figure 11.9 Vector representation of pseudocolor contrast transforms.

The scalar transfer functions $\mathcal{N}_C(\ldots)$ are individually designed for every color component C according to the required appearance of the display. The resulting transform can be described by a vector function $\boldsymbol{\mathcal{N}}_{RGB}(\ldots)$, as depicted in Figure 11.9. Obviously, three lookup tables with a common input are then needed, possibly realized by three parallel and identically addressed memories as a generalization of Figure 11.1c. The possibilities of color scale design are very rich; however, some selected scales are generally accepted as intuitive. The examples may be the "spectral" scale, which might start at black and continue through dark violet and blue via green, yellow, and red, reaching white on the other side of the input gray scale, or the "heat" scale, from black and brown via red, orange, and yellow to greenish white, etc.

A brief note on realistic color images met in medical imaging, e.g., in optical microscopy, ophtalmology, or endoscopy: they may be considered vector-valued images (Section 10.4.2) and analyzed as such. This may substantially improve the results of analysis in comparison with analyzing every color component separately, e.g., in segmentation. The reader interested in realistic color image processing and displaying is referred to the specialized literature listed in the references, e.g., [76].

11.2 SHARPENING AND EDGE ENHANCEMENT

The subjective quality of an image depends to a great extent on the sharpness (visibility of details at the limits of the sight resolution) and explicitness of edges. This is related to human sight properties — in the lack of sharpness, the sight strains to achieve a better focus; also, as recognition of shapes or objects starts from identification of sharp edges, their weak presentation is felt as lack of details. Enhancing the sharpness by accentuating the edges may thus contribute to a more pleasing subjective appearance of an image (or even, via better visibility of details, to its diagnostic utility).

The lack of sharpness may be inherent to the imaging method or it may be due to bad quality of imaging equipment (e.g., lens), wrong focus, motion blur, etc. Compensation of a blur of known properties, given, e.g., by the point-spread function (PSF) of the blur, is a subject of image restoration (Section 12.3); the desired result is then a good estimate of the original image. Here, we follow a different purpose — the enhancement of sharpness should lead to a better-looking image, in which details may be better visible. The evaluation of the enhancement success is therefore fully on subjective appraisal, and usually a satisfactory effect is a result of a (possibly iterative) trial-and-error approach.

The degree of low sharpness may be formulated either in the spatial domain, as too slowly varying intensities, namely across object borders, or in the frequency domain, as a low share of (or even missing) high-frequency components. The sharpness enhancement may therefore be interpreted as increasing the slope of intensity profiles, or — in the frequency domain — as augmentation of higher-frequency components.

The important difference between resolution and sharpness should be stressed: any sharpening cannot improve the real resolution of the image (in resolvable lines per length unit); resolution improvement would need to add the missing high-frequency components*. Sharpening can only enhance the details already present; in other words, it can only improve the steepness of edges and

*However, when the image is sampled with an aperture of a certain width, some frequencies above a real resolution may appear due to high-frequency side lobes of the aperture spectrum. This phenomenon is called *pseudoresolution* and its (mostly undesirable) effects may be enhanced by sharpening as well.

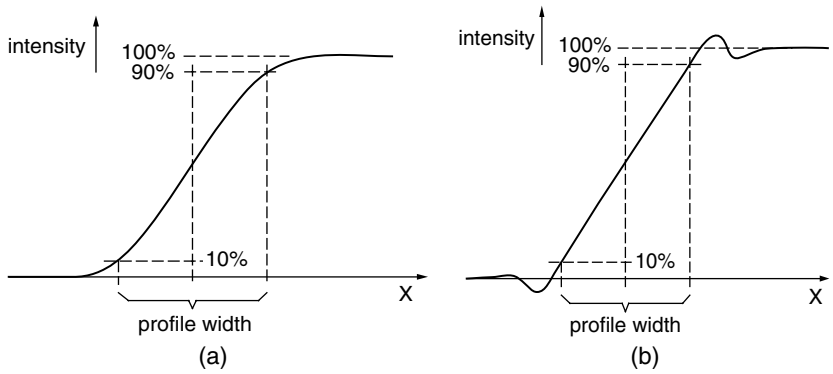


Figure 11.10 Edge profiles schematically: (left) before and (right) after sharpening.

possibly add certain overshoot contours (Figure 11.10; see also, Figure 11.14), thus making the borders and edges better visible. The quantitative evaluation of sharpening is usually based on comparison of the edge profile width (between 10 and 90% of the amplitude) before and after processing, as indicated in the figure.

11.2.1 Discrete Difference Operators

Blur may be interpreted as a result of certain integration, or averaging, in the spatial domain. It can therefore be expected that the inverse operation — differentiation — may improve the sharpness of the image. Which sharpening operator is suitable depends on the character of the blur, which may be directional or isotropic. The sharpening may require a simple or higher-order differentiation, a combination with the original image content, etc., depending on the image character and also on features that should be enhanced. Let us briefly review the properties of differential operators in continuous two-dimensional space.

The partial derivatives $\partial f(x,y)/\partial x, \partial f(x,y)/\partial y$ are obviously directional or *anisotropic operators*, as they react predominantly to intensity variations along the respective axis, and thus, as far as edges are concerned, primarily to vertical or horizontal edges, respectively. The otherwise directed edges lead to a response decreasing with the cosine of the angle difference from the optimum. A differentiating

operator maximally sensitive to edges perpendicular to a unit vector **s** is obviously the directional partial derivative

$$\frac{\partial f}{\partial s} = \frac{\partial f}{\partial x}\cos\vartheta + \frac{\partial f}{\partial y}\sin\vartheta \tag{11.13}$$

where ϑ is the angle between **s** and the x-axis. Similarly, higher-order derivatives are anisotropic operators as well. Profiles of output images derived by the first- and second-derivative operators from an image containing an edge are schematically shown in Figure 11.11. All the directional operators and operators related to them are suitable when

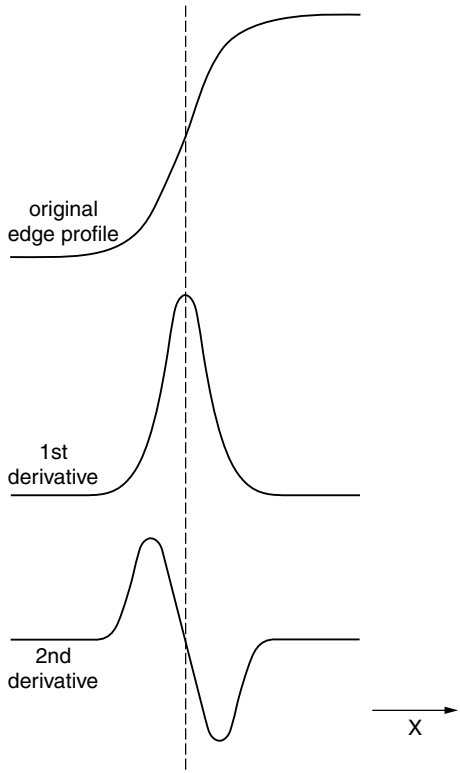


Figure 11.11 Schematic profiles of responses of the first- and second-derivative operators when applied to an edge ramp. Upper line: Original profile perpendicular to the edge. Middle line: Response to the first derivative operator. Lower line: Response to the second-derivative operator.

enhancement of edges of certain direction are preferred, while the perpendicular edges will be completely omitted in the response. Directional sharpening is also suitable when it is obvious that the blur is anisotropic, as, e.g., blur due to a linear motion.

More often, isotropic sharpening or edge enhancement is required. It can be shown that combining the partial derivatives as

$$\left(\frac{\partial^m f}{\partial x^m}\right)^k + \left(\frac{\partial^m f}{\partial y^m}\right)^k \tag{11.14}$$

defines *isotropic operators*, providing that the derivatives of odd orders are in even powers or vice versa, i.e., even-order derivatives in odd powers. Of these isotropic operators, two are used most often, the *magnitude of gradient*,

$$|\nabla f(x,y)| = \sqrt{\left(\frac{\partial f}{\partial x}\right)^2 + \left(\frac{\partial f}{\partial y}\right)^2}\ , \tag{11.15}$$

often though improperly called simply "gradient," and the *Laplace operator* (*Laplacian*),

$$\mathcal{L}(f(x,y)) = \nabla^2 f(x,y) = \frac{\partial^2 f}{\partial x^2} + \frac{\partial^2 f}{\partial y^2}. \tag{11.16}$$

Later, the *gradient* operator will be needed, which is the vector

$$\nabla f(x,y) = \frac{\partial f}{\partial x}\mathbf{j} + \frac{\partial f}{\partial y}\mathbf{i}, \tag{11.17}$$

where $\mathbf{i}, \mathbf{j}$ are unit vectors along axes directions; its direction ϑ with respect to the x-axis is

$$\vartheta(x,y) = \arctan\left(\frac{\partial f}{\partial x} \Big/ \frac{\partial f}{\partial y}\right). \tag{11.18}$$

In the discrete environment of digital images, the differential operators must be approximated by *difference operators* based on finite differences among pixels,

$$\frac{\partial f(x_i, y_k)}{\partial x} \approx \frac{\Delta_x f_{i,k}}{\Delta x}, \qquad \frac{\partial f(x_i, y_k)}{\partial y} \approx \frac{\Delta_y f_{i,k}}{\Delta y}, \tag{11.19}$$

where the *first differences* are

$$\Delta_x f_{i,k} = f_{i,k} - f_{i-1,k}, \qquad \Delta_y f_{i,k} = f_{i,k} - f_{i,k-1}. \tag{11.20}$$

Providing equidistant image sampling with $\Delta x = \Delta y$, the approximations of the derivatives are proportional to the corresponding differences between pixel values; as the operators based on them are linear and must be normalized anyway with respect to preservation of the available extent of intensity values, usually the division by Δx or Δy is omitted. Thus, instead of derivatives, the differences are used directly. The second partial derivatives are then replaced by the *second differences*:

$$\begin{aligned} \Delta_x^2 f_{i,k} &= \Delta_x f_{i+1,k} - \Delta_x f_{i,k} = f_{i+1,k} + f_{i-1,k} - 2f_{i,k}, \\ \Delta_y^2 f_{i,k} &= \Delta_y f_{i,k+1} - \Delta_y f_{i,k} = f_{i,k+1} + f_{i,k-1} - 2f_{i,k}. \end{aligned} \tag{11.21}$$

The difference operators can be realized as local (mask) operators (Section 2.2.1); by means of them, images describing the differences as functions of position are generated. The masks for the first and second differences may be as follows:

$$\Delta_x f_{i,k} : \begin{bmatrix} 0 & -1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \text{ or } \frac{1}{2}\begin{bmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \Delta_y f_{i,k} : \begin{bmatrix} 0 & 0 & 0 \\ -1 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \text{ etc.,}$$
$$\Delta_x^2 f_{i,k} : \begin{bmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{bmatrix}, \quad \Delta_y^2 f_{i,k} : \begin{bmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{bmatrix}. \tag{11.22}$$

As the values of differences are generally in the range $<-f_{max}, f_{max}>$, while the available extent of image value representation is only $<0, f_{max}>$, a normalization is necessary, when the output should fit in an image matrix; from now on, we shall omit mentioning the need of normalization. Examples of images derived by both first difference masks are shown in Figure 11.12. Also, the directional differences under ±45° may be similarly expressed this way,

$$\Delta_s f_{i,k} : \frac{1}{2\sqrt{2}}\begin{bmatrix} 0 & 0 & -1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \text{ and } \frac{1}{2\sqrt{2}}\begin{bmatrix} -1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{11.23}$$

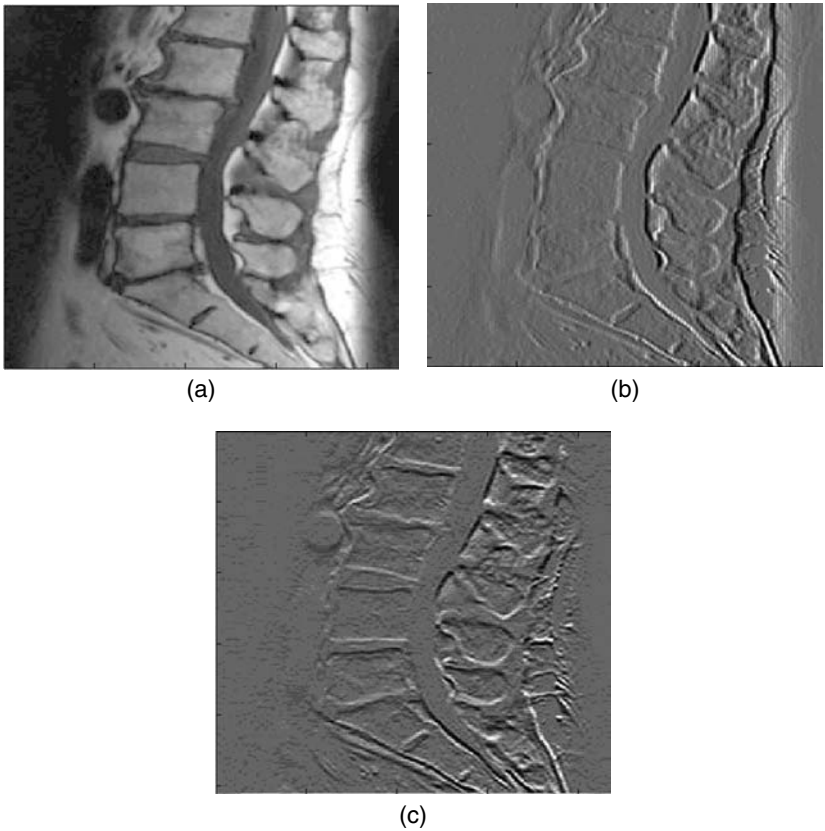


Figure 11.12 Images derived by first difference operators: (a) original, (b) via x-differences, and (c) via y-differences.

Note that the spatial diagonal increment is $\sqrt{2}$ times greater, which should be taken into account when comparing the mask operator results. However, it is usually neglected in favor of computational simplicity.

The *discrete Laplace operator* may be expressed as the sum of the last two masks in Equation 11.22 by the following mask:

$$\mathcal{L}\{f_{i,k}\}: \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{11.24}$$

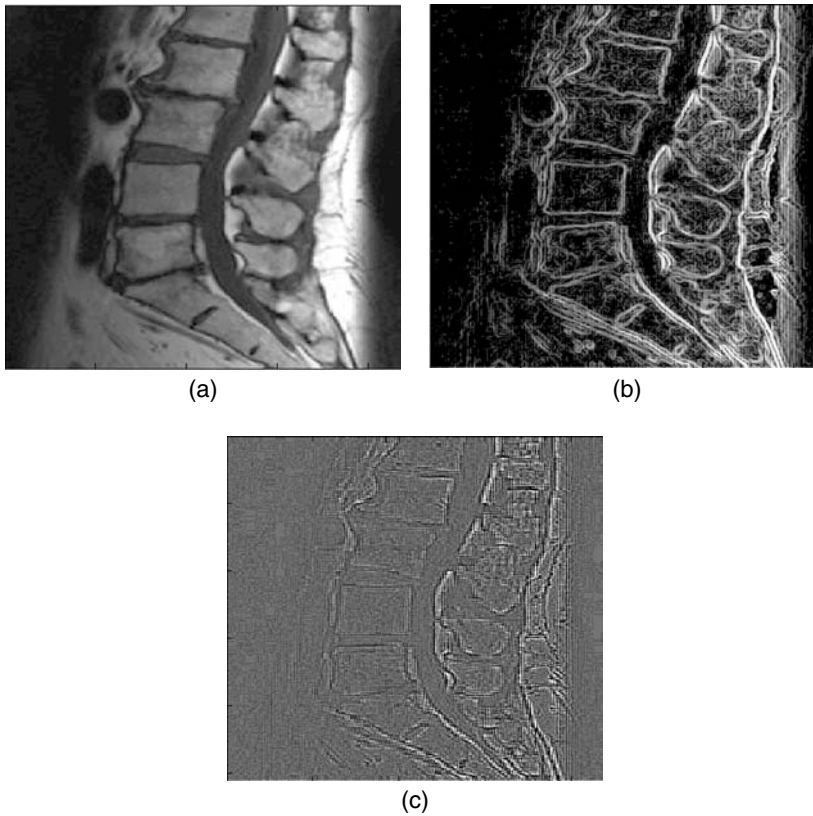


Figure 11.13 Images derived by difference operators: (a) original image, (b) magnitude of gradient, and (c) image derived via Laplacian.

An example of the images derived by the digital gradient magnitude operator and by the digital Laplacian are in Figure 11.13. The gradient magnitude cannot be directly expressed by a mask as there are nonlinear operations involved, as visible from Equation 11.15. However, the needed first partial differences may be provided by the masks (Equation 11.22) and the results pixel-wise combined. Because the nonlinear square and square-root operations are computationally demanding, the expression is often approximated as $\sqrt{a^2+b^2} \approx |a| + |b|$, or $\max(|a|, |b|)$. The second approximation is also used in *Robert's operator*, the gradient magnitude approximation based on skew (±45°) differences.

The difference operators are sensitive to noise in the processed image; the noise influence may be partly suppressed by averaging,

possibly weighted. Thus, the first difference operators may become, e.g.,

$$\Delta_x f_{i,k}: \quad \frac{1}{6}\begin{bmatrix} -1 & -1 & -1 \\ 0 & 0 & 0 \\ 1 & 1 & 1 \end{bmatrix}, \quad \text{or} \quad \Delta_y f_{i,k}: \frac{1}{8}\begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix} \text{ etc.,} \tag{11.25}$$

where the second matrix is an example of an unevenly weighted average. The Laplace operator may be averaged from two 45° rotated operators as either of the masks

$$\mathcal{L}\{f_{i,k}\}: \begin{bmatrix} 1 & 1 & 1 \\ 1 & -8 & 1 \\ 1 & 1 & 1 \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} 1 & 2 & 1 \\ 2 & -12 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{11.26}$$

Note that the vertical–horizontal operator predominates slightly (by the factor $\sqrt{2}$) in the second mask, while the situation in the first mask is the opposite.

11.2.2 Local Sharpening Operators

The difference operators all have zero mean, thus losing substantially from the original image information. This leads to a suggestion to include a certain share of the original into the output. The most commonly used sharpening operator simulates the old photographic technique of unsharp negative masking — subtracting a blurred version of the input from the original image. Intuitively, it can be estimated that it would lead to crispening, as subtracting low-frequency components contained in the blurred image B means emphasizing relatively the detail-carrying high frequencies. The enhanced image G is thus in principle calculated as

$$G = aF - L \tag{11.27}$$

where the chosen weight a determines the degree of sharpening and L is the blurred (low-pass) version of F. Choosing $a = 10/9$ and L provided by convolution with the smoothing mask (see next section) sized 3×3 containing all elements equal to 1/9, and realizing that the original image would be obtained by a mask of

all zeros except the central 1, we obtain by subtraction a common sharpening mask:

$$\frac{10}{9}\begin{bmatrix}0&0&0\\0&1&0\\0&0&0\end{bmatrix}-\frac{1}{9}\begin{bmatrix}1&1&1\\1&1&1\\1&1&1\end{bmatrix}=\frac{1}{9}\begin{bmatrix}-1&-1&-1\\-1&9&-1\\-1&-1&-1\end{bmatrix}. \tag{11.28}$$

Note that if $a = 1$ and disregarding the multiplicative factor 1/9, we would obtain the first Laplacian mask in Equation 11.26, thus gaining the maximum sharpness at the cost of losing completely the area intensity information (low-frequency components) of the image. On the other hand, a value of $a >> 1$ emphasizes the original and diminishes the sharpening effect.

The sharpening mask in Equation 11.28 (disregarding the factor 1/9) may be alternatively interpreted as the Laplacian (Equation 11.26) subtracted from the original. Using the Laplacian mask as in Equation 11.24 in the same construction leads to the alternative equally popular sharpening mask

$$\begin{bmatrix}0&-1&0\\-1&5&-1\\0&-1&0\end{bmatrix} \tag{11.29}$$

In both sharpening masks, the sharpening effect is controlled by varying the central element around the indicated values in the range from about –1 to +2. The schematic plots in Figure 11.14 demonstrate the sharpening effect on an edge profile. Examples of practical sharpening using the masks in Equations 11.28 and 11.29, and the same masks with the central element modified, can be found in Figure 11.15 and Figure 11.16.

A formally more convincing derivation of this method is as follows [64]: Let the image $f(x, y)$ be represented by an initial concentration $g(x, y, 0)$ of a dye dissolved in a thin layer of fluid. Due to diffusion, the image is gradually blurred according to the diffusion equation

$$\frac{\partial g(x,y,t)}{\partial t}=k\nabla^2 g(x,y,t), \tag{11.30}$$

so that $g(x, y, t)$ represents the blurred image with the blur increasing with t. It is known that the diffusion leads to blurring each

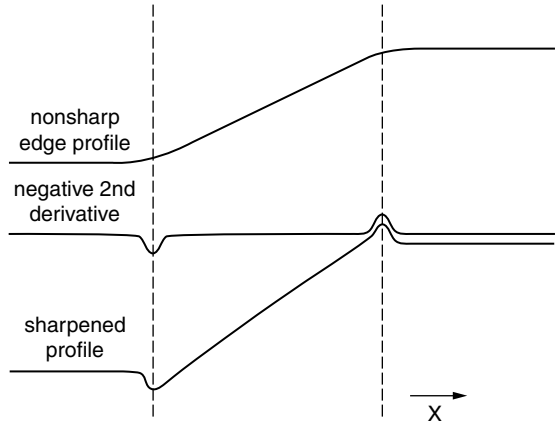


Figure 11.14 Explanation of sharpening edge profile by subtracting the Laplacian.

initial local (point) concentration of the dye to a "blob" described by the two-dimensional Gaussian function, the variance (width) of which increases with time. This corresponds well with most types of natural image blurring that might be considered for improvement by sharpening. Expanding the function $g(x, y, t)$ into the Taylor series around $t = \tau$ yields

$$f(x,y) = g(x,y,0) = g(x,y,\tau) - \tau \frac{\partial g}{\partial t}(x,y,\tau) + \cdots. \qquad (11.31)$$

When substituting for the time derivative from Equation 11.30 and neglecting the higher-order terms, we obtain

$$f(x,y) \approx g(x,y,\tau) - k\tau \nabla^2 g(x,y,\tau), \qquad (11.32)$$

which may be interpreted as the difference between the blurred image and the Laplacian of it, weighted by $k\tau$. Neither of these two parameters is known; however, their product that controls the degree of sharpening may be determined experimentally as the weight yielding the optimum output image.

11.2.3 Sharpening via Frequency Domain

The relative accentuation of high-frequency components, generally required by sharpening, can also be naturally done via frequency domain,

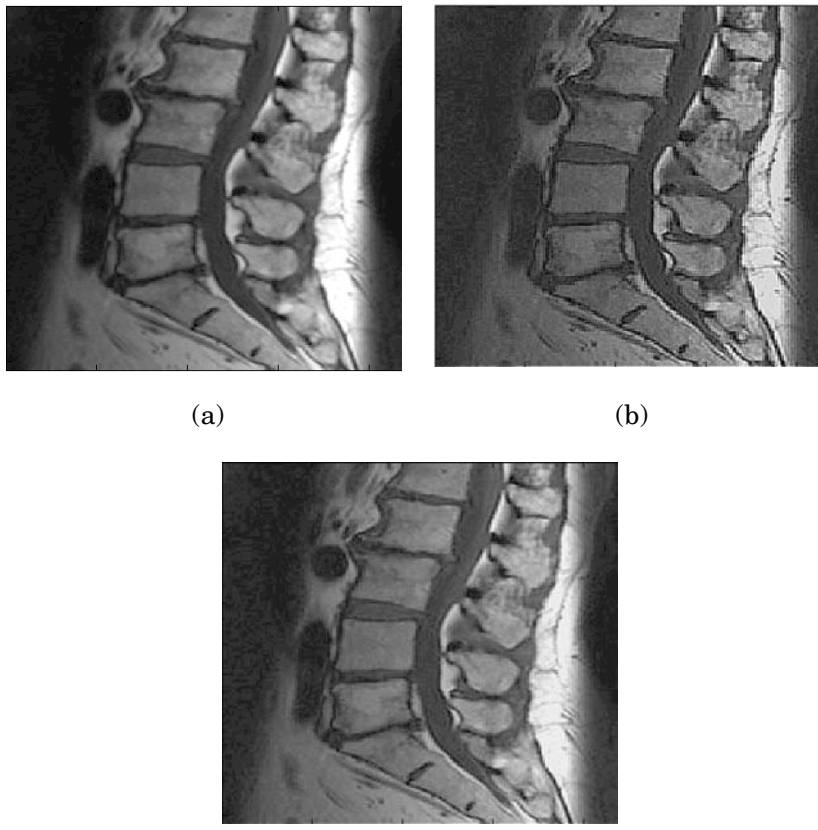


Figure 11.15 Examples of sharpening the (a) original by the operator in Equation 11.28 (b), and (c) by a similar mask with the central element equal to 12.

as schematically depicted in Figure 11.17 (see also, Section 1.3.3). The discrete two-dimensional spectrum of the input image is modified so that the part of upper frequencies is relatively emphasized; the resulting spectrum is then inverse transformed, thus obtaining the sharpened image.

The modification should be rather mild in the lack of information on the input image properties, in order not to introduce artifacts. The frequency responses that are used to multiply the input spectrum are typically rotationally symmetric, as the sharpening should usually be isotropic. They also should be real-valued (zero

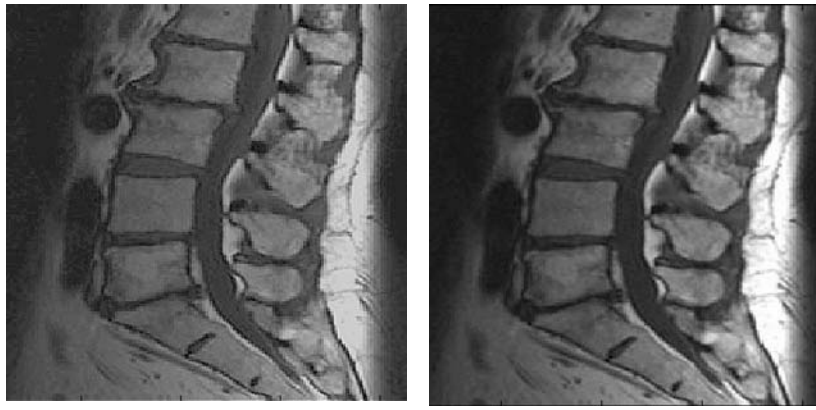


Figure 11.16 (Left) Example of sharpening by the operator in Equation 11.29 and (right) by a similar mask with the central element equal to 7.

phase) in order not to shift mutually the individual spectral components; this way, the crucially important phase information on the image is preserved. Obviously, only a one-dimensional description of the frequency response of the filter (its radial profile) is sufficient. Some typical curves are shown in Figure 11.18. The simplest is the linear enhancement (dashed line), possibly approximated as

$$H(u,v) = w = \sqrt{(u^2 + v^2)} \approx |u| + |v|. \tag{11.33}$$

However, such curves lead to the excessive loss of low-frequency components, namely of the zero-frequency component, i.e., of the overall image brightness and regional intensities; therefore, the characteristic is usually kept constant in a certain range of low frequencies (solid line). Another possibility is to shift the line upward by an

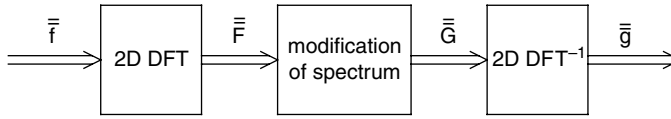


Figure 11.17 Frequency-domain filtering schematically.

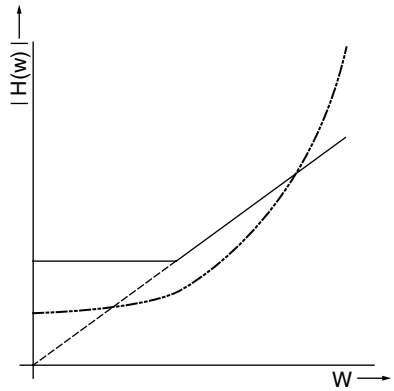


Figure 11.18 Typical curves describing the rotationally symmetric frequency response of sharpening filters.

additive constant (dotted line). Also, exponential enhancement of the type

$$H(u,v) = e^{\alpha w} \tag{11.34}$$

may turn out useful (dotted curve, see Figure 11.19). Naturally, all these curves must be approximated by their samples at the nodes of the sampling grid in the frequency domain.

While the above approach is a linear filtering, the following nonlinear modification of the input magnitude spectrum (sc., *root filtering*),

$$G(u,v) = |F(u,v)|^{\alpha}\, e^{j\varphi(u,v)}, \quad \alpha \in \langle 0,1 \rangle \tag{11.35}$$

has another philosophy. It again preserves the phase spectrum $\varphi(u, v)$ of the input image, but the amplitude is compressed; the closer α is to zero the higher the compression. (When $\alpha = 0$, the magnitude spectrum is equalized; surprisingly, the image after the inverse FT remains recognizable as the spatial information is mostly contained in the phase spectrum.) The basic idea is that the detail-carrying components are usually weaker than the low-frequency components; the (partial) equalization thus leads to a relative detail enhancement. Again, determining a proper value of α is usually a matter of experiment; an example is shown in Figure 11.20.

General observation seems to show that unless there is at least some preliminary spectral information on the character of blur that should be compensated (and then the problem belongs more to

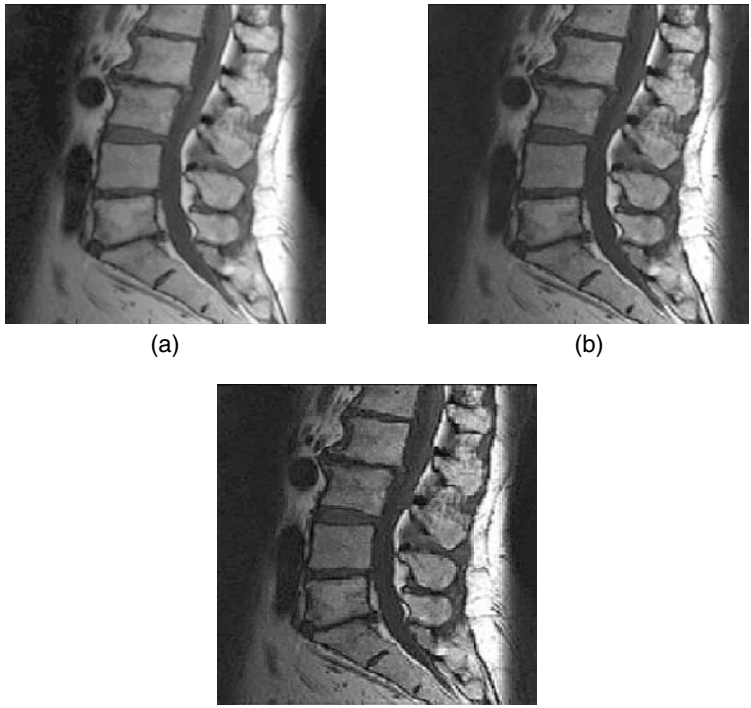


Figure 11.19 Sharpening via frequency domain — exponential enhancement of higher frequencies. (a) original, $\alpha = 0$; (b) enhanced images with, $\alpha = 1.5$; and (c) $\alpha = 3$.

the restoration field), the results obtained by local original-domain filtering (mask processing) are comparable to those obtained via frequency domain, but with a lower computational expenditure.

11.2.4 Adaptive Sharpening

Any sharpening inevitably increases the image noise, which has a relatively higher share at high frequencies. Sharpening thus further decreases the already lower signal-to-noise ratio (SNR) at these frequencies, and this phenomenon limits the degree in which the enhancement may be applied. Particularly disturbing is the noise in flat intensity regions, while at highly structured areas (edges, lines, small details, or even textured areas), human sight does not perceive the noise as disturbing. This leads to the idea to apply the sharpening algorithms only in the structured areas, while leaving

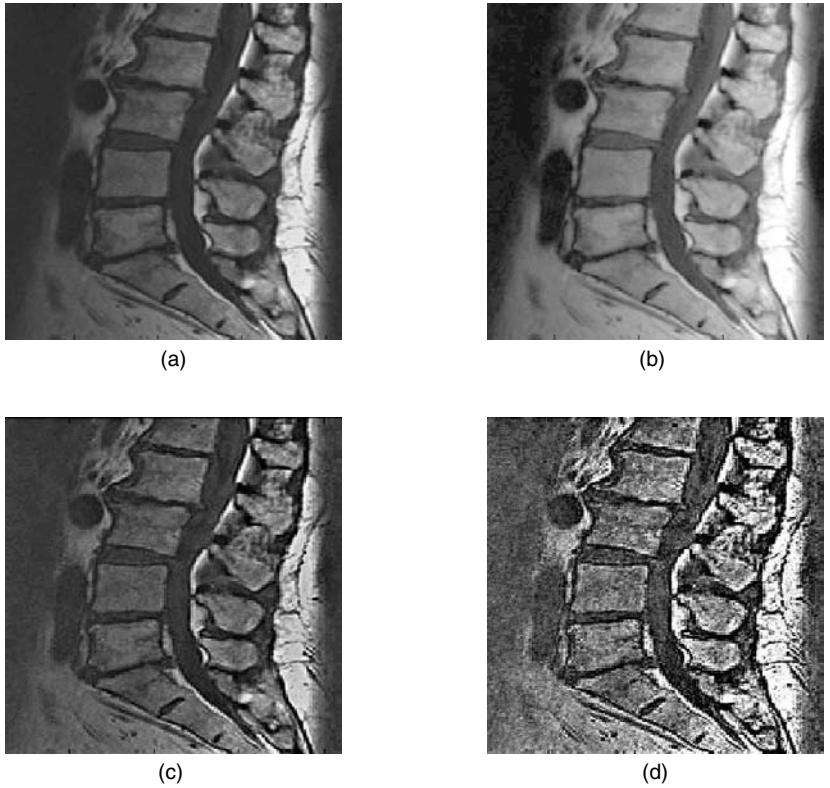


Figure 11.20 Sharpening by compressing the dynamics of the amplitude spectrum: (a) original image and (b–d) sharpened images with $\alpha = 0.9$, 0.7, and 0.5, respectively.

the flat regions untouched (also because there are no details to be enhanced). Such an approach taking into account the local properties of the image is a version of *adaptive sharpening*. It requires analyzing the image content locally in order to decide whether to apply the sharpening, and if so, with what intensity. The used local criteria on which the decision may be based are manifold, some quite generic, while others adapted to a particular image type. The simplest and most commonly used criterion is the local variance determined from a small neighborhood of the pixel in question: a high variance means a detailed area where a strong sharpening can be applied and vice versa. The threshold with which the local variance should be compared in case of binary decision is one of the selectable parameters of the method.

Figure 11.21 The result of adaptive sharpening. Note the pronounced sharpening effect at edges while preserving the smooth character of greater areas.

This approach (see example in Figure 11.21) may be generalized so that instead of the binary decision, the degree of sharpening may be controlled (e.g., by the central element value of the masks in Equations 11.28 and 11.29, together with the mask normalization by its sum of elements), depending on the value of the local criterion, e.g., the local variance. A similar approach, though with inverse decision, may be applied in adaptive noise suppression (Section 11.3).

11.3 NOISE SUPPRESSION

Noise of different origins is inevitably present in any image. Here, we shall deal with the situation that not much is known about its character; however, some properties may be apparent or easily identifiable, and these should be utilized when selecting a suitable noise-suppressing algorithm.

A simple classification of noise is thus needed:

- According to dependence on the image content:
  - *Image-independent noise* (e.g., communication interference or thermal noise).
  - *Image-dependent noise* (e.g., photographic grain); usually difficult to deal with and rarely a subject of image enhancement.

- According to its amplitude distribution and spatial distribution:
  - *Gray noise*, the values of which cover a certain continuous interval (e.g., Gaussian noise); this type of noise usually afflicts all image pixels, but the intensity values of the pixels remain basically preserved (or are supposed to); i.e., the typical noise amplitude is usually distinctly smaller than the intensity values.
  - Large amplitude *impulse noise* ("salt and pepper") having (almost) binary distribution of nearly extreme values; only isolated pixels (or small groups of them) are usually affected, but the intensity information of affected pixels must be considered lost.
- According to the noise relation to the image content:
  - *Additive noise*, simply added to the original pixel values; most common case, easiest to deal with; this type is primarily subject to suppression in the frame of image enhancement.
  - *Multiplicative noise* — each pixel intensity has been multiplied by the noise amplitude at this pixel; such a noise may even originate from additive noise linearly dependent on the image content.
  - *Other types*, e.g., convolutional noise.
- According to noise character in the frequency domain:
  - *Wideband noise* — most of the previous types of noise.
  - *Narrowband noise* — combination of a few narrowband additive or multiplicative signals manifesting themselves as stripe structures (possibly multiple), or moiré patterns.

The basic problem with all simple noise-suppressing methods is how to prevent a substantial loss of sharpness (or even detail). As the noise occupies the same frequency range as the detail-carrying spectral components, it must be expected that noise suppression unavoidably leads to a certain blur and loss of detail. With the lack of *a priori* information on spectral SNR distribution (as used extensively in image restoration, e.g., via Wiener filtering), the noise is suppressed independently of its power spectrum, and also regardless of the importance of a particular spectral component for the image content. The degree with which the *blind noise suppression* may be applied is then a compromise between the desirable image smoothing, leading to an improvement of the overall SNR on one hand and, on the other hand, the losing of the useful details and possibly the pleasing feeling of sharpness.

11.3.1 Narrowband Noise Suppression

Narrowband noise is characterized by its point-wise representation in the frequency domain: only a few possibly strong frequency components form the noise. It is then mostly possible to remove these obviously false frequency coefficients from the discrete two-dimensional spectrum and to reconstruct the image from the remaining majority of spectral information. This may be interpreted as a frequency-domain filtering, as in Figure 11.17; this approach implies that the noise should be additive. In this case, the filter is obviously of a stop-band type, usually rejecting several narrowbands (mostly so narrow that they are represented by isolated frequency samples or small groups of them). An example is shown in Figure 11.22.

The above procedure may be improved by replacing the false spectral coefficients by values interpolated from neighboring coefficients of the spectrum. This is justified by the experience that the

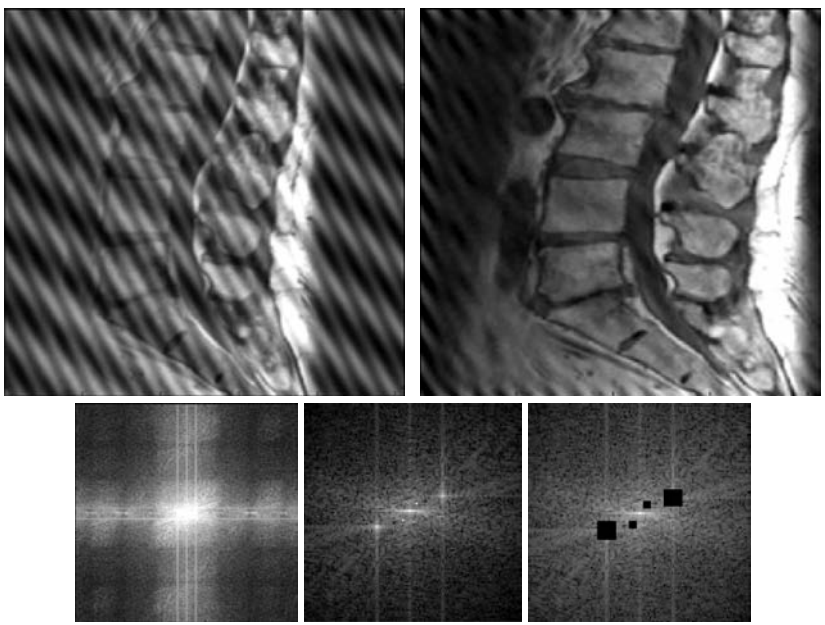


Figure 11.22 Upper row: (Left) Image distorted by two narrowband interfering signals and (right) filtered version of the image. Lower row: (Left) Its spectrum, (center) detail of the spectrum, and (right) detail of the spectrum modified by deletion of suspicious disturbing components.

natural image spectra are usually rather smooth; i.e., the closely situated coefficients in the discrete spectrum may be considered samples of a smooth, e.g., bilinear or bicubic, surface.

The situation is slightly more complicated when the narrowband noise is multiplicative, i.e., when the image is modulated by (or it modulates) a narrowband disturbing signal. This is, e.g., the case of periodic interference of moiré patterns when the amplitude of the image signal may change around the proper values in the range of up to <0, 200%>, or the disturbance by the line structure of a TV scan, which may be understood as multiplying the original continuous image with a black-and-white (0 and 100%) stripe structure. The multiplicative image mixture must first be transformed into the additive one by logarithmic transform and consequently linearly filtered as above; finally, the inverse (exponential) nonlinear transform must be applied in order to remove the intensity distortion. This way, homomorphic filtering (Section 1.3.4) is effectively applied, with both the input and output operations being multiplication. It may be objected that the spectrum of the mixture is changed by the logarithmic point transform so that even the noise is partly distorted and its spectrum would be richer in harmonics. Though this is theoretically right, the distortion is rather mild, as the logarithmic curve is smooth and may often be closely approximated by a line in the extent of the signal amplitudes; thus, the influence of the harmonics can often be neglected.

11.3.2 Wideband "Gray" Noise Suppression

This is the noise usually caused by imperfections of the imaging system and its circuitry, the communication channels via which the image is transported (in all these cases mostly thermal noise), or by random errors due to quantization during analogue-to-digital (A/D) conversion, or due to rounding of the results of computation. Such noise typically affects all pixels in the image, often but not necessarily all of them to the uniform extent, as to the noise power concerns. The noise value at a particular pixel of a concrete image may then be considered a realization of a random variable with zero mean and a smooth* probability distribution; the most frequently met distributions are Gaussian characterized by a certain variance (e.g., in case

Smooth is meant here in the sense that neighboring probabilities do not differ wildly and may be considered samples of a smooth curve.

of thermal noise), or uniform in a certain interval of values (e.g., due to rounding errors). The noise variables at different pixels may be considered in most cases (but not always) independent, meaning spectral whiteness, and usually all with the same distribution; these properties apply often not only in the frame of an individual image, but also for all images of a certain group (type, class).

A pixel value is given by the sum of the proper image intensity and the noise realization. As the concrete noise values are not known, averaging a number of noise realizations is the only way to suppress it. The basic idea is that while the image intensity is supposed to be identical (at least approximately) in all averaged samples, and therefore remains unchanged after averaging, the noise power is suppressed. Really, in the case of a zero-mean noise, the noise power is given by its variance σ_v^2. When a number N of identically distributed noise variables v_i are averaged, the variance of the resulting variable can be shown

$$\sigma_{v\,aver}^2 = \mathrm{var}\left\{\sum_{i=1}^{N} \frac{1}{N} v_i\right\} = \sum_{i=1}^{N} \frac{1}{N^2} \mathrm{var}\{v_i\} = \frac{\sigma_v^2}{N}, \qquad (11.36)$$

thus decreasing linearly with the number of averaged values. As long as the useful signal is identical in all considered samples, the improvement in the power signal-to-noise ratio is proportional to the number N of averaged samples. As the typical amplitude of noise is given by the standard deviation $\sigma_{v\ aver}$, the amplitude-SNR improves with the square root of N.

However, this conclusion applies only to averaging under ideal conditions, i.e., for averaging of N identical images differing only in noise realizations, as mentioned in Section 10.4.1 on fusion. A similar situation may arise in a single image with a periodical structure, where the individual periods may be averaged as well (e.g., in crystalography). Under such circumstances, each pixel is averaged over the ensemble of images (or subimages) independently of its neighbors or, more generally, differently situated pixels, and therefore no spatial blur appears under such *ensemble averaging*.

When just a single image is available, different realizations of noise can be obtained only from neighboring pixels of the processed pixel, which leads to *local averaging*. The condition of a fixed image value is then fulfilled only approximately, i.e., when the image intensity is about constant in the chosen neighborhood. This may be true inside flat image areas, but obviously it is violated near edges, small details,

and object borders. Here, the averaging leads to mixing of different image values and consequently results in blur. This is the price for the (partial) noise suppression. Obviously, the degree of SNR improvement (*image smoothing*) depends on the size of the neighborhood, providing that the noise is spatially homogeneous; unfortunately, the blur increases with this size as well. The selection of the averaged area is thus a compromise, usually determined by a just acceptable blur.

The spatial averaging may be generally expressed as

$$g_{i,k} = \sum_{\mathrm{A}_{i,k}} c_{m,n} f_{m,n}, \quad (m,n) \in \mathrm{A}_{i,k}, \tag{11.37}$$

where $\mathrm{A}_{i,k}$ is the chosen neighborhood of the pixel position (i, k) and $c_{m,n}$ are weights given to individual pixels contributing to the new value $g_{i,k}$. The shape and size of the neighborhood $\mathrm{A}_{i,k}$ is usually chosen by the operator and invariable; however, in adaptive smoothing it may depend on the position (i, k), as we shall see later. Similarly, the weights need not all be identical, so that a weighted average may be used, and the set of weights may also depend on (i, k). However, it can be shown that the best improvement of SNR for a given N is achieved when the weights are uniform $(1/N)$; unfortunately, it is the worst case as to blur concerns.

Most commonly, the simple smoothing in the 3×3 neighborhood of the calculated pixel is used. As the averaging is obviously a linear local operation, it may be realized by mask operators (Section 2.2.1), with the masks, e.g.,

$$\frac{1}{9}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \frac{1}{10}\begin{bmatrix} 1 & 1 & 1 \\ 1 & 2 & 1 \\ 1 & 1 & 1 \end{bmatrix}, \quad \frac{1}{16}\begin{bmatrix} 1 & 2 & 1 \\ 2 & 4 & 2 \\ 1 & 2 & 1 \end{bmatrix}. \tag{11.38}$$

While the first mask represents the plain average (Figure 11.23), the other two gradually emphasize the influence of the central element (i.e., of the original $f_{i,k}$ in the average); the SNR improvement is best for the left mask, and the least blur is induced by the rightmost mask. The last mask approximates weighting by a bell window, e.g., Gaussian with its certain good properties. As the averaging may obviously also be interpreted as convolution of the image with the mask*, representing PSF, the frequency response of the averaging filter may be

*The 180° rotation of PSF is clearly irrelevant due to symmetry of the masks.

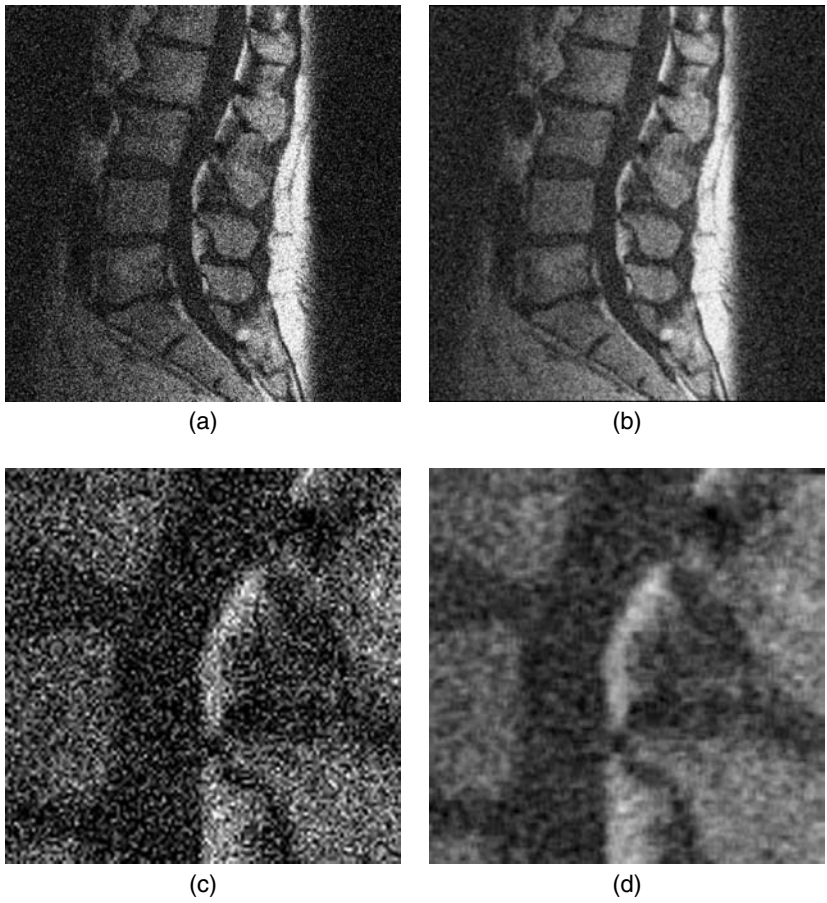


Figure 11.23 Upper row: (Left) Example of image distorted by Gaussian noise and (right) its plain convolutional smoothing. Lower row: Details of the corresponding images.

considered. All of the operators suppress high-frequency components to an extent. The frequency response of the left mask suffers naturally with side lobes, as the mask is a rectangular window; the other masks have less pronounced side lobes, namely the rightmost one, which may be an advantage in a more complex chain of processing. Greater square matrices (up to about 9×9) are occasionally used as well when a high noise suppression is needed; nevertheless, it is achieved at the cost of substantially more severe blur.

11.3.2.1 Adaptive Wideband Noise Smoothing

The blur appearing as a by-product of noise suppression is particularly annoying in the highly structured image areas, where it leads to loss of details and smearing of edges. On the other hand, the noise in these areas is less disturbing to human sight, so that the noise suppression may be lower here, or omitted. This leads to the idea of *adaptive noise smoothing*—higher smoothing in flat image areas where the noise is disturbing and the blur is harmless, and lower or no smoothing in the structured areas. The extent of smoothing (and blur) may be influenced by the size of the mask, which in turn should be controlled by the local content of detail.

The criterion evaluating the local detail content may be formulated most simply, based on the local variance of image intensities, as

$$\sigma_{i,k}^2 = \frac{1}{\mathrm{count}(\mathrm{B}_{i,k}) - 1} \sum_{\mathrm{B}_{i,k}} (f_{m,n} - \bar{f}_{i,k})^2, \quad (m,n) \in \mathrm{B}_{i,k}$$
$$\bar{f}_{i,k} = \frac{1}{\mathrm{count}(\mathrm{B}_{i,k})} \sum_{\mathrm{B}_{i,k}} f_{m,n}. \tag{11.39}$$

The size (and perhaps even shape) of $\mathrm{B}_{i,k}$ is then influenced by $\sigma_{i,k}^2$; a higher variance reduces the size of $\mathrm{B}_{i,k}$. The size of $\mathrm{B}_{i,k}$, the weights $c_{m,n}$, and the dependence on $\sigma_{i,k}$ are all parameters of the method that allow adaptation of the processing to the character of the image type. Obviously, this approach is in a sense the counterpart of the local variance–controlled adaptive sharpening, as mentioned in the previous section.

Another approach to adaptation of spatial smoothing (Equation 11.37) to the local image properties consists of *averaging with neighborhood-growing*. The area $\mathrm{A}_{i,k}$ is initially formed by the single pixel (i,k) only; then it grows gradually, taking in the neighboring pixels on the condition that they are of similar intensity as $f_{i,k}$. This way, the neighborhood grows in the directions where the intensity is flat, while a nearby edge or other detail of significantly different intensity is not included; the shape of the neighborhood is thus space variant. This way, the average does not contain pixel values differing by more than a certain small difference (see below); consequently, the blur of structures is prevented. The neighborhood may grow iteratively until a fixed size (the chosen pixel count N) is reached; then the

improvement of SNR would be uniform on the whole image area. When the closely valued pixels within a chosen maximum distance are exhausted before reaching N (as in highly structured image areas), the neighborhood remains at the reached smaller size, thus providing a lower degree of smoothing desirable in these areas. Alternatively, the growing may be stopped when reaching a given sum S of the included pixel values; in this case, a higher SNR improvement is achieved for lower-level pixels, as is useful, e.g., for gammagraphic images.

The similarity condition may be formulated as

$$|f_{i,k} - f_{i+j,k+l}| < s\,, \quad j,l \in \langle -L, L \rangle, \tag{11.40}$$

where L determines the maximum directional distance of the neighborhood element from the pixel (i, k). The limit s of absolute difference serves as the threshold for the decision whether to include the pixel $(i+j, k+l)$ into the averaged neighborhood. It should be related to the standard deviation of noise (e.g., $s = b\sigma_v$), may be dependent on $f_{i,k}$, and/or be explicitly spatially variant. Again, the parameters N or S, L, s (or b) and possibly the spatial variability of the similarity criterion determine the exact result of smoothing.

The adaptive smoothing methods (often improperly called intelligent smoothing) may provide a visible noise suppression without an important deterioration of sharpness, naturally at the cost of substantially higher computational effort (Figure 11.24).

Figure 11.24 Example of adaptive noise smoothing in the original image of Figure 11.23 using neighborhood growing with; $N = 50$, $L = 11$, (left) $s = 0.3$; and (right) $s = 1.0$.

While the isoplanar smoothing can be done alternatively via frequency domain (though it pays off only when at least some information is available on the spectral properties of noise and image), the adaptive methods, being space variant, obviously do not have direct frequency-domain counterparts.

Numerous experiments on suppressing gray noise in the scale-space domain of wavelet transform by modifying the spectral coefficients have appeared in recent literature. Some of them provided useful results; however, the reasoning by which coefficients should be modified (and how) relies mostly on experiment and thus does not seem very convincing so far; therefore, there is still a lack of generally applicable rules.

11.3.3 Impulse Noise Suppression

The *impulse noise* ("pepper-and-salt" noise) afflicts only isolated, usually randomly positioned pixels; due to the randomness, it may happen that a few neighboring pixels are hit. The noise amplitudes are typically extreme so that the affected pixel also acquires mostly extreme intensity — near to white or black. Obviously, the information carried originally by such a pixel is lost. The only remedy is then to replace such a pixel with a new one derived somehow from its neighborhood.

There are thus two problems, the first of which is more demanding:

1. How to detect the false pixels
2. How to calculate a suitable substitute

The detection of a false pixel can only be based on its distinct dissimilarity with the neighborhood; locally inconsistent pixels are sought. The most reliable is still the interactive detection by a human operator easily finding the false pixels based on experience and intuitive decision, taking into account context information; this approach is typical for retouching photographs.

The automatic *false-pixel detection* may be based on artificial intelligence methods that would also take into account the contextual information, image probability description, etc. However, this is a rather demanding approach and usually much simpler criteria, based only on differences between the intensity $f_{i,k}$ of the pixel in question and the intensities of the surrounding pixels, are used in the frame of image enhancement.

Some typical criteria of inconsistency are listed below:

- Limit on the sum of absolute differences

$$\sum_j \sum_l |f_{i,k} - f_{i+j,k+l}| > S, \quad (j,l) \in \mathrm{A}, \tag{11.41}$$

  where A defines the considered neighborhood of usually a constant number N of pixels. The limit S determines the sensitivity of the detector: the lower it is, the more pixels will be detected as false. The proper value minimizing both probabilities of false decision, falsely positive and falsely negative, might be found using Bayesian criteria when the probability distributions of the image intensities and noise amplitudes are known. However, these characteristics are usually unavailable in image enhancement, and the suitable parameter S has to be determined by experiment. Similarly, the size (and perhaps even shape) of the neighborhood A is also a parameter to be suggested or determined by a trial-and-error approach; commonly, A may be a square 3×3 or 5×5. With the increasing size of A, the sensitivity to noise in averages is reduced and the ability to detect even groups of false pixels improves, but the detection becomes less locally determined.
- Limit on the number of dissimilar pixels in the neighborhood:

$$\mathrm{count}\{(j,l) \in \mathrm{A} : |f_{i,k} - f_{i+j,k+l}| > s\} > M. \tag{11.42}$$

  Besides A, this detector has two parameters: s determines the sensitivity to individual differences and M defines how many distinctly different neighbors indicate a false pixel. The last parameter influences particularly the behavior of the detector in the vicinity of edges.
- Limit on the absolute difference to the average of the neighborhood:

$$\left| f_{i,k} - \frac{1}{N} \sum_j \sum_l f_{i+j,k+l} \right| > s; \quad (j,l) \in \mathrm{A} \tag{11.43}$$

  The parameter s again determines properties of the detector, together with the size and shape of A.

When other properties of false pixels are known, as, e.g., some geometrical regularity in the probability of their occurrence, they should be included into the decision algorithm.

The intensities of the pixels detected as false are consequently to be replaced with intensities interpolated somehow from a neighborhood that may be similarly shaped as the detection neighborhood. The simple possibilities of the estimation are:

- The nearest unaffected neighbor.
- The average of the neighborhood, not including the pixels affected by the noise.
- Interpolation from the unaffected neighboring pixels, e.g., linear, bilinear, or bicubic (see Section 10.1.2), rarely more complex in the frame of enhancement.
- In binary images, the value prevailing in the neighborhood would be taken, possibly constrained by requirements on connectivity of areas.

Note that the described detection-and-replacement method means nonlinear processing, even though the interpolation itself may be linear. The decision making and the following spatially irregular replacement obviously violate the conditions of linearity.

The above method, which either leaves the original intensity or replaces it with the interpolated value, can be further refined by combining both possibilities. The underlying idea is that the detection provides only probable decision on the consistency of the pixel $f_{i,k}$, with a possibility of error. The credibility of the decision depends obviously on the value of similarity criterion; when the criterion is close to the threshold, the probability of erroneous decision is high. This way, it is possible to define a simple function

$$p(\text{criterion value} - \text{threshold}) \in \langle 0,1 \rangle, \tag{11.44}$$

which evaluates the estimated probability of $f_{i,k}$ being a proper value. Then it is possible, instead of switching, to interpolate between the two available values when calculating the new value $g_{i,k}$ of the pixel,

$$g_{i,k} = p f_{i,k} + (1-p)\bar{f}_{i,k}, \tag{11.45}$$

where $\bar{f}_{i,k}$ denotes the replacement value interpolated by some of the methods described above.

Results of similar quality as those with the detection-and-replacement method can be obtained by another nonlinear method — *median filtering* or similar order statistics operators (see

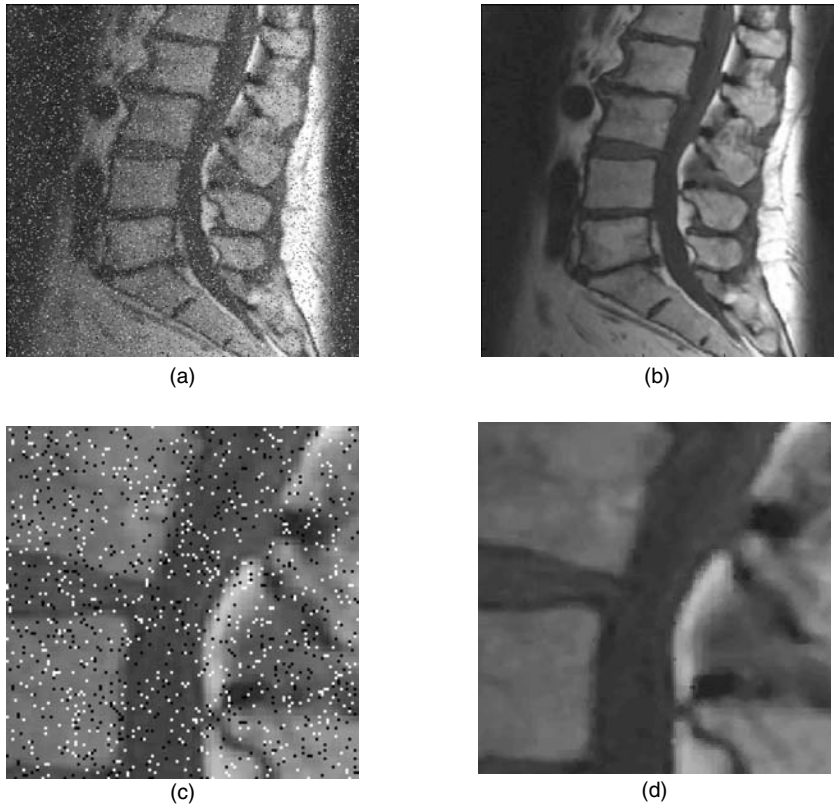


Figure 11.25 Suppression of impulse noise by median filtering: (left) noisy image and (right) the same image after median filtering. Below: Corresponding details.

Section 2.2.2). In comparison with the above method, the explanation of its effects is less transparent, but the results in the case of impulse noise interference are convincing (Figure 11.25). It has been observed that the median filter removes all small distinctly differing objects up to about the size of the used mask; greater objects are preserved without their borders and edges being smeared.

It is obvious that the methods, designed to suppress the impulse noise, are in principle unsuitable for suppression of the gray type of noise, which afflicts all pixels.

11.4 GEOMETRICAL DISTORTION CORRECTION

Restitution of simple geometrical distortion may also constitute a part of the image enhancement field, as long as the type and parameters of the applied geometrical transform are estimated by the user, usually in an interactive manner. Commonly, visible deformations due to imaging geometry, namely perspective distortions, or consequences of imperfect lens systems, like pincushion or barrel distortion, are considered

In principle, the geometrical transforms used for restitution are the same, as mentioned in Section 10.1.1; however, in the frame of enhancement, the choice of the transform type is usually very limited. Primarily, only a single type of transform is normally offered to the user for a particular task, and moreover, the parameters of the transform are usually fixed, except for a single one (exceptionally a few), which is to be determined by the user interactively, based on the visual appearance of the image being corrected. The interaction may incorporate manual definition of landmarks to be put to predefined positions by the restitution procedure, which would allow algorithmic derivation of the needed transform; however, exact restitution of carefully identified distortion is a subject of image restoration (Chapter 12) rather than enhancement.

12

Image Restoration

The concept of image restoration differs substantially from the idea of image enhancement. While enhancement aims at improving the appearance of the image or its properties with respect to the following analysis (by a human operator or even automatic), the goal of restoration is to remove an identified distortion from the observed image $\hat{\mathbf{g}}$, thus providing (in a defined sense) the best possible estimate $\hat{\mathbf{f}}$ of the original undistorted image $\mathbf{f}$. The observed image may be distorted by blur, geometrical deformation, nonlinear contrast transfer, etc., and is usually further degraded by additive or otherwise related noise $\boldsymbol{\nu}$. The identification of the properties of distortion (i.e., of the distorting system, the disturbing noise, etc.) therefore forms an essential part of the restoration process. Having described the distortion formally by a mathematical model with measured or estimated parameters, we can try to invert the model and obtain the restored image (estimate of the original) as the result of applying the inverse procedure to the observed (measured, received) image. The schematic representation of the distortion and restoration process is depicted in Figure 12.1.

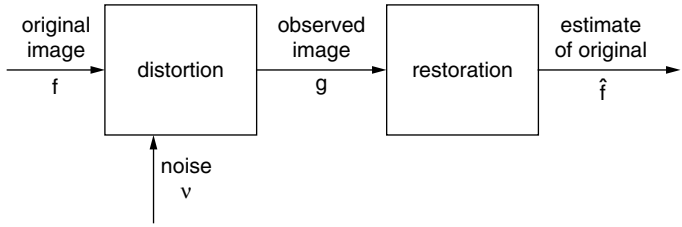


Figure 12.1 The chain of distortion and restoration of images.

The images may be distorted in many directions. In this chapter, we shall deal primarily with the following types of distortion, which cover most of those met in practice:

- Intensity distortion, global or space variable
- Geometrical distortion
- Blur, e.g., due to defocus or motion
- Interference by noise of certain properties

The methods of identification of distortion models and parameters are specific to individual types of distortion. Therefore, the identification will be discussed in the individual sections, usually before the actual methods of restoration.

Basically, two approaches are used in restoration. The conceptually simpler of them means formulating distortion models that can be directly inverted (as, e.g., for purely linear distortion); solving the equation systems or using closed formulae obtained by the inversion then provides straightforwardly the estimate of the original. When noise cannot be neglected, the exact inversion is impossible, as the noise represents an unknown stochastic component. In these cases, an approximate inversion minimizing the noise influence must be sought; the commonly used approach is the least mean square (LMS) error approach, which may lead to closed formulae as well (e.g., Wiener type filtering). Often, however, the realistic distortion models are so complex (structurally and/or mathematically) that the direct inversion is not feasible or suffers with too high errors. In such a case, the way that proved successful is gradual optimization aiming at an extreme of a criterion function derived from the distortion model. The rich body of optimization theory and iterative algorithms combined with the signal-theoretical concepts thus provide a powerful tool that often enables the recovering of useful information even from heavily distorted images. Different concepts of this kind will be briefly discussed in Section 12.4, though they are still not routinely used in

medical imaging. However, their inspiring potential may be considerable for an interested reader. A deep yet comprehensible treatment of restoration problems is not frequently met in image processing literature. Of the referenced literature in Part III, the main sources to this chapter, [26], [63], [64], are inspiring and can be used to find further details, as well as [11], [12], [59]. References [1], [39], [37], [23], [76], [40] can also be consulted for further reading and, e.g., [44], [62] for the relevant mathematics.

It should be mentioned that the reconstruction from projections and some other image reconstruction techniques used in medical imaging belong, in fact, to the field of restoration as well. The measured data may be considered heavily distorted (even unintelligible) observed images, on the basis of which the proper image slice or three-dimensional image data are obtained by inverting the transform that has yielded the data (e.g., Radon transform). However, these algorithms, crucially important primarily in medical imaging, have already been discussed in Chapter 10 and thus will not be treated here.

12.1 CORRECTION OF INTENSITY DISTORTIONS

The correction of intensity distortion is identical to the contrast enhancement procedures, as to the realization of the contrast transform concerns (Section 11.1). Once the transform is defined — for the digital environment in the form of a lookup table (LUT) — the problem may thus be considered solved.

What distinguishes the restoration from contrast enhancement is the goal: the correction seeks to restore the linear (or another clearly defined) relation between the imaged parameter, e.g., the reflectivity at a spatial position on the image field, and the response, i.e., the intensity value of the corresponding pixel. The assessment of the proper contrast transform requires a preliminary identification of the distortion, usually using a phantom — an object of known spatial distribution of the imaged parameter — for calibration of the imaging system.

Whether the correction is global, i.e., identical for all image pixels, or if it has to be done individually for separate image areas or even individual pixels, depends on the character of the imaging system and on the kind of distortion.

12.1.1 Global Corrections

Global correction concerns the cases when the relation between the measured parameter and the pixel response is space invariant, i.e.,

identical for all image pixels. Here belong primarily the imaging systems using a single sensor scanning, in a way, the complete image, as, e.g., in rotational drum scanners, or bolometric infrared cameras with mirror-based scanning. The modern integrated field sensors (e.g., CCD matrices) may have so uniform properties of individual pixel sensors that the same restoration approach may be applicable as well, if needed.

The identification then consists of a series of measurements of phantoms of different values of measurement parameters, e.g., for an optical system, of a set of gray plates with different shades of gray. With respect to the spatial invariance of the input–output relation to be identified, it is possible to also measure a single plate with a known distribution of gray degrees, e.g., an optical wedge or an image of gradually darkening stripes. This way, a set of measurements may be obtained defining the input–output transform in a point-wise manner, as in Figure 12.2. The complete nonlinear relation $g_{i,k} = \mathcal{N}(f_{i,k})$ between the original parameter value $f_{i,k}$ and the sensor output $g_{i,k}$ may then be obtained by interpolation. Inverting the function leads to the required contrast transform

$$\hat{f}_{i,k} = \mathcal{N}^{-1}(g_{i,k}), \tag{12.1}$$

yielding the estimated parameter value based on the observed value contained in the processed image. Applying the same inverse transform to all image pixels provides the (estimated) image of linearized intensities. When constructing the inverse function, its discrete character should be observed and possible ambiguities prevented by proper treatment.

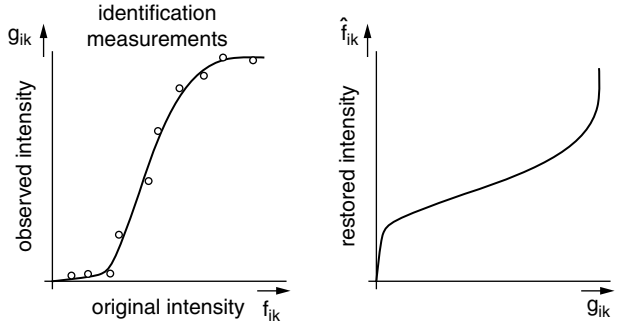


Figure 12.2 Schematic example of identified nonlinear curve of intensity transform and its inversion to be used in restoration.

12.1.2 Field Homogenization

The measured intensity at an individual optical-image pixel depends, besides on a measured parameter such as reflectivity, on the illumination at the pixel position and on the sensitivity of the particular sensor that measures the reflected or transmitted light at that position. If any of the last two fields (illumination, sensitivity) are uneven, i.e., the illumination distribution on the image area and/or the sensitivity of the sensor(s) is spatially variable, the field of image becomes inhomogeneous in the sense that a uniform subject (e.g., a gray plane of a constant reflectivity) is not represented by an image of uniform intensity. It is the purpose of the field homogenization to remedy the situation.

The identification as well as restoration of a single pixel value is in principle the same as in the global case described above. However, the inverse function generally varies for different pixels, which makes the problem much more demanding as to both measurement and memory requirement concerns (a complete individual lookup table would have to be formed and saved for each pixel). Similarly, the restoration of measured images, based on this set of LUTs, would be correspondingly complicated.

This is why the identified functions (and consequently also the inverse transform functions) are often approximated by linear ones, as in Figure 12.3. The plain proportionality as in the left plot has a single parameter—the slope describing the product of local illumination and local sensitivity. The restoration, based on this model,

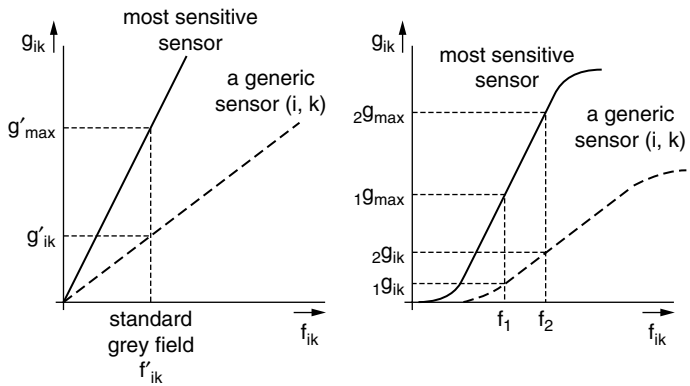


Figure 12.3 Simplified linear distortions of intensity as used in field homogenization: (left) proportionality with differing sensitivity and (right) linear dependence respecting different dead bands and sensitivity.

thus enables identification and compensation for the unevenness in both fields, providing that the characteristics of the sensing elements are all linear, though with different sensitivities. This is reasonably fulfilled, e.g., in modern semiconductor elements. In this case, a single measurement is needed for each pixel that can be accomplished by imaging a phantom with a constant gray shade $f'_{i,k} = \text{const.}, \forall i,k$ on the whole image area, and measuring the individual responses $g'_{i,k}$. To equalize the sensitivity and illumination to the level of the most sensitive (and/or illuminated) pixel, the response of which is $g'_{\max}$, each pixel value of an observed image to be corrected must be multiplied by an individual constant $K_{i,k}$,

$$\hat{f}_{i,k} = g_{i,k} K_{i,k}, \qquad K_{i,k} = \frac{g'_{\max}}{g'_{i,k}}. \tag{12.2}$$

The matrix $\mathbf{K}$ of the correction coefficients $K_{i,k}$ is obviously of the same size as the image field. It may be saved as a parameter matrix of the imaging system should the sensitivities and illumination be time invariable and applied to every measured image.

The vacuum sensors (and some others) have the characteristics of the type indicated in the right plot by dotted curves, with a certain insensitive range in low intensities, a working range that is approximately linear, and the saturation region at high intensities (dotted curves). The intensities in a real image should be constrained to the working range by a proper exposure; anyway, the almost zero and maximum values would not be restorable even if the model were complicated this way. It is therefore sufficient to model the linear part of the characteristics; such a model allows the taking into account of two individual parameters for each pixel: the sensitivity (perhaps combined with illumination) and the range of insensitivity. The identification of the general linear function requires two points to be measured for each pixel. Two evenly gray phantoms are needed with differing shades of gray, f_1, f_2, constant on the field of view; measuring each of them provides two sets of values, ${}_1g'_{i,k}$ and ${}_2g'_{i,k}$. The linearized characteristic of an arbitrary pixel (i, j) is then characterized by the slope $K_{i,k}$ and the intercept $c_{i,k}$; the corrected estimate of the intensity $f_{i,k}$ in an observed image can easily be shown

$$\begin{aligned} \hat{f}_{i,k} &= K_{i,k} g_{i,k} + c_{i,k}, \\ K_{i,k} &= \frac{\Delta g'_{\max}}{\Delta g'_{i,k}}, \quad c_{i,k} = {}_1g'_{\max} + {}_2g'_{\max} - K_{i,k}({}_1g'_{i,k} + {}_2g'_{i,k}), \end{aligned} \tag{12.3}$$

where the meaning of symbols is visible from Figure 12.3b. Thus, two matrices **K** and **C** of the correction coefficients, sized as the processed image, are needed, again being characteristics of a particular imaging system (possibly under a particular illumination).

Naturally, the model need not be only linear. Higher-polynomial approximations or point-wise defined curves may model the individual sensor (pixel) characteristics; however, the complexity of both identification and correcting calculation increases substantially, together with increased memory requirements, enabling the saving of many coefficient matrices.

12.1.2.1 Homomorphic Illumination Correction

Though it is at the border between restoration and enhancement, a method to suppress the influence of the uneven illumination via homomorphic filtering should be mentioned. It is a blind method assuming only that the illumination is varying slowly in space, thus being represented by low-frequency spectral components. On the other hand, the reflectivity distribution forming the useful image content is supposedly formed primarily by details that occupy a higher-frequency range. However, simple suppression of the lower end of the spectrum would not lead to the required (rough) equalization of the field as to the illumination concerns, because the disturbance is multiplicative. It is then necessary to convert the multiplicative mixture of illumination and reflectance values into an additive mixture by a logarithmic point-wise transform. The lower-frequency components are consequently partly suppressed by linear filtering (in either the original or frequency domain) to obtain the image hopefully without (or rather with a limited) illumination influence. The obtained image must be point-wise transformed by the inverse (exponential) transform, in order that the resulting contrast is not deformed. The obtained homomorphic filter (Section 1.3.4, Figures 1.12 and 1.13) may provide substantial improvement, but obviously not a complete correction, as the illumination distribution has not been identified in detail.

12.2 GEOMETRICAL RESTITUTION

The exact geometrical restitution requires complete information on the distorting transform, i.e., its type, formula, and parameters, should the restitution be formulated as global. This is only rarely available; rather, the information on the amount of local distortion may be available in

the form of the disparity map between the observed image and the ideal nondistorted image. Identification of the distortion thus requires imaging of a geometrically standard phantom, e.g., a planar scene with known geometry and distinctive details to serve as landmarks. The disparities for individual pairs of landmarks in the distorted and undistorted images, together with their coordinates, provide the information needed to design the correcting transform. The types of common geometrical transforms, including the piece-wise transform often used in the more complicated cases of formalized restitution, and methods to identify the transform parameters, as well as needed interpolation methods, were discussed in Section 10.1.

12.3 INVERSE FILTERING

As the name suggests, inverse filtering serves to restore images that may be interpreted as outputs of two-dimensional linear (mostly space-invariant) systems, e.g., imaging or communication systems accepting the original undistorted image as input. In other words, the subjects of inverse filtering are observed images distorted by convolution with a certain point-spread function (PSF). The practical consequence of such convolution is usually blurring of the image, as the imaging and transfer imperfections act as low-pass (or high-frequency-suppressing) filters. Seemingly, when the distortion may be described by a concrete frequency transfer function, it should suffice to design the filter with the inverse frequency response and apply it on the distorted image. Unfortunately, the unavoidable presence of noise often complicates the action of the inverse filter to the extent that the result of the plain inverse filtering is useless. It is then necessary to modify the filtering even substantially to prevent or suppress the adverse impact of noise.

The identification of a problem leading to inverse filtering should thus embrace the identification of both the PSF (or alternatively the corresponding frequency-domain description) and the properties of noise, to the extent as required by the individual methods.

12.3.1 Blur Estimation

In case of convolutional (space-invariant) blur, the identification concerns the PSF of the imaging system or other source of distortion. This may be either derived in the form of an expression using the physical description of the imaging or, more commonly in the field of image processing, experimentally measured directly or indirectly.

12.3.1.1 Analytical Derivation of PSF

Exceptionally, the PSF can be determined analytically, from a mathematical model based on the physical description of the mechanism of blur. A few examples may be given. One is the optical diffraction limiting the angular resolution due to the finite aperture of the lens that might be derived based on wave (Fourier) optics; another example is the blur of distant objects observed through turbulent atmosphere, where the refraction statistics of a turbulent medium with a randomly distributed refraction index is involved.

Motion is a common type of blur in medical imaging. We shall show now that the blur due to arbitrary translational parallel movement of the complete image $f(x', y')$ with respect to the sensor field (or film) during the exposition $t \in \langle 0, T \rangle$ is a convolutional distortion. If the origin of image coordinates (x', y') moves in the sensor coordinates (x, y) according to

$$x = X(t), \quad y = Y(t), \tag{12.4}$$

the obtained smeared image obviously is

$$g(x,y) = \int_0^T f(x - X(t), y - Y(t))\,\mathrm{d}t. \tag{12.5}$$

The spectrum of this image is

$$\begin{aligned} G(u,v) = \mathrm{FT}_{2D}\{g(x,y)\} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left(\int_0^T f(x - X(t), y - Y(t))\,\mathrm{d}t\right) e^{-j(ux+vy)}\mathrm{d}x\,\mathrm{d}y \\ &= \int_0^T e^{-j(uX(t)+vY(t))}\mathrm{d}t \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi,\eta) e^{-j(u\xi+v\eta)}\mathrm{d}\xi\,\mathrm{d}\eta \\ &= H(u,v)\mathrm{FT}_{2D}\{f(u,v)\}. \end{aligned} \tag{12.6}$$

We utilized here the interchange of integration order and the substitution $\xi = x - X(t)$, $\eta = y - Y(t)$; then the integrand could be factored into two expressions depending on only ξ, η and t, respectively. Thus, the spectrum of the smeared image is a product of the original image

spectrum with some frequency transfer function $H(u, v)$, which is fully determined by the motion during the exposition,

$$H(u,v) = \int_0^T e^{-j(uX(t)+vY(t))} \mathrm{d}t, \qquad (12.7)$$

so that the blur is really convolutional. Knowing its frequency transfer function, we can in principle easily design the corresponding inverse filter. Note that when the motion is unidirectional with a constant speed V, the PSF is a short abscissa of length VT with the same direction, and the corresponding frequency response has the sinc(…) shape (Figure 12.4), which is rather unpleasant in inverse filtering because of the presence of zeros (see below).

12.3.1.2 Experimental PSF Identification

In practice, often little or no information is available on the distorting mechanism, and then the PSF or transfer function of the distortion must be determined experimentally. The identification can be done *a priori* by imaging a suitable phantom, or it is possible to attempt a posterior identification based on the observed image. In both cases, the image would have to contain small bright points or thin lines on a dark background, or at least sufficiently distinct long linear edges.

The simplest is the first case of point sources: then the spots resulting as the convolution of the sources with the system PSF are directly approximating the PSF; the smaller are the source bright areas (points) in the phantom image, the better the approximation. The ideal situation arises when the size of the points is far smaller than the resolution of the system (as typical in astronomic images of stars, less commonly in medical images). Except for the astronomical or night-shot applications, the point sources can be expected only in phantom images. In a natural image, lines or edges can be found. The next paragraphs show how to derive the posterior estimate of PSF from such images (Figure 12.5).

The relation between the point-spread function $h(x, y)$ and the response of the system to an ideal infinitely long line — the *line-spread function* (LSF) $h_l(x, y)$ — follows from the convolution of the line with the PSF,

$$h_l(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-x', y-y')\, \delta(x'\cos\vartheta + y'\sin\vartheta - \tau)\, \mathrm{d}x'\mathrm{d}y'. \qquad (12.8)$$

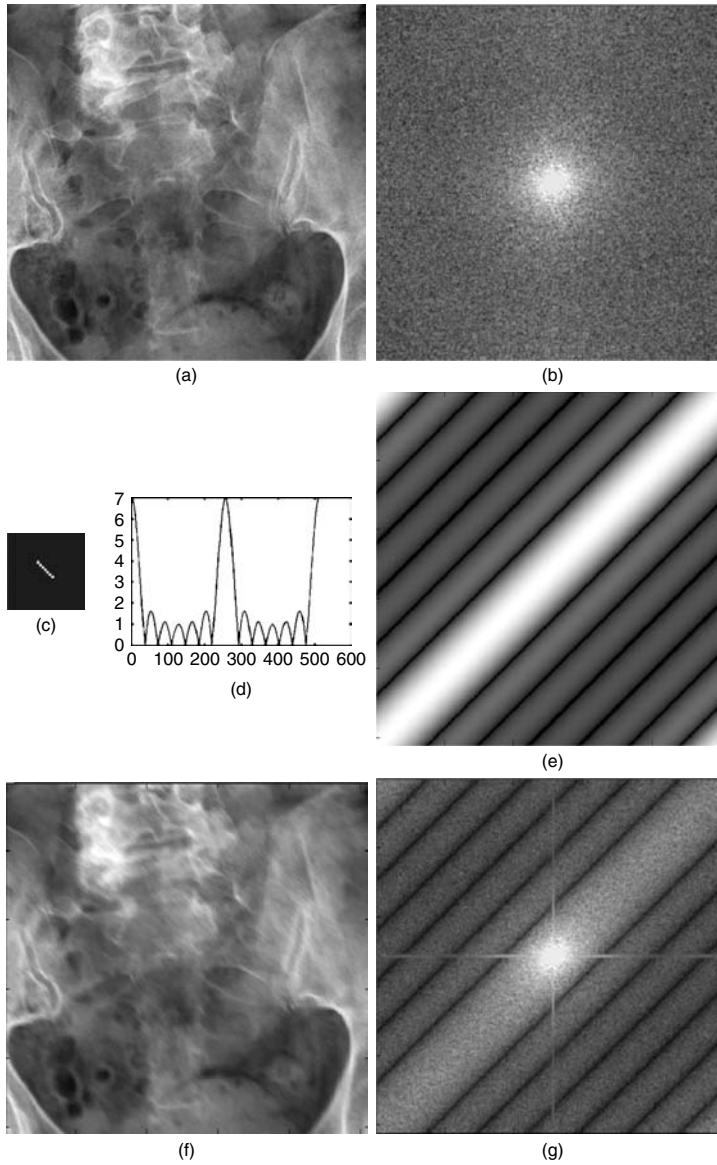


Figure 12.4 Blur due to global linear steady motion during exposition: (a) original image, (b) its amplitude spectrum, (c) blur PSF, (d) oblique profile of panel e, (e) frequency response of blur, (f) distorted image, and (g) spectrum of the distorted image (notice the zero stripes due to zeros in the frequency response of the blur).

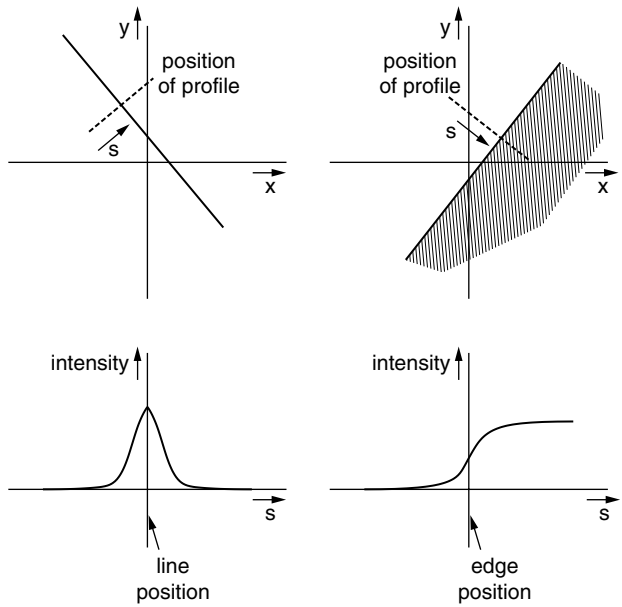


Figure 12.5 Posterior PSF identification. Above: (Left) A line object and (right) an edge object. Below: The schematic profiles of blurred responses as measured in the observed image.

Here, the infinitesimally thin line inclined by ϑ to the x-axis and with the distance τ from the origin is expressed by the impulse function. Without losing on generality with respect to the result interpretation, we may choose $\vartheta = \pi/2$ so that the expression simplifies to

$$\begin{aligned} h_l(x,y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x-x',y-y')\,\delta(x'-\tau)\,\mathrm{d}x'\mathrm{d}y' \\ &= \int_{-\infty}^{\infty} h(x-\tau,y-y')\mathrm{d}y' \\ &= \int_{-\infty}^{\infty} h(x-\tau,y'')\mathrm{d}y'' = h_l(x-\tau). \end{aligned} \tag{12.9}$$

The last integral is the projection of the PSF, centered on the line, along the y-axis as defined in Section 9.1.1. Thus, the line-spread

function is a function of only a single variable, perpendicular to the line; the function describes the profile of the response to the line. This conclusion is obviously independent of the orientation of the line with respect to coordinates that were chosen arbitrarily. The profile that can be measured on the observed image therefore determines a single projection of PSF that would be theoretically sufficient to reconstruct an isotropic PSF. If the PSF is anisotropic, more projections are needed that would have to be provided by measuring the response profiles of differently oriented lines. When the projections have been measured, any of the methods of reconstruction from projections (Chapter 9) may be used to calculate the PSF.

From a practical point of view, some comments should be made. Primarily, no line in a real image is infinitely long; often the lines are even shorter than the image extent would allow. However, the influence of distant parts of the line to the response is small, and when the profile is measured at distances from the line ends longer than the practical extent of PSF, the end effects may be neglected. Further, the measurement is naturally always affected by noise; averaging more profiles of the same line, measured at different positions, may diminish its influence. In the common case of isotropic PSF, a single line may suffice to provide the information needed for estimation; this may enable posterior restoration of images where such a line object has been imaged.

However, often no line objects can be found in the image. Still, there is a possibility to recover the blur PSF when there are some clear, sufficiently long borders between areas of sufficiently constant gray shades. Theoretically, the *step-spread function* (SSF) $h_s(x, y)$ is the response of the imaging system to the linear border between the areas of 0 and 1 levels, i.e., to the *step function*

$$s(x,y) = \begin{cases} 1 & \text{if} \quad x\cos\vartheta + y\sin\vartheta - \tau \geq 0 \\ 0 & \text{otherwise} \end{cases}. \tag{12.10}$$

When again simplifying the expressions by arbitrarily choosing the vertical edge at the position $x = \tau$, and utilizing the commutativity of convolution, we obtain

$$h_s(x,y) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x',y')\, s((x-\tau)-x')\,\mathrm{d}x'\mathrm{d}y' = h_s(x). \tag{12.11}$$

Differentiating the SSF with respect to the variable perpendicular to the edge yields the LSF,

$$\begin{aligned}\frac{\partial h_s(x,y)}{\partial x} &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x',y')\,\frac{\partial}{\partial x}s((x-\tau)-x')\,\mathrm{d}x'\mathrm{d}y' \\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} h(x',y')\,\delta((x-\tau)-x')\,\mathrm{d}x'\mathrm{d}y' \\ &= \int_{-\infty}^{\infty} h((x-\tau),y')\,\mathrm{d}y' = h_l(x-\tau),\end{aligned} \tag{12.12}$$

and the problem is thus converted to the previous one: having measured the SSF, the LSF is obtained by differentiating perpendicularly to the edge.

Though this approach is theoretically elegant, in practice a difficulty arises in the high sensitivity of the differentiation result to the image noise. Again, averaging more profiles of the same edge (though sufficiently distant from the ends of the edge) may supply more stable results. An edge provides only a single profile, sufficient for estimation of an isotropic PSF; several sufficiently straight and regular edges would be needed for estimating the PSF in the anisotropic case.

12.3.2 Identification of Noise Properties

Image noise, as a random two-dimensional signal generated by a stochastic field (see Section 2.4), has to be described by its statistical properties; its concrete values are naturally unknown in restoration. It is generally supposed that the noise has zero mean; thus, its mean power is given by the noise variance. Even with this simplification, the complete description of noise as a random signal generated by a stochastic field would require estimation of the multidimensional joint probabilities or distribution functions; an almost infeasible task. Fortunately, many restoration methods do not require knowledge that detailed, and only the relatively simple parameters and functions, as mentioned below, must be estimated.

The statistical properties may vary over the image area — the stochastic field may be generally inhomogeneous. However, taking into account the noise-generating mechanisms, it is mostly reasonable to expect that all the individual stochastic variables, describing noise of

concrete pixels, have the same probabilistic properties, i.e., that the field is homogeneous. In such a case, the ergodicity may usually be supposed as well, which allows estimation of the statistical characteristics and parameters even by spatial averages over the image area, and the ensemble averages over a group of similarly generated images are not necessarily required. Both approaches are often combined, if more than a single image is available — the averages are taken spatially based on individual images and the reliability of the estimates checked by the variance, derived by comparing the individual results. Finally, averaging over the ensemble may improve the estimates.

Practically, the experimental identification should be arranged so that only the noise is (possibly repeatedly) measured. It is often possible, e.g., when identifying the noise added to the image signal in a communication channel or in a processing system: the input image to such a system should be zero; then, the output becomes a realization of pure noise. Sometimes, namely when the noise is dependent on image content, only the observed noisy images are available, without a possibility to get rid of the image content. The noise estimate must then be based on some *a priori* assumptions — the most natural being the zero mean of the noise. This enables subtraction of the average value in image areas, the useful content of which might be considered flat (of a constant intensity), and consequently, consideration of the remaining data as noise.

The most important noise parameter is its power proportional to (usually even defined as equal to) its variance σ_v^2; the average amplitude of noise is then defined as the standard deviation σ_v. Some restoration methods require knowledge of the two-dimensional autocorrelation function or two-dimensional power spectrum of the noise. Providing the noise is homogeneous and ergodic, all these quantities or functions (in the discrete matrix form) may be estimated as explained in Sections 2.4.3 and 2.4.4. Using Equation 2.109, the estimates of variance and the autocorrelation function may be determined, while Equation 2.110 provides an estimate of the autocorrelation function weighted by a triangular window. The sequence expressed by Equations 2.111 to 2.113 offers an estimate of the power spectrum via the periodogram approach, while the estimate given by Equation 2.114, based on the Wiener–Khintchin theorem, is derived from the weighted autocorrelation function.

Sometimes, the homogeneity of noise cannot be assumed, which considerably complicates the identification, as well as the following restoration. In these cases, the identification of spatially dependent characteristics (e.g., four-dimensional autocorrelation matrices)

requires estimates based on ensemble averages that are much more demanding to provide — many realizations of noise must be acquired and intricately processed.

12.3.3 Actual Inverse Filtering

12.3.3.1 Plain Inverse Filtering

In this and the next section, we shall base our considerations on the distortion model consisting of a linear distortion (generally space variant) and additive zero-mean random noise $v(\ldots)$, represented in continuous space as

$$g(\mathbf{r}) = \iint_{\mathbf{r}' \in \mathrm{A}} h(\mathbf{r}, \mathbf{r}')\, f(\mathbf{r}')\, \mathrm{d}S_{\mathbf{r}'} + v(\mathbf{r}), \quad \mathbf{r} = [x, y]^T, \tag{12.13}$$

where A is the area of the image and $\mathrm{d}S_{\mathbf{r}'}$ is the area element at position $\mathbf{r}'$. Often, the linear distortion is space invariant (i.e., convolutional) and the expression may be simplified as

$$g(\mathbf{r}) = \iint_{\mathbf{r}' \in \mathrm{A}} h(\mathbf{r} - \mathbf{r}')\, f(\mathbf{r}')\, \mathrm{d}S_{\mathbf{r}'} + v(\mathbf{r}), \tag{12.14}$$

where the function $h(\mathbf{r}) = h(x, y)$ is the PSF of the linear distortion. Though simple, this model (Figure 12.6) covers many practical restoration problems with a reasonable degree of accuracy.

The distortion can be described equivalently in the frequency domain by the equation obtained by Fourier transforming Equation 12.14,

$$G(u,v) = H(u,v)\, F(u,v) + N(u,v), \tag{12.15}$$

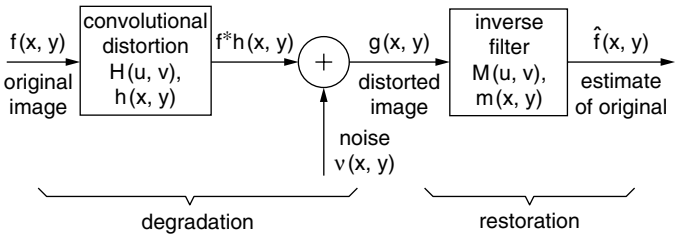


Figure 12.6 Distortion consisting of convolutional blur and additive noise followed by restoration.

where $G(u, v)$, $F(u, v)$, and $N(u, v)$ are spectra of the observed image, original image, and noise, respectively, and $H(u, v)$ is the frequency transfer function of the linear distorting system, $H(u, v) = \mathrm{FT}\{h(x, y)\}$. Thanks to the convolution property of the Fourier transform, the frequency-domain equation can be easily inverted as

$$\hat{F}(u,v) = \frac{1}{H(u,v)}(G(u,v) + N(u,v)) = F(u,v) + \frac{1}{H(u,v)}N(u,v). \quad (12.16)$$

This is the frequency-domain representation of action of a convolutional restoration filter, the frequency response of which is

$$M(u,v) = \frac{1}{H(u,v)}; \quad (12.17)$$

it is called the *inverse filter*. Equation 12.16 thus represents the simplest approach to this restoration problem—*plain inverse filtering*. When the noise may be neglected, the second term in the last expression vanishes and the output of the filter will obviously be exactly the required original. Figure 12.7 shows an example of such filtering.

However, the noise can be neglected only occasionally, and only on the condition that $H(u, v)$ has no zeros in the complete range of frequencies; still, this is a necessary but not sufficient condition for a reasonably successful restoration. Clearly, there is no way to separate the image and noise components, and to submit only the useful part of the signal to the filtering. As it is seen, the output spectrum differs from the proper result by the term $N(u, v)/H(u, v)$, which may become very large at frequencies where $H(u, v)$ is small or approaches zero. Physical interpretation of this phenomenon is obvious: the inverse filter tries to recover the frequency components of the original attenuated by the distorting system; the stronger the attenuation, the higher the restoration amplification, which may even grow over any limits for frequencies (u, v) for which $H(u, v) = 0$. As the useful image signal components are very weak in such frequency regions, the SNR is low (or even approaching zero), and only the noise is practically amplified. It may then easily happen that the noise prevails; obviously, such restoration is useless (Figure 12.8).

12.3.3.2 Modified Inverse Filtering

Based on this analysis, *intuitive modifications* of the inverse filter, hopefully leading to an improvement, may be attempted. Obviously, the

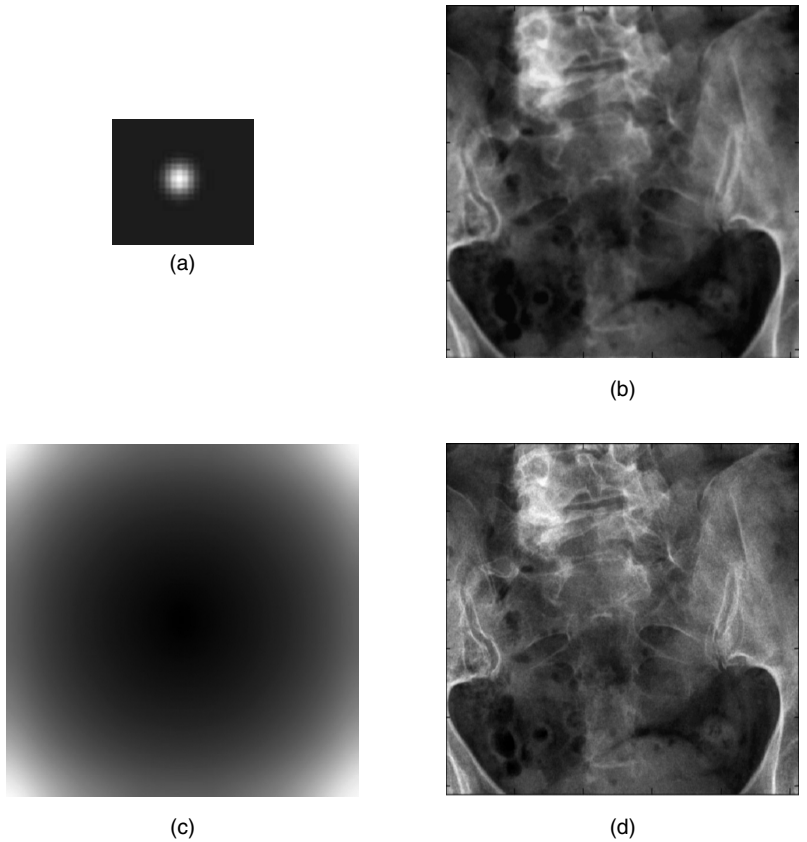


Figure 12.7 Relatively successful plain inverse restoration of the image blurred by heavily, but nowhere completely, suppressed higher-frequency components: (a) PSF of the blur, (b) blurred image, (c) frequency response of the inverse filter, and (d) the restored image.

result of such subjective modifications will only be a compromise: it will be necessary to limit in a way the extreme transfer values of the restoration filter amplifying mainly noise and, at the same time, to try minimizing the losses due to insufficiently amplified useful components.

The simplest classical modification is to lower the filter transfer at higher frequencies, where the SNR is usually deteriorating. Another but similarly acting approach is to limit the frequency range of the filter to only lower frequencies so that the extreme transfer values are excluded. In both cases, the net effect is a substantial loss of details. Further, when only zero-phase filters are used (as in the

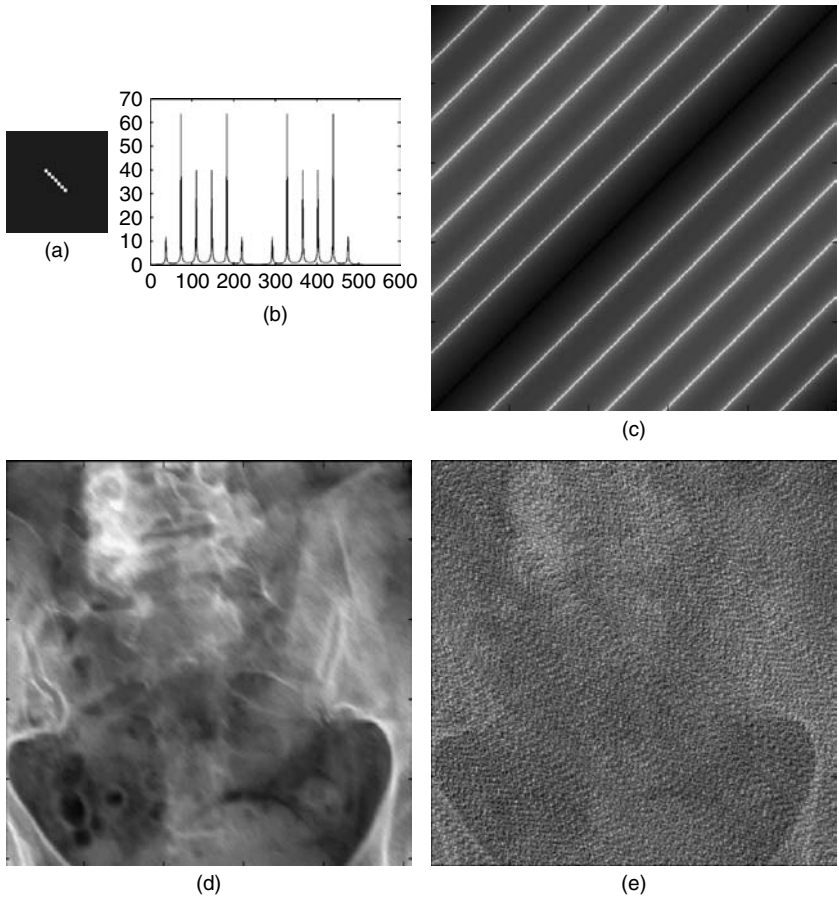


Figure 12.8 Unsuccessful plain inverse restoration of the blurred image from Figure 12.4: (a) PSF of the blur, (b) oblique profile of panel c, (c) amplitude frequency response of the plain inverse filter, (d) blurred image, and (e) "restored" image.

case of simply emphasizing frequency bands known to be suppressed by the distortion), the phase distortion of the observed image is not respected, which may also lead to unsatisfactory restoration.

Of the intuitive inverse filter modifications, simple clipping of the amplitude frequency response as in [28] seems most effective and reasonably justifiable. It utilizes the known fact that a substantial part of the image information, namely the space relations of edges, is carried by the phase spectrum. Therefore, clipping of the

peaks of the amplitude response only lowers the frequency components with a low SNR, without completely throwing away the respective information, as is the case with the above-mentioned filters assigning stop-bands in the sensitive spectral areas. The application of the inverse filter modified this way to the case in Figure 12.8 is presented in Figure 12.9.

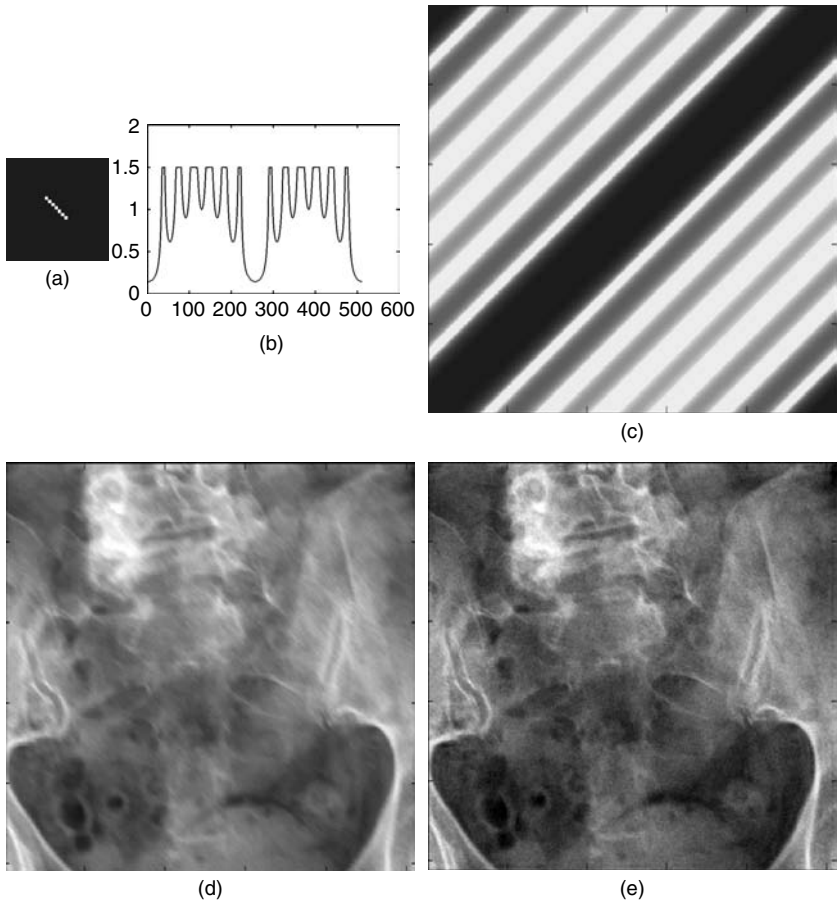


Figure 12.9 Approximate restoration of the blured image from Figure 12.4 by modified inverse filter with limited amplitude response (compare with Figure 12.8): (a) PSF of the blur, (b) oblique profile of panel c, (c) amplitude frequency response modified from the plain inverse filter, (d) blurred image, and (e) roughly restored image.

12.4 RESTORATION METHODS BASED ON OPTIMIZATION

12.4.1 Image Restoration as Constrained Optimization

A majority of advanced methods of signal and image restoration are formulated as *optimization* problems. The philosophy of this approach is the following: an optimization criterion is found that can be considered a reasonable measure of the restoration quality*; i.e., it expresses in an intuitively or even formally justifiable way the similarity of the restored image with the original. The restoration then consists in principle of finding the optimum (maximum or minimum) of the criterion function as dependent on the restoration parameters (e.g., parameters of the restoration filter or algorithm, or directly the values of the restored image). The restoration can be visualized as looking for the optimum point in a multidimensional space of the restoration parameters (or directly of the restored image intensities).

Some criteria (e.g., the LMS criterion) define such a situation uniquely; a scalar field is defined by the chosen criterion in the parameter space and the restoration problem is transformed to finding the position of the zero gradient in the field. In other cases, the criterial function offers more or even infinitely many solutions, or the absolute optimum is not acceptable for reasons that may not be reflected in the criterion (e.g., the noise influence). It is then necessary to add other conditions (constraints, bindings) to be fulfilled in restoration that choose the proper parameters out of the possible solutions satisfying the optimal criterion, and consequently the restoration result. This requires a more complex procedure of *constrained optimization*, which may be visualized as looking not for the absolute extreme of the criterion, but for an extreme on a certain hypercurve or hypersurface in the parameter space.

Some of the criteria, namely of the first group, allow the derivation of an equation or a set of equations that can be solved analytically, thus yielding the solution (e.g., the frequency response of the restoration filter) in a closed form. The restoration is then straightforward and reaches the desired result — the restored image — in a single step (though possibly quite complex). With more complex or sophisticated criteria, particularly when constraints are imposed, the corresponding equation system is nonlinear or even cannot be

*Do not confuse with the image quality; this is not in question in restoration.

explicitly expressed; in this case, it becomes necessary to reach the result by gradual approximations. The visualization in the parameter space then corresponds to starting at a position given by the initial guess (more or less arbitrary) and approaching the solution in iterative steps, each of which leads to a (hopefully, but not necessarily) better approximation of the desired result. The path in the space, possibly quite complex, thus starts in the initial guess and should terminate at the point of optimal solution.

An important issue in the restoration field is the computational complexity and memory requirements. Both may easily become prohibitive for practical purposes with the available contemporary hardware; however, many methods considered infeasible several years ago are now used routinely thanks to fast development in the available computing power. In this book, we shall occasionally mention the problems of the complexity and indicate possible simplifications, but with respect to the aim of introducing the restoration concepts, these will be only marginal comments.

Let us try to roughly classify the restoration methods:

- According to linearity of image processing:
  - *Linear* (fulfilling the superposition principle, thus enabling arbitrary cascading of procedures, e.g., Wiener filtering and all LMS filters)
  - *Nonlinear* (most other methods; the order of partial procedures matters)
- According to applicability of the method:
  - *Restoration of a class* of images (a common approach to all images of a class with the same expression or algorithm and parameters, the class often being defined as a stochastic field with certain probability characteristics, as in LMS methods)
  - *Image-specific methods* (individually adjusted to a particular image, not applicable to other images without at least a change in restoration parameters, e.g., constrained deconvolution)
- According to acceptance of nonlinear distortion:
  - Not allowed (inverse filtering)
  - Not considered (e.g., Wiener filtering)
  - Allowed, to be identified (e.g., maximum aposteriori probability (MAP) method)
- According to identification requirements:
  - Identification of blur:
    - Only isoplanar allowed (most methods)

  - Space-variant blur accepted (e.g., in generalized LMS deconvolution)
- Identification of probabilistic characteristics:
  - Only noise variance required (e.g., in constrained deconvolution)
  - Power spectra of noise and of distorted image (in narrow-sense Wiener filtering), cross-spectrum between original and observed images (generic Wiener filter); alternatively, the corresponding correlation functions or matrices
  - Joint probability distributions and conditional distributions — Bayesian approaches (e.g., MAP method)

The grouping of the methods in the following sections, though it is partly unconventional, reflects this classification. In compliance with the character of the book, the emphasis is put on the approaches and concepts of the methods.

12.4.2 Least Mean Square Error Restoration

12.4.2.1 Formalized Concept of LMS Image Estimation

A formalized approach to image restoration in the presence of noise is based on the idea of the processed images being realizations of stochastic fields: the field $\tilde{f}(\mathbf{r})$* generating the original images, another field $\tilde{v}(\mathbf{r})$ producing noise, and the derived field $\tilde{g}(\mathbf{r})$ encompassing all possible realizations of the observed images. Formulated this way, another view of optimality may be adopted. Following Wiener's idea (about 1944), a method is sought to derive the best possible estimate $\hat{f}(\mathbf{r})$ of the original, not in a particular case of a concrete given observed image $g(\mathbf{r})$, but on average, when the restoration is performed many times with different realizations $g(\mathbf{r})$ derived from different originals $f(\mathbf{r})$ and noise realizations $v(\mathbf{r})$. Though it might not be optimum in a particular case, the approach formulated this way enables the design of a *universal* method that is common to the complete class of images defined by the stochastic fields.

---

*In this section, the position vector $\mathbf{r}$ in image A may be interpreted either in continuous space or on the discrete grid of sampled space.

The criterion of optimality is defined based on the notion of *error image*,

$$e(\mathbf{r}) = f(\mathbf{r}) - \hat{f}(\mathbf{r}), \tag{12.18}$$

which is naturally random as well and may be considered a realization of another stochastic field $\tilde{e}(\mathbf{r})$. The criterion to be minimized will be defined, with respect to the requirement of the average optimum result, as the mean value over all possible realizations,

$$\varepsilon^2(\mathbf{r}) = \mathbf{E}_\omega\{e^2(\mathbf{r})\} = \mathbf{E}_\omega\{f(\mathbf{r}) - \hat{f}(\mathbf{r})\} \to \min, \quad \forall \mathbf{r} \in \mathrm{A}. \tag{12.19}$$

In fact, this is not a single criterion, but a set of criteria, each valid for a particular position $\mathbf{r}$ in the image (thus an infinite set in the continuous-space formulation). Using the squared error values serves to prevent cancelling of errors with opposite signs in the mean; the square has better analytic properties than the absolute value alternative. This optimizing *least mean square criterion* depends on errors of all possible restorations; however, the influence of a particular case depends, according to the definition of mean, on the probability of that case. Therefore, the restoration of images that appear rarely, so that their probability is low, might be unsatisfactory in an extreme theoretical case, without influencing the mean substantially. As the probabilities of concrete processed images are usually not known, it is impossible to predict the cases of good or worse restoration. This is the price for the universality of the derived restoration method or algorithm; on the other hand, most of the processed images are well restored, and in practice, the results are usually satisfactory, should the images fulfill some basic requirements, as will be discussed below.

As every possible original image $f(\mathbf{r})$ has its *a priori* probability* of appearance, the mean value $\bar{f}(\mathbf{r})$ of the original may thus be defined. It can be proved that this is the best estimate of the original (in sense of the criterion in Equation 12.19) in the complete lack of any other information, i.e., before the observed image is available. Once the distorted image $g(\mathbf{r})$ is received, its content, similar to some of the possible originals, while dissimilar to others, gives a clue as to which images might be considered the concrete original; this

*The probability concerns fully discrete images; in the case of continuous-space continuous-intensity images, this would have to be modified to the respective probability density.

changes the probabilities of the individual originals to become the proper estimate of $f(\mathbf{r})$. The new probabilities, forming the *posterior distribution* of probability, define a new mean image, conditioned by the concrete observed image $g(\mathbf{r})$. The corresponding *conditional mean* $\bar{f}(\mathbf{r})\big|_{g(\mathbf{r})}$ is the best estimate of the original (in the LMS sense) that can be derived from the observed image $g(\mathbf{r})$, as it may be proved in a nontrivial way that will be omitted here. We have therefore arrived at the theoretical *golden standard* of the restored image; nevertheless, it is usually infeasible to calculate the restored image this way for two reasons. Primarily, the probability of every image is in fact determined by a complicated system of joint probability distributions; such a probabilistic model is very difficult to identify in practice. Second, even if those distributions were known, the estimated (restored) image would be a complicated nonlinear function of the observed image, requiring extremely demanding computation. For these reasons, some constraints are usually introduced that limit the class of images to which the derived restoration method applies, together with introducing simplified *suboptimum methods* (e.g., linear ones) instead of using the above-mentioned nonlinear optimal approach.

Notice that *no* limitations were introduced so far; the generic LMS approach thus applies to any type of images distorted by any (even nonlinear and random) mechanism. In the following three sections, we shall derive the formulae that enable the obtaining of the restored image based on the observed distorted image. The constraints that will have to be imposed, in consequence limiting the class of processed images, or limitations concerning the class of the used restoration systems will be gradually introduced, with detailed discussion, in order to understand the real applicability of different formulae or algorithms.

12.4.2.2 Classical Formulation of Wiener Filtering for Continuous-Space Images

So far, no limitation was imposed on the type of system used for restoration. The first important restriction concerns the type of restoring systems: only linear filters, generally space variant, will be used, as

$$\hat{f}(\mathbf{r}) = \iint_{\mathbf{r}' \in \mathrm{A}} m(\mathbf{r}, \mathbf{r}')\, g(\mathbf{r}')\, \mathrm{d}S_{\mathbf{r}'}, \quad \mathbf{r} \in \mathrm{A}', \tag{12.20}$$

where the area A′ has been augmented from A due to convolution. Further on, the infinite two-dimensional area $\mathrm{B} \equiv \{\mathbf{r} = (x, y) : x, y \in \langle -\infty, \infty \rangle\}$ will be considered in order to include A′ automatically. Introducing this limitation — the restricted form of the restoring system in Equation 12.20 — obviously means that, except for quite specific cases, the filtering will be suboptimum, as the optimal system may be generally nonlinear.

Substituting Equation 12.20 for the estimated restored image $\hat{f}(\mathbf{r})$ (or briefly, "restoration") into Equation 12.19 gives

$$\varepsilon^2(\mathbf{r}) = \mathbf{E}\left\{\left(f(\mathbf{r}) - \iint_{\mathbf{r}' \in \mathrm{B}} m(\mathbf{r}, \mathbf{r}') g(\mathbf{r}') \mathrm{d}S_{\mathbf{r}'}\right)^2\right\} \to \min, \quad \forall \mathbf{r} \in \mathrm{B}. \quad (12.21)$$

The integral in Equation 12.21 is the limit case of a sum, so that $\hat{f}(\mathbf{r})$ may be considered a linear combination of values of a realization of the stochastic field $\tilde{g}(\mathbf{r})$, i.e., of discrete random variables $g(\mathbf{r}')$, $\forall \mathbf{r}' \in \mathrm{A} \in \mathrm{B}$. Then we have the situation described in Section 1.4.6 — every unknown stochastic variable $\tilde{f}(\mathbf{r})$ for arbitrary $\mathbf{r} \in \mathrm{B}$ is estimated linearly based on a set (infinite set in continuous space, finite set in the discrete case) of known stochastic variables $g(\mathbf{r}')$, $\forall \mathbf{r}' \in \mathrm{B}$. According to the principle of orthogonality, the error of the LMS estimate must be uncorrelated with (statistically orthogonal to) all values $g(\mathbf{s})$, $\forall \mathbf{s} \in \mathrm{B}$; thus,

$$\mathbf{E}\{e(\mathbf{r}) g(\mathbf{s})\} = \mathbf{E}\left\{\left(f(\mathbf{r}) - \iint_{\mathbf{r}' \in \mathrm{B}} m(\mathbf{r}, \mathbf{r}') g(\mathbf{r}') \, \mathrm{d}S_{\mathbf{r}'}\right) g(\mathbf{s})\right\} \equiv 0 \quad (12.22)$$

$\forall \mathbf{r}, \mathbf{s} \in \mathrm{B}.$

This equation — an equivalent alternative to Equation 12.21 — may be visualized as a system of linear equations for the (generally infinite) set of coefficients $m(\mathbf{r}, \mathbf{r}')$, one equation for each $\mathbf{r}$ (obviously an infinite system as far as the space is continuous). However, in the continuous-space representation, we will not try to solve this linear system; rather, the valid Equation 12.22 will be transformed into the form leading to the desired four-dimensional impulse response $m(\mathbf{r}, \mathbf{r}')$ of the restoration filter. Because both the mean and integral operators in the previous equation are linear, they may be interchanged, which yields

$$\iint_{\mathbf{r}' \in \mathrm{B}} m(\mathbf{r}, \mathbf{r}') \, \mathbf{E}\{g(\mathbf{r}') g(\mathbf{s})\} \, \mathrm{d}S_{\mathbf{r}'} = \mathbf{E}\{f(\mathbf{r}) g(\mathbf{s})\}, \quad \forall \mathbf{r}, \mathbf{s} \in \mathrm{B}. \quad (12.23)$$

Here, we can recognize that the mean values of products of stochastic variables are the respective correlation functions — the term in the integral is the autocorrelation function $R_{gg}(\mathbf{r}', \mathbf{s})$ of the field $g(\mathbf{r})$, while the term on the right side is the cross-correlation function $R_{fg}(\mathbf{r}, \mathbf{s})$. The equation can therefore be rewritten as

$$\iint_{\mathbf{r}' \in \mathrm{B}} m(\mathbf{r}, \mathbf{r}')\, R_{gg}(\mathbf{r}', \mathbf{s})\, \mathrm{d}S_{\mathbf{r}'} = R_{fg}(\mathbf{r}, \mathbf{s}), \quad \forall \mathbf{r}, \mathbf{s} \in \mathrm{B}. \tag{12.24}$$

In order to enable useful simplification of this equation, we shall introduce an important limitation on the restoration system, namely, that the system should be space invariant (convolutional),

$$\widehat{f}(\mathbf{r}) = \iint_{\mathbf{r}' \in \mathrm{A}} m(\mathbf{r} - \mathbf{r}')\, g(\mathbf{r}')\, \mathrm{d}S_{\mathbf{r}'}, \quad \mathbf{r} \in \mathrm{B}. \tag{12.25}$$

This means that a plain linear filter has been chosen as the restoration system; deriving its PSF $m(\Delta\mathbf{r})$ or its frequency response $M(u, v)$ based on properties of the processed images and noise is the aim of the following paragraphs (more precisely, the design is based on properties of the stochastic fields generating the images and the noise). The spatial invariance of the filter implies that the statistical properties of the fields are also space independent, i.e., that the *involved stochastic fields are homogeneous*. It is the first and fundamental limitation imposed on the class of processed images.

On the other hand, the severely limiting assumption of homogeneity, which allows the convolutional restoration filtering, enables the rewriting of Equation 12.24 as

$$\iint_{\mathbf{r}' \in \mathrm{B}} m(\mathbf{r} - \mathbf{r}')\, R_{gg}(\mathbf{r}' - \mathbf{s})\, \mathrm{d}S_{\mathbf{r}'} = R_{fg}(\mathbf{r} - \mathbf{s}), \quad \forall \mathbf{r}, \mathbf{s} \in \mathrm{B}. \tag{12.26}$$

Introducing the substitutions $\mathbf{r}_s' = \mathbf{r}' - \mathbf{s}$, $\mathbf{r}_s = \mathbf{r} - \mathbf{s}$ yields

$$\iint_{\mathbf{r}_s' \in \mathrm{B}} m(\mathbf{r}_s - \mathbf{r}_s')\, R_{gg}(\mathbf{r}_s')\, \mathrm{d}S_{\mathbf{r}_s'} = R_{fg}(\mathbf{r}_s), \quad \forall \mathbf{r}_s, \mathbf{r}_s' \in \mathrm{B}, \tag{12.27}$$

where the left-hand side is a convolution of two deterministic two-dimensional functions; the right-hand side is also a deterministic function. This deterministic integral equation can easily be solved

in the Fourier domain. Utilizing the convolutional property of FT and the Wiener–Khintchin theorem, we obtain

$$M(u,v)\,S_{gg}(u,v) = S_{fg}(u,v), \tag{12.28}$$

where $M(u,v) = \mathrm{FT}\{m(\mathbf{r})\}, S_{fg}(u,v) = \mathrm{FT}\{R_{fg}(\mathbf{r})\}, S_{gg}(u,v) = \mathrm{FT}\{R_{gg}(\mathbf{r})\}$, so that finally

$$M(u,v) = \frac{S_{fg}(u,v)}{S_{gg}(u,v)}\,. \tag{12.29}$$

This is a fundamental result providing the frequency response of the optimal linear restoration filter in the LMS sense — the *generic Wiener filter*. It is expressed by means of two two-dimensional spectra, both deterministic functions: the mutual spectrum $S_{fg}(u, v)$ between the original field $\tilde{f}(\mathbf{r})$ and the distorted image field $\tilde{g}(\mathbf{r})$ and the power spectrum $S_{gg}(u, v)$ of $\tilde{g}(\mathbf{r})$. Thus, the result is deterministic and unique, although the images to be processed may be different realizations of the underlying stochastic fields. In other words, the optimal filter is determined by the statistics of the classes of processed images.

It is important to realize that the filter applies to all images of a given class. Although the character of distortion influences the spectra and thus even the filter, no constraints are so far imposed on the model of distortion. Therefore, the filter (Equation 12.29) is applicable even in cases of nonlinear distortion, multiplicative noise, etc., and it is the best *linear convolutional* filter in the LMS sense. It is well possible that the linear restoration may not be very successful in such cases and a nonlinear procedure might provide much better results; however, the frequently met statement that the Wiener filter does not apply for nonlinearly distorted images is incorrect (for this generic filter form).

From the practical viewpoint, nevertheless, the filter design based on Equation 12.29 may not be available. While the distorted images are always available and the power spectrum $S_{gg}(u, v)$ can thus be estimated, the estimate of the cross-spectrum $S_{fg}(u, v)$ (or equivalently of the cross-correlation function $R_{fg}(\mathbf{r})$) becomes infeasible when there is no access to the original images $f(\mathbf{r})$, as is frequent in practice.

In order to obtain a more practical formula, we shall introduce the second limitation by imposing a *restriction to the model of*

distortion; the common model (Equation 12.14), consisting of linear convolutional blurring and additive noise, will be supposed. Moreover, it will be supposed that the stochastic fields $\tilde{f}(\mathbf{r})$ and $\tilde{v}(\mathbf{r})$ are independent (or at least uncorrelated) and that the noise has zero mean, so that

$$\mathbf{E}\{f(\mathbf{r})\, v(\mathbf{s})\} = \mathbf{E}\{f(\mathbf{r})\}\, \mathbf{E}\{v(\mathbf{s})\} = 0, \quad \forall \mathbf{r}, \mathbf{s} \in \mathrm{B}. \tag{12.30}$$

When all the terms in Equation 12.14 are multiplied by $f(\mathbf{r})$, we obtain

$$f(\mathbf{r})g(\mathbf{s}) = f(\mathbf{r}) \iint_{\mathbf{r}' \in \mathrm{B}} h(\mathbf{s} - \mathbf{r}')\, f(\mathbf{r}')\, \mathrm{d}S_{\mathbf{r}'} + f(\mathbf{r})\, v(\mathbf{s}), \tag{12.31}$$

which applies to any pairs of $f(\mathbf{r})$ and corresponding $g(\mathbf{r})$. As these are realizations of respective joint stochastic fields, the mean value operator may be applied to both sides, yielding

$$\mathbf{E}\{f(\mathbf{r})g(\mathbf{s})\} = \iint_{\mathbf{r}' \in \mathrm{B}} h(\mathbf{s} - \mathbf{r}')\, \mathbf{E}\{f(\mathbf{r})f(\mathbf{r}')\}\, \mathrm{d}S_{\mathbf{r}'} + \mathbf{E}\{f(\mathbf{r})\, v(\mathbf{s})\} \tag{12.32}$$

when realizing that the function $h(\mathbf{r})$ is deterministic. Here, the means of products of stochastic variables are correlation functions that only depend on difference vectors, as the fields were previously restricted to homogeneous. The last term may be omitted because the image and noise are supposed uncorrelated. Therefore,

$$R_{fg}(\mathbf{r} - \mathbf{s}) = \iint_{\mathbf{r}' \in \mathrm{B}} h(\mathbf{s} - \mathbf{r}')\, R_{ff}(\mathbf{r} - \mathbf{r}')\, \mathrm{d}S_{\mathbf{r}'}. \tag{12.33}$$

The result is the correlation integral of the form in Equation 1.42; utilizing the property (Equation 1.41) of the Fourier transform, we have

$$S_{fg}(u,v) = H^{*}(u,v)\, S_{ff}(u,v). \tag{12.34}$$

This expression for $S_{fg}(u, v)$ will be later substituted for the numerator in Equation 12.29.

The power spectrum $S_{gg}(u,v)$ of the observed image may be expressed in terms of the original and noise power spectra, using the independence of noise on the image, and the result (Equation 1.98) on the spectrum of a field processed by a linear system, as

$$S_{gg}(u,v)=|H(u,v)|^2 S_{ff}(u,v)+S_{vv}(u,v). \tag{12.35}$$

When substituting $S_{ff}(u,v)$ from this into Equation 12.34, it becomes

$$S_{fg}(u,v)=\frac{H^*(u,v)}{|H(u,v)|^2}S_{ff}(u,v) \tag{12.36}$$

and substituting into Equation 12.29 finally yields

$$M(u,v)=\frac{1}{H(u,v)}\frac{S_{gg}(u,v)-S_{vv}(u,v)}{S_{gg}(u,v)}. \tag{12.37}$$

This is a practical formula for designing a (less generic*) Wiener filter, as all the involved functions can be identified. The frequency response of the blurring system $H(u,v)$ can be derived from the impulse response that has to be determined *a priori* or as *a posteriori* estimate (Section 12.3.1). Further, both the observed images and noise realizations are available so that the needed power spectra may be estimated (Sections 12.3.2 and 2.4.4).

Equation 12.37 has an interesting interpretation: it consists of two factors, the frequency transfer of the plain inverse filter (Equation 12.17) and the *Wiener correction factor* (WCF). It is immediately seen that the WCF is a real function, with its values in the range <0, 1>, as the power spectra are real non-negative functions and, for obvious physical reasons,

$$S_{gg}(u,v)\geq S_{vv}(u,v), \quad \forall u,v. \tag{12.38}$$

The WFC therefore does not change the phase frequency response of the inverse filter, and it only decreases the amplitude transfer at certain frequencies. Notice that the concept of inverse filter was not involved in the formalized LMS-based derivation of the Wiener filter; it appeared rather unexpectedly. It is also interesting to compare this formally derived result with the "brutal" intuitive

*In comparison with the generic form (Equation 12.29).

modification of the inverse filter as introduced in Section 12.3.3 (Figure 12.9); it also preserved the phase and decreased the amplification in some frequency areas, though naturally not optimally.

An interesting physically plausible interpretation of WCF is possible when an alternative formula for the Wiener filter is derived, substituting Equations 12.34 and 12.35 into Equation 12.29, which gives

$$M(u,v) = \frac{1}{H(u,v)} \frac{|H(u,v)|^2}{|H(u,v)|^2 + \frac{S_{vv}(u,v)}{S_{ff}(u,v)}}. \tag{12.39}$$

The WCF is expressed here in terms of the blur transfer function and power spectra of noise and of the *original* image field. It is again seen that the WCF is a real function with values in the range <0, 1>. The fraction $S_{vv}(u,v)/S_{ff}(u,v)$ may be interpreted as the frequency-dependent *a priori* noise-to-signal ratio (NSR), which comes close to zero when the noise is negligible in comparison with the useful image signal in certain frequency areas, while it approaches infinity when the original frequency components are vanishing at certain frequencies. Consequently, the WCF may be close to 1 for frequencies where the noise is negligible, and will be substantially diminished at frequencies where the useful signal is *a priori* weak. Moreover, the SNR in the *observed* image is further decreased if the transfer $H(u,v)$ of the blurring system is low for some frequencies. Then the term $|H(u,v)|^2$ may become small or negligible compared to the *a priori* NSR, and the whole WCF is thus smaller than 1, or even approaching zero, for frequencies u, v, where $|H(u,v)|^2 \to 0$.

As a continuation to already presented attempts of restoration of the motion blurred image, application of the Wiener filter is presented in Figure 12.10. By direct comparison, it can be seen that the sharpness of the restored image is slightly better than that in Figure 12.9, without noise level being increased. Another example of restoration via Wiener filtering based on the known blur PSF and on the estimated constant ratio of power spectra of the observed image and noise is in Figure 12.11. Here, the ultrasonographic radio frequency (RF) data are restored; hence, also the PSF was identified in the RF domain. Both the original and restored RF image data have been consequently envelope detected and format transformed, this way obtaining the video image data. Comparison of the video domain images shows a visible improvement of details in the restored image.

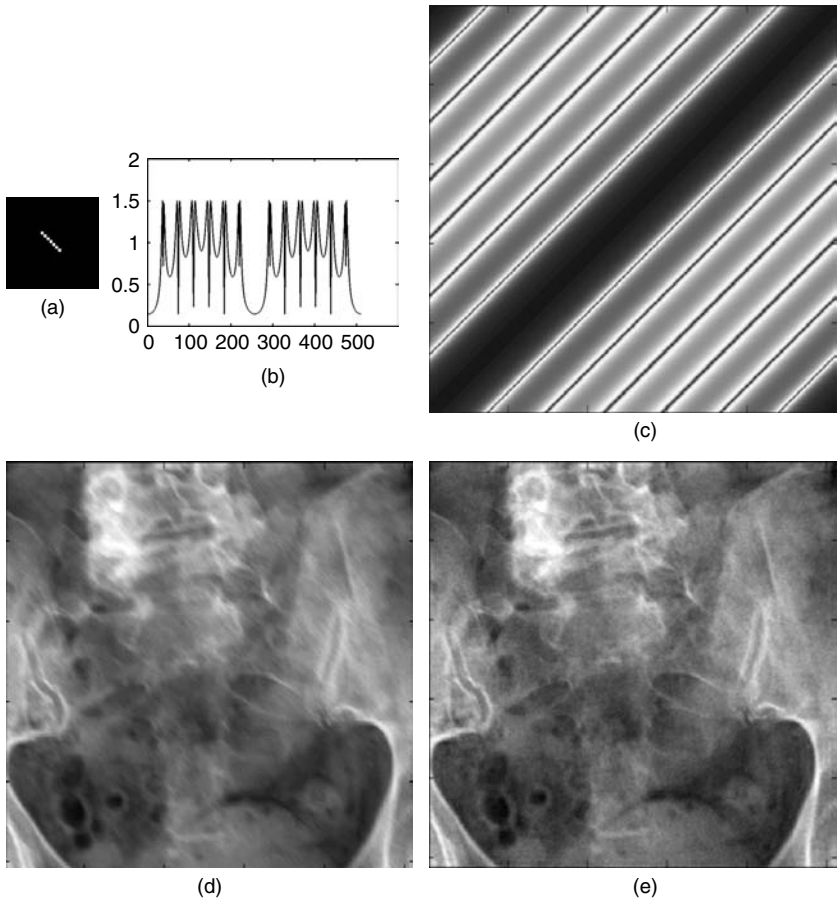


Figure 12.10 Motion blur restored by simplified Wiener filter: (a) PSF of the blur, (b) oblique profile of panel c, (c) amplitude frequency response of the Wiener filter, (d) blurred image, and (e) restored image.

Summarizing, the Wiener filter is basically the modified inverse filter, which behaves as the plain inverse filter when the noise is negligible. Otherwise, its amplitude transfer is reduced when the SNR in the observed image is low, either due to already *a priori* low signal levels in the original image or because of a loss in some image frequency components due to low values or zeros in the frequency response of blur. The amplitude frequency response

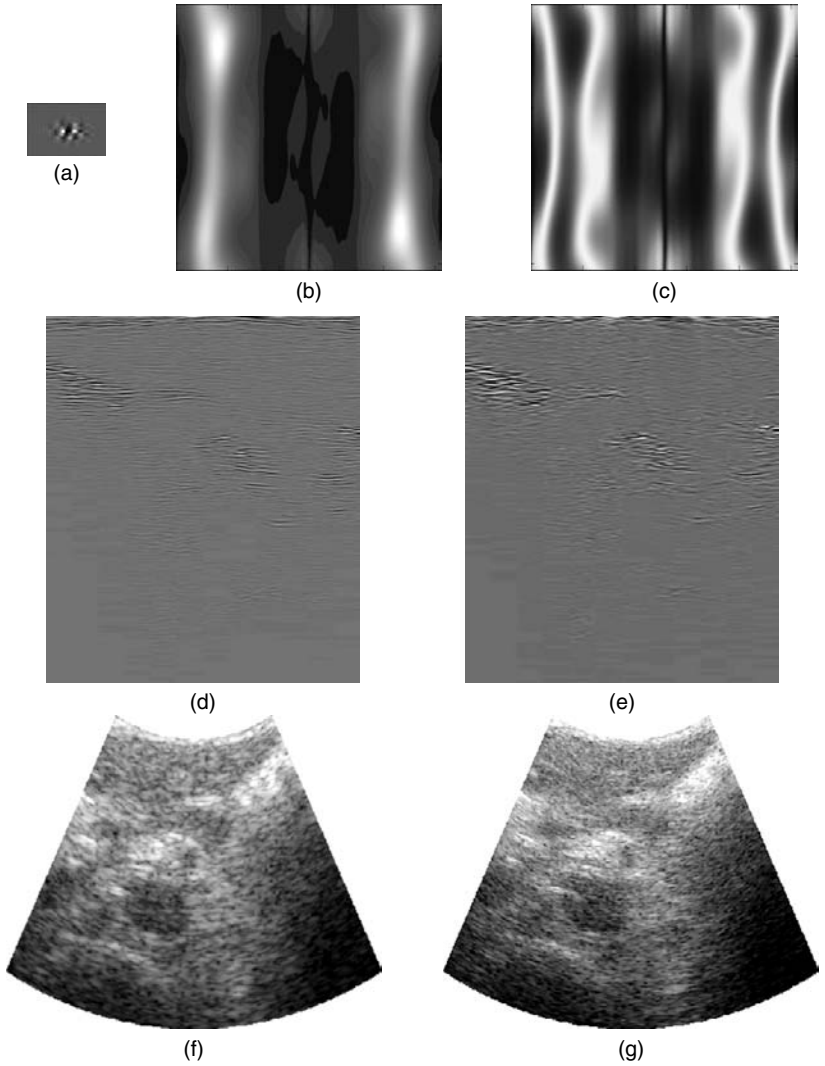


Figure 12.11 Restoration of an ultrasonographic image by Wiener filtering based on the known radio frequency PSF of blur: (a) magnified radio frequency PSF of the distortion, (b) corresponding frequency response of the distortion, (c) Wiener filter amplitude frequency response, (d) original RF data, (e) restored RF data, (f) video image from original data, and (g) video image from restored data. (From T. Taxt and R. Jirik, *IEEE Trans. Ultrason. Ferroelec. Freq. Cont.*, 51(2), 163–175, 2004. With permission.)

of the Wiener filter is thus influenced by both the transfer function of blur and the spectrum of noise related to the spectrum of the original images.

From the practical point of view, the usually mentioned formula (Equation 12.39) is hardly applicable, as estimating the power spectrum of an unavailable original is usually infeasible. The more practical, though rarely cited version (Equation 12.37), suffers, however, with the commonly imperfect estimates of the power spectra. Though it is physically clear that Equation 12.38 must be valid, due to estimate errors it may happen that the numerator in Equation 12.37 is negative at a frequency. Such a value must be rejected and some substitute provided at this frequency; the most likely choice is then zero.

The frequency response according to Equation 12.37, designed for continuous images, remains valid even for digital images, as long as the passband of frequencies does not exceed the Nyquist limits — the half of sampling frequencies in each dimension. It is then obviously possible to realize the filtering straightforwardly as two-dimensional or three-dimensional fast convolution in the frequency domain; alternatively, standard techniques to design two-dimensional or three-dimensional filters given the frequency response can be used.

12.4.2.3 Discrete Formulation of the Wiener Filter

The best linear restoration filter in the LMS sense, derived in the previous paragraph, may be of any space-invariant type; generally, filters with infinitely long impulse responses must be considered. We shall now limit ourselves to filters with PSF of finite size (finite impulse response (FIR) filters) that may be realized, at least in principle, also in the original domain by mask operators.

As it has been shown in Section 2.2.1, the two-dimensional (and even three-dimensional) operators can be expressed as one-dimensional, as in Equation 2.22, when the matrices to be operated on are expressed by vectors via column scanning. In order to simplify the notation and to make the principles clear, we shall use this approach in the rest of the chapter. The reader should realize that the two-dimensional version is considered throughout, although the two-dimensional $\rightarrow$ one-dimensional conversion and vice versa may be nontrivial.

One of the problems in the two-dimensional $\rightarrow$ one-dimensional conversion is to prevent interperiod interference that might appear

due to the circular character of discrete convolution. The discrete version of the distortion model (Equation 12.14) is

$$g_{i,k} = \sum_{m=0}^{i} \sum_{n=0}^{k} h_{i-m,k-n} f_{m,n} + \nu_{i,k}. \tag{12.40}$$

When the original image matrix has the format $K \times K$ and the PSF of the blur is sized $M \times M$, the size of the degraded image and consequently also of the noise matrix is $(K + M - 1) \times (K + M - 1)$. This is the minimum size of the matrices to be used in the algorithms. Prevention of interference in case of double convolution, as will be needed in some of the methods discussed in the next sections, requires increasing the matrices to the size $N \times N$, $N = (K + 2M - 2)$ or greater; N is usually chosen as the nearest higher integer power of 2 in order to allow use of the most effective DFT algorithms. The valid input data $g_{i,k}$, $\nu_{i,k}$ should be set to the greater matrices according to indices (i.e., in the left upper corner) and the remaining positions padded with zeros. The extended image matrices will be denoted **G**, **F**, and **N**; the corresponding vectors, obtained by column scanning and sized $N^2 \times 1$, are then **g**, **f**, and $\boldsymbol{\nu}$, respectively.

Discrete convolution realized by a FIR filter may be described, as in Equation 2.22, by the vector equation

$$\hat{\mathbf{f}} = \mathbf{M}\,\mathbf{g} \tag{12.41}$$

where the (block-circulant) matrix **M** expressing the two-dimensional convolution has the size $N^2 \times N^2$. Alternatively, a single element of the estimated original may be expressed as

$$\hat{f}_n = \mathbf{g}^T{}_n\mathbf{m} \tag{12.42}$$

where ${}_n\mathbf{m} : [{}_n m_k = M_{n,k}]$ is the transposed n-th row of **M** that carries the complete information on the PSF of the filter.* The stochastic variable $\hat{f}_n$ is thus estimated as a linear combination of the elements of **g** weighted by elements of **m**. According to the principle of

*n should be chosen so that f_n is the valid element of **f**, i.e., calculated based on only valid elements of **g**, not on padded zeros.

orthogonality, the error of the LMS estimate must be uncorrelated with all the input variables,

$$\mathbf{E}\{\mathbf{g}(f_n - \hat{f}_n)\} = \mathbf{E}\left\{\mathbf{g}\left(f_n - \mathbf{g}_n^T\mathbf{m}\right)\right\} = \mathbf{E}\{f_n\mathbf{g}\} - \mathbf{E}\{\mathbf{g}\,\mathbf{g}^T\}_n\mathbf{m} = \mathbf{0}. \quad (12.43)$$

The first mean in the last expression is the vector ${}_n\mathbf{\Phi}_{fg}$ of cross-correlations between the original value f_n and each of the elements of $\mathbf{g}$. The other mean is applied to the matrix of pair-wise products of elements of $\mathbf{g}$; the resulting matrix $\mathbf{\Phi}_{fg}$ is the *autocorrelation matrix* of the field generating the observed images. The last equation can be rewritten as

$$\mathbf{\Phi}_{gg\,n}\mathbf{m}_{opt} = {}_n\mathbf{\Phi}_{fg}, \quad (12.44)$$

which is a system of linear equations for the elements of $\mathbf{m}$. The coefficients of the PFS of the optimal restoration filter can be obtained by inversion,

$${}_n\mathbf{m}_{opt} = \mathbf{\Phi}_{gg}^{-1}\,{}_n\mathbf{\Phi}_{fg}. \quad (12.45)$$

The matrix $\mathbf{M}$ can be constructed based on ${}_n\mathbf{m}$ according to the block-circulant structure.

The characteristics of the restoration filter (Equation 12.45) are, similarly as in the classical formulation, determined by the statistical properties of the fields forming the original and distorted images; however, here both the characteristic of the filter (PSF) and the statistical characteristics of the fields (the correlations) are expressed in the original domain. The restoration system described by Equation 12.45 is called the *Wiener–Levinson filter*.

The formulation is very generic; the only assumption (made implicitly by not distinguishing to which n Equation 12.45 belongs) is that the involved stochastic fields are homogeneous, as in the classical formulation. The correlation vector and matrix should be estimated based on ensemble averages; i.e., the matrices for a particular n should be averaged over a set of realizations of the involved fields. As this is a rather demanding approach, often the ergodicity is supposed, which allows use of the spatial averages based on a single or a few realizations; however, reshuffling of elements of the correlation vectors for different n before averaging is needed. The sizes $N^2 \times N^2, N^2 \times 1$ of the correlation matrix and vector are unnecessarily

large when $M << K$, as common; a much smaller equation system may then be used thanks to the sparsity of **M**, but at the cost of further reshuffling of the elements before averaging. The computational details not influencing the restoration principle will not be treated here.

It should be noted that no assumptions were made about the character of distortion and noise, so that the filter is the generic optimal FIR linear filter of the given size for any particular restoration problem, including possibility of nonlinear distortions. Nevertheless, this generality is paid for by difficulties with providing the cross-correlation vector ${}_n\Phi_{\mathbf{fg}}$, which would require access to the values of the original images. Obviously, the efficacy of Wiener–Levinson filters is theoretically inferior to that of the generic infinite impulse response (IIR) Wiener filter (Equation 12.29) described in the previous section; the limit on the size of PSF might cause the filter to only approximate the optimum behavior and thus provide suboptimum results (though optimal in the class of FIR filters).

12.4.2.4 Generalized LMS Filtering

Generalized LMS filtering may be considered an extension of Wiener FIR filtering, without the most limiting requirement of homogeneity of the involved fields. Its explanation starts with the idea of realization in the "spectral" domain; however, the filtering may be obviously performed in the spatial domain as well. The method will be presented in the one-dimensional version, to which the problem of image restoration may be converted by standard scanning of image matrices (see the discussion following Equation 12.40).

In the "spectral" domain, the generalization is reflected by two properties. Primarily, the used forward and inverse discrete transforms need not be of the Fourier type (hence the quotation marks), as the convolution theorem is not utilized. Moreover, the spectral representation of the filter is not a vector of spectral coefficients (i.e., sampled frequency response of the filter), but rather a *matrix* **M** even in the one-dimensional case.

When the transform is defined by a matrix **A**, the "spectral" domain filtering providing the estimate of the original image can be described as the matrix product

$$\hat{\mathbf{f}} = \mathbf{A}^{-1}(\mathbf{M}(\mathbf{A}\mathbf{g})). \tag{12.46}$$

As the product $\mathcal{M} = \mathbf{A}^{-1}\mathbf{M}\mathbf{A}$ is the matrix of a proper size, the relation between the input and output may also be expressed in the spatial domain, as in Equation 12.41,

$$\hat{\mathbf{f}} = \mathcal{M}\mathbf{g}. \tag{12.47}$$

The estimate should be made in the LMS sense, so that the orthogonality principle applies; thus,

$$\mathsf{E}\{(\mathbf{f} - \hat{\mathbf{f}})\mathbf{g}^T\} = \mathsf{E}\{(\mathbf{f} - \mathbf{A}^{-1}\mathbf{M}\mathbf{A}\mathbf{g})\mathbf{g}^T\} = \mathbf{0}. \tag{12.48}$$

Let us define the correlation matrices:

$$\begin{aligned} \mathbf{R}_{ff} &= \mathsf{E}\{\mathbf{f}\mathbf{f}^T\}, \quad \mathbf{R}_{gg} = \mathsf{E}\{\mathbf{g}\mathbf{g}^T\}, \\ \mathbf{R}_{vv} &= \mathsf{E}\{\boldsymbol{v}\boldsymbol{v}^T\}, \quad \mathbf{R}_{fg} = \mathsf{E}\{\mathbf{f}\mathbf{g}^T\}. \end{aligned} \tag{12.49}$$

The condition (Equation 12.48) can then be rewritten as

$$\mathbf{R}_{fg} = \mathbf{A}^{-1}\mathbf{M}\mathbf{A}\,\mathbf{R}_{gg} = \mathcal{M}\mathbf{R}_{gg}, \tag{12.50}$$

so that the spatial-domain transform matrix can be expressed as

$$\mathcal{M} = \mathbf{R}_{fg}\mathbf{R}_{gg}^{-1}. \tag{12.51}$$

This may be considered a generalization of the Wiener–Levinson filter for nonhomogeneous image fields; naturally, the elements of rows of $\mathcal{M}$ may differ arbitrarily among different rows, while in the Wiener–Levinson filter individual rows differ only by elements being reshuffled. Obviously, the filter applies to restoration of any distortion, including nonlinear, as no restriction in this direction has been introduced so far.

The problem with the unavailability of the cross-correlation matrix in many cases reappears here as well. Similarly, as in the classical development, the more restricted model of distortion is introduced,

$$\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{v}, \tag{12.52}$$

which is the discrete form of Equation 12.14, but enabling a space-variant PSF (**H** is not block-circulant then). Providing that **f** and $\boldsymbol{\nu}$ are independent (or at least uncorrelated), it may be shown, utilizing the generalized transfer formula of a stochastic field by a linear system and additivity of correlation of the independent noise, that

$$\mathbf{R}_{gg} = \mathbf{H}\mathbf{R}_{ff}\mathbf{H}^T + \mathbf{R}_{\nu\nu}, \quad \text{and also} \quad \mathbf{R}_{fg} = \mathbf{R}_{ff}\mathbf{H}^T, \tag{12.53}$$

which may be considered a generalization of the Wiener–Lee theorem. Substituting the last expression into Equation 12.50 and solving for the "spectral" transform matrix **M**, we obtain

$$\begin{aligned} \mathbf{M} &= \mathbf{A}\mathbf{R}_{ff}\mathbf{H}^T(\mathbf{H}\mathbf{R}_{ff}\mathbf{H}^T + \mathbf{R}_{\nu\nu})^{-1}\mathbf{A}^{-1}, \quad \text{or} \\ \mathcal{M} &= \mathbf{R}_{ff}\mathbf{H}^T(\mathbf{H}\mathbf{R}_{ff}\mathbf{H}^T + \mathbf{R}_{\nu\nu})^{-1}. \end{aligned} \tag{12.54}$$

The difficult cross-correlation has been removed at the cost of narrowing the class of accepted distortions. The result is the generalized version of the discrete LMS restoration filter. The formal similarity with the classical Wiener filter is even more evident in the "spectral" domain, when all the matrices are appropriately transformed into their "spectral" counterparts; however, this will be omitted here.

When interpreting the result, it should be remembered that the one-dimensional formulation represents a two-dimensional problem. Thus, the generic (noncirculant) correlation matrices indicate the nonhomogeneity of the involved fields; the lack of homogeneity means that the correlations can be estimated only by using more demanding ensemble averages. Similarly, the generic matrix **H** of the linear distortion enables the encompassing of a spatial-variant distortion, the identification of which may also be difficult. Correspondingly, the restoration filter is generally space variant as well. The formulations in the original domain and in the "spectral" domain are equivalent not only as to the result concerns, but (unfortunately) also as for the computational complexity. Due to the nonconvolutional character of the problem, the transformation into the "spectral" domain does not bring any reduction in the problem dimension. The purpose of introducing the "spectral" domain formulation thus may only be to show the similarity with the frequency-domain formulation of the Wiener filter. The size of matrices remains enormous and is at the edge of the present realization possibilities.

12.4.3 Methods Based on Constrained Deconvolution

12.4.3.1 Classical Constrained Deconvolution

Although the final formula in the frequency domain resembles the Wiener filter, constrained deconvolution is a conceptually completely different approach from LMS filtering. The supposed model of distortion is again formulated in discrete space as Equation 12.40; e.g., when using the extended (zero-padded) matrices discussed in the paragraph following that equation, the sum limits are modified as

$$g_{i,k} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} h_{i-m,k-n} f_{m,n} + \nu_{i,k} = h * f\,|_{i,k} + \nu_{i,k} \,. \qquad (12.55)$$

When using the vector representation of images yielded via column scanning of image matrices, it is then possible to express the model as Equation 12.52, $\mathbf{g} = \mathbf{H}\mathbf{f} + \boldsymbol{\nu}$, where $\mathbf{H}$ is a block-circulant matrix describing the convolutional blur. Basically, we would like to invert this equation to obtain the original $\mathbf{f}$ based on the observation $\mathbf{g}$; it is, however, not exactly possible due to the presence of the unknown noise vector. It is nevertheless possible to find an estimate $\hat{\mathbf{f}}$ of $\mathbf{f}$, when applying some precautions against the increase of noise that would otherwise degrade the result.

The key notion of the method is the *residual* vector $\mathbf{r}$,

$$\mathbf{r} = \mathbf{g} - \mathbf{H}\hat{\mathbf{f}} \,, \qquad (12.56)$$

which evaluates the difference between the observation and the estimation of the original distorted by the known blur. Obviously, the optimal residual corresponding to the perfect estimate $\hat{\mathbf{f}} = \mathbf{f}$ is the noise $\boldsymbol{\nu}$. Not knowing $\boldsymbol{\nu}$, it is not possible to determine this perfect estimate; however, when another estimate can be found that would lead to a residual about comparable in overall intensity with $\boldsymbol{\nu}$, it might be a good solution of the restoration problem.

The intensity of noise is defined via its energy,

$$\varepsilon = \sum_{i=0}^{N^2-1} \nu_i^2 = \boldsymbol{\nu}^T \boldsymbol{\nu}; \qquad (12.57)$$

similarly, the energy of the residual, which is obviously a function of the estimate, is

$$E(\hat{\mathbf{f}}) = \mathbf{r}^T \mathbf{r} = (\mathbf{g} - \mathbf{H}\hat{\mathbf{f}})^T (\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}). \tag{12.58}$$

The above requirement of preservation of residual intensity may then be formulated as

$$E(\hat{\mathbf{f}}) = (\mathbf{g} - \mathbf{H}\hat{\mathbf{f}})^T (\mathbf{g} - \mathbf{H}\hat{\mathbf{f}}) = \varepsilon; \tag{12.59}$$

the energy of the residual of the estimate $\hat{\mathbf{f}}$ is constrained to equal the known noise energy ε. This is the first condition to be fulfilled by the solution $\hat{\mathbf{F}}$. Nevertheless, this equation does not determine the solution uniquely.

In order to find the best estimate, an additional condition must be imposed. A reasonable requirement is that the solution should be smooth in the sense of its Laplacian $\mathbf{l}$ having a minimum energy,

$$\Lambda(\hat{\mathbf{f}}) = \sum_{i=0}^{N^2-1} l_i^2 = \mathbf{l}^T \mathbf{l} = (\mathbf{L}\hat{\mathbf{f}})^T (\mathbf{L}\hat{\mathbf{f}}) \rightarrow \min, \tag{12.60}$$

where $\mathbf{L}$ is the transform matrix expressing the discrete Laplacian operator, defined by either of the masks, Equation 11.24 or Equation 11.26. This is the second condition determining the solution.

We have arrived at the constrained optimization formulation of the restoration problem: the minimum of the Laplacian energy (Equation 12.60) is sought in the space of pixel intensities of the restored image; the search is limited to the hypersurface defined by the constraint (Equation 12.59).

The direct solution of this problem in the original domain is in principle possible but practically very complex due to the extreme size of the involved matrices and the necessity to solve iteratively a complicated nonlinear equation system. Alternatively, the restoration process can be shown to be equivalent to two-dimensional linear filtering with a certain frequency response that can be derived, using the special properties of the matrices [64]. Later, the same result was derived directly [63], utilizing the circular convolution property of two-dimensional DFT (Equation 2.66) and the discrete form of Parseval's theorem.

Like the model of distortion (Equation 12.55), the noise energy may obviously be expressed as

$$\varepsilon = \sum_i \sum_k v_{i,k}^2, \tag{12.61}$$

the smoothness criterion (Equation 12.60) as

$$\Lambda(\hat{f}) = \sum_i \sum_k (l * \hat{f}\,|_{i,k})^2 \to \min, \tag{12.62}$$

and the residual constraint (Equation 12.59) as

$$E(\hat{f}) = \sum_i \sum_k (g_{i,k} - h * \hat{f}\,|_{i,k})^2 = \varepsilon. \tag{12.63}$$

Here, the lowercase letters mean the image matrices or their elements, when the indices are indicated.

The problem can be transformed into the frequency domain in a straightforward manner, using Equation 2.66 and the equality

$$\sum_i \sum_k |\, f_{i,k}\,|^2 = \sum_m \sum_n |\, F_{m,n}\,|^2, \tag{12.64}$$

following from the discrete Parseval's theorem. Here, as well as in the following equations, the capital letters denote the spectral matrices (in two-dimensional DFT) or their elements. Thus, we obtain

$$\sum_m \sum_n |\, L_{m,n} \hat{F}_{m,n}\,|^2 \to \min \tag{12.65}$$

and

$$\sum_m \sum_n |\, G_{m,n} - H_{m,n} \hat{F}_{mn}\,|^2 = \varepsilon. \tag{12.66}$$

Thus, the optimum estimate of the restored image spectrum $\hat{F}$ in the sense of Equation 12.65 has to be found, constrained by Equation 12.66. As the spectral coefficients $\hat{F}_{m,n} = A_{m,n} + jB_{m,n}$ are generally complex, $2N^2$ unknowns are sought. Using the method of Lagrange's coefficient, we have to find the extreme of the functional

$$U = \sum_m \sum_n (|L_{m,n}\hat{F}_{m,n}|^2 + \lambda |G_{m,n} - H_{m,n}\hat{F}_{m,n}|^2 - \lambda\varepsilon) \to \min, \quad (12.67)$$

where λ is Lagrange's coefficient, to be determined during solution, besides the matrix $\hat{\mathbf{F}}$ of the original spectrum estimate. The advantage of this formulation is that the individual partial derivatives, as needed in the condition of zero gradient of U, are separated, and each may thus be solved independently of others in a closed form. This way, the following is obtained:

$$\begin{aligned} \frac{\partial U}{\partial A_{m,n}} = 0 \quad &\Rightarrow \quad A_{m,n} = \frac{\lambda \,\mathrm{Re}\{G_{m,n}H^*_{m,n}\}}{|L_{m,n}|^2 + \lambda |H_{m,n}|^2}, \\ \frac{\partial U}{\partial B_{m,n}} = 0 \quad &\Rightarrow \quad B_{m,n} = \frac{\lambda \,\mathrm{Im}\{G_{m,n}H^*_{m,n}\}}{|L_{m,n}|^2 + \lambda |H_{m,n}|^2}. \end{aligned} \quad (12.68)$$

Thus, the spectral elements of the estimate are

$$\begin{aligned} \hat{F}_{m,n} &= \frac{H^*_{m,n}}{|H_{m,n}|^2 + \frac{1}{\lambda}|L_{m,n}|^2} G_{m,n} = M_{m,n}G_{m,n}, \quad \text{i.e.,} \\ M_{m,n} &= \frac{1}{H_{m,n}} \; \frac{|H_{m,n}|^2}{|H_{m,n}|^2 + \frac{1}{\lambda}|L_{m,n}|^2}. \end{aligned} \quad (12.69)$$

It is therefore obvious that the restoration by constrained deconvolution, though formulated initially differently, also leads to modified space-invariant inverse filtering.

However, Lagrange's coefficient λ is still to be determined. As the term, containing ε in Equation 12.67, has been omitted due to differentiation, the condition (Equation 12.66) must be observed and

can be utilized for determination of λ. Substituting the result (Equation 12.69) into the condition yields

$$\sum_m \sum_n \frac{\frac{1}{\lambda}|G_{m,n}|^2 |L_{m,n}|^4}{\left(|H_{m,n}|^2 + \frac{1}{\lambda}|L_{m,n}|^2\right)^2} = \varepsilon, \tag{12.70}$$

which is a nonlinear equation in λ with all the other quantities *a priori* known. Thus the nonlinearity, inherent in the filter design method, requires solving only this single equation, while the other unknowns need not be included in the iterative loop. This turns out to be a substantial advantage of the frequency-domain formulation.

The discrete frequency response of the restoration filter $M_{m,n}$, though it looks similar to the Wiener filter response (Equation 12.39), has a different character. The similarity is in the pure inverse filter that is again modified by a real-valued correction factor with the values in the range <0, 1>. However, this correction filter is designed specially for a particular processed image, as can be seen in Equation 12.70, where $G_{m,n}$ is explicitly present and thus influences the value of λ.

The λ-parameter then carries the specificity of the filter for a particular image g. In comparison with the Wiener filter, which works for a class of images, it is obviously a limitation; on the other hand, the constrained deconvolution does not require any restrictions concerning the statistical properties (namely, homogeneity of fields, as no fields are considered). Consequently, no power spectra of the noise field and the field of the observed images need to be identified; the needed characteristic of the noise is only its total energy (for homogeneous zero-mean noise, it is proportional to its variance).

The requirement of smoothness of the restored image does not cause blurring of the solution; it just prevents the enormous increase in noise component to which the restoration procedures tend. The resulting sharpness is dependent on the initial SNR in the observed image. The used criterion of smoothness (Laplacian) has been chosen rather arbitrarily; though it proved useful, there might be other possibilities as well.

12.4.3.2 Maximum Entropy Restoration

Maximum entropy–based restoration is conceptually related to the constrained deconvolution. It expects the same model of distortion (Equation 12.55), which is to be approximately inverted (via observing

a set of constraints formed by the model), and also optimizes a criterion of smoothness preventing the explosive increase of noise, though it is defined differently. The smoothness criterion to be maximized uses the formula of entropy*,

$$H(\hat{f}) = \sum_m \sum_n \hat{f}_{m,n} \ln(\hat{f}_{m,n}) \to \max; \tag{12.71}$$

however, here only its monotonous dependence on the smoothness of the image is utilized. The criterion is maximized for the smoothest possible image $f_{i,k} = \text{const.}, \forall i,k$; as the pixel intensities are nonnegative, it is always defined, though a small positive quantity may need to be added to prevent zero entries.

The method aims at estimating not only the image values $\hat{f}_{m,n}$, but also the values of noise $v_{i,k}$. The model of distortion applied to the estimates,

$$g_{i,k} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} h_{i-m,k-n} \hat{f}_{m,n} + \hat{v}_{i,k} = h * \hat{f}\,|_{i,k} + \hat{v}_{i,k}, \tag{12.72}$$

provides first N^2 constraints. An additional constraint is the requirement of preserving the constant sum of image elements,

$$\sum_{m=0}^{N-1} \sum_{n=0}^{N-1} \hat{f}_{m,n} = \sum_{m=0}^{N-1} \sum_{n=0}^{N-1} g_{m,n} = s = \text{const}, \tag{12.73}$$

as a calculation stabilizing element. As the noise values are also estimated, the smoothness criterion of the type (Equation 12.72) will be applied to the noise as well; however, as noise is not nonnegative, the argument of the criterion will be the sum of the noise and a quantity $a > -v_{\max-}$. The overall smoothness criterion is then a linear combination of both partial criteria,

$$H = H(\hat{f}) + cH(\hat{v} + a) \to \max, \tag{12.74}$$

where the chosen weight c determines the relative suppression of noise compared to the image (usually chosen ~20).

Thus, the problem is again converted to searching a constrained extreme by the method of Lagrange's coefficients. The corresponding

*On the interpretation of entropy for stochastic ensembles, see Section 10.1.3.

functional reflects the numerous conditions: N^2 of them formed by Equation 12.72, plus that formulated by Equation 12.73,

$$U = H(\hat{f}) + cH(\hat{v} + a) + \sum_i \sum_k \lambda_{i,k}(h * \hat{f}\,|_{i,k} + \hat{v}_{i,k} - g_{i,k}) + \\ + \lambda_0 \sum_i \sum_k \hat{f}_{i,k} - s \quad \rightarrow \quad \max. \tag{12.75}$$

There are thus $N^2 + 1$ coefficients to be determined, plus $2N^2$ estimated values of the restored image and noise. Maximizing U requires zero gradient; therefore, $2N^2$ following equations are formed that are separable and yield

$$\left.\frac{\partial U}{\partial \hat{f}_{i,k}}\right|_{\hat{f}} = 0, \quad \Rightarrow \quad \hat{f}_{i,k} = \exp(-1 + \lambda_0 + h^* \lambda\,|_{i,k}), \\ \left.\frac{\partial U}{\partial \hat{v}_{i,k}}\right|_{\hat{v}} = 0, \quad \Rightarrow \quad \hat{v}_{i,k} = \exp(-1 + \lambda_{i,k}/\lambda_0). \tag{12.76}$$

This way, the sought quantities are expressed in terms of Lagrange's coefficients. These are provided by solving a nonlinear system of $N^2 + 1$ equations, obtained by substituting the expressions (Equation 12.76) into Equations 12.72 and 12.73. The system is rather complex, but numerically solvable.

The basic advantage of the maximum entropy method is its universality; actually, no conditions are imposed on the restored image and noise. It is obviously a (complicated) nonlinear estimate. The concrete procedure is specific *for a particular image* and must thus be completely repeated in every new case.

12.4.4 Constrained Optimization of Resulting PSF

This method represents an approach different from the previous ones: instead of considering the differences between the original and the restored image, it tries to improve the overall impulse response $c(x, y)$ of the system combined of distortion and restoration, as may be vaguely formulated in continuous space,

$$c(x,y) = h^* m\,|(x,y) \quad \rightarrow \quad \delta(x,y), \tag{12.77}$$

where $h(x, y)$ and $m(x, y)$ are PSFs of the distorting blur and restoration system, respectively. It is requested that the noise level should not increase during restoration. Only the basic idea will be presented, for simplicity in the continuous space.

The optimization criterion S is evaluating the slenderness of the resulting PSF by comparing the energy of $c(x, y)$ with the weighted version of it, when the weight $w(x, y)$ defines a penalty for the improper shape of $c(x, y)$,

$$S\{c(x,y)\} = \frac{\iint_A w(x,y)c^2(x,y)\mathrm{d}x\mathrm{d}y}{\iint_A c^2(x,y)\mathrm{d}x\mathrm{d}y} \quad \rightarrow \quad \min. \tag{12.78}$$

The weight may penalize the undesirable values, e.g., quadratically with the distance from the origin, $w(x, y) = x^2 + y^2$, or it may tolerate a certain extent without penalty and penalize the rest equally,

$$w(x,y) = \begin{cases} 0 & \text{when } (x^2+y^2) \in D \\ 1 & \text{outside of } D \end{cases}, \tag{12.79}$$

where D is a suitably defined neighborhood of the origin (e.g., circular).

The noise may reasonably be considered homogeneous. After restoration filtering, the zero-mean noise and its power given by the variance become

$$v'(x,y) = m * v\big|(x,y), \quad \sigma'^2 = \mathbf{E}\{v'^2(x,y)\} = \sigma_0^2 \tag{12.80}$$

(the position may be omitted thanks to homogeneity). The last equality expresses the constraint to the noise; σ_0^2 is the noise power in the observed image.

An additional constraint is provided by the natural requirement that the energy of $c(x, y)$ corresponding to the overall gain over distortion and restoration is unitary,

$$E\{c(x,y)\} = \iint_A c^2(x,y)\mathrm{d}x\mathrm{d}y = 1. \tag{12.81}$$

Obviously, all the involved equations contain the PSF of the restoration filter $m(x, y)$, the shape of which is to be determined via

optimization. Thus, the problem is converted to the constrained optimization of Equation 12.78 in the space of coefficients of the discrete version of $m(x, y)$, while the constraints (Equations 12.80 and 12.81) are observed.

The required inputs to the filter design procedure are the impulse response of distortion $h(x, y)$, power and the autocorrelation function (matrix) of the input noise; the choice of the weight function $w(x, y)$ influences the properties of the method. A detailed description of how the impulse response $m(x, y)$ of the restoration filter is computed is beyond the frame of this book; the design of the filter is rather complex.

The actual restoration procedure, on the other hand, is quite simple — a linear space-invariant filtering. The approach is attractive thanks to the unlimited class of processed images; the observed images do not enter the design process at all. The mild limitations concern only the noise, which should be homogeneous, and the distortion that is supposedly convolutional. The identification is standard and not very demanding: it concerns just the distortion PSF and the two-dimensional autocorrelation matrix of noise, including the noise power as its component.

12.4.5 Bayesian Approaches

We have already introduced (in Section 12.4.2.1) the *a priori* and *a posteriori* probabilities of the original image $f(\mathbf{r})$; they differ in availability of the observed (distorted) image $g(\mathbf{r})$. When considering the continuous image intensities in a spatially discrete image, the corresponding probability densities must be used. Such a density for a complete discrete image f, considered a stochastic vector, may be visualized as a scalar field in the multidimensional space of pixel intensities. This density determines the mean value of $\bar{f}$,

$$\bar{f} = \iint_V \ldots \int f\, p(f)\, \mathrm{d}V_f, \tag{12.82}$$

where V is the support space of the image. The *a priori* density distribution $p(f)$ is changed to the conditional density $p(f|g)$ once the observed image g is available; roughly said, the density usually becomes higher than $p(f)$ in the neighborhood of g, while it is decreased elsewhere. The conditional density $p(f|g)$ defines the conditional mean $\bar{f}\,|_g$; it can be proved that this represents the best estimate $\hat{f} = \bar{f}\,|_g$ in the

LMS sense, based on the knowledge of g. Unfortunately, it is usually infeasible to find the conditional mean, as unavailable knowledge of probabilistic characteristics of the image-generating fields would be needed. An alternative may be the suboptimal LMS approaches as described in Section 12.4.2. Another alternative of the estimate is the modus f_{MAP} of the density $p(f|g)$ given g,

$$f_{\mathrm{MAP}} = \arg\max_{f} p(f \mid g), \tag{12.83}$$

which may be considered a newly defined different estimate $\hat{f} = f_{\mathrm{MAP}}$. It is usually a good approximation of $\bar{f}\,|_g$ (as the maximum and mean of monomodal distributions are usually close to each other). The name — *MAP estimate* — describes its character.

Bayes' rule relates the two above distributions with another distribution pair: the marginal (unconditional) probability density $p(g)$ of the observed image g, and the conditional density $p(g|f)$ of the observed image, provided that the concrete f was the original; Bayes' relation is

$$p(f \mid g)p(g) = p(g \mid f)p(f). \tag{12.84}$$

This equality forms a basis of *Bayesian methods* and is utilized when deriving the statistically satisfactory estimates of f given g.

The shape of the conditional probability density $p(g|f)$ depends obviously on the original f, which may be considered a (vector-valued) parameter of this distribution. The maximum of $p(g|f)$ determines its modus, which is the most probable observed image g given the original f. Conversely, when a set of mutually independent observations g_j, $j = 1, 2, \ldots N$ is available for a particular but unknown original f, then the likelihood function L may be defined as

$$L(f) = \prod_{j=1}^{N} p(g_j \mid f), \tag{12.85}$$

which describes the probability distribution of f given the concrete observed set of g_j. The modus of this distribution is the maximally probable original from which the concrete observation set could have been derived. The guess $\hat{f} = f_{\mathrm{ML}}$ based on this idea,

$$f_{\mathrm{ML}} = \arg\max_{f} L(f), \tag{12.86}$$

is called the *maximum-likelihood estimate* of the original image.

Both estimates f_{MAP} and f_{ML} are valid in the sense of their definitions and may provide practically useful estimates of f based on the given g; however, generally they are not identical. It is usually easier to provide the information needed to estimate either f_{MAP} or f_{ML} (usually one of them is sought) than the more complex knowledge required to estimate $\bar{f}|_g$. The Bayesian methods can cope with a more generic nonlinear model of distortion, in the spatially discrete version

$$g_{i,k} = s(h*\hat{f}|_{i,k}) + \hat{v}_{i,k}, \tag{12.87}$$

where $s(\ldots)$ is a generally nonlinear function (a point-wise intensity transform). Except for special cases, the Bayesian estimates of f are nonlinear functions of g, so that they cannot be realized by linear filters.

12.4.5.1 Maximum *a Posteriori* Probability Restoration

The MAP method looks for the maximum of $p(f|g)$, which is usually directly unavailable. According to Equation 12.84, Equation 12.83 can be expressed as

$$f_{\mathrm{MAP}} = \arg\max_f p(f \mid g) = \arg\max_f p(g \mid f)p(f), \tag{12.88}$$

because the factor $1/p(g)$ may be omitted on the right-hand side as independent of f. This is an important simplification: the *a priori* probability distribution $p(g)$ may be difficult to obtain due to distortion nonlinearity even if $p(f)$ and $p(v)$ are known.

The *a priori* distribution $p(f)$ may be estimated either theoretically, based on the model of image generation, or experimentally, when there is (even temporary) access to the originals. Further, the conditional distribution $p(g|f)$ can be determined when the function $p(v) = p_v(v)$ is known (which is typically the case; $p_v(\ldots)$ may be multidimensional Gaussian, uniform, or any experimentally determined shape of distribution). Really, as $v = g - s(h*f)$ according to Equation 12.87, the distribution $p(g|f)$ for a given f is described by the function $p_v(\ldots)$ shifted to the position of the image $s(h*f)$. Thus,

$$p(g \mid f) = p_v(g - s(h*f)) \tag{12.89}$$

and the estimate f_{MAP} is then, according to Equation 12.88,

$$f_{\mathrm{MAP}} = \arg\max_{f} p_v(g - s(h*f))p(f). \tag{12.90}$$

The necessary condition for optimum is the zero gradient, in components

$$\frac{\partial}{\partial f_{i,k}}[p_v(g - s(h*f))p(f)]\Big|_{f_{\mathrm{MAP}}} = 0. \tag{12.91}$$

This is a system of N^2 nonlinear equations for the same number of unknown intensities of the estimated image f_{MAP}. Note that all functions in the equations are known and the derivatives thus can be determined for any given f; nevertheless, the formulation of the equations and their solution in practical cases is rather complex. Simplifying approximations have been suggested, providing that both the distributions $p(f)$ and $p(v)$ are Gaussian; in this case, the mean of $f_{i,k}$ may be space variable, thus removing the condition of homogeneity with respect to mean, although the autocorrelation matrices are supposedly two-dimensional (Hunt, see [64]).

12.4.5.2 Maximum-Likelihood Restoration

Generally, an ML estimate is based on a set of observations (realizations) of the same stochastic variable. In the image restoration problems, only a single, though vector-valued, observation g is usually available; thus, it is a special case in this respect. Therefore, Equation 12.85 simplifies to

$$L(f) = p(g \mid f) \tag{12.92}$$

and the components of f may be found by solving the system of nonlinear equations

$$\frac{\partial}{\partial f_{i,k}} L(f) = \frac{\partial\, p(g \mid f)}{\partial f_{i,k}} = 0\,, \quad \forall i,k\,, \tag{12.93}$$

following from the condition of zero gradient of L in the space of pixel intensities. Instead of solving the difficult equation system, iterative optimization approaches may be used.

An example of a *de facto* restoration problem solved this way is the ML approach to reconstruction from noisy projections in the second part of Section 9.2.2.

12.5 HOMOMORPHIC FILTERING AND DECONVOLUTION

Homomorphic processing is usually rated as a restoration method, though it is often closer to image enhancement when little or only rather vague identification is involved. The reason is in its relative conceptual and computational complexity. The principles of homomorphic filtering and deconvolution have been explained in the second part of Section 1.3.4, and the discrete homomorphic operators are mentioned in Section 2.2.2.

The homomorphic approach is used when a mixture of nonadditive components forms the observed image. It must be possible to determine the operation that connects the image components. Most often, it is multiplication (namely, disturbance by multiplicative noise or modulation interference caused by undesired signals); the filtering of multiplicative mixtures is usually understood under *homomorphic filtering* (see example in Figure 1.13). When the mixture has convolutional character, but not enough is known about the characteristics of the distortion (e.g., its PSF cannot be well determined), the blind *homomorphic deconvolution* may be attempted via converting the convolution into addition, as explained in Section 1.3.4.

In any case, the identification must determine the operation among the components of the observed image. Consequently, the conversion subsystem (the input homomorphic system), with the input operation equal to the identified one and the output operation of addition, must be available or newly designed. Similarly, the inverse output homomorphic subsystem must be provided. The design of the inner linear filter of the homomorphic system should be based on as good an identification of the signal properties as possible; namely, it is necessary to analyze the influence of the input-decomposing transformation on the spectral components. When homomorphic filtering is considered, the information available is most often rather limited; therefore, design of the linear part of the system should be constrained so that the filter only mildly influences the spectrum (e.g., smoothly enhancing the upper part of the spectrum, etc.). Experimentation with the parameters of the linear filter may be needed in the lack of a sufficient knowledge on image properties.

12.5.1 Restoration of Speckled Images

An important use of homomorphic filtering has been described concerning the restoration of *speckled images*. Such images are generated by imaging systems using coherent radiation — in medical applications, primarily ultrasound and laser light. The radiation components scattered by the structure of the imaged objects are mutually interfering; the wave interference leads to local enhancement or suppression of the measured response. The phases of the individual components are randomly determined by the spatial distribution of the elementary scatterers; however, their mutual relations are fixed as long as the geometry of imaging is fixed. This leads to randomly distributed small image areas of intensity varying between high and low. Such a texture degrades the image quality, decreases the resolution, and may obscure important details, as may be well seen, e.g., on ultrasonograms. This distortion can be described approximately as multiplicative, the image being a product of the useful image content and the speckle texture. Notice that the speckle texture depends on the PSF of the imaging system; it thus carries primarily the information on the imaging system and less on the imaged object. Roughly said, the mentioned factorization is only valid when details of the useful image are greater than the system resolution. Obviously, the multiplicative mixture is a suitable object of homomorphic filtering with the logarithmic input transform; using the Wiener filter as the linear part of the system enables suppression of the noise component optimally providing that spectral properties of both components after input transformation have been properly identified.

Alternatively, the speckle disturbance may be suppressed by averaging more images of the same scene with the speckle texture diversified, e.g., by differently positioning the illuminating source or working with different wavelengths of the radiation. This is demanding as to the imaging system concerns and often cannot be done in real time when the image synthesis should be done automatically. However, the same approach is performed routinely: a mental averaging of similar images is done when examining an organ with the hand-carried ultrasonic probe so that each image frame is provided from a slightly different position, thus diversifying the speckled texture.

13

Image Analysis

The procedures described in this chapter differ from the previous ones substantially: they provide, as their results, a *description* of the input image rather than a processed image as such. We shall begin with the analytic methods that are determining local properties of the image or spatially limited relations between these local properties, as used in characterizing textures. Subsequently, segmentation approaches will be presented that allow dividing the image area into meaningful parts as required by a particular purpose of imaging. Finally, as a part of this chapter, we shall introduce the strongly nonlinear morphological operators that belong to both image processing and analysis; however, they may significantly contribute to image examination and recognition (e.g., by thinning, connecting or disconnecting objects, or providing their structural description) and are thus usually considered analytic methods.

Of the literature cited in the references to Part III, the most relevant sources to this chapter are [5], [80], [13], [70], [72], [25], [79], [59], [17], as well as [26], [6], [7], where numerous additional references are cited.

Though some of the involved methods are quite sophisticated as to the conceptual and computational means concerned, this type

of examination is often denoted as the low-level analysis to distinguish it from a higher-level evaluation of shapes, spatial relations among objects, and automatic scene recognition. It would be perhaps useful to subdivide the low-level analysis into the lowest-level evaluation of local properties and textures, and the mid-level analysis enabling segmentation based on the local features, and elementary size and volume measurements. In the field of medical imaging, the low- and mid-level analyses are prevailing so far. The higher analysis tasks are still reserved mainly for the medical staff evaluating the images; probably, they will remain so into the foreseeable future, not only because of the technical difficulty of the analytic tasks, but also due to ethical and legislative issues. The reader interested in principles of the automation of the higher-level procedures, e.g., three-dimensional scene recognition, which may find use, e.g., in interventional radiology, should refer to the literature on computer vision, e.g., [17], [72].

13.1 LOCAL FEATURE ANALYSIS

Local features are the simplest description of image properties. They may concern each single pixel individually (e.g., the pixel intensity or color, possibly also vector pixel values in fused images), but more important is the analysis of local properties based on some neighborhood pertaining to a particular pixel, to which the parameter obtained by the analysis is then attributed. The magnitude of local gradient, determined for each image pixel that is based on intensities of the neighboring pixels, may serve as an example. As far as such a feature or parameter can be basically determined (derived from the original image) for every pixel on the image grid, a new image, denoted as the *parametric map* or *parametric image*, may be defined by the parameter values. The further step of analysis—the segmentation—is in most of its numerous forms based on such parametric images, explicitly or implicitly, when the feature values determine the membership of a concrete pixel in an object or a segmented area. This classification should be based on the knowledge that a particular feature is (or may be) characteristic for a particular type of object (e.g., a type of tissue). Another use of the maps of local features is in adaptive image processing, where the spatially variable parameter(s) may influence the character of local processing (Sections 11.2.4 and 11.3.2).

Most of the features are defined by means of some local parameters expressed either in the original domain or in a transform

domain (e.g., in Fourier or wavelet space); the term *local* means that the analytic statistics or transform concern a certain, usually small, area surrounding the position of the feature-carrying pixel. A generic problem common to all analyses of this kind is the choice of the size and shape of the neighborhood. On one hand, the local statistics has a lower variance (hence, it is more reliable) when there are more data, i.e., when the area determining the parameter is larger. On the other hand, the larger are the local analytic areas, the coarser is the spatial resolution of the parametric image because the uncertainty as to where exactly the derived parameter belongs is higher. Thus, the choice of the analyzed local neighborhood is always a compromise depending on the character of the image and the purpose of a particular analysis. Practical values may start at the minimum of 2×1 or 1×2 pixels on one side and may end at areas of many tens of pixels in diameter on the other side. In addition, the shape of the neighborhood is important and may influence the isotropicity of the analysis.

13.1.1 Local Features

Local features are determined for each pixel by its small neighborhood of a chosen size. A trivial case is the 1×1 neighborhood: the original image becomes its own parametric image; however, when a fused vector-valued image is considered, the concept of interpreting such vectors as descriptive parameters may have its foundation. To substantiate the definition of features as local, the neighborhood size should not exceed a fraction of the size of the smallest important objects in the image; mostly, the size measures in units, at maximum in tens, of pixels.

13.1.1.1 Parameters Provided by Local Operators

All linear and nonlinear local operators (particularly those discussed in Sections 11.2 and 11.3) may be understood as defining certain local features, as, e.g., the local derivative or Laplacian. The classification of a pixel, as needed in similarity evaluation or in segmentation, may then be based on vectors consisting of several such local parameters derived for each pixel. This way, a vector-valued parametric image is formed.

13.1.1.2 Parameters of Local Statistics

The simplest local features are based on local statistics of a neighborhood $S_{i,k}$ of a pixel at i, k. The neighborhood contains M pixels

indexed $i + s, k + r$; the limits of s and r may be mutually dependent so that the shape of $S_{i,k}$ need not be rectangular (for a simple 3×3 neighborhood, obviously both $s, r = -1, 0, 1$, and consequently $M = 9$). The statistical parameters used may be divided into two groups:

- *Local moments* of n-th order,

$$ {}_n m_{i,k} = \frac{1}{M} \sum_{S_{i,k}} f^n_{i+s,\,k+r}, \tag{13.1} $$

  of which the *local mean* ($n = 1$) and *local power* ($n = 2$) are most frequently used
- *Local central moments* of n-th order,

$$ {}_n \mu_{i,k} = \frac{1}{M} \sum_{S_{i,k}} (f_{i+s,k+r} - m_{i,k})^n, \tag{13.2} $$

  of which the *local variance* $\sigma^2_{i,k} = {}_2\mu_{i,k}$ and standard deviation $\sigma_{i,k} = \sqrt{{}_2\mu_{i,k}}$ are most commonly used. Other statistical parameters that can be derived from the moments (e.g., *local skewness*) may prove useful for characterizing a particular image area. Note that for greater object areas, moments taking into account the mutual pixel positions may be derived and that some of their combinations form features that are invariant under geometrical transforms (Hu, Suk, and Flusser, as cited in [50]).

All these (and also other) parameters may be derived from the local histogram. Generally, the calculations of these features (except the mean) involve nonlinearities and thus cannot be realized by linear mask operators.

13.1.1.3 Local Histogram Evaluation

Local histogram $\mathbf{h}^S_{i,k}$ of the image with q levels of gray is counted from the neighborhood $S_{i,k}$ according to definition (Equation 2.12),

$$ \begin{aligned} &\mathbf{h}^{S_{i,k}} : {}_{i,k}h_l^{S_{i,k}} = \mathrm{count}({}_{i,k}S_l) = \sum_{{}_{i,k}S_l} 1, \quad l = 0,1,\ldots,q-1, \\ &{}_{i,k}S_l = \left\{ f_{i,k} : ((f_{i,k} \in S_{i,k}) = l) \right\}. \end{aligned} \tag{13.3} $$

Often it is useful to provide a courser gray scale by grouping close gray levels, thus obtaining, e.g., 16 instead of 256 levels. This way, the histogram is much smaller and its classes acquire higher counts

that are statistically more reliable. The local histogram itself may be considered a vector-valued local parameter of q components that may be used independently, or grouped in a way.

The normalized local histogram $\mathbf{h}^S_{i,k}/M$ provides rough estimates of the gray-level probabilities in the neighborhood of the pixel (i, k). This justifies deriving such parameters of the probability distribution (besides the already mentioned moments) as the *local median*, *local mode*, and also the *local entropy*,

$$H_{i,k} = -\sum_{l=0}^{q-1} \frac{{}_{i,k}h_l^{S_{i,k}}}{M} \log_2 \frac{{}_{i,k}h_l^{S_{i,k}}}{M} \quad \text{[bits]}. \tag{13.4}$$

On the physical interpretation of the entropy parameter, see Section 10.1.3 concerning the information-based similarity criteria.

13.1.1.4 Frequency-Domain Features

Locally applied unitary transforms may also give relevant information on the local character of the image. Applying the transforms on small local neighborhoods prevents achieving high spectral resolution (which is usually not needed or may even be undesirable in local description), but enables localization of spectral properties with the resolution given by the neighborhood size. Naturally, it is usually not useful to calculate the spectral features for each pixel as the center of a neighborhood because the results for closely spaced central pixels would be almost identical due to prevailing overlap of the adjacent neighborhoods. The density of such centers—samples of spectral-domain description—should correspond to the chosen neighborhood size; reasonable overlapping would be some 1/3 to 2/3 in each dimension. Consequently, the *spectral parametric image* would usually be more coarsely sampled than the original image.

The two-dimensional discrete Fourier transform (DFT) can provide the information in the classical spatial frequency domain; however, the individual spectral coefficients usually have too high variance to be considered reliable parameters, and the complete spectrum may form a too numerous set of local parameters. Therefore, some grouping is often used; e.g., the energy of a selected frequency band (sum of the respective squared spectral coefficients) may be utilized as a local parameter (Figure 13.1). The ring grouping separates the different absolute frequencies and may provide information on the prevailing spatial frequencies. Alternatively, the wedge grouping determines total energies for particular directions (or rather direction ranges), thus enabling evaluation of the main

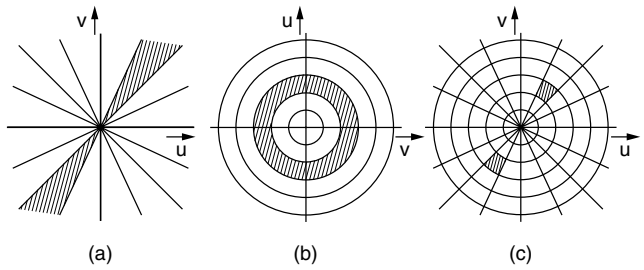


Figure 13.1 Partitioning of local spectra to provide a low number of spectral parameters: (a) wedge filter, (b) ring filter, and (c) combined filter.

spatial orientation of the local intensity variations. When both partitions are combined, a rough approximation of the spectrum is obtained, providing combined information on both direction and absolute frequency distribution.

It is further possible to integrate these energies to a smaller number of combined parameters or even to a single parameter. For instance, the variance of the set of ring energies provides the information on size-of-detail distribution, while the variance of wedge energies describes the degree of (an)isotropy. On the other hand, complete power spectra estimates may also be considered vectors of parameters; they correspond via the Wiener–Khintchin theorem to the local autocorrelation function of the inspected neighborhood, which is the equivalent original-domain description.

13.1.2 Edge Detection

An important local property that can be attributed to a pixel (or group of pixels) is whether there is an edge, i.e., a sufficiently important and fast change of intensity in a small neighborhood. The edge may be a border of a greater object in the image, or a feature of an object—a rim or an angle—or it may also indicate a self-standing object, such as a line or a point. Unfortunately, quite often, local intensity changes due to noise are also detected as edges; these should be excluded by subsequent processing taking image context into consideration. Global edge detection applied to an image provides a derived parametric binary image, sc., *raw edge representation* with ones (white) at the pixels where the edge has been detected, on the black (zeros) background or vice versa.

Most edge detectors are based on difference operators, as described in Section 11.2.1. We shall refer to the descriptions where needed.

13.1.2.1 Gradient-Based Detectors

The edge detection may be based on evaluating the rate of change of intensities in the pixel neighborhood; this is given by the absolute value of gradient according to Equation 11.15. In the discrete image representation, the partial derivatives are approximated by directional differences (Equation 11.22 or 11.25), and consequently should be combined as

$$g_{i,k} = \sqrt{(\Delta_x f_{i,k})^2 + (\Delta_y f_{i,k})^2} \ ; \tag{13.5}$$

such an operator is approximately isotropic. As squaring and square-rooting are demanding operations, the discrete absolute gradient is usually further approximated by either of the following formulae,

$$g_{i,k} \approx \max(\Delta_x f_{i,k},\ \Delta_y f_{i,k}), \quad g_{i,k} \approx |\Delta_x f_{i,k}| + |\Delta_y f_{i,k}|. \tag{13.6}$$

Obviously, these operators are partly anisotropic: the first operator emphasizes horizontal and vertical edges, while the second operator emphasizes the inclined ones, in both cases by the factor of $\sqrt{2}$. The decision on whether the pixel belongs to an edge is given by comparison of the local gradient value with a chosen threshold, which is the only parameter of this edge detection approach, given the absolute gradient formula. Clearly, as for any binary decision, the threshold determines the compromise between the false positive and false negative responses; a higher threshold diminishes the spurious edges at the cost of possibly missing some existing ones. In the absence of any statistics, the optimal threshold for a particular image or image type should be chosen experimentally based on visual evaluation of the edge representation (Figure 13.2).

A similar result provides the *Roberts* operator, which calculates the maximum of oblique differences obtained by convolution with the masks

$$h_1 = \begin{bmatrix} 1 & 0 \\ 0 & -1 \end{bmatrix}, \qquad h_2 = \begin{bmatrix} 0 & 1 \\ -1 & 0 \end{bmatrix}. \tag{13.7}$$

The absolute gradient provides usually relatively thick edges that should be postprocessed by thinning.

As the partial differences are available, it is possible, as an addition, to determine the local direction of the edge (if the edge is locally detected) as

$$\theta_{i,k} = \arctan \frac{\Delta_y f_{i,k}}{\Delta_x f_{i,k}}. \tag{13.8}$$

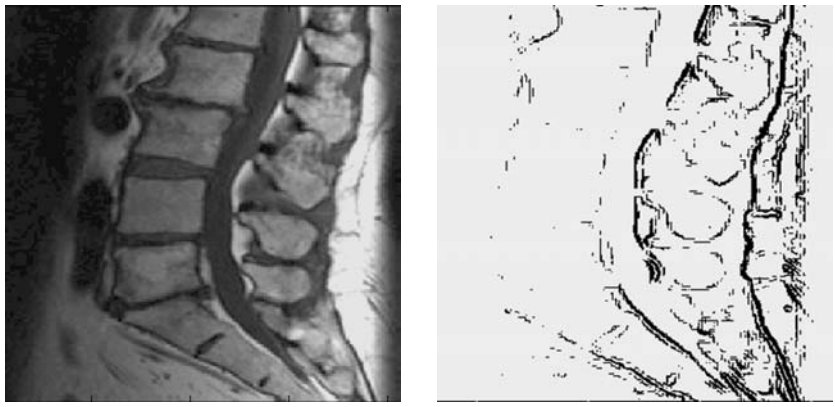


Figure 13.2 An image and its absolute gradient-based raw edge representation.

The demanding calculation of the arctan(…) function may be approximated by searching in a two-dimensional lookup table or more roughly by a conditionally calculated value. This way, the parametric edge image becomes vector-valued: for each positively detected edge pixel, the local direction of the edge is available, which may be advantageous on a higher level of analysis where edge consistency would be tested based on context.

Edge detection, including the rough edge direction estimation, may be alternatively performed by *compass detectors*. These are based on repeated convolution of the image with all eight masks of a set of *directional masks*, approximating directional derivatives by the weighted averages of differences, e.g.,

$$h_0 = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix}, \quad h_1 = \begin{bmatrix} 0 & 1 & 2 \\ -1 & 0 & 1 \\ -2 & -1 & 0 \end{bmatrix},$$
$$h_2 = \begin{bmatrix} -1 & 0 & 1 \\ -2 & 0 & 2 \\ -1 & 0 & 1 \end{bmatrix}, \quad h_3 = \begin{bmatrix} -2 & -1 & 0 \\ -1 & 0 & 1 \\ 0 & 1 & 2 \end{bmatrix}, \quad \text{etc.} \tag{13.9}$$

Each mask obviously provides a different parametric image $\{_j g_{i,k}\}$ that emphasizes a different edge orientation. Note that these four

masks are sufficient, as the remaining four are only negatives of the first group. The edge detection and direction estimation for each pixel (i, k) is then provided by finding (and consequently thresholding) the maximum,

$$g_{i,k} = (\max_j(|_j g_{i,k}|) \geq T), \quad \theta_{i,k} = j_{\max}\, 45° \tag{13.10}$$

where $j_{\max}$ is the order of the mask that gave the maximum absolute response and T is the chosen threshold. The masks in Equation 13.9 represent the *Sobel* operator; when all nonzero elements are ±1, the similarly acting operator is called *Prewitt*. A comparable function has the *Kirsch* compass operator

$$h_0 = \begin{bmatrix} 3 & 3 & 3 \\ 3 & 0 & 3 \\ -5 & -5 & -5 \end{bmatrix}, \quad h_1 = \begin{bmatrix} 3 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & -5 & 3 \end{bmatrix}, \quad h_2 = \begin{bmatrix} -5 & 3 & 3 \\ -5 & 0 & 3 \\ -5 & 3 & 3 \end{bmatrix}, \quad \text{etc.} \tag{13.11}$$

The advantage of compass masks is in removing the nonlinear operation of arctan(…) in direction estimation; however, it is at the cost of calculating eight* complete intermediate parametric images by convolution, instead of a mere two.

13.1.2.2 Laplacian-Based Zero-Crossing Detectors

As visible in Figure 11.11, the Laplacian (Equation 11.24 or 11.26) reacts to an edge by zero value between values of different signs. Therefore, instead of using the Laplacian itself (or absolute Laplacian) as the edge indicator, which would be imprecise and possibly ambiguous, the zero-crossings are searched in the Laplacian image. This approach provides thin and usually more precise edge positions than the gradient-based detectors, although the resulting edge representation tends to be noisier (Figure 13.3).

The search of zero-crossings requires a more detailed treatment. As it is unlikely that the zeros would be exactly at the sampling positions, looking for zeros in the Laplacian image is useless. Rather, locations where the neighboring pixels of important intensities have

*Or four, but with subsequent slightly more complex evaluation.

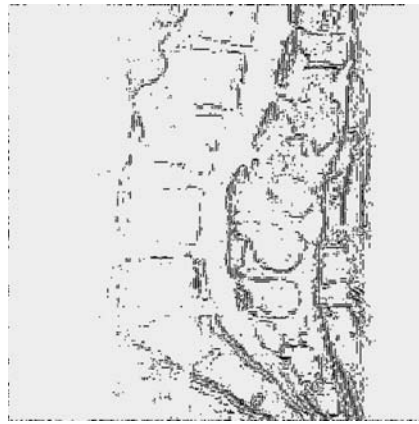


Figure 13.3 Raw edge representation based on detection of zero-crossings in the Laplacian of the image in Figure 13.2.

opposite signs are to be found. This can be performed by a nonlinear mask operation, using a 3×3 "cross" mask,

$$m = \begin{bmatrix} 0 & 1 & 0 \\ 1 & \times & 1 \\ 0 & 1 & 0 \end{bmatrix}, \tag{13.12}$$

which selects four closest neighbors of each analyzed pixel. The further operations are logical:

- if at least one of the neighbors has a different sign than others, and
- if the difference between the extreme differently signed values of neighbors is greater than a chosen threshold T, and
- if the central value at the cross position lies in between the extreme differently signed values of neighbors,

then the pixel (in a corresponding new matrix of the edge representation) is marked as an edge (e.g., 1). Once any of the conditions is not fulfilled, the pixel is abandoned as a nonedge (0 marked). The procedure again has a single threshold parameter.

This decision may be further checked by sufficiently high magnitude of gradient in the original image, should the gradient image be available. Another efficient, though perhaps demanding, test consists of checking the spatial extent of the central slope

between the neighboring extremes of the Laplacian image with respect to the expected sharpness of edges to be detected.

13.1.2.3 Laplacian-of-Gaussian-Based Detectors

The above Laplacian-based zero-crossings suffer with a high sensitivity to image noise, as the Laplacian itself is highly noise sensitive. It may be intuitively expected that the situation would improve when the image is preprocessed by a smoothing operator, such as convolution with a Gaussian mask h_G, so that the investigated image g becomes

$$g = \nabla^2(h_G * f) = (\nabla^2 h_G) * f. \tag{13.13}$$

The linearity of both cascaded operators allows joining of the Laplacian with the Gaussian, thus limiting the computation of g to a single convolution. The newly derived operator $\nabla^2 h_G$–*Laplacian of Gaussian* (LoG) may be expressed by its point-spread function (PSF). In the continuous version, the PSF of Gaussian is given by $h_G(x,y) = \exp((x^2+y^2)/2\sigma^2)$, so that, disregarding a multiplicative constant,

$$\nabla^2 h_G \propto (x^2+y^2-\sigma^2)\exp((x^2+y^2)/2\sigma^2). \tag{13.14}$$

This can be discretely approximated by a mask, the size of which depends on the chosen variance σ^2 of the Gaussian; the 5×5 mask can be shown to be

$$\begin{bmatrix} 0 & 0 & -1 & 0 & 0 \\ 0 & -1 & -2 & -1 & 0 \\ -1 & -2 & 16 & -2 & -1 \\ 0 & -1 & -2 & -1 & 0 \\ 0 & 0 & -1 & 0 & 0 \end{bmatrix}. \tag{13.15}$$

For higher σ^2, i.e., for greater mask operators, it is advantageous to utilize the separability of the operator (see [59]) and to process the columns and rows in sequence by the corresponding one-dimensional operators. The higher the variance, naturally the smoother is the resulting image g; however, a further inspection shows that though the position of a straight edge is not influenced by higher smoothing, the precision of edge positioning in corners suffers.

The zero-crossing detection in the image g may be organized equally as in the previous paragraph. An example can be seen in Figure 13.4.

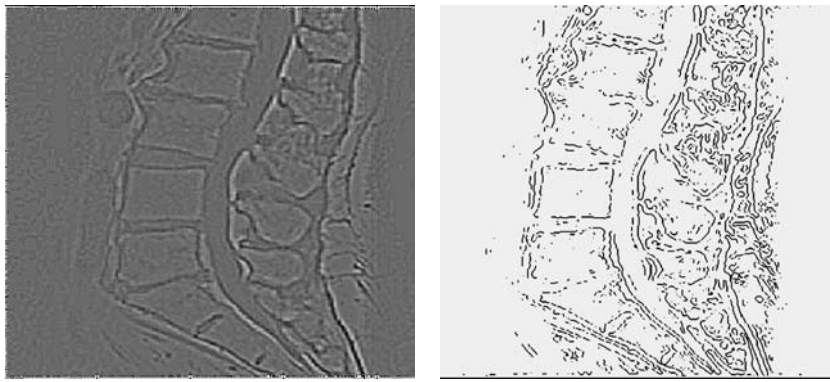


Figure 13.4 (Left) Laplacian of Gaussian of the image in Figure 13.2 and (right) the corresponding raw edge representation based on zero-crossings.

13.1.2.4 Other Approaches to Edge and Corner Detection

Some authors basically utilized the LoG approach and improved the procedure by adding some further features and conditions; details, which can be found in references, go beyond the frame of principles.

Canny's detector, which is often used as a rather reliable routine tool, is also based on the image smoothed by Gaussian. It improves the detection by taking directional derivatives and utilizing gradient orientation, introduces some additional constraints that improve the robustness with respect to noise, and conditionally adds weaker edges that do not reach the chosen limit (but are stronger than an auxiliary lower limit) only if they are connected with already detected stronger edges. This way, spurious edges are suppressed while the continuity of edges is improved. The detector may also introduce an intrinsic gradual change of scale via altering the variance of the used Gaussian and then combine the results achieved under different scales, thus improving the localization of edges without losing on robustness (see example in Figure 13.5).

The *area border–based approach* to edge finding consists of first delimiting an area in the image by some of the segmentation techniques and then determining its borders that may be considered edges. This could be substantially more robust than the local edge detection in noisy and textured situations; however, such methods rely on previous segmentation and therefore do not belong among the local feature determination.

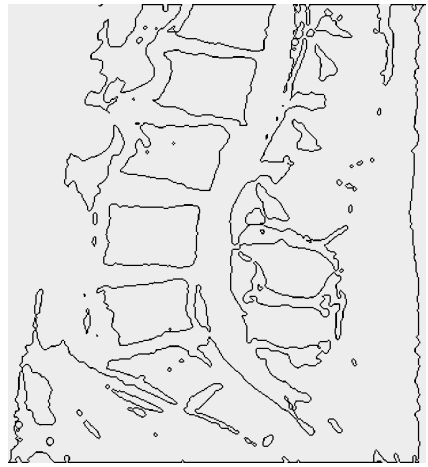


Figure 13.5 The edge representation of the above image as provided by Canny's detector.

In industrial applications, an important role belongs to *corner detectors* that enable the finding of image features in man-made scenes where rectangular and similar objects are common. In medical images, the corners may appear, e.g., at a branching or projected crossing of arteries. The corner detectors may be based on finding the edge meetings or crossings, utilizing the edges previously detected by the methods described above. Another, more reliable method follows the direction of the detected edge and detects the corners where the edge direction changes rapidly [17]. Alternatively, corner detectors may search for image locations where there is a high variance in the direction of an important gradient (i.e., with a high enough magnitude to prevent detection of noisy areas; in fact, this condition is equivalent to preliminary edge finding). These methods are in principle based on use of the gradient detectors providing both the magnitude and orientation; analysis of the orientation image with respect to its variance would then provide the indication of corner positions.

13.1.2.5 Line Detectors

Lines in the image may be defined as very narrow prolonged objects of the width of about one to two pixels. The edges of such objects cannot be well detected by common edge detectors, as there is not

sufficient area inside the object, which is expected to be constant by the edge detectors. Such lines can be detected by *ad hoc* designed convolutional mask operators, the basic idea of which is visible from the following examples:

$$\begin{bmatrix} -1 & 2 & -1 \\ -1 & 2 & -1 \\ -1 & 2 & -1 \end{bmatrix}, \quad \begin{bmatrix} -1 & -1 & -1 \\ 2 & 2 & 2 \\ -1 & -1 & -1 \end{bmatrix}, \quad \begin{bmatrix} -1 & -1 & 2 \\ -1 & 2 & -1 \\ 2 & -1 & -1 \end{bmatrix}, \quad \text{etc.} \tag{13.16}$$

The principle of the line detection is clear: when a mask is situated on the image so that the line is covered by the twos, the operator gives the extreme response (maximum or minimum, depending on whether the line is lighter or darker than the surroundings). The first mask obviously detects vertical line segments; the other masks, horizontal and inclined segments, respectively. The content of the remaining mask for perpendicular incline can easily be found. When the curved (broken) lines should be considered, the same principle leads to the masks sized 5×5, although segments only three pixels long are detected. There are thus still only nine nonzero elements, but they may be shifted along rows or columns in order to detect lines, parts of which are inclined by 45° or even 90°. Some of these masks are

$$\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & -1 & 2 & -1 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 & -1 & 0 \\ 0 & -1 & -1 & 2 & 0 \\ 0 & 2 & 2 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & -1 & 2 & -1 & 0 \\ -1 & 2 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \text{etc.} \tag{13.17}$$

The shape of the detected broken segment can easily be recognized by the layout of twos in the masks; finding the layout of other masks for differently shaped segments is straightforward. These masks for 45°-broken segments can also detect reasonably well the straight segments inclined by about 22° should they be slightly thicker than one pixel.

13.1.3 Texture Analysis

Textures are frequent in natural images, including medical ones, and consequently, they are well perceived and recognized by human observers; they serve as an important clue to recognizing borders and the character of image areas, types of surfaces or structures, shapes of objects, and sometimes also their spatial placement. In spite of it, it is difficult to define a texture, and a precise formal definition does not exist. In this situation, only rather vague descriptions are used.

Texture may thus be described as an image area where the variation of intensity has some characteristic features, perceived by humans (or by an analyzing algorithm) as uniform. The uniformity may even include slow changes in the texture appearance, e.g., on a curved surface due to differing perspective and illumination (see Section 13.1.3.7 on textural gradient); such variations would not preclude classifying the texture as constant if the *a priori* knowledge (experience) is embedded in the recognition.

A texture may be either approximately regular or even more or less periodical (then it is sometimes called *strong texture*), or it may be rather stochastic, with characteristic statistics but without expressed spatial regularity (*weak textures*). A texture may have a certain prevailing orientation (e.g., of edges or elongated cells) or, on the contrary, be essentially isotropic; it may be coarse or fine (in the chosen scale), be felt as rough or smooth, grained, cellular or amorphous, etc. Several examples of textures common in biomedical images are in Figure 13.6. Because of such a wide variety of different properties, of which only certain feature(s) may be important in a particular recognition task, many different approaches for texture classification have been suggested, tested, and combined.

Many textures are (or appear to be) formed of mutually similar little objects, usually consisting of small groups of pixels, sc., *primitives* (or *texels*); a texture may contain one or more kinds of primitives (e.g., fibers, grains, cells, etc.). The size of primitives then determines whether the texture is felt as fine or coarse. These can be arranged either randomly (weak textures) or more or less regularly (strong textures).

Obviously, texture analysis always concerns a certain area or even the whole image if it depicts a uniform texture (as, e.g., in tissue or material analysis). In such cases, it is advantageous to take as large an area as possible into account when determining the textural parameters because it allows a higher accuracy of the classification. However, the area may also be chosen small to allow the building of a parametric image of the local texture classifications

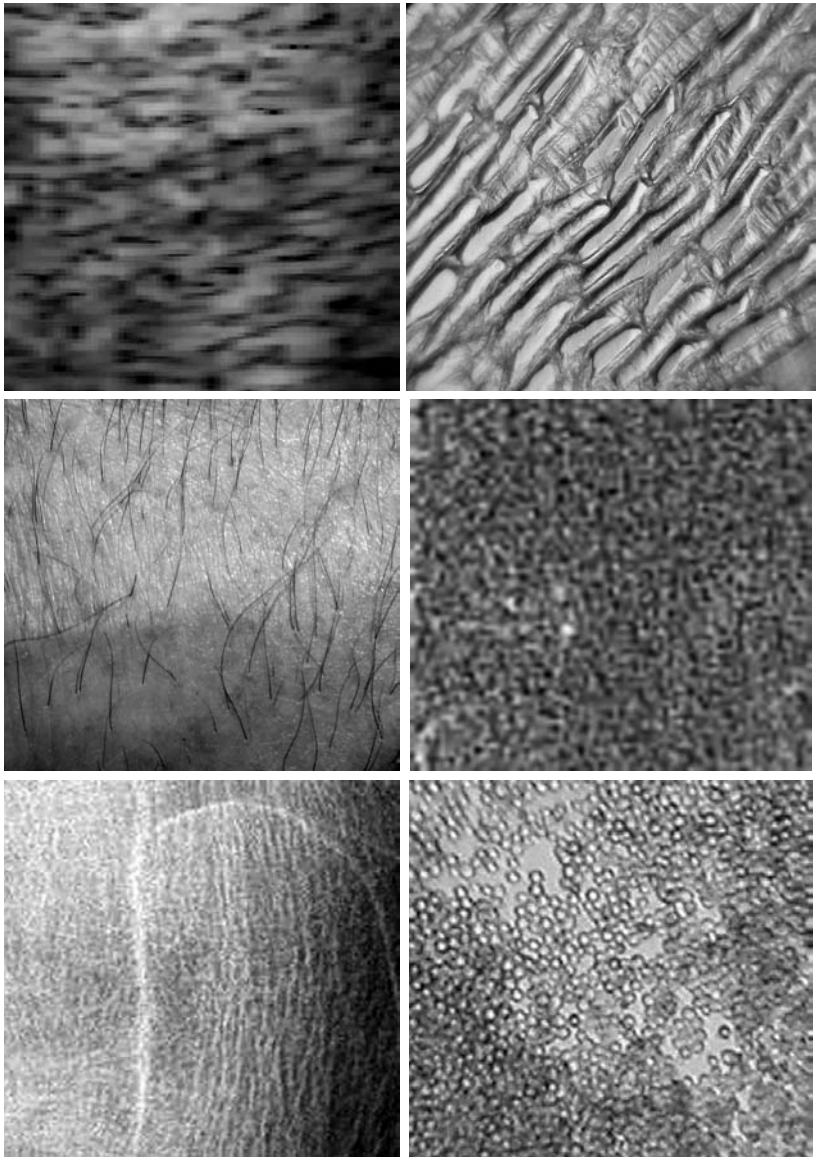


Figure 13.6 Typical textures in biomedical images (ultrasonic speckles, cell tissue, skin surface, MRI brain texture, x-ray bone structure, cell population; left to right, top to bottom).

and, consequently, to perform the image segmentation based on the textures. The choice of the analyzed area size is then a compromise between the spatial resolution and the accuracy of the texture classification. In this sense, the texture analysis also provides some local features and may thus provide (usually vector-valued) parametric images. A similar comment applies which has been mentioned in Section 13.1.1.4 on local spectral parameters: the density of samples of the parametric image should reasonably correspond to the size of the analyzed areas (neighborhoods of the pixel for which the texture-characterizing vector is calculated).

All the methods of texture analysis described below (except for the syntactic methods mentioned in Section 13.1.3.6) belong to *feature-based texture classification*. The following list of methods (and features that are mostly *ad hoc* designed) is by far not exhaustive; only those that illustrate a certain approach and are commonly used have been included.

13.1.3.1 Local Features as Texture Descriptors

Of the local parameters mentioned in the previous sections, the parameters of the *local first-order statistics*, including, e.g., local mean or variance, may be quite descriptive for certain types of textures, though the spatial relations in the local area are not included. Because many textures, namely the stronger ones, are characterized by important spatial relations among their primitives, higher-order statistics (describing probability distributions for couples or groups of somehow spatially distributed pixels) should be involved. One of them, the *frequency-domain description* (frequency-band power coefficients), has also been mentioned in the previous section; due to the Wiener–Khintchin theorem, the spectral-domain power description is closely related to the autocorrelation-based description (see below). Thus, both the locally determined first-order statistics and spectral parameters should be considered suitable means of texture analysis, besides the approaches discussed below.

13.1.3.2 Co-Occurrence Matrices

Co-occurrence matrices evaluate rates of repeated combinations of pixel intensities f, f' for pairs of pixels that are situated in a defined mutual position—at a distance Δr and direction φ (or with Cartesian differences $\Delta x, \Delta y$).

A simple yet precise definition of the co-occurrence matrix may be based on the concept of joint histograms (Equation 10.67,

Figure 10.10). Let us consider, in the analyzed image having q levels of gray, a neighborhood of a pixel (x, y) and a neighborhood of the same shape and size belonging to the pixel $(x + \Delta x, y + \Delta y)$. These two areas may be considered two equally sized images, for which the joint histogram may be constructed; this two-dimensional histogram, formed by a $q \times q$-sized matrix, is called the *co-occurrence matrix* for the shift parameter $(\Delta x, \Delta y)$. The matrix thus carries, in its elements, the counts of particular combinations of gray levels at corresponding pixels regardless of where the combinations are situated in the images.

A texture is described by a set of co-occurrence matrices for selected values of the shift parameter. The parameters are usually expressed in polar notation as absolute distance Δr and direction φ. Obviously, only discrete values of both parameter components are admissible, that is, are allowed by the discrete sampling grid of the image; mostly only four directions differing by 45° are utilized ($\varphi = 0, 45°, 90°$, and 135°), and correspondingly, Δr is an integer multiple of the sampling distance (vertically and horizontally), or $\sqrt{2}$ times greater in the inclined directions. Often, only eight neighbors are considered.

Six (or more) characteristics are usually derived from each of the co-occurrence matrices taken as normalized histograms containing approximate probabilities. The characteristics are, e.g., entropy, energy (sum of squares), mode (position of maximum probability) contrast $\sum_{f,f'}(f - f')^2 p_{f,f'}^2$, inverse difference moment $\sum_{f,f'} p_{f,f'}/(1+(f - f')^2)$, and correlation $(1/\sigma_x\sigma_y)\sum_{f,f'} f f' p_{f,f'} - \mu_x\mu_y$, where σ^2, μ are marginal variances and means, respectively. This way, many features are derived for each analyzed neighborhood of a pixel (x, y), which may be considered elements of a local feature vector; the vectors for all positions (x, y) then form the *textural parametric image*.

The co-occurrence matrices have proved successful in allowing the classification of different types of textures in numerous applications.

13.1.3.3 Run-Length Matrices

Runs are defined as continuous sequences of pixels of identical gray level in a certain direction (usually in the directions $\varphi = 0, 45°, 90°$, and 135°). The *run-length matrix* can also be explained in terms of the two-dimensional histogram: it is a two-dimensional histogram in dimensions of gray level vs. run length, containing the counts of appearances of runs of particular lengths and gray levels. For each

direction, a matrix is provided. Of these matrices, some suitable characteristics (short-run emphasis, long-run emphasis, gray-level nonuniformity, run-length nonuniformity, and run percentage) are calculated as elements of the local feature vectors. As may easily be seen, the run-length approach enables distinction between texture with long and short primitives, and also discriminates among different prevailing orientations of the primitives.

13.1.3.4 Autocorrelation Evaluators

The (discrete) *local autocorrelation function* (see Section 2.4.2),

$$_{x,y}R_{ff}(m,n) = \sum_i \sum_k f_{i,k}\, f_{i-m,k-n} \tag{13.18}$$

calculated on a small area around the position (x, y) can also reveal spatial properties of the analyzed texture. Basically, the coarseness influences the width of the main lobe of the function—the bigger are the primitives, the wider is the main lobe. Besides that, longer-distance relations that may resemble some periodicities are well revealed by periodical components of the correlation function visible outside of the main lobe. The information contained in the autocorrelation function is equivalent to the power spectrum derived from the same area.

However, the complete function is too large a texture descriptor, which would be difficult to interpret. In order to decrease the dimension of the local feature vector, some integrating descriptors derived from the autocorrelation function are used, construed again in the form of moments of the first, second, or even higher and mixed orders, which in turn serve as elements of the feature vector.

13.1.3.5 Texture Models

A different approach to texture analysis is based on the notion of modeling the texture generation. A simple idea of the *texture model* is depicted in Figure 13.7. Two-dimensional homogeneous white noise (i.e., an ensemble of independent stochastic variables of individual pixel intensities) is input to a linear or nonlinear two-dimensional system (filter) that produces the two-dimensional texture. Depending on the character and parameters of the filter, different textures may be obtained. Besides the filter properties, the texture is also influenced by the amplitude probability distribution of the input noise. Usually, it is supposed that the input white noise is Gaussian,

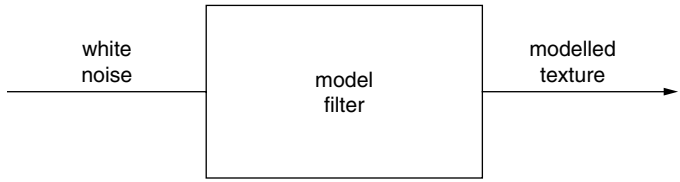


Figure 13.7 Model of texture generation.

and that the filter is linear. Often, autoregressive models are used because, in this case, the inverse filter needed in texture parameter estimation is purely nonrecursive (two-dimensional finite impulse response (FIR) type).

The methodology of describing the texture properties (for a linear model) is based on the notion of a *whitening filter*, which is a system converting the intrinsically correlated texture image back to a white noise image. Obviously, the whitening filter must have a transfer function that is the inversion of the model filter transfer. As it is shown in signal theory, the whitening filter can be derived based on the autocorrelation function of the received image, i.e., from the measured image of the texture. Once we have the inverse filter, its parameters (directly related to the model filter parameters) may be used to characterize the texture. Besides that, the whitening filter, when applied to the textured image, provides an estimate of the (decorrelated) input noise. It is then possible to analyze the amplitude distribution of the noise that also influences the texture, by means of constructing the histogram of the whitening filter output (Figure 13.8). This way the description of the texture

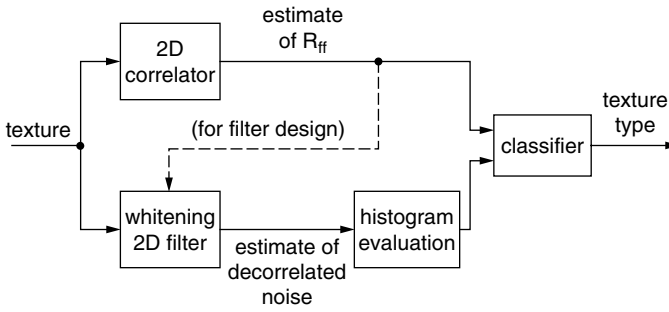


Figure 13.8 Model-based texture identification.

by autocorrelation-based features, as well as the features derived from the histogram of the recovered noise, are available; the additional information may improve the texture classification reliability.

A similar approach is based on Markov chains as texture-generating models. These are two-dimensional stochastic processes generating a new pixel value randomly with a probability distribution influenced by one or several previously determined pixel values of the generated image. It can be shown that for Gaussian distribution, the previous autorecursive model is a special case of the Markov model, so that the Markov approach is more generic. Again, the parameters of the Markov model may be used as textural features in its classification. However, it is not easy to design the model for a particular type of texture: the proper structure and order of the model and the probabilistic characteristics must be determined based on the received texture. Interested readers are referred to, e.g., [15].

13.1.3.6 Syntactic Texture Analysis

When the primitives of the texture are well identifiable and spatially clearly organized, it is possible to describe the texture alternatively by listing the primitive types and describing the spatial arrangement of the primitives by means of mathematical linguistics, i.e., by a suitable grammar. This interesting approach—*syntactic texture analysis*—requires first identifying all possible types of primitives appearing in the analyzed image: the primitives must be described using some suitably chosen features, the classes of primitives defined, and the primitives classified. As the next preparatory step, the grammars describing all possible spatial arrangement of primitives (and allowing not only for regular structures, but also for all possible irregularities) are to be found and defined, thus establishing classes of spatial structures. Each of the grammars is expressed either as a (two-dimensional or three-dimensional) graph or as a regular syntactic expression, usually recursive to include all the possible hierarchies, insertions, and side or error structures in the spatial structure. Finding the generic syntactic description for a concrete texture is the most difficult step of this type of texture analysis. Analyzing a concrete image area then means first classifying the primitives and then determining the grammar, which best describes the spatial structure formed by the primitives; the types of primitives, together with the optimal syntax of the structure, then determine the classification of the texture.

Though interesting, this approach is rather demanding, and it seems that after initial enthusiasm, it is now used only occasionally, in specialized cases. More detailed discussion is beyond the scope of this book; relevant information and further references can be found in books on computer vision, e.g., [17], [72].

13.1.3.7 Textural Parametric Images and Textural Gradient

Before concluding the section on textures, it should be stressed once more that textural parameters derived using small neighborhoods are just some of many possible local features of the image that may be derived for a particular position in the image. This way, the local texture parameter vectors may be used as the values of a parametric image, obtaining a textural image, belonging to the original image. This may be utilized in segmentation, the naturally following phase of image analysis.

As we shall see, a segmented image area, determined by equally classified texture, should have basically identical texture vectors on the whole extent of an area. However, slow and smooth variations of the texture vectors need not mean a change of the corresponding object or surface, but rather only the consequences of differing illumination or perspective distortion. Similarly, as with the intensity images, the parametric images have defined their gradients by the differences between corresponding elements of feature vectors in neighboring pixels. (This means that there are so many generally differing gradients, such as the dimensionality of feature vectors; they may, in turn, form a corresponding higher-dimension vector-valued gradient image.) The continuity and low amplitude of such gradients (here the *textural gradients*) then may mean continuity of the area to be considered in the following segmentation.

On the other hand, the texture analysis may also be applied to an already segmented area (or even whole image) of a supposedly single texture; the goal is then to classify the texture as precisely and reliably as possible, without a need for determining its extent. In this case, there is no constraint on the size of the analyzed area dictated by the requirement of high spatial resolution of the textural parametric image. The analyzed area should then be as large as possible in order to offer more information with less variance to the classification procedure. In medical imaging, this situation is met typically when tissue analysis is performed.

13.2 IMAGE SEGMENTATION

Image segmentation is the fundamental step in image analysis. It should delimit the image areas representing different objects, e.g., organs, bones, different tissue types, vasculature, etc. The word *segmentation* has two meanings: the procedure leading to the segmented image, and the result of such a procedure. Formally, a *concrete segmentation* can be described as splitting the complete image area R into a set of partial areas (segments) $\{R_1, R_2, \ldots, R_s\}$ that are disjoint, $R_i \cap R_j = \emptyset$, $i \neq j$, and cover the image area completely, $R = \bigcup_{i=1}^{s} R_i$. There is no unified approach to achieve the best segmentation, nor is the optimum segmentation well defined in most practical cases. This is the consequence of the large variability of image analysis tasks.

Consequently, many different segmentation methods and approaches have been applied in different situations, mostly *ad hoc* formulated, many with intuitive steps and heuristics.* In spite of the enormous amount of independently invented and applied procedures that have been published, certain common features enable the classification of the segmentation methodology into several classes, the principles of which are described in the following sections.

Segmentation is usually an iterative process progressively improving the resulting segmentation so that it gradually corresponds better to a certain *a priori* image interpretation. In the course of such iterative procedure, a number of different approaches may be applied in order to utilize as much information concerning the segmentation as possible. Some of the methods are applied globally, to the complete image; others are better applied only locally, in predetermined image areas where they serve to elaborate local segmentation.

The methods will be mostly presented only in their two-dimensional version in order to simplify the explanation. Nevertheless, most of the described approaches can (and have been) applied also to segmentation of three-dimensional image data; the generalization is usually straightforward.

13.2.1 Parametric Image–Based Segmentation

The methods of parameter-based segmentation are based on the intuitive idea of *homogeneity of areas*. An image segment (an area delimited by the segmentation process) is supposed to be homogeneous

*Heuristics is a technique that lacks formal substantiation; the algorithm or procedure is usually based on intuition or experience.

with respect to the parameter by which it is characterized. In the simplest case, the parameter (feature) may be the intensity (brightness or gray shade) of the area pixels; however, the parameter may be generally described by a complex vector of a high dimensionality, as in the case of texture-based segmentation. Thus, in the scalar parameter case, the parameter value should remain inside a certain tolerance interval on the whole area; similarly, the vector parameter of an image segment should belong to a certain spatial area in the corresponding feature space, which need not be rectangular; i.e., the ranges of individual features forming the parameter vectors may be mutually dependent.

The segmentation based on the assumption of homogeneity of segments naturally has its limitations—it would not work when there are changes in the parameters due to varying illumination, field inhomogeneity of the image sensor, or properties of the imaged objects (e.g., variations of tissue properties inside an organ that should be considered a single segment). In these cases, region-based segmentation that allows the toleration of such spatial variations may be more suitable.

13.2.1.1 Intensity-Based Segmentation

Segmentation based on intensity simply sets the upper and lower limits l_u, l_d (thresholds) of pixel intensity for each class of objects (or an object); thus, a segment is automatically homogeneous in the sense that all its pixels have intensities in the given range. Obviously, such an approach is only justified when a particular gray level uniquely defines the type of object; this may be, more or less, the situation of imaging the attenuation coefficient in the X-ray CT modality, where different types of tissues are marked by characteristic attenuation (though not quite uniquely; therefore, the segmentation should be checked or modified using other criteria as well).

More classes can be defined at a time; as every pixel should be sorted to a single class, the ranges of individual classes must be disjoint, though often immediately adjacent on the intensity axis. The segmented image areas may be marked by different colors, with the same color used only for areas of the same class, though perhaps not connected. Alternatively, the individual classes may be distinguished by different grades of gray in a gray-scale image (only a few classes can then be recognized easily on a display). The conversion from the original gray-scale image into the segmented one is straightforward: a simple stair-like contrast transform (Section 11.1.2) expressed by a lookup table would do.

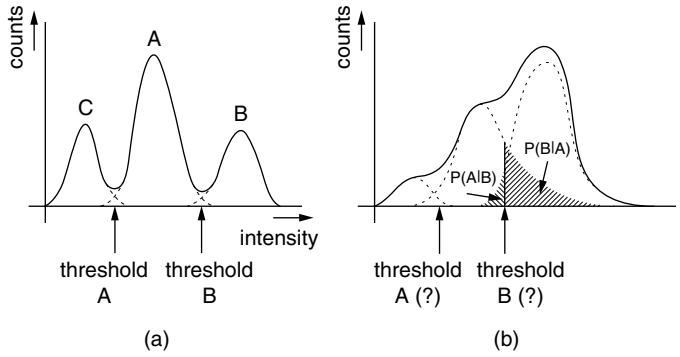


Figure 13.9 Three object classes defined by different pixel-intensity clusters: (a) a well-defined case and (b) a weakly defined case.

However, the real problem in this type of segmentation is how to adjust the limits of intensity for individual classes. This can be done either interactively, with the feedback via an immediately displayed result of segmentation, or algorithmically and more formally, based on the gray-scale histogram. When the classes are well separated on the intensity scale, as in the trimodal histogram of Figure 13.9a, the limits should be simply situated in the minima between the classes and a low rate of false classification of pixels can be expected.

If, on the contrary, the modes of individual classes overlap substantially (Figure 13.9b, dashed curves), a higher error rate must be expected and admitted. The position of an intensity limit that serves here as a one-dimensional separator between classes with neighboring intensities obviously influences the error rate, as can be seen in detail in Figure 13.9b. Here, the probabilities of pixel intensities in individual classes are sketched and, being summed, form approximately the resulting histogram envelope. Obviously, the shaded area under the curve A to the right of the threshold B means the total probability of A pixels being erroneously classified as B; the other shaded area under the B curve to the left of the limit similarly means the probability of the other error, i.e.,

$$P(B|A) = \int_{l_{dB}}^{l_{uB}} p(A)\mathrm{d}f, \quad P(A|B) = \int_{l_{dA}}^{l_{uA}=l_{dB}} p(B)\mathrm{d}f. \qquad (13.19)$$

The optimum may be defined at the minimum total probability of error, or one type of error may be more acceptable than the other (e.g., a non-negligible probability of classifying a pathological tissue

as healthy may not be acceptable). Nevertheless, the information on the class distributions is usually not available in practice, and some compromise or trial-and-error approach to the threshold determination must be taken. When assigned algorithmically, the limits are usually set to the minima on the histogram, which obviously need not be the optimum solution, namely, when the variances and/or total counts of neighboring classes differ substantially, even if both are basically Gaussian. The situation is even more complicated when both distributions have different characters. Setting of the limits belongs among problems of decision making that are studied in artificial intelligence or, more classically, in communication detection theory; however, these approaches can hardly be used in image segmentation due to the lack of the necessary statistical information. Two examples, each segmenting a class of areas defined by a couple of pixel intensity limits in an X-CT image, are shown in Figure 13.10; the two ranges for both segmented images are disjoint.

The simplest case of intensity-based segmentation is *thresholding*, where only a single limit separating two classes is defined. This way, the original gray-scale image is converted into a binary one. Though often used in computer vision, e.g., for recognizing objects against a darker or lighter background, it finds only limited use in medical image analysis (e.g., when analyzing microscopic images of a tissue or blood). An interesting, but in medical imaging hardly applicable, method of the single-threshold determination is *percentual thresholding*. When the percentage of black in the image is *a priori* known, as, e.g., in a textual image with a certain font, it is possible to adjust the threshold using the histogram so that the normalized sum of histogram counts to the left from the threshold gives the desired deal of all pixels.

The choice of the limits (thresholds) based on the global histogram may be unsuitable for the whole image. Similarly, as described in Section 11.1 for adaptive contrast adjustment, the *adaptive segmentation* may be based on regional histograms derived from some subimage areas, thus respecting, e.g., varying illumination. The thresholds must vary smoothly on the total image area; therefore, they should be interpolated from the available locally derived values.

The intensity-based segmentation is naturally suitable only when the object intensities are smooth and almost constant inside the segments; otherwise, it fails or at least produces rough and imprecise borders with many erratically classified areas. This would obviously be the case of most ultrasonographic or gammagraphic images due to a high level of inherent texture or noise.

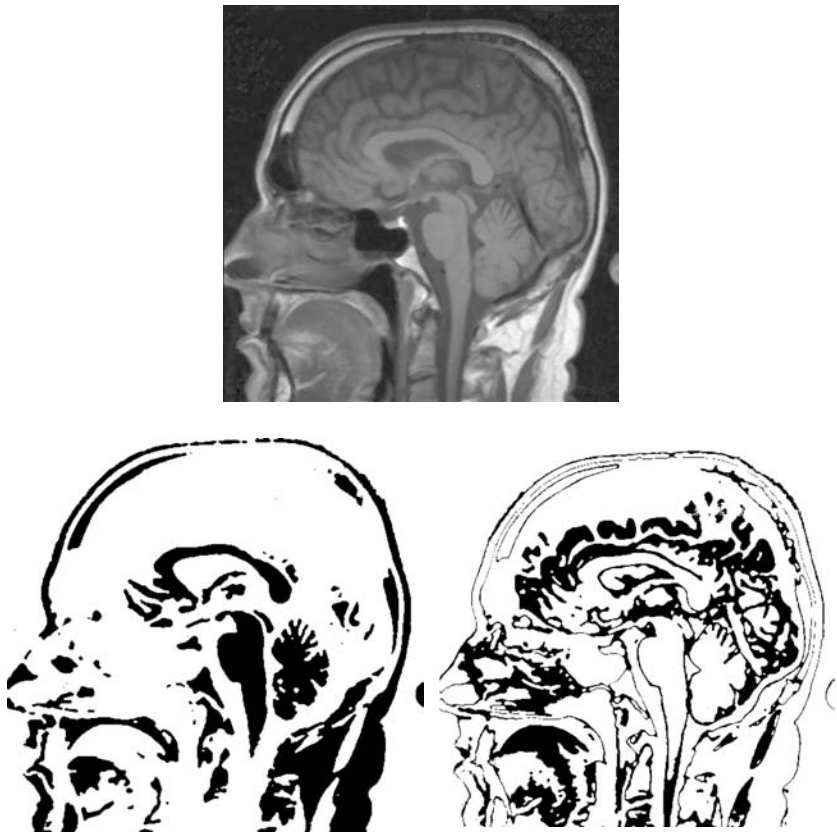


Figure 13.10 Segmentation of an X-CT image with two differently defined disjoint ranges of pixel intensities (upper, original image).

13.2.1.2 Segmentation of Vector-Valued Parametric, Color, or Multimodal Images

The pixels of color images or fused multimodal images are vector-valued. The previous one-dimensional intensity-based segmentation can then be generalized using joint multidimensional histograms. It can most easily be imagined when the vectors are two-dimensional, as in case of fused bimodal images: the standard two-dimensional joint histogram (as in Equation 10.67) can then be used to visualize the classes in the two-dimensional intensity scale. The individual modes in the histogram (two-dimensional clusters of high counts,

visible as peaks in three-dimensional histogram representation, or as high-intensity areas in two-dimensional gray-scale representation) would be mutually separated by suitably chosen curvilinear borders. The borders may again follow the bottoms of the respective valleys (in analogy to minima in the previous one-dimensional case) or be designed more formally, when the probability distributions of individual classes are available (which is rather seldom; however, knowledge on typical situations in medical imaging is gradually gained and utilized). Providing the marked segmented image is then in principle identical as in the one-dimensional case, only the contrast-converting (or false color-generating) lookup table is represented by a two-dimensional matrix, and thus has a two-dimensional input.

This segmentation method relies again on the homogeneity of the segments; the requirements are now more stringent: both components of the parametric vector are taken into account and the vector must belong to a cluster in the histogram. This means that cases when a component is constant on a segment, while the other would have different values, are excluded; this can allow a better discrimination of classes. However, when a certain variability inside a class should be expected, as may be the case when X-CT and magnetic resonance (MR) images have been fused, more sophisticated segmentation methods must be applied.

Generalization of the above approach to a multidimensional case of more complex parameter vectors is straightforward, though perhaps not as easy to imagine: a joint multidimensional histogram replaces simply the two-dimensional one. Obviously, the histogram may be interpreted also as a multidimensional discrete space in which the individual axes correspond to individual components of the parameter vectors and the content of individual spatial cells in the parametric space is the absolute or normalized count of the corresponding parameter vectors in the image. Then, finding the proper border hyperplanes or hypersurfaces may be interpreted as the classical classification problem based on cluster analysis (here simply realized by constructing the multidimensional histogram).

13.2.1.3 Texture-Based Segmentation

The texture description of an image, as far as it is done with an appreciable spatial resolution, is nothing other than a parametric image with a multidimensional vector-valued parameter; the texture-based segmentation is done basically by the methods of the previous paragraph. Naturally, the confusing influence of border areas, where

two or more textures enter the analyzed neighborhoods, thus generating nonclassifiable vectors, and the limited spatial resolution should be taken into account.

13.2.2 Region-Based Segmentation

Segmentation based on the notion of regions—image areas of certain common property—also utilizes the concept of homogeneity, though it may be applied more locally, thus providing a higher flexibility in a segment definition. All these methods may be explained based on the concept of parametric images. We shall use this approach, although the parametric image values need not necessarily be precomputed; often, they are computed during the segmentation process (or they are *a priori* available, as when the parameter is simply the pixel intensity).

13.2.2.1 Segmentation via Region Growing

Region growing constitutes the conceptually simplest of the region-based methods and probably also one of the historically oldest approaches to segmentation. Its principle is quite simple: a *seed* is determined in each potential region as a pixel that belongs to that region and has, in a defined sense, properties typical for the region, usually given by a parameter p like intensity, local mean, local variance, etc. Other surrounding pixels will then be aggregated to the region if they fulfill a certain criterion of homogeneity. Determining the seeds is a separate task, often performed interactively, or it may be stochastic, or even based on some preliminary analysis.

In its simplest form, the following iterative procedure is then executed: For each pixel already recognized as a region member (initially for the seed), check all pixels in its neighborhood (4- or 8-connected, Figure 13.20) and compare the parameters p_j of these candidates with that of the seed, p_s. If the condition

$$|p_s - p_j| \leq T \tag{13.20}$$

for a particular neighboring pixel is fulfilled, add the pixel as a new member to the region; otherwise, mark it as unacceptable.

The algorithm ends when there are no pixels to be added, i.e., when, for all pixels of the region, their neighbors are marked as either belonging to this region, unacceptable, or belonging to other already established regions. The result of such a region growing is obviously identical with that of segmentation based on the parametric image

of p with the tolerance for the segment given by Equation 13.20, should the ranges for different classes be disjoint. The advantage is in the possibility to choose the seed manually; otherwise, the parametric image–based algorithm is more effective. Obviously, when the parameter ranges of neighboring segments overlap, the borders will depend on the sequence in which the segments are treated.

The situation changes dramatically when the criterion is modified to

$$|p_i - p_j| \leq T, \tag{13.21}$$

where p_i is the parameter value of the already accepted pixel central to the tested neighborhood. Then, a (slow) increase or decrease of the parameter may be followed inside the region, as the reference value p_i is allowed to change gradually. The growing in a particular direction stops when there is an abrupt change indicating a high gradient, possibly due to a border edge. This way, not only homogeneous regions in the sense of invariability of the parameter are allowed; the homogeneity notion is now generalized in the sense of accepting regions with slowly and smoothly varying parameter(s) (e.g., the smoothly curved and unevenly illuminated surfaces). This may substantially improve the segmentation in certain cases.

On the other hand, the resulting segment shape and size then depends not only on the position of the seed, but also on the order in which the pixels are processed. An edge impenetrable after a certain path may become weak enough when arriving to it from another side via a path leading to a different value of p_i, and vice versa.

13.2.2.2 Segmentation via Region Merging

Region-merging segmentation starts with some small elementary areas that may be considered homogeneous regions—possibly and most naturally with individual pixels. Two adjacent regions may be merged when a certain criterion of homogeneity is fulfilled. The process of merging then continues until all regions are surrounded only by regions with markedly different criterion so that no further merging is possible.

The criterion of homogeneity may again be formulated statically, similarly as in Equation 13.21, i.e., by a fixed range for a certain parameter of all pixels of the region to be formed by merging,

$$p_{i,m,n} \in \langle p_0 - \Delta p, p_0 + \Delta p \rangle, \tag{13.22}$$

where p_0 is the (mean) parameter value of the initial part of the region relaxed by a chosen tolerance interval Δp. The results are then again identical with those described in Section 13.2.1, should the intervals be disjoint. A more flexible definition of homogenity is the dynamic definition based on the somehow defined *similarity* of the regions to be merged, e.g., on the absolute difference of mean values of the criterion parameter for both areas. As the mean value is influenced by gradually aggregated areas, its value changes in the course of the procedure, which may allow the merging of more differing regions; consequently, larger and more meaningful regions may be achieved in comparison with the static criterion.

A similar (and even more) dynamic region merging may be based on the notion of the *border strength* between regions considered for merging. This method visualizes the local differences between the border pixels on both sides as elementary boundaries ("crack edges")—the *strength* $s_{i,j}$ of the *local boundary* separating the regions R_i and R_j at a particular pixel pair is given by the absolute difference between the parameter values of these adjacent pixels belonging to the two classes. The crack boundary is considered weak if $s_{i,j} \leq T_1$; otherwise, it is strong. Using this concept, the integral strength $S_{i,j}$ of the common border consisting of N pixels can be defined as the share of strong-boundary pixels. The border is taken as weak when

$$S_{i,j} = \frac{N_s}{N} \leq T_2, \tag{13.23}$$

where N_s is the number of the strong crack edges on the common border. When a border is found weak, it is removed (melted) and both regions merge. It is clear that the static homogeneity condition (Equation 13.22) need not be preserved, which may be advantageous for the reasons explained above.

The resulting segmentation depends on the choice of the elementary regions and on the order in which the regions are processed. Obviously, the regions treated first have a higher chance to aggregate other regions; once a region has been merged with another, it is marked as such and cannot be merged to any region processed later unless this region merges to a complete previously formed one.

13.2.2.3 Segmentation via Region Splitting and Merging

Splitting regions is in a sense a complementary procedure to region merging. The basic idea of region splitting is that a region that is

not uniform (homogeneous) in the sense of a chosen criterion must be split into smaller regions; this is repeated with gradually smaller regions until every region is uniform. Naturally, the already uniform regions would not be split. However, the splitting itself does not provide a satisfactory segmentation, as adjacent regions may remain split even if they fulfill the criterion for merging. Thus, the split-and-merge approach is usually used in segmentation. It should be mentioned that the segmentation results of both methods, region merging and split-and-merge, might differ substantially.

The most straightforward way of splitting consists of subdividing the image into quarters (three-dimensional data into eight equally sized spatial segments), which is recursively repeated on as many levels as needed to achieve the uniformity inside regions. It can be easily seen that for a square image of $N \times N$ pixels, at maximum $\text{int}(\log_2 N + 1)$ levels are needed to achieve the absolutely uniform areas formed by individual pixels. On each level, the uniformity of a region is evaluated; should it be insufficient, the region is split further into quarters, thus entering a new level. When the region is homogeneous according to the criterion, no further splitting is performed. This way, the hierarchical *quadtree representation* of the image is obtained: the image can be described by a tree graph with either four or zero down-leading branches originating in each node. The supreme (root) node corresponds to the complete image, a node on the immediately lower level to its quarter regions, etc. Each node thus represents the image area whose size and location follow from the node position in the graph; the node is marked by the mean value of the parameter for this area. When the region is not homogeneous, four additional branches lead from this parent node to the next daughter level. Note that such an economic image description—a *pyramid representation*—may also be used as a simple compression method of image data. An example of such splitting is in Figure 13.11.

Though the image is subdivided into uniform regions by the above-described region-splitting algorithm, the result is usually not an acceptable segmentation, as there may be neighboring regions fulfilling the homogeneity criterion that are not merged. This may be partly improved already during the split phase: when a region is split, some of the adjacent daughter quarters may fulfill the criterion of homogeneity and therefore be immediately merged, this way lowering the number of descendants. Such merging steps may alternate with splitting and can be embedded into the split algorithm. However, the adjacent areas that are not descendants of the same parent node remain split even if they fulfill the criterion of

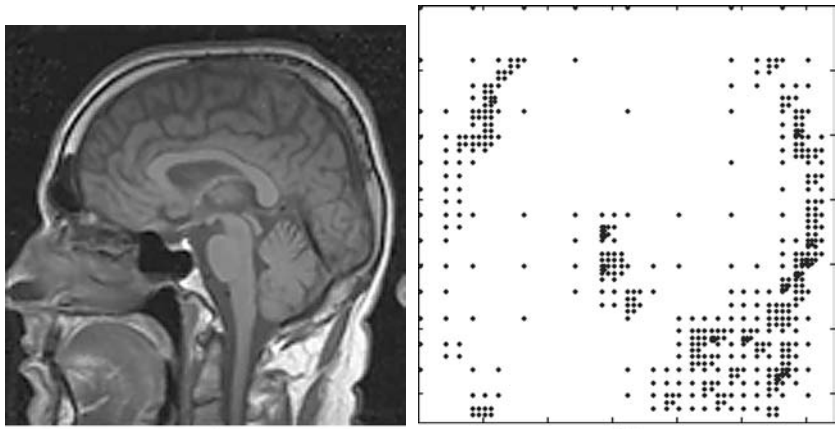


Figure 13.11 Example of region splitting (without merging) applied to a natural image: (a) the input image and (b) the found left upper corners of square areas homogeneous in the sense of a chosen tolerance.

homogeneity, as it is not possible to merge across the daughters of different parents during the split phase.

Therefore, when the split phase is over, the final segmented image should be submitted to the region-merging algorithm as described in the previous section. The splitting phase then serves as preparation of a good initial set of regions to be considered for merging. Obviously, this approach would work equally well when the local merging steps in the splitting phase are omitted; only the number of the initial regions for merging would be higher. The region split-and-merge-based segmentation is rather robust, though it naturally tends to cornered (i.e., piece-wise rectangular) regions (see example in Figure 13.12).

13.2.2.4 Watershed-Based Segmentation

Though it belongs to region-based methods, the background idea of *watershed segmentation* differs from the above homogeneity concept (even from the generalized nonzero gradient homogeneity concept). In a sense, it may be understood as a border case between homogeneity-based and edge-based segmentation.

The watershed method takes, as segments, the areas formed by defined neighborhoods of each local minimum of the image intensities (or another parameter). The name of the method comes from the similarity of the surface description of image intensities, as can be seen

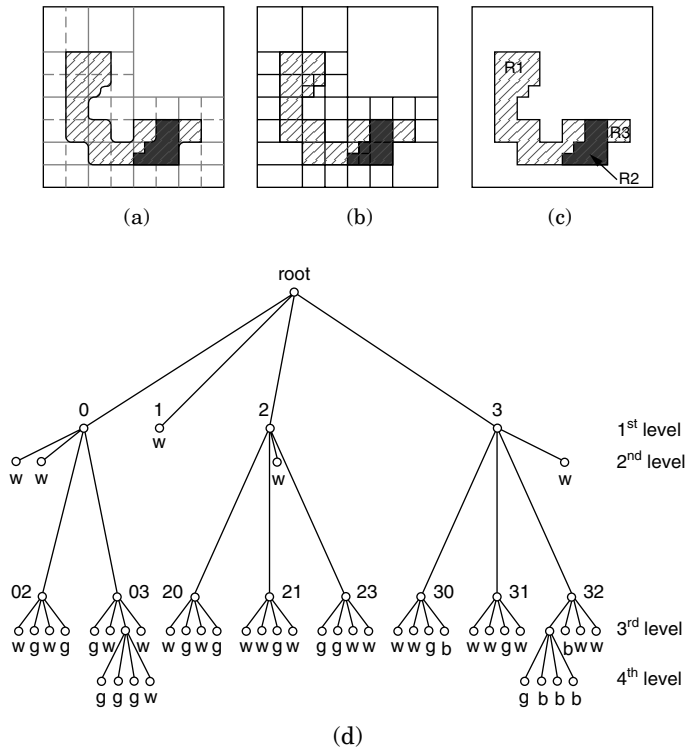


Figure 13.12 Split-and-merge segmentation: (a) original image, (b) split regions, (c) result of final merging, and (d) tree representation of the split image.

in Figure 1.1b, with topographical altitudinal models. Certain local minima can always be identified in such a surface (including those at the image border), which are then marked uniquely as the representatives of the individual segments. The region (segment) belonging to a minimum is defined as the locus of points such that a drop of water falling to a point of the region would flow along the steepest descent path to the location of the minimum. The region, usually called a *catchment basin*, is uniquely determined by the definition, should the surface be continuous.

The definition formulates a hypothesis that the catchment basins correspond to meaningful image regions; this is an alternative to the above-mentioned homogeneity assumption. Whether this hypothesis is valid depends on the character of the image and of the imaged object or scene. Often, the watershed segmentation is applied to a parameter image obtained by application of the absolute

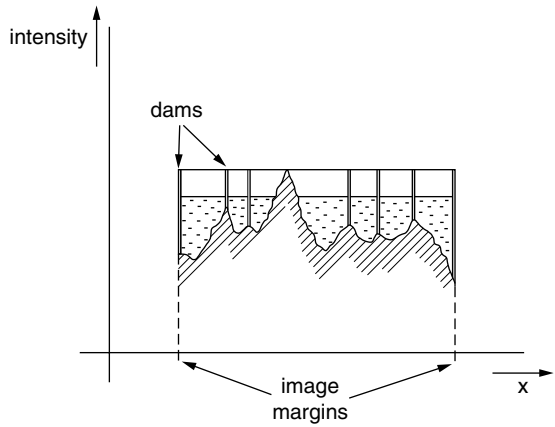


Figure 13.13 Profile of the partly flooded intensity surface with the mental dams.

gradient operator; it is obvious that in this case, the crest lines forming the watersheds would mostly correspond to the object borders, and the result might be similar to segmentation based on thresholded edge representation (Section 13.2.3). The advantage of the watershed segmentation is, as might be expected, its substantially higher robustness with respect to noise.

The above definition determines the area of each basin uniquely if the surface were continuous; in the discrete environment, the path of the drop need not be unique if the neighboring pixel values are identical. In addition, determining the downstream paths from all image points in order to determine to which minimum each concrete point belongs is computationally very intensive. An equivalent definition of the regions has therefore been formulated, unique even in discrete formulation and substantially easier computationally. The definition is based on the mental *flooding simulation*: if at each local minimum location a small hole was punched to the surface and the complete surface was slowly immersed in water, the water would fill the basins gradually. So that the water could reach the highest altitudes on the ridges while the basins remain separated, vertical dams must be erected on the ridgelines; these lines then determine the segment borders (Figure 13.13).

The computational realization of the flooding is in principle simple: the first step is to sort all image pixel values (with the spatial coordinates attached) in ascending order. In the following

phase, the image matrix is gradually filled with labels* (see below), while the ordered list of pixels is emptied starting from the lowest minimum. The basins on the surface are thus gradually filled from the bottoms, while the concerned pixels are marked with the labels of the corresponding minima: at a certain instant, the level of water reaches the k-th level of altitude (i.e., of pixel intensity) and all pixels processed so far are appropriately marked. To add all the pixels of the $(k+1)$-th level, each such pixel is assigned the label of an already marked pixel from its immediate neighborhood if there is such; should there be two differently labeled pixels in the neighborhood, the pixel under discussion is obviously the border pixel (a part of the dam support).

Some of the pixels will not find any marked neighbors; this occurs for two reasons. The pixel either really cannot be connected with an existing basin, and then it should be marked with a new label as a representative minimum of a new basin, or there are more pixels of the $(k+1)$-th level in the neighborhood, so far unmarked, that separate the pixel from the marked lower-level pixels. The connectivity test must be performed on the set(s) of $(k+1)$-th level pixels; the subsets that are disconnected from all existing basins are of the first kind, the connected ones of the second. In the latter case, when a subset is connected only to a single basin, all its pixels will be marked correspondingly. If, however, the subset touches more than one existing basin, the pixels may be marked according to the closer representative. This provides a unique border even if the intensity itself does not; naturally, the border shape is somewhat arbitrary, as there is no exact clue on how to situate it. A concrete example is shown in Figure 13.14.

13.2.3 Edge-Based Segmentation

Another approach to segmentation is based on edge representation of the segmented image; this type of segmentation aims at providing the borders of image segments. When the boundaries are determined properly, they are formed by closed curves and the interior of such a curve is then the sought area of a segment. In a sense, this approach also utilizes the idea of homogeneity of segmented regions; however, the requirements are much weaker: the edge is detected at locations where there is an abrupt change in the image properties, thus indicating a probable region border.

*The labels are replacing the pixel intensities or are being put in an auxiliary matrix of the image size.

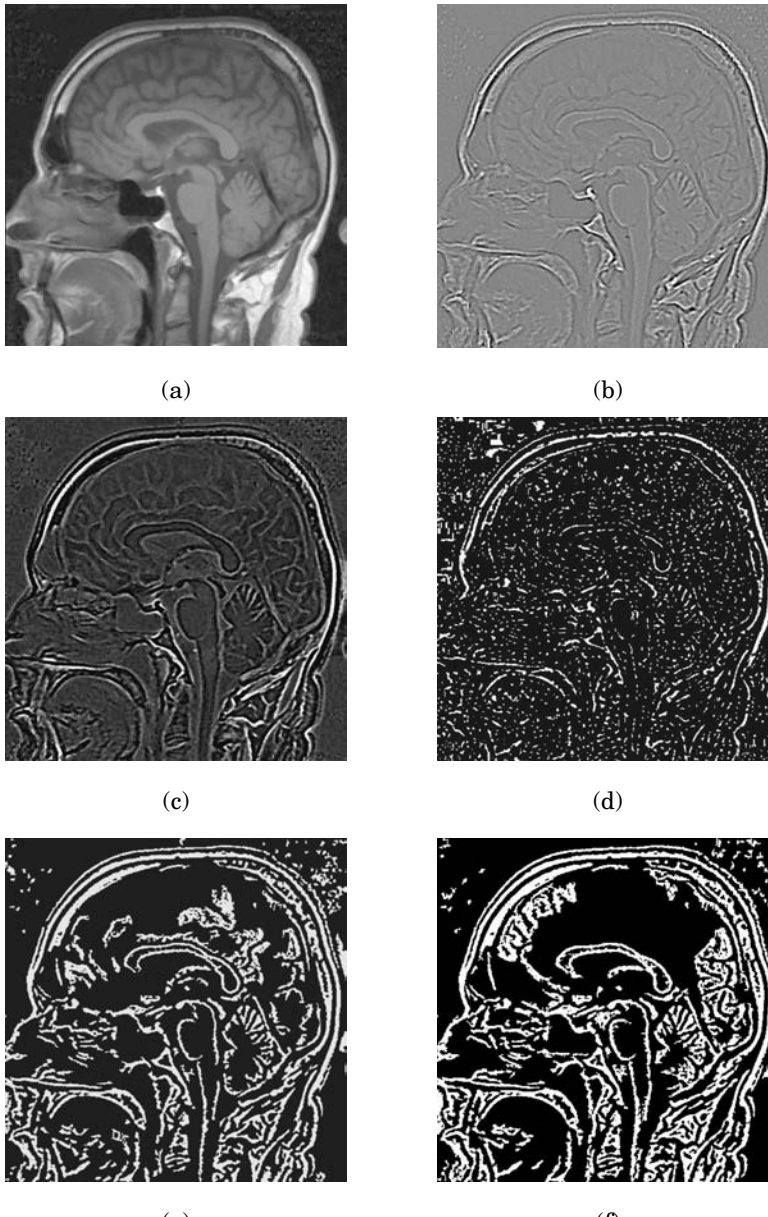
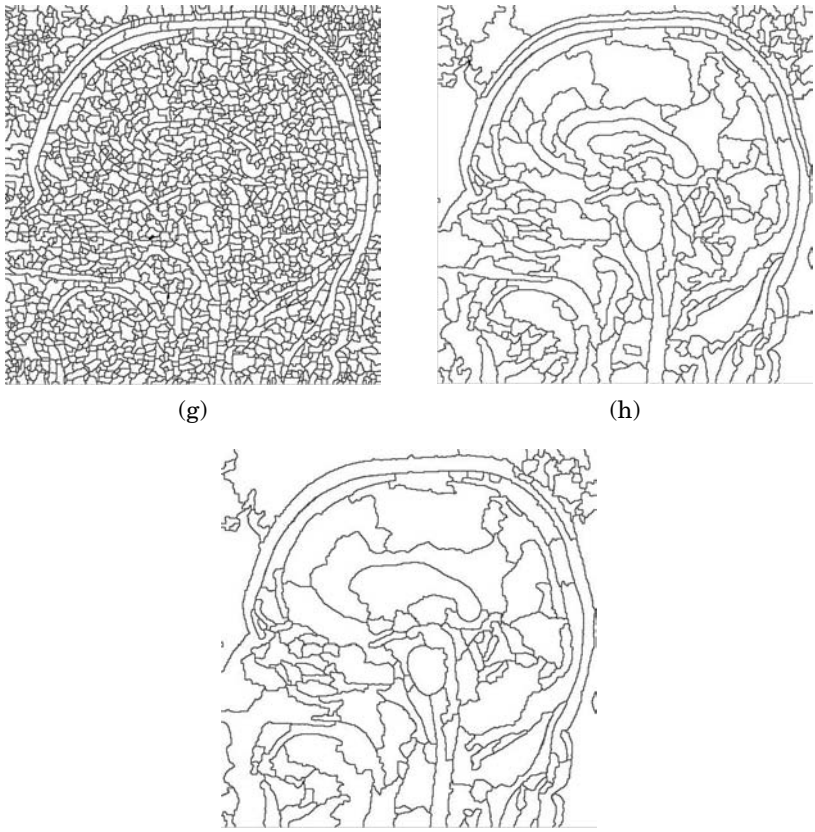


Figure 13.14 Example of watershed segmentation: (a) original image, (b) Laplacian of Gaussian, (c) LoG inverted and enhanced, (d–f) morphologically processed minima (seeds of valleys), and (g–i) resulting segmentation for different parameters of preprocessing, corresponding to d–f.



(g) (h)

(i)

Figure 13.14 (Continued).

No strict requirements are applied to the homogeneity of the regions, except that changes inside the regions are considerably weaker (i.e., primarily slower).

The detected region borders should be continuous and closed, reliable, and should describe the perimeters of regions with a good accuracy of localization. Unfortunately, the raw edge representation of an image as obtained by methods described in Section 13.1.2 is usually far from this ideal. The obtained edges are usually disconnected, sometimes multiple or excessively thick, and many of the detected edges do not correspond to segment borders, but rather to spurious rims due to noise. The following section deals with two approaches to derive a reasonable border representation from the raw edge image.

13.2.3.1 Borders via Modified Edge Representation

The modification of the raw edge representation has the following purposes:

- Thinning of edges; ideally, the edge should have a thickness of just a single pixel.
- Removing spurious edges: the edges due to noise that are not parts of a real segment border should be removed.
- Connecting the edges belonging to a border so that segment borders become continuous and, when completely inside the image, also closed curves.

It is desirable that the obtained borders are really the object perimeters. This is not always completely achieved at this level; the ultimate proper segmentation is often possible only when some *a priori* knowledge, including image context, is taken into account in the frame of higher-level analysis. An example of a border representation obtained in this way is shown in Figure 13.15.

The edge representation is a binary image and many of the above-mentioned tasks (thinning, cleaning, connecting) can be done using morphological transforms (see the following section). Although in some cases these operations may be based entirely on elementary morphological operations, it is often necessary to use more complicated conditional transforms that prevent undesirable side effects, such as merging edges that belong to different borders, disconnecting thin borders, etc.

Figure 13.15 Continuous border representation obtained via modification of raw edge representation.

Let us mention first the simplest means of edge modification. Primarily, the spurious edges due to noise are mostly rather short in comparison with proper edges. Thus, the cleaning as the first step may be based on simply removing all edges whose length is under a chosen limit. Morphological transforms (see Section 13.3), designed *ad hoc* with typical edge situations in mind, then may be used for thinning and connecting edges. The conditional erosion masks, examples of which may be

$$\begin{bmatrix} 1 & x & 0 \\ 1 & 1 & 0 \\ x & 0 & 0 \end{bmatrix} \begin{bmatrix} x & 1 & 1 \\ 0 & 1 & x \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} x & 1 & x \\ 1 & 1 & x \\ x & x & 0 \end{bmatrix} \quad \text{etc.,} \tag{13.24}$$

(altogether eight masks) enable thinning of edges. Each of the masks is gradually shifted above the image so that its central element is identified with a pixel of the binary image. If the image pixels under ones of the masks are 1 and the pixels under zeros are 0, the image pixel under the central element of the mask is set to zero regardless of the image content under the x's. This way, the thickness of the touched edge is hopefully lowered. Consequently, the conditional dilation, expressed by masks, e.g.,

$$\begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix} \quad \text{etc.,} \tag{13.25}$$

(altogether eight masks), tries to connect the edges by setting the image element under the mask center to 1, when the conditions as above are fulfilled. Repeated use of the masks may lead to the desired edge representation with thin and connected edges.

As an example of a more sophisticated approach to modification of edge representations, the *edge relaxation* method as cited in [72] will be briefly described. This approach defines an elementary edge as the connection between a pair of vertically or horizontally neighboring pixels that have been marked in the raw edge representation. Obviously, not all elementary edges defined this way are really parts of borders of any object. A variable confidence $c(t) \in \langle 0,1 \rangle, t = 1,2,\ldots$ of being a boundary is attached to each elementary edge (t denotes the step number in iteration). These confidences are iteratively reevaluated during the relaxation procedure until all the confidences are close to either 0 or 1. Only the elementary edges with the unit final confidence will be preserved; the others will be discarded.

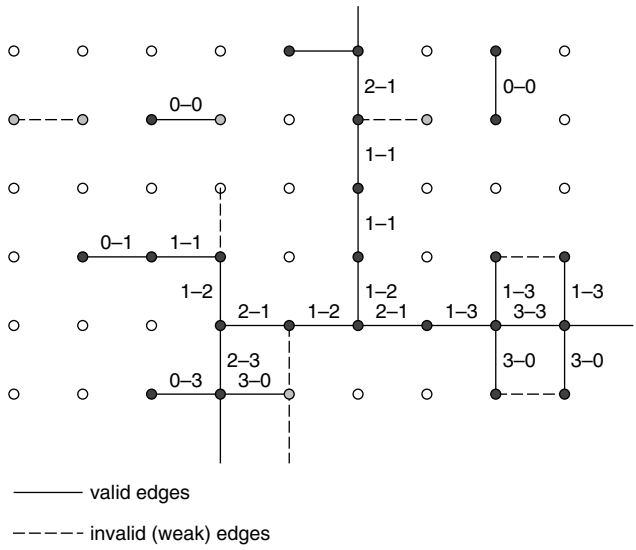


Figure 13.16 Elementary edges and their types.

The initial confidence in the range <0, 1> should be somehow assigned to each elementary edge, e.g., proportionally to the length of the detected edge to which it belongs or, if the intensities of the previously detected edges are available, according to the edge intensity in the nodes (pixels) defining the elementary edge. In each iterative step, all confidences are modified according to the current type of each elementary edge. The type is defined by the continuation of the edge at both its ends, as visible in Figure 13.16.

The actual type of the elementary edge influences the modification of its confidence in each iterative step, according to the following reasoning:

- Type 0–0: Isolated edge; strongly lower the confidence (suggested correction, $-2d$)
- Types 0–2, 0–3, 2–0, 3–0: Protrusion; lower the confidence ($-d$)
- Types 2–2, 2–3, 3–2, 3–3: Possibly a bridge between borders; leave the confidence unchanged (0)
- Type 0–1: Uncertain continuation; leave or slightly increase the confidence (0 or $+d/2$)
- Types 1–2, 1–3, 2–1, 3–1: Continuation to branching or intersection; increase the confidence ($+d$)
- Type 1–1: Sure continuation; strongly increase the confidence ($+2d$)

The new confidence is calculated based on the previous one, as

$$c(t+1) = \max[0, \min(1, c(t) + correction)] \tag{13.26}$$

so that the result is always in the range <0, 1>; d may be chosen on the order of ~0.1. The elementary edges are classified into two classes in each iterative step: those with the confidence above a chosen threshold are considered valid, while the edges, the confidence of which dropped below the threshold, are not considered in the next step of determining the edge types.

The algorithm thus consists of the following steps:

1. Initiate the confidence set.
2. Classify the elementary edges as valid or nonvalid.
3. Determine the type of each elementary edge, based on currently valid edges.
4. Update the confidences.
5. If all the confidences are close to either 1 or 0, terminate the calculation; otherwise, continue from step 2.

The method can be tuned by selecting many parameters: choice of the correction strategy (the above values are just a possibility), the confidence threshold determining the validity of edges, the method of initial confidence determination, etc. It has been proved that the method converges; details and further references can be found in [72]. However, the convergence is fast only in a few initial steps; later on, it slows down. This may be improved by introducing some nonlinearity in the confidence correction.

13.2.3.2 Borders via Hough Transform

Sometimes a border of a certain known shape (i.e., piece-wise linear, circular, elliptical, or even more complex) is sought in the edge representation of an image, where it is incomplete and imprecise, i.e., consists of disconnected and possibly also thick edge elements. The goal then is to fit the expected curve optimally to the available edge representation by adjusting the parameters of the curve. Such problems may be effectively solved by the *Hough transform*, which enables utilization of the incomplete and noisy information contained in the raw edges.

The concrete size, position, and possibly distortion of the curve can be described by a set of parameters, e.g., slope and y-intercept for a line, radius and x-y center coordinates for a circle, etc. Hough transform–based curve fitting is a kind of cluster analysis in the

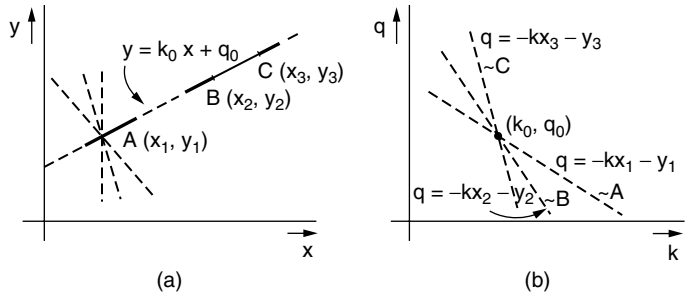


Figure 13.17 Principle of Hough transform in case of a line fitting: (a) original space and (b) parameter (Hough) space.

parameter space: this space is filled with points representing all possible curves crossing individual edge points in the edge representation. The locations of maximum density of such points are then considered representatives of the proper curves. Let us explain the method first concretely on the case of line fitting.

A pixel at the position (x_1, y_1), marked as an edge in the edge representation (Figure 13.17a), may potentially be a point of a line passing through this point; the so far unknown parameters of the line are the slope k and the intercept q; thus,

$$y_1 = kx_1 + q. \tag{13.27}$$

As x_1, y_1 are fixed, while unknown k and q may be considered variable, the equation

$$q = -x_1 k + y_1 \tag{13.28}$$

may be interpreted as an equation of a line in the parameter space, the coordinates of which are k, q (Figure 13.17b). Such a line in the parameter space represents all the lines in the original space that pass through the point (x_1, y_1), i.e., a bundle of lines in the original space.

To enable selecting the proper parameters k, q, other points of the original line must be investigated. As indicated in the figure, each point on the original-space line $y = k_0 x + q_0$ is represented in the parameter space by a different line. All the parameter-space lines intersect at the point k_0, q_0, which can thus be precisely determined providing that the positions (x_1, y_1), (x_2, y_2), (x_3, y_3) all lie exactly on the original line. Should the positions be imprecise, the parameter-space lines would not intersect at a single point, but for

many such lines derived from as many approximate points of the original line as available in the edge representation, the density of points in the parameter space would obviously reach maximum at or close to the point k_0, q_0.

The transform is naturally realized in the discrete original space; also, the parameter space must be discretized adequately to the desired precision of the parameter determination. For the line identification, the Hough space would be formed by a two-dimensional field of counters; for each edge point in the original space, all the counters located on the corresponding line in the Hough space are incremented. After processing the complete edge representation point by point this way, the coordinates of the maximally filled counter determine the parameters of the line.

For other simple curves (circle, ellipse, parabola, etc.), the approach would remain basically identical; the higher number of the curve parameters would be reflected by the correspondingly higher dimensionality of the counter field. The curve can be described either by an explicit equation as above or by an implicit equation of the type $f(x, y, \mathbf{a}) = 0$, $\mathbf{a}$ being the parameter vector of the concrete curve; each of the counters then corresponds to a particular value of $\mathbf{a}$. In every step, all counters will be incremented, for which the curve equation is fulfilled (with the precision given by the density of Hough-space sampling). It is important to realize that the finally obtained parameters determine the complete (unlimited) curve even in cases when only a section of the curve appears in the original image and in its edge representation (as, e.g., when the lines form a polygon border). Determining the limits is then a separate problem that must be solved utilizing image context or some *a priori* information. An example of a circle fitting is illustrated in Figure 13.18.

In a similar way, the *generalized Hough transform* enables identification of even complex shapes that cannot be described by simple equations. The generalization requires that the edge representation also includes the directions of edges, not only the detected edge points. Thus, the edge representation should be a vector-valued image with the first standard binary component and the second component being the direction of the detected edge for each edge point, possibly estimated via gradient components or by means of compass masks.

The curve to be identified will now be described by a s.c. R-table (see Figure 13.19); inside the curve, a reference point R must be defined to which the description data are related. The curve is described by a set of sample points, dense enough to express all

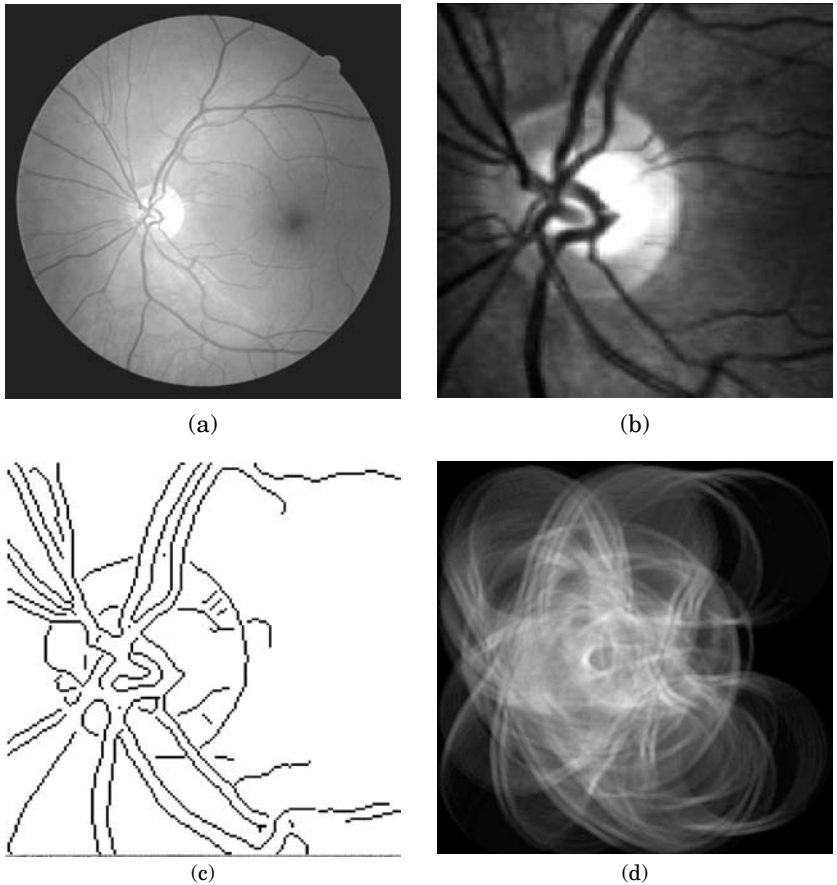
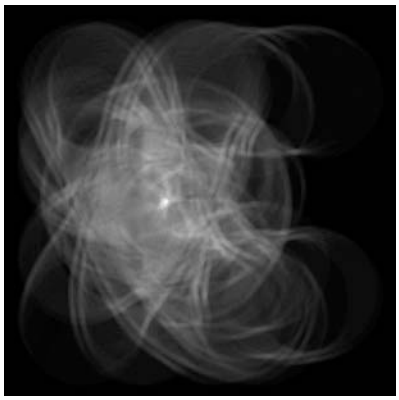
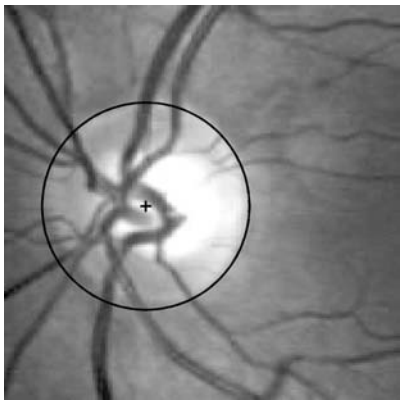


Figure 13.18 Detection of s.c. optical disc on a retina image: (a) original image, (b) region of interest (ROI) in the original image, (c) edge representation (via Canny detection), (d) two-dimensional profile of three-dimensional Hough-space image (the plane for a constant imprecise radius; the variables are the coordinates of the circle center), (e) two-dimensional profile of three-dimensional Hough space for the precise radius, and (f) the found circle together with its center imposed on the original. (Parts a–c and (f) from Chrastek et al. in *Bildverarbeitung für die Medizin 2002: Algorithmen–System–Anwendungen*, pp. 265, 266. Springer-Verlag, Heidelberg. With permission.)

details of the curve. For each k-th point, its location in polar coordinates (r_k, α_k) relative to the reference point R is given, together with the tangent direction Φ_k. As all the parameters are again discretized, there is only a finite number n of different tangent directions.



(e)



(f)

Figure 13.18 (Continued).

The curve-point data are then sorted according to Φ, which groups the data of points with parallel tangents, so that the R-table looks like this:

Edge Direction	Polar Coordinates of Curve Points			
Φ_1	r_{k1}, α_{k1}	r_{k2}, α_{k2}	r_{k3}, α_{k3}	…
Φ_2	r_{k4}, α_{k4}	r_{k5}, α_{k5}	…	…
⋮				
Φ_n	r_{k6}, α_{k6}	r_{k7}, α_{k7}	r_{k8}, α_{k8}	…

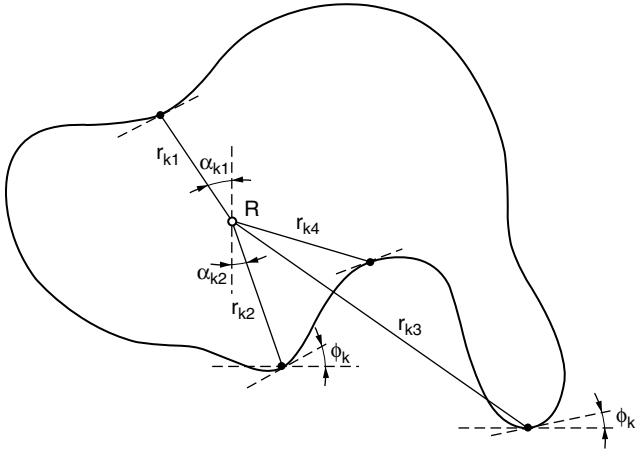


Figure 13.19 A generic closed curve to be described by an R-table.

When searching for such a curve in the given raw edge representation, only two parameters have to be determined: the coordinates of the reference point R. It is relatively easy to generalize the search even for the resized and rotated version of the same curve. Introducing a single parameter s with which all the radial distances r would be multiplied enables representation of differently scaled curves by a single R-table. Rotation can easily be enabled by adding the rotation angle φ to all directional data, Φ and α. There are then four parameters to be determined: two positional coordinates of R, rotation φ, and scale s. This means a substantial reduction in Hough-space dimension in comparison with the classical approach, which would require in such cases using curves with a high number of parameters.

Searching a curve described by a previously derived R-table via the generalized Hough method thus consists of the following steps:

- Provide the field of counters (two- to four-dimensional as required); to each of them, a particular vector parameter (x, y, s, φ) belongs.
- For every edge pixel (x, y) of the input image:
  - Find the edge direction $\Phi(x, y)$ as the second component of the pixel value of the input edge image.
  - Repeat for all considered scales and rotations:
    - Calculate the parameter coordinates (x, y, s, φ) of all points that are potential reference points, and increment all relevant counters.

The maxima of counts, or the centers of gravity of clusters, in the Hough space determine the parameter vectors of all appearances of the curve in the image. Notice that there may be more appearances; therefore, a careful inspection of the Hough space is needed in order not to miss a weaker object.

13.2.3.3 Boundary Tracking

Boundary tracking is an intuitively natural approach to be applied on the edge representation where the boundaries are nonconnected and multiple. However, some additional image information is needed in order to enable tracking based on some similarity among pixels belonging locally to a border. Therefore, the binary edge representation b (1 for edge pixel, 0 otherwise) would be complemented by both edge intensity e and direction information φ, thus using a

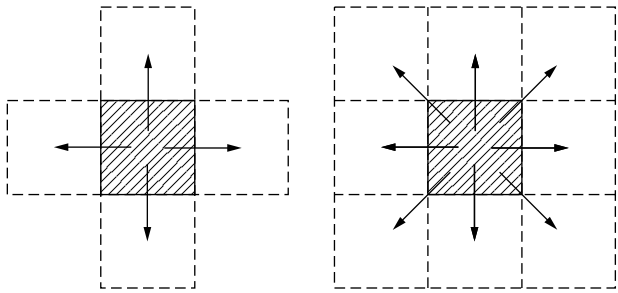


Figure 13.20 Four-connected and eight-connected neighborhoods.

vector-valued edge representation as input ((b, e, φ)-vector per pixel); the additional two components are only needed for positively marked pixels. Obviously, such information may only be obtained from edge detectors based on some gradient estimation; zero-crossing Laplacian or Laplacian-of-Gaussian detectors do not provide this input. Also, to prevent tracking of a totally false boundary starting from a false edge point marked due to noise, some *a priori* information about the boundaries or objects is needed, perhaps in the form of initiating the tracking interactively or based on previous rough segmentation utilizing other methods.

Before describing the algorithm, the used neighborhood notion as well as the similarity measure should be defined. Either a 4-connected or 8-connected neighborhood is considered to define the connectivity, as depicted in Figure 13.20. The measure of similarity between the i-th starting pixel and the $(i + 1)$-th (continuing) pixel may be defined quantitatively by the length L of the difference vector,

$$L = \sqrt{(a\Delta e)^2 + (\Delta\varphi)^2}, \quad \Delta e = |e_i - e_{i+1}|, \quad \Delta\varphi = |\varphi_i - \varphi_{i+1}| \bmod 2\pi, \tag{13.29}$$

where the chosen coefficient a balances the influence of both differences.

When the initial pixel of the border is chosen, the idea of tracking is simple: the b-marked pixels in the chosen neighborhood around the already accepted point are tested on similarity with this pixel. The most similar neighbor is then tested on absolute similarity, and if this is above a chosen threshold, the new pixel is accepted as the continuation of the border. Consecutively, the algorithm is repeated with the new pixel as the starting point until the path terminates,

not having a continuation (there are no marked pixels in the neighborhood, or none fulfills the minimum similarity requirement), the boundary closes, or the border reaches the image margin. Several initial starting points may be chosen successively to provide either boundaries of several objects or a number of partial boundaries, or to offer alternatives to be chosen from on a higher level of analysis.

13.2.3.4 Graph Searching Methods

The sampled image area may be adjoined to a two-dimensional *graph*, in which the sample points are considered *nodes*, while *links* are formed by connections of the samples. A *path* between two nodes in the graph is a set of links commencing in the start node and terminating in the end node. Such a path in the adjoined graph may represent a boundary in an image; path-finding algorithms, as discussed in the graph theory, thus may be applied to this purpose. However, even when both the start and end nodes are known, there are usually many paths and obviously only one or some of them are acceptable as the boundary. Therefore, the *optimal path* should be found in the sense of following the previously determined edge locations. The evaluation of optimality is generally based on the cost of the path consisting of the costs of touched nodes and the transition costs associated with the links. The node cost can be derived from the edge intensity previously determined during the edge detection procedure—the higher the intensity, the lower the cost. The transition costs may be given, depending on the character of the image, by the length of the path, or by the variances of the edge intensity, and namely of edge direction among closely linked nodes. Even when the successive pixel is always chosen approximately in the direction dictated by the gradient (i.e., roughly perpendicularly to it), there are usually more or many possible paths, and the optimization aiming at cost minimization is necessary.

Relatively complex graph-theoretical approaches to boundary-tracking border determination are commonly used, either heuristic or based on dynamic programming. However, these are already beyond the frame of principles treated here; the details can be found, e.g., in [5] or [72].

13.2.4 Segmentation by Pattern Comparison

In industrial applications, searching for a concrete pattern in the image is a frequent task (e.g., a particular mechanical part, a tool, or a letter of a concrete font). As the exact fit of an image area with

the sought pattern cannot be expected, the search is controlled by some local similarity criteria similar to those discussed in Section 10.1.3. The found object then may be labeled as a segment; this way, the pattern detection and localization may be considered a segmentation method.

Often only a shift of the sought pattern on the image plane is expected; inclusion of any further degree of freedom (rotation, scaling, perspective) complicates the detection extremely. In the medical imaging field, this approach may find its application in searching for a position and shape of an organ, or of a marker affixed, e.g., to a bone. Rather complex nonrigid geometrical transforms of the pattern representation, as well as a need for recognizing different two-dimensional projections of a three-dimensional organ, should be supposed. In such a complicated case, the procedures as used in flexible image registration (Section 10.3)—including complex similarity criteria and determination of geometrical transform parameters (i.e., position and size, rotation, distortion, etc.) via optimization—would have to be used.

13.2.5 Segmentation via Flexible Contour Optimization

When images are noisy, textured, or the imaging method provides spurious details, the contours of the areas obtained by the above-described standard methods would be too structured or rough, or the proper segmentation might be impossible, as small crumbled segments corresponding to the texture or noise would be provided instead of meaningful areas. This is a typical situation in medical imaging, namely in ultrasonography and gammagraphy, but not exclusively. The modern approach to this problem is based on *flexible or deformable contours*—the concept based on suitably defined curves that may be deformed to fit reasonably well the real image contours, while their definition prevents too detailed following of small details and noise patterns. Development of such a contour should take into account the image properties, the chosen degree of the curve flexibility, and possibly also the interactive interventions of the operator wishing to keep the contour close to some landmarks or to prevent it from making undesirable inlets or protrusions.

The research field of flexible contour segmentation is very active, and the present state may be considered provisional. Principles of three differing approaches will be explained here briefly in order to present the concepts; more details and further modifications,

together with examples of medical applications, can be found in the detailed overview [80], references listed therein, and current journals. Although the methods are again presented in two-dimensional formulation, all of them can be rather straightforwardly generalized to the three-dimensional case.

13.2.5.1 Parametric Flexible Contours

A parametric flexible contour is defined as a parametrically described curve, possibly but not necessarily closed,

$$\mathbf{X}(s,\mathbf{p}) = [X(s,\mathbf{p}), Y(s,\mathbf{p})]^T, \quad s \in \langle 0,1\rangle, \tag{13.30}$$

the character of which is given by the chosen functions $X(s, \mathbf{p})$, $Y(s, \mathbf{p})$; $\mathbf{p}$ is a parameter vector influencing the shape of the functions and therefore also the shape of the contour in the (x, y)-plane. The problem can be in principle formulated as searching for the optimal vector $\mathbf{p}$ that constrains the curve to the desired shape.

However, it is rarely possible to find closed formulae of the functions; instead, the vector function $\mathbf{X}(s)$ is usually described by a table of pairs $X(s_i)$, $Y(s_i)$ for a discrete sequence $\{s_i\}$, $i = 0, 1, \ldots, n - 1$; the continuous curve is then provided by suitable interpolation. The task is to find the table that optimally describes the desired contour. Although the formulation is thus discrete, the explanation will be presented in the more transparent continuous formulation, and only the final step of solving the resulting partial differential equation will be converted to the discrete problem via method of finite differences.

The method is based on minimization of energy, which can more vividly be formulated as the equilibrium of forces. The curve is interpreted as a string that can be both stretched (or compressed) and bent, which causes certain internal forces resulting in the total internal force (dependent on the position on the curve),

$$\mathbf{F}_{\text{int}}(\mathbf{X}(s)) = \frac{\partial}{\partial s}\left(E\frac{\partial \mathbf{X}(s)}{\partial s}\right) - \frac{\partial^2}{\partial s^2}\left(R\frac{\partial^2 \mathbf{X}(s)}{\partial s^2}\right), \quad s \in \langle 0,1\rangle. \tag{13.31}$$

Here, the first term is inversely proportional to elasticity of the string (i.e., suppresses its stretching), while the second term corresponds to rigidity and thus suppresses its bending. These string properties can be controlled by the coefficients E and R (which may be variable along the string, but for simplicity we shall omit this

possibility). Obviously, the higher are E and R, the stiffer and less inclined to tracking of details is the contour.

Without external forces, the (homogeneous) string would become a circle when closed, or a line otherwise. The external forces acting on the string should try to deform the contour maximally close to the actual borders. Two types of external forces are considered: the forces derived from the image content, and other (auxiliary) forces that help to find the proper shape. We shall return to the second group later; now, let us define the forces from the image. The forces should obviously be zero at the proper borders and should increase with the distance from the proper position (at least until a certain distance). Forces suitable for attracting the contour to edges can be derived from the potential function

$$P_{\text{edge}}(x, y) = -W\left|\nabla(G_\sigma * I|(x, y))\right|^2, \tag{13.32}$$

the squared absolute gradient of the smoothed image produced by preprocessing with the Gaussian operator of a suitable variance σ, weighted by a chosen W. As this potential function has extremes at the edges, the *potential forces* proportional to the gradient of $P_{\text{edge}}(x, y)$,

$$\mathbf{F}_{\text{pot}}(\mathbf{X}(s)) = -\nabla P_{\text{edge}}(\mathbf{X}(s)), \tag{13.33}$$

have the required properties: the edge attracts the contour as far as it is in a close neighborhood.

When thin lines (possibly curved) are to be approximated by the contour, the potential function simplifies to

$$P_{\text{line}}(x, y) = W(G_\sigma * I|(x, y)), \tag{13.34}$$

as in this case, the extremes of the smoothed image are situated at the contour; the sign of the weight W depends on whether a light line on dark background, or vice versa, is sought. The potential forces are then again given by Equation 13.33, with P_{line} replacing P_{edge}.

The *auxiliary forces* $\mathbf{F}_{\text{aux}}(\mathbf{X}(s))$ serve to help deforming the contour properly; they may either be independent of the image content or depend on it vicariously. A common example of the first kind is applying a *pressure*, i.e., forces, perpendicular to the string tangents, trying to expand the curve; this may allow for reaching of even

distant borders starting from a small initially estimated contour that may be out of reach of the edge or line potential forces.

The forces of the second kind are usually dependent on a preliminary image analysis. When the position of the borders has been somehow preliminarily (and approximately) determined, it is possible to define the auxiliary potential field based on the *distance map* of distances $d(x, y)$ to the nearest border point,

$$P_{\text{aux}}(x, y) = W_a \, e^{(-d[x,y])}, \tag{13.35}$$

the gradient of which would yield the auxiliary attractive force acting to longer distances from the estimated border than $\mathbf{F}_{\text{pot}}(\mathbf{X})$.

Another possibility is to introduce some auxiliary forces *interactively.* They may be chosen as either attracting to or repulsing from manually selected control points in the image, thus again forming the contour according to visual inspection or *a priori* information. The former forces should be proportional to the distance d from the control point and directed toward it, while the latter should somehow decrease with the distance and aim radially from the point. The inverse proportionality would lead to a singularity for $d = 0$, and therefore, e.g., a Gaussian function of d is used instead.

The force equilibrium is then expressed as

$$\begin{aligned} &\mathbf{F}_{\text{int}}(\mathbf{X}(s)) + \mathbf{F}_{\text{ext}}(\mathbf{X}(s)) = \mathbf{0}\text{ , i.e.,} \\ &\mathbf{F}_{\text{int}}(\mathbf{X}(s)) + \mathbf{F}_{\text{pot}}(\mathbf{X}(s)) + \mathbf{F}_{\text{aux}}(\mathbf{X}(s)) = \mathbf{0}. \end{aligned} \tag{13.36}$$

When substituting from respective equations (perhaps summing when there are more components of the forces), we obtain a (static) differential equation that is difficult to solve directly. Adding dynamic forces—the Newton's inertial force and the viscous force—can modify the equation so that it describes not only the final state, but also the complete course of deformation starting from the initial state; all the involved functions thus become dependent also on time t. The inertial force given by the second derivative with respect to time is often neglected, as the inertia may cause overshoots behind the proper positions; thus, only the viscosity term remains, giving

$$\begin{aligned} &\mathbf{F}_{\text{vis}}(\mathbf{X}(s,t)) + \mathbf{F}_{\text{int}}(\mathbf{X}(s,t)) + \mathbf{F}_{\text{pot}}(\mathbf{X}(s,t)) + \mathbf{F}_{\text{aux}}(\mathbf{X}(s,t)) = \mathbf{0}, \\ &\text{where} \quad \mathbf{F}_{\text{vis}}(\mathbf{X}(s)) = -V \frac{\partial \mathbf{X}(s,t)}{\partial t}. \end{aligned} \tag{13.37}$$

This (dynamic) partial differential equation can be solved by the finite-difference method as follows. Both variables are discretized equidistantly, so that $\mathbf{X}_m^n = \mathbf{X}(m\Delta s, n\Delta t)$. Setting the viscous force to the other side of the equation, the difference approximation of Equation 13.37 becomes

$$V\frac{\mathbf{X}_m^n - \mathbf{X}_m^{n-1}}{\Delta t} = \frac{E}{(\Delta s)^2}\left(\mathbf{X}_{m+1}^n + \mathbf{X}_{m-1}^n - 2\mathbf{X}_m^n\right) - \frac{R}{(\Delta s)^4}\left(\mathbf{X}_{m-2}^n - 4\mathbf{X}_{m-1}^n + 6\mathbf{X}_m^n - 4\mathbf{X}_{m+1}^n + \mathbf{X}_{m+2}^n\right) + \mathbf{F}_{\text{aux}}\left(\mathbf{X}_m^{n-1}\right) \tag{13.38}$$

When denoting $\mathbf{X}^n = [\mathbf{X}_1^n, \mathbf{X}_2^n, \ldots, \mathbf{X}_M^n]^T$, where M is the number of points on the contour, and forming the corresponding matrices $\mathbf{X}^{n-1}$, the coefficient matrix $\mathbf{A}$ (sized $M \times M$), and $\bar{\mathbf{F}}_{\text{aux}}(\mathbf{X}^{n-1}) = [\mathbf{F}_{\text{aux}}(\mathbf{X}_1^{n-1}), \mathbf{F}_{\text{aux}}(\mathbf{X}_2^{n-1}), \ldots, \mathbf{F}_{\text{aux}}(\mathbf{X}_M^{n-1})]^T$, the equation can be rewritten as an iterative formula,

$$\mathbf{X}^n = (\mathbf{I} - c\mathbf{A})^{-1}(\mathbf{X}^{n-1} + c\bar{\mathbf{F}}_{\text{aux}}(\mathbf{X}^{n-1})), \quad c = \Delta t/V, \tag{13.39}$$

enabling calculation of the sequence of gradually improving contours.

An example of a lesion border approximated in an ultrasonographic speckled image using deformable contour is in Figure 13.21.

13.2.5.2 Geometric Flexible Contours

Geometric flexible contours are defined using the concept of level sets of a scalar function. Let a scalar function $\Phi(x, y, t)$ be defined

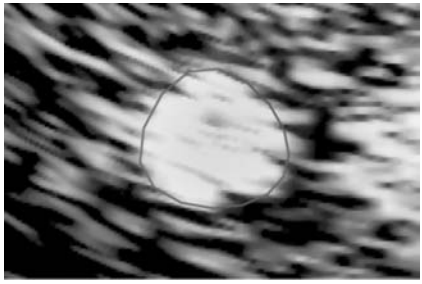


Figure 13.21 An example of using a deformable contour as an approximation of a lesion border in a speckled ultrasonographic image. (Courtesy of Brno University of Technology, R. Kurecka, M.Sc.)

in the two-dimensional coordinate space x, y (usually in the same space as the image to be segmented). A *level set* is the locus of points $\{(x, y)\colon \Phi(x, y, t) = a\}$, where a is an arbitrary constant; therefore, the set is formed by all (possibly nonconnected) isolines of Φ for a particular level a. In the context of segmentation, usually $a = 0$ is chosen; the corresponding curves are called *zero-level set*. These curves will be used as the deformable contours. The concrete form of function $\Phi(x, y, t)$ determines the zero-level set uniquely for each time instant t; on the other hand, it means that the temporal variability of Φ also allows development of the contour (or nonconnected contours) $\mathbf{X}(s, t) = [X(s, t), Y(s, t)]$, similarly as in the previous section.

There are now two problems: how to formulate the function $\Phi(x, y, t = 0)$ so that its zero-level set would be the required initial form of the contour(s) $\mathbf{X}(s, t = 0)$, and how to arrange the temporal variability of $\Phi(x, y, t)$ for $t > 0$ so that the time development of the zero level would fulfill the expectations of approaching gradually the proper border(s).

A common choice for $\Phi(x, y, 0)$ is the *distance function*,

$$\Phi(x, y, 0) = \text{dist}(x, y), \tag{13.40}$$

which is defined for each point (x, y) as the signed distance to the nearest point of the zero-level curve $\mathbf{X}(s, t = 0)$. Obviously, the distance function is zero everywhere on the initial contour; let us define it as negative inside while positive outside of the (closed) contour.

The suitable time development of $\Phi(x, y, t)$ can be described by a differential equation that comes from the definition of the contour(s) as the zero-level set,

$$\Phi(x, y, t) = \Phi(\mathbf{X}(s, t), t) = 0, \tag{13.41}$$

which after differentiation with respect to time yields

$$\frac{\partial \Phi(\mathbf{X}(s, t), t)}{\partial t} + (\nabla \Phi(\mathbf{X}(s, t), t))^T \frac{\partial \mathbf{X}(s, t)}{\partial t} = 0 \tag{13.42}$$

When the above sign convention on $\Phi(x, y, t)$ applies, the inward normal vector $\mathbf{N}$ is given by the negative gradient $-\nabla\Phi$.

Let us now formulate the time development just for the contour $\mathbf{X}(s, t)$. Its evolution concerning the shape is given only by the motion of contour elements perpendicularly to the tangent;

therefore, the curve development may be described by the differential equation

$$\frac{\partial \mathbf{X}(s,t)}{\partial t} = V(\ldots)\,\mathbf{n}. \tag{13.43}$$

$V(\ldots)$ is a factor influencing the evolution: as $\mathbf{n}$ is the inward normal unit vector, a positive value of $V(\ldots)$ means that the element tends to move inside and vice versa; the speed of the movement is proportional to this factor. The function $V(\ldots)$ thus may be called the *speed function*. Two types of this function are commonly used: linear function of the local curvature κ of $\mathbf{X}(s, t)$ or a constant V_0. It can be shown that using the first speed function type leads to gradual smoothing of the curve; the other type acts in the opposite sense. Moreover, positive V_0 leads to shrinking the curve, while negative values lead to expansion; the choice depends on whether the proper border is basically expected inside or outside of the initial contour. Both types of function are often combined, and the development of the contour may be influenced by the relation of both components of $V(\ldots)$.

So far, only the general character of the contour development was considered, disregarding the need for convergence to the proper border. There must be a constraint preventing the contour from passing the proper locations. It is the *stopping factor* $f(I(x, y))$, dependent on the image content,

$$f(I(x, y)) = \frac{1}{\left|\nabla(G_\sigma * I|(x, y))\right| + 1}, \tag{13.44}$$

which obviously slows down the curve development when the absolute gradient of the smoothed image is high (i.e., in the neighborhood of edges). Though it may be sufficient in images with a high contrast, it turns out that in noisy or low-contrast images, or at weak boundaries, it may fail; therefore, more sophisticated stopping functions have been suggested and tested.

When the speed function includes all the mentioned elements, Equation 13.43 becomes

$$\frac{\partial \mathbf{X}(s,t)}{\partial t} = f(I(x, y))(\kappa + V_0)\,\mathbf{n} = \frac{\kappa + V_0}{|\nabla(G_\sigma * I|(x, y))| + 1}\,\mathbf{n}. \tag{13.45}$$

As the unit vector is $\mathbf{n} = \mathbf{N}/|\nabla\Phi|$, substituting from the last equation to Equation 13.42 gives

$$\frac{\partial\Phi}{\partial t} = \frac{\kappa(x, y) + V_0}{|\nabla(G_\sigma * I|(x, y))| + 1} |\nabla\Phi|. \tag{13.46}$$

Although this equation has been derived for the zero level, obviously its validity will not change, when the complete function will behave accordingly. This is the reason why the specification of the parameter $\mathbf{X}(s, t)$ (which is implicitly given by the discrete formulation of the initial contour) has been omitted. We thus arrived at a partial differential equation describing the behavior of the function $\Phi(x, y, t)$ in space and time, while respecting the condition for the initial contour.

The segmentation via geometrical deformable contours thus consists of choosing the initial contour(s), determining the initial values of $\Phi(x, y, t = 0)$ as the distance function, and then developing $\Phi(x, y, t)$ in time. Consequently, the contour is completely given by the gradual solution of the differential Equation 13.46. Naturally, in the discrete environment, the method of finite differences will be used.

One of the advantages of the method is its ability to change automatically and consistently the topology of the contour—it may be split during the iteration or several partial contours may merge automatically, when required by the image character; this is something not easily possible with parametric contours. According to published results, the method gives good results when applied appropriately; however, it should be used with caution as the topology of derived contours may be inconsistent with the image content when noisy images with partially weak borders are processed. Only basic ideas have been presented here, primarily adapted from [80]. Further developments of the method can be found in this overview, together with references to original papers.

13.2.5.3 Active Shape Contours

So far, we discussed deformable contour methods where the types of flexible curves were analytically specified; subsequently, their shape was adapted to the image properties starting from certain initial estimates by varying some parameters or functions. The choice of the initial contours that may substantially influence the success of segmentation has not been discussed yet. It may be

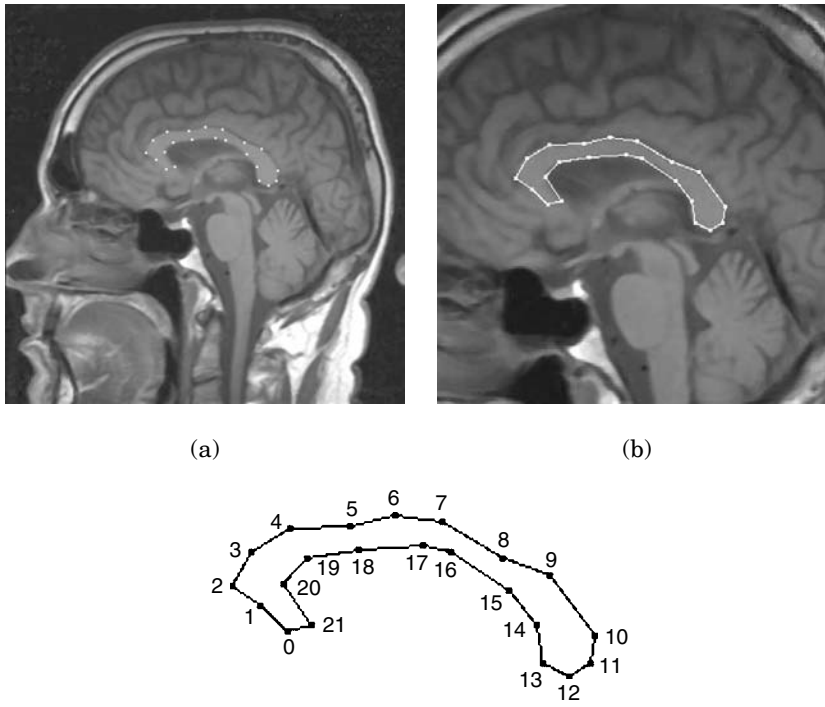


Figure 13.22 (a, b) Identified features in an image and (c) the obtained contour, as a member of the measuring set to be averaged.

formulated by manually specifying the initial contour via a pointing device; alternatively, some *a priori* knowledge may be utilized to form the initial contour automatically.

The *method of active shape contours* specifies the curve shape for a particular type of imaged object (a particular anatomical structure or organ) *a priori*, based on a set of measurements, thus enabling natural inclusion of anatomical knowledge into the segmentation process. The borders in a particular anatomical scene are characterized by discrete samples at the contours (Figure 13.22a); these points are situated at selected landmarks characteristic for every image of the same scene (typical corners, bays or protrusions, holes, blood vessel branching, etc.). The selection of a set of such landmarks (features) typical for a concrete scene is thus the first step in preparation of the segmentation procedure. Depending on the image character, the feature points in the typical image may

form one or more closed borders surrounding anatomically meaningful areas.

A training set of roughly registered images of the same scene is manually processed, the feature points identified, and their positions $\mathbf{X} = (x, y)$ measured; consequently, the measurements from the k-th image forming a closed contour (Figure 13.22c) are arranged into an $N \times 2$ matrix $\mathbf{C}_k$,

$$\mathbf{C}_k = [\mathbf{X}_{k,0}, \mathbf{X}_{k,1}, \ldots, \mathbf{X}_{k,n_k-1}]^T, \quad \mathbf{X}_{k,j} = [x_{k,j}, y_{k,j}], \quad k = 1, 2, \ldots, N. \tag{13.47}$$

When the complete training set has been measured, the mean value of the contour matrices

$$\bar{\mathbf{C}} = \frac{1}{N} \sum_{k=1}^{N} \mathbf{C}_k \tag{13.48}$$

is computed, thus obtaining a representation of the average shape of the contour.

Any particular contour image of the training set (or any other similar image) may be approximated by modifying the average contour according to the method of principal components, as

$$\mathbf{C} \approx \bar{\mathbf{C}} + \mathbf{E}\mathbf{p}, \tag{13.49}$$

where $\mathbf{E}$ is the matrix of the first m eigenvectors of the covariance matrix $\mathbf{R}_C$,

$$\mathbf{R}_C = \frac{1}{N-1} \sum_{k=1}^{N} (\mathbf{C}_k - \bar{\mathbf{C}})(\mathbf{C}_k - \bar{\mathbf{C}})^T, \tag{13.50}$$

and $\mathbf{p} = (p_1, p_2, \ldots, p_m)^T$ is a weight vector to be determined in the fitting procedure. The point is that, as it has turned out, the necessary variability is described by only a few eigenvectors, and therefore a correspondingly low number of parameters are then needed to modify the shape as necessary.

The contour determined this way might then be adapted to a concrete image by geometrical transforms and possibly serve as the initial estimate to be further elaborated as a flexible contour by the parametric method presented above. Deeper analysis and further references can be found, e.g., in the overview chapter [80].

13.3 GENERALIZED MORPHOLOGICAL TRANSFORMS

Morphological approaches are methodologically different from the mainstream of the methods transferred from the signal processing field that after generalization form the backbone of this book. The background theory—mathematical morphology based on topological and discrete algebra concepts—is rather intricate and not easy to follow for a non-mathematician. However, modern interpretations, relying on lucid theoretical background, enable explanation that in spite of its simplicity remains reasonably exact, consistent, and even intuitively applicable. We shall follow this way, loosely paraphrasing recent sources, namely, [70] and [13]; further useful sources are [72], [7], [17], [26], [12], [59], and many others mentioned therein.

The morphological transforms were initially applied to binary images as an *ad hoc* postprocessing method helping to remove spurious image elements and to provide elementary analysis of size, shape, and count of objects. Nowadays, the modern generalized morphological methods form a wide class of nonlinear methods applicable not only to binary images, but also to multidimensional gray-scale data; they enable novel interesting approaches to image processing and analysis. The generalization of the original binary image–processing methods to the gray-scale images is now straightforward and seamless, thanks to the theoretical development; therefore, we shall treat both cases together, considering the binary transforms as a special case. Most morphological transforms are image-to-image transforms and might therefore be classified as image processing. Nevertheless, the output images are often substantially modified and thus may be considered parametric images in a sense, or even—because they are often also substantially simplified—a description of the input image; therefore, the morphological methods are usually classified as analytic.

13.3.1 Basic Notions

13.3.1.1 Image Sets and Threshold Decomposition

A discrete gray-scale image $\{f_{i,k}\}$ may be interpreted as a mapping of a finite two-dimensional discrete subset D_f of the index space Z^2 into a set of non-negative integers representing the gray shades,

$$f: D_f \subset Z^2 \rightarrow \{0,1,2,\ldots,q\}, \tag{13.51}$$

where $q + 1$ is the number of gray levels, e.g., $q = 2^8 - 1$ for a single-byte intensity representation. Obviously, a binary image is a special case for $q = 1$. As it will be shown, any morphological transform (operator) yields an output image $\{g_{i,k}\}$ of the same size as the input image and in a given range of integer values, so that it can be described by a form similar to Equation 13.51.

The image may also be represented by a surface above the support D_f, as in Figure 1.1b. As both the spatial coordinates and pixel intensities are discrete, the discrete volume under the surface is filled by a finite number of discrete points that (including the points on the surface and on the support plane) form a set called *subgraph* G_f of f. Note that this way we obtain a *set representation* of the gray-scale image. Subsets of it, called *cross-section sets*, can be formulated: cross-section set S_m at a level m is formed by the points of the subgraph at the level m above the base,

$$S_m = \{(i,k,m) \subset G_f\}. \tag{13.52}$$

The cross-section set can also be formulated by means of thresholding. The *thresholding operator* $T_{t,s}(f)$ is defined here so that it sets to the pixel, whose input intensity is at or between two specified levels t and s, thus yielding the output

$$g = T_{t,s}(f), \quad g_{i,k} = T_{t,s}(f_{i,k}) = \begin{cases} 1 & \text{if } f_{i,k} \in \langle t,s \rangle \\ 0 & \text{otherwise} \end{cases}. \tag{13.53}$$

The single-limit thresholding at the level m can then be obviously described by the operator $T_{m,q}(f)$; let us call the pixels that it sets to 1 active pixels. The cross-section set at m can then be described as the subset formed of active pixels of the binary image obtained from f by thresholding at the level m,

$$S_m = \{(i,k) \subset D_f : f_{i,k} \geq m\}. \tag{13.54}$$

This thresholded image is called the *cross-section*. It can be easily recognized that the set (Equation 13.54) when elevated to the level m above the base, is identical with the above-defined set (Equation 13.52). To construct the subgraph of the original image using thresholded images, it suffices to stack these images for $m = 1, 2, \ldots, q$; as each of them has the unit height of active pixels, the subgraph will be built. Therefore, the original image may be composed of all the

thresholded images, as

$$f = \sum_{m=1}^{q} T_{m,q}(f). \tag{13.55}$$

Conversely, it may be said that a gray-scale image can be decomposed into its cross-sections. The formula (Equation 13.55) is called the *threshold decomposition* of image *f*.

13.3.1.2 Generalized Set Operators and Relations

In the field of mathematical morphology, the set operators of union and intersection are generalized for morphological transforms as follows.

The *intersection* of two images *f*, *h* is defined as a new image *g*,

$$g_{i,k} = f \wedge h|_{i,k} = \min(f_{i,k}, h_{i,k}), \tag{13.56}$$

while the *union* is defined as

$$g_{i,k} = f \vee h|_{i,k} = \max(f_{i,k}, h_{i,k}). \tag{13.57}$$

It can be easily checked that when object sets in binary images are defined as sets of active pixels equal to 1, the above definitions formulate the standard set operators $\cap$ and $\cup$, respectively.

Similarly, the operator of *complementation* is generalized as

$$f^C = C(f), \quad f^C_{i,k} = q - f_{i,k}. \tag{13.58}$$

Again, this definition applies to binary images in the standard sense.

The remaining definitions concern standard set operations: the *set difference* is

$$X \backslash Y = X \cap Y^C; \tag{13.59}$$

the *transposition* of a set X is the set symmetrical with respect to the origin,

$$\breve{X} = \{-\mathbf{x}, \quad \forall \mathbf{x} = (i,k) \in X\}. \tag{13.60}$$

When the set contains the origin of coordinates, it is symmetric only if $\breve{X} = X$.

The set relation of *inclusion* ($X \supset Y$) can be extended to gray-scale images as the "magnitude" relation. The image *f* is then said to be less than or equal to an equally sized image *g*; the simple

definition is as follows:

$$f \le g \quad \Leftrightarrow \quad f_{i,k} \le g_{i,k}, \forall i,k \in D_f \quad \Leftrightarrow \quad \forall m, S_m(f) \subseteq S_m(g). \qquad (13.61)$$

Obviously, when applied to binary images, $m = 1$ and the relation reduces to the standard set inclusion.

This relation between images can be extended to image transforms Φ, Ψ. The transform (operator) Φ is said to be less than or equal to the transform Ψ, if for all images belonging to their (common) definition range is valid:

$$\Phi \le \Psi \quad \Leftrightarrow \quad \forall f, \Phi(f) \le \Psi(f). \qquad (13.62)$$

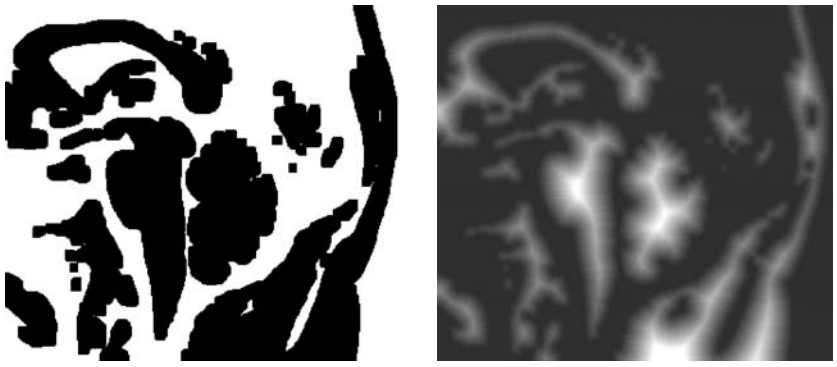
13.3.1.3 Distance Function

We shall consider only the rectangular regular grid of image samples. The sample positions may be regarded as nodes of a *graph*, the vertices of which connect the neighboring nodes. Both the 4-connectivity and 8-connectivity may be considered; in the former case, only horizontal and vertical vertices exist, while in the latter case, the skew ones are also present. It is then possible to define a *path*, connecting two nodes A, B of the graph, as a sequence of vertices leading from the first node to the second; its *length* is measured by the number of involved vertices. Mostly, there are more paths than a single one; the length of the shortest path is denoted as the *discrete distance* d_{D} between A and B. The metric depends on the selected type of connectivity; for 8-connectivity the distance is obviously shorter than or equal to that for 4-connectivity. A subset of nodes (e.g., carrying active pixels) is called *connected* if a path exists between any couple of nodes of the subset that leads only through the nodes belonging to the subset.

Another metric may be based on the *Euclidean distance* d_{E} between the nodes.

The *distance function* (or map) refers to a binary image defined above the node set D_f. The function is defined for the same node set and is nonzero only for the elements of the binary sets represented in the image by active pixels. The value of the distance function in each node is the distance d (of any chosen definition d_{D} or d_{E}) from this node to the nearest zero-valued pixel,

$$D(f)\Big|_{i,k} = \min_{m,n}[d((i,k),(m,n)), f_{m,n} = 0]. \qquad (13.63)$$



(a) (b)

Figure 13.23 (a) A binary image and (b) its distance map.

The distance function finds important application in morphological image processing (see below). An example can be seen in Figure 13.23.

13.3.2 Morphological Operators

The basic morphological operators are erosion and dilation. Combining them, other operators can be derived, as opening and closing, hit or miss (fit and miss), etc. The operation of all of them is based on using a *structuring element*, a matrix mask, usually much smaller than the processed image, that is gradually shifted above the image, such as in the local operators described in Chapter 2. In this sense, the morphological operators can be classified as *local nonlinear operators*.

Because the operators are nonlinear, characteristics such as impulse response or frequency transfer cannot be used; however, some other properties may be defined. It is useful to analyze the operators from this viewpoint and to utilize the properties when designing image processing procedures. The most important properties are:

- *Invariance with respect to threshold decomposition*: An operator Φ is denoted as invariant to threshold decomposition (Equation 13.55) when

$$\Phi(f) = \sum_{m=1}^{q} \Phi(T_{m,q}(f)). \tag{13.64}$$

  Note that this property applies to all linear operators, but some nonlinear operators fulfill it as well.

- *Invariance with respect to translation*: An operator Φ is denoted as invariant to shift, when a shift $\mathbf{s}$ of the input image only shifts correspondingly the output,

$$\forall f, \quad \Phi_{\mathbf{s}}(f(\mathbf{x})) = \Phi(f_{\mathbf{s}}(\mathbf{x})). \tag{13.65}$$

  Here, $g_{\mathbf{s}}$ means the image g shifted by a vector shift $\mathbf{s}$ with respect to the coordinates.
- *Idempotence*: An operator (a transform) Φ is idempotent, when applying it a second time does not change the result,

$$\Phi\{\Phi(f)\} = \Phi(f). \tag{13.66}$$

- *Increasingness*: An operator Φ is increasing if it preserves the "magnitude" relation (Equation 13.61) (i.e., inclusion in the binary case),

$$\forall f, g, \quad f \le g \Rightarrow \Phi(f) \le \Phi(g). \tag{13.67}$$

- *Extensivity*: An operator Φ is extensive if its output is greater than or equal to its input,

$$\forall f, \quad f \le \Phi(f). \tag{13.68}$$

  An *antiextensive* operator is characterized by the property

$$\forall f, \quad f \ge \Phi(f). \tag{13.69}$$

- *Duality*: Two operators are dual with respect to complementation when applying Ψ to an image f gives the same output image as applying Φ to the complement f^C and complementing the result,

$$\forall f, \quad \Psi(f) = (\Phi(f^C))^C = C\{\Phi(C\{f\})\}. \tag{13.70}$$

  An operator Φ is *self-dual* if

$$\forall f, \quad \Phi(f) = (\Phi(f^C))^C. \tag{13.71}$$

The mask (structuring element) of a morphological operator is usually binary, but it might acquire gray-level values as well, though it is not utilized in the definitions mentioned below. Obviously, the binary structuring element is a special case of the gray-level mask; we shall limit ourselves to binary masks only. The mask is specified, in addition to by its (active) set, also by the position of the reference

pixel ("origin") in the mask. The most common size of the mask is 3×3, but greater masks of possibly anisotropic and irregular shape also may be used. As for the mask content, the two isotropic binary masks, formed by neighboring pixels, as in Figure 13.20, are most frequently used; however, greater square, cross, line, or irregularly shaped masks are commonly used as well. In each mask, a cross marks the origin. Composite structuring elements (a couple of non-overlapping masks) are used in fit-and-miss (hit-or-miss) transform.

Following both the historical development and a natural way of explanation, we shall present each morphological operator first in its binary form (binary input and output images, binary mask), which will be consequently generalized to the gray-scale type. It should be understood that the binary version is always a special case of the more generic operator. To prevent confusion, we shall formulate the output image g in a new matrix of the same size as the input f, thus preventing any recursion. Further on, we shall call the set of active (equal to 1) pixels in a binary image simply the set of the image.

13.3.2.1 Erosion

Binary erosion provides, for each spatial position (i, k), the binary answer to the question: Does the mask overlaid on the image, with its origin placed at (i, k), fit the set in f in the sense that under all ones of the mask there are ones in the image? When denoting the erosion operator using a mask H as $\mathbf{E}_H(\ldots)$, the input (or output) image set as X (or Y), the position (i, k) as $\mathbf{x}$, the mask set as H, and the mask set shifted so that its origin is placed at $\mathbf{x}$ as $H_{\mathbf{x}}$, the set of the output image is

$$Y = \mathbf{E}_H(X) = \{\mathbf{x} | H_{\mathbf{x}} \subseteq X\}. \tag{13.72}$$

When further denoting a vector leading from the mask origin to one of its active elements as $\mathbf{h}$, and the input set shifted by $-\mathbf{h}$ as $X_{-\mathbf{h}}$, the erosion result may be equivalently expressed as the intersection of mutually shifted versions of X,

$$Y = \mathbf{E}_H(X) = \bigcap_{\mathbf{h} \in H} X_{-\mathbf{h}}. \tag{13.73}$$

The binary erosion, as the name suggests, erodes the input set—some border pixels (usually perimeter "stripes") are set to zero,

Figure 13.24 (Above) Example of two binary images and (below) their erosion by the centered 7×7 mask.

thus reducing the set size. This change is irreversible, as small objects and thin protrusions that cannot accommodate the structuring element have been removed completely. Thus, the erosion may cause noise suppression and (sometimes undesired) disconnection at the cost of diminishing the object perimeters (Figure 13.24).

The definition (Equation 13.73) may be directly extended to the definition of *gray-scale image erosion* via replacing the set intersection by the generalized definition (Equation 13.56),

$$g = \mathbf{E}_H(f) = \bigwedge_{\mathbf{h} \in H} f_{-\mathbf{h}}, \tag{13.74}$$

Figure 13.25 (Above) Gray-scale original and (below) its erosion by the 7×7 mask.

where $f_{-\mathbf{h}}$ is the input image shifted by $-\mathbf{h}$. Notice that the mask remains binary. Thus, the result is, for each pixel, the minimum of the pixel values among the shifted images, which may be expressed equivalently as

$$g_{i,k} = \mathbf{E}_H(f)|_{\mathbf{x}} = \min_{\mathbf{h} \in H} f(\mathbf{x} + \mathbf{h}), \tag{13.75}$$

So that the lack of valid image values behind the margins of the shifted images would not influence the result, the image is, for this purpose, supposed to acquire the intensity q outside of its definition range; thus, only valid values influence the minimum operator. An example of gray-scale erosion can be seen in Figure 13.25.

Erosion is invariant to threshold decomposition and to shift; it is an increasing operation and is dual to complementation. It is distributive with respect to the minimum operator,

$$\mathbf{E}_H(\wedge_k f_k) = \wedge_k \mathbf{E}_H(f_k)\,. \tag{13.76}$$

A practically important property of erosion is that the cascade of two erosions can be replaced by a single erosion with the mask expanded by dilation (see below),

$$\mathbf{E}_{H_2}(\mathbf{E}_{H_1}(f)) = \mathbf{E}_{[\mathbf{D}_{H_2}(H_1)]}(f). \tag{13.77}$$

Conversely said, erosion with a large mask may be decomposed into a chain of erosions with smaller masks when the large mask can be expressed as $\mathbf{D}_{H_2}(H_1)$.

13.3.2.2 Dilation

Binary dilation provides, for each spatial position $\mathbf{x} = (i, k)$, the binary answer to the question: Does the mask overlaid on the image, with its origin placed at $\mathbf{x}$, hit the set in f in the sense that under at least one 1 of the mask there is a 1 in the image? When using notation as before, the set of the output image can be expressed as

$$Y = \mathbf{D}_H(X) = \{\mathbf{x} | H_{\mathbf{x}} \cap X \neq \emptyset\}. \tag{13.78}$$

The dilation result may be equivalently expressed as the union of mutually shifted versions of X,

$$Y = \mathbf{D}_H(X) = \bigcup_{\mathbf{h} \in H} X_{-\mathbf{h}}. \tag{13.79}$$

In accordance with its name, the dilation increases the input set—some originally zero (background) pixels at the object margins (usually "stripes" around the perimeters) are set to 1, thus increasing the set size. Due to the nonlinearity, this change is irreversible, as small holes and thin gaps (that cannot contain the structuring element without one or more of its elements intersecting with the image set) have been changed to set pixels. Thus, the dilation may be used for background-level noise suppression, the filling of small holes, and may cause the connecting (possibly undesirably) of disconnected objects, at the cost of increasing the object perimeters.

The definition (Equation 13.79) may be directly extended to the definition of *gray-scale image dilation* via replacing the set union by the generalized definition (Equation 13.57),

$$g = \mathbf{D}_H(f) = \bigvee_{\mathbf{h} \in H} f_{-\mathbf{h}}, \tag{13.80}$$

where $f_{-\mathbf{h}}$ is the input image shifted by $-\mathbf{h}$. Notice that the mask remains binary. Thus, the result is, for each pixel, the maximum of

Figure 13.26 Dilation of upper images of Figure 13.24 by the centered 7×7 mask.

the pixel values among the shifted images, which may be expressed equivalently as

$$g_{i,k} = \mathbf{E}_H(f)|_{\mathbf{x}} = \max_{\mathbf{h} \in H} f(\mathbf{x} + \mathbf{h}). \tag{13.81}$$

Dilated binary images are shown in Figure 13.26; a dilation of a grey-scale image is shown in Figure 13.27. So that the lack of valid image values behind the margins of the shifted images would not influence the result, the input image is, for this purpose, supposed to acquire the intensity 0 outside of its definition range; thus, only valid values influence the maximum operator.

Like erosion, dilation is invariant to threshold decomposition and to shift; it is an increasing operation and is dual to complementation.

Figure 13.27 Dilation of the original image of Figure 13.25.

It is distributive with respect to the maximum operator,

$$\mathbf{D}_H(\vee_k f_k) = \vee_k \mathbf{D}_H(f_k). \tag{13.82}$$

Dilation has a similar practically important property as erosion: the cascade of two dilations can be replaced by single dilation with the mask expanded by dilation,

$$\mathbf{D}_{H_2}(\mathbf{D}_{H_1}(f)) = \mathbf{D}_{[\mathbf{D}_{H_2}(H_1)]}(f). \tag{13.83}$$

Therefore, dilation with a large mask may also be decomposed into a chain of dilations with smaller masks, should the large mask be expressible as $\mathbf{D}_{H_2}(H_1)$.

13.3.2.3 Opening and Closing

Erosion removes some of the object noise, small objects and protrusions, while dilation is used to filter out the background noise, small holes and gaps. The usually undesirable side effect of both operations is the change of size of the objects. The dilation is partly complementary to erosion, in the sense that it restores approximately the original size of objects diminished by erosion, when the same mask (structuring element) is used in both successive operations, though transposed in the second operation with respect to the first. Similarly, the cascade of dilation–erosion also restores basically the original shapes and sizes of greater objects. These two combined operations are therefore most often used in morphological image filtering. Thanks to the nonlinearity of both, the objects (or holes) that have been removed in the first phase of processing will not reappear after the second shape-restoring phase. The filtering effect is thus preserved without substantially influencing the useful objects. These conclusions apply to the gray-scale images as well.

Morphological opening $\mathbf{O}(\ldots)$ is an operator formed by the cascade of erosion and dilation,

$$g = \mathbf{O}_H(f) = \mathbf{D}_{\check{H}}(\mathbf{E}_H(f)). \tag{13.84}$$

Though usually not mentioned, as most of the masks are symmetrical, the condition of using the transposed structuring element in the second operation is substantial for obvious reasons. It can easily be seen that the opening is then independent of the position of the mask origin. Opening is an antiextensive operation—active pixels may only remain or be removed. The binary opening can also be

interpreted differently, based on the "question" approach. The question remains the same as for erosion: Does the mask fit the set in f? However, the response is different: while for erosion only the pixel under the origin of the mask is set, now the complete mask set range must be assigned with ones,

$$\mathbf{O}_H(X) = \bigcup \{H_{\mathbf{x}} | H_{\mathbf{x}} \subseteq X\}. \tag{13.85}$$

This visual description may be helpful when considering use of opening; it says that the opening filters (selects) the objects from inside of the objects. Opening is an idempotent operation—after the first use, further applications do not change the result; no iteration thus makes sense.

Morphological closing $\mathbf{C}(\ldots)$ is an operator formed by the cascade of dilation and erosion,

$$g = \mathbf{C}_H(f) = \mathbf{E}_{\check{H}}(\mathbf{D}_H(f)). \tag{13.86}$$

The condition of using the transposed structuring element in the second operation is again substantial, for the same reasons. Closing is extensive operation—active pixels remain and some may be added. The closing is also invariant to the position of the mask origin (i.e., translation of the structuring element), and can also be interpreted based on the "question" approach. The question is now different: Does the mask fit the *background* set in f? The positive answer leads to setting all pixels under the mask to 0; i.e., add them to the background X^C,

$$g = \mathbf{C}(f) = \left[\bigcup \{H_{\mathbf{x}} | H_{\mathbf{x}} \subseteq X^C\} \right]^C. \tag{13.87}$$

This visual description says that closing, contrary to opening, filters the image set from outside, i.e., from the background. Closing is also an idempotent operation—multiple applications have identical results as a single use.

It may be summarized that the opening removes all objects (or their parts) that cannot accommodate the mask set, while the closing assigns ones to all background structures that cannot accept the mask set, i.e., adds the background pixels to the objects (Figure 13.28). This statement also applies to gray-scale images; the relative intensity then decides if a particular image area should be considered an object (higher intensity) or the background (lower intensities). Therefore, should dark noisy pixels be removed from a brighter object, closing may be the solution and vice versa. It can be

Figure 13.28 Binary (above) opening and (below) closing of images of Figure 13.24.

proved that opening and closing are dual operators with respect to set complementation.

13.3.2.4 Fit-and-Miss Operator

Traditionally called *hit-or-miss* operator*, this operator is defined only for binary images. The operator is defined by a composite mask containing two disjoint structural elements H_F and H_M with a common origin. It may be imagined that both sets H_F and H_M are fixed to the mask and are gradually moving above the input image in the standard way. To provide a unit (active) output at a position **x**, two

*The name *fit-and-miss*, suggested in [70], is more descriptive and is therefore preferred here.

conditions must be simultaneously fulfilled: the set H_F being accommodated by the input set X, while the set H_M must not hit X; i.e., it must be accommodated by the background X^C,

$$g = \mathbf{F}(X) = \left\{ \mathbf{x} | (H_F)_{\mathbf{x}} \subseteq X \wedge (H_M)_{\mathbf{x}} \subseteq X^C \right\}. \quad (13.88)$$

An alternative definition expresses the *fit-and-miss transformation* (FMT) as the intersection of respective erosions,

$$g = \mathbf{F}(X) = \mathbf{E}_F(X) \cap \mathbf{E}_M(X^C). \quad (13.89)$$

The operator may look for an occurrence of a particular local binary image, expressed by the mask, in X. In this case, the union of the sets H_F and H_M fills the mask completely. However, more generic use is possible, when the union need not be continuous, namely in thinning and thickening algorithms.

13.3.2.5 Derived Operators

Top-hats $\mathbf{W}(f)$ and *bottom-hats* $\mathbf{B}(f)$ are operators simply defined as differences between the original image and one of the morphologically derived images,

$$g = \mathbf{W}_H(f) = f - \mathbf{O}_H(f), \quad \text{and} \quad g = \mathbf{B}_H(f) = \mathbf{C}_H(f) - f. \quad (13.90)$$

Obviously, the top-hats (or bottom-hats) are a set of subsets that were removed from the image by the operation of opening (or closing), respectively. The first operator, also called white top-hats, selects the high-intensity details refused by the opening; the other, black top-hats, offers the details with the underaverage intensity. In both cases, the detection of small objects based on the negative definition—the refusal by the method of opening or closing, the properties of which are well known—might be easier than formulating the properties of these small objects positively.

Thinning and thickening are operations aiming at improving the shape of thin objects by adding or subtracting some pixels at the object perimeter, perhaps in a certain direction only. Both operators, determined only for binary images, use the result of fit-and-miss transform; a direct generalization of the procedures for gray-scale images is not possible as the FMT itself is defined only for binary sets.

Thinning is given by subtracting the set obtained by FMT from the original image set,

$$g = \mathbf{T}_{H_F, H_M}(X) = X \setminus \mathbf{F}_{H_F, H_M}(X); \quad (13.91)$$

the structuring sets H_F and H_M have to be designed so that pixels at proper positions are removed. The origin of the mask must be a member of H_F. Several sets of masks (containing both H_F expressed by ones and H_M expressed by zeros, stars are nonchecked positions), called Golay alphabet, have been suggested (see [72]); examples that may be extended by rotating the content of the mask are

$$\begin{bmatrix} 0 & 0 & 0 \\ * & 1 & * \\ 1 & 1 & 1 \end{bmatrix}, \begin{bmatrix} * & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 1 & * \end{bmatrix}, \cdots \quad \text{or} \quad \begin{bmatrix} * & 1 & * \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \begin{bmatrix} 0 & * & 1 \\ 0 & 1 & * \\ 0 & 0 & 0 \end{bmatrix}. \tag{13.92}$$

Obviously, there is a close relation with the *ad hoc* designed line-thinning algorithms in Section 13.1.2. One step of thinning with the alphabet consists of applying sequentially all eight masks of a type. Such steps may be repeated, ultimately leading to an approximation of the skeleton (see below). Thinning may be roughly interpreted as a kind of conditioned erosion.

Thickening $\mathbf{R}(\ldots)$ is a dual transformation to thinning,

$$\mathbf{R}_{H_F, H_M}(X) = \left(\mathbf{T}_{(H_F, H_M)^C}(X^C) \right)^C; \tag{13.93}$$

it is the union of the original set and the result of FMT,

$$g = \mathbf{R}_{H_F, H_M}(X) = X \cup \mathbf{F}_{H_F, H_M}(X). \tag{13.94}$$

The structuring sets H_F and H_M are to be designed so that pixels in the resulting image at proper positions are added; obviously, the mask origin must be a part of the set H_M.

A note on *thinning and thickening of gray-scale images*: Conditional expressions based on dilation and erosion of f have been designed (e.g., Beucher, see [70]) that are declared to reduce to formulae (Equations 13.91 and 13.93) for binary images. Only the formulae without derivation will be presented here. For thinning, it is

$$\mathbf{T}_{H_F, H_M}(f)|\mathbf{x} = \begin{cases} \mathbf{D}_{H_M}(f)|\mathbf{x} & \text{if } \mathbf{D}_{H_M}(f)|\mathbf{x} < f(\mathbf{x}) \ \wedge f(\mathbf{x}) = \mathbf{E}_{H_F}(f)|\mathbf{x} \\ f(\mathbf{x}) & \text{otherwise} \end{cases}, \tag{13.95}$$

while for thickening, it is

$$\mathbf{R}_{H_F,H_M}(f)|\mathbf{x} = \begin{cases} \mathbf{E}_{H_M}(f)|\mathbf{x} & \text{if } \mathbf{D}_{H_F}(f)|\mathbf{x} = f(\mathbf{x}) \wedge f(\mathbf{x}) < \mathbf{E}_{H_M}(f)|\mathbf{x} \\ f(\mathbf{x}) & \text{otherwise} \end{cases}. \tag{13.96}$$

13.3.2.6 Geodesic Operators

A new area inside the field of morphological processing has been opened with the advent of *geodesic transformations*. These transforms are different from the standard morphological operators in several respects. They do not use specified structuring elements (except for elementary isotropic masks of unit size); instead, the resulting (effective) spatially variable structuring elements are locally formed during the transformation. The geodesic transformations are based on a couple of images, sc. *marker* and *mask* images, rather than on a single image and a structuring element. Both images are of the same geometrical size. The marker image is processed in a specified manner, while the mask image serves as a certain guide to control the processing of the marker. Geodesic operators (transforms) are defined for gray-scale images; binary sets interpreted as binary images are included as a special case. The basic operators are geodesic dilation and geodesic erosion. Based on them, other operators, namely iterative transforms, are formulated.

Geodesic dilation ($\mathbf{D}_G$) may be simply defined as follows: the marker image f must be smaller than or equal to the mask image g, $f \le g$, in the sense of the definition (Equation 13.61). A single step of the $\mathbf{D}_G$ (or $\mathbf{D}_G$ of size 1) consists in processing the marker by the isotropic dilation of size 1, while the result is constrained to remain under the surface of g, i.e.,

$$\mathbf{D}_G^{(1)}(f) = \mathbf{D}^{(1)}(f) \wedge g. \tag{13.97}$$

This is realized as a two-step procedure: a simple dilation followed by point-wise minimum operation between the mask image and the marker image. The geodesic dilation of size n, of a marker image, is provided by repeating the size 1 $\mathbf{D}_G$ steps iteratively,

$$\mathbf{D}_G^{(n)}(f) = \mathbf{D}_G^{(1)}\left(\mathbf{D}_G^{(n-1)}(f)\right). \tag{13.98}$$

It should be noted that this geodesic dilation of size n is not identical to the constrained (conditional) standard dilation of the same size; it can be shown that $\mathbf{D}_g^{(n)}(f) \leq \mathbf{D}^{(n)}(f) \wedge g$.

The *geodesic erosion* ($\mathbf{E}_G$) is defined as a dual transformation to $\mathbf{D}_G$: the marker f must be greater than or equal to the mask image g, $f \geq g$. A single step of $\mathbf{E}_G$ is the size 1 erosion constrained to remain above the g surface,

$$\mathbf{E}_G^{(1)}(f) = \mathbf{E}^{(1)}(f) \vee g\,, \tag{13.99}$$

which is again realized as a sequence of the standard size 1 erosion and the subsequent point-wise maximum operation. The size n $\mathbf{E}_G$ is then provided by the iteration

$$\mathbf{E}_G^{(n)}(f) = \mathbf{E}_G^{(1)}\left(\mathbf{E}_G^{(n-1)}(f)\right). \tag{13.100}$$

It can be proved that both geodesic dilation and geodesic erosion, when applied repeatedly in the iterative manner, finally reach a stable state when further application does not lead to any change. This leads to the definition of *morphological reconstructions*.

Reconstruction by dilation $\mathbf{R}_G(\ldots)$ is the operator trying to reconstruct the mask image g from the marker image f as well as possible by repeating the geodesic dilation until stability,

$$\mathbf{R}_G(f) = \mathbf{D}_G^{(N)}(f), \qquad \text{where} \qquad N:\ \mathbf{D}_G^{(N)}(f) = \mathbf{D}_G^{(N+1)}(f)\,. \tag{13.101}$$

The effect of the transform on f may be alternatively (and vaguely) described as filling the space under the g surface (the subgraph of g) by the subgraph of f as much as allowed by the g surface character with respect to the dilation properties.

Similarly, the *reconstruction by erosion* $\mathbf{R}_G^*(\ldots)$ is the operator reconstructing the mask image g from above via processing the marker image f by geodesic erosion until stability,

$$\mathbf{R}_G^*(f) = \mathbf{E}_G^{(N)}(f), \qquad \text{where} \qquad N:\ \mathbf{E}_G^{(N)}(f) = \mathbf{E}_G^{(N+1)}(f)\,. \tag{13.102}$$

Greatly varied and surprisingly effective operators can be designed using the geodesic reconstruction transforms that may replace more demanding or conceptually complex linear processing. Suggestions for the detailed design of such filters and analyzers are discussed and illustrated by examples in [13] and also in [70] and [72].

13.3.3 Some Applications

Though the morphological transforms are mathematically well formulated so that they can be deeply analyzed, the design of the morphological operators as nonlinear systems for a concrete task is still rather difficult, as formalized synthesis procedures are more or less lacking, as they are for most nonlinear processing. It is therefore necessary to rely on intuitive and *ad hoc* design that can be analytically evaluated; despite this, the designed algorithms have to be verified experimentally and usually modifications introduced, or parameters adjusted. More than classical approaches, the morphological operators require a high level of experience and intuition from their applicators, should the operators be well adapted to a particular problem and yield the best possible results. Nevertheless, the morphological operators are now firmly positioned in the image processing field thanks to their effectiveness both in the sense of low computational requirements and as to the image processing effects that may even supersede those of classical linear local operators.

We shall only briefly mention several possibilities for using the morphological operators for particular tasks in image processing. Though effective low-level image processing applications like filtering or edge revealing are frequent, most uses of morphological transforms belong rather to image analysis (like segmentation applications), including higher-level analysis tasks. The latter are already mostly beyond the frame of this book. Interested readers should refer to the above-mentioned literature, where many commented examples and further references can be found.

Numerous simple applications in image filtering and edge enhancement can be found in literature, mostly *ad hoc* adapted to a particular type of images. Other frequent applications of morphological methods are counting and labeling of objects (possibly via ultimate erosion of objects until each collapses into an isolated point), measuring total object area, evaluating shape and orientation of objects by iterative directional erosion, etc. Utilizing repeated erosion by a mask, consisting of two distant points with variable distance and direction, periodicities may be revealed.

An interesting problem is to find a *skeleton* of an object. A skeleton is vaguely defined as a "central line" representation of the object—the curve that goes in the middle of each branch of the object, thus also indicating holes, branching, disconnections and gaps, etc. A slightly more precise and vivid definition formulates

the skeleton utilizing the vision of the (binary) object as a field of grass that is simultaneously set on fire at its complete boundary. The grass fire is then supposed to propagate inside with a constant speed, thus forming the respective fire fronts. The point locus where the fronts meet and the fire thus extinguishes is considered the skeleton. Still another formulation defines the skeleton as the point locus of centers of all possible circles ("maximal balls") just touching the opposite (in a very wide sense) sides; this is obviously connected with the definition formulating the skeleton as the point set, the members of which have identical distances to (at least) two different border points. The last formulation is directly connected with the above-mentioned distance transform: the ridges of this transform correspond to the skeleton curves. Probably the most effective contemporary algorithms of skeleton analysis are based on this idea and utilize preliminarily determined distance maps.

However, many other algorithms exist, among them conceptually simple iterative thinning, as described by Equation 13.91, which in turn is based on fit-and-miss transform. It can be shown that under certain conditions concerning the used masks, the iterative use of the thinning algorithm finally leads to stability (further thinning does not change the image). As the thinning does not remove thin lines that ultimately remain as the fixed residua after thinning all object branches, the line image obtained this way may also be considered a skeleton. In Figure 13.29, the skeletons of two binary images are presented. The input image on the right-hand side is derived by closing from the left image and therefore has smoother borders of the objects. Notice how small bays, gaps, and holes substantially influence the topography of the skeletons.

Morphological operations are also used in different stages of *image segmentation*; e.g., watershed segmentation as formulated in Section 13.2.2 can be performed alternatively by relying on morphological operators. The geodesic transforms and the distance transform find effective applications in this area.

Morphological gradients are operators that detect abrupt changes in pixel intensities by comparing two of the three images: the original image, its erosion and dilation. They utilize the property of the erosion or dilation, locally producing minima or maxima that may be compared mutually or with the local values of the original. Obviously, the operator reacts to the range of values in the neighborhood, the size of which is defined by the mask, rather than to the slope of intensities, as is done by the linear-gradient type detectors.

Figure 13.29 Below: Skeletons of a binary image and its closing (above). Note the influence of border complexity on the skeletons.

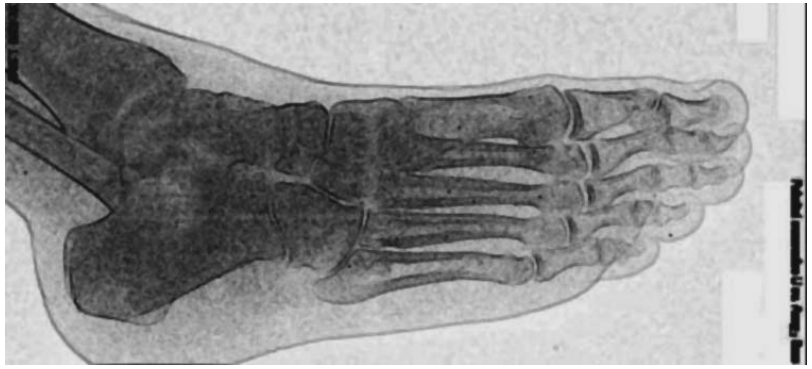


Figure 13.30 Morphological (negative) gradient of the original image of Figure 13.25.

The basic *gradient operator* subtracts the erosion image from the dilation image,

$$g = \mathbf{G}_H(f) = \mathbf{D}_H(f) - \mathbf{E}_H(f). \tag{13.103}$$

The thickness of the detected (originally step-formed) borders is two times the depth of erosion/dilation—at best two pixels; the contour is to be located in the middle of the stripe. Alternately, the *morphological half-gradients* may be used,

$$g = \mathbf{G}_{\mathbf{H}}^{-}(f) = f - \mathbf{E}_H(f), \qquad \text{or} \qquad g = \mathbf{G}_{\mathbf{H}}^{+}(f) = \mathbf{D}_H(f) - f. \tag{13.104}$$

They both provide a single-pixel-wide border each, the first (*internal gradient*) at the internal margin of the object (Figure 13.30) and the other (*external gradient*) from outside, in the background region.

A *granulometric sequence* is a sequence of openings with gradually increasing size of H, with the aim to determine the content of objects of different sizes. The philosophy behind this approach is as follows: the sequence of operators acts in analogy with sorting grains of a grainy material (e.g., rocks or potatoes) by sifting with sieves with differently dense mesh. Initially, the densest mesh is used to separate the finest grain; gradually, the mesh used is coarser, and the last sieve allows the passing of the greatest acceptable objects, while the rejected large objects form the rest. The analogically organized processing sequence is as follows:

$$\begin{aligned} g_1 &= \mathbf{O}_{H_1}(f), \\ g_{i+1} &= \mathbf{O}_{H_i}(g_i), \quad i = 1, 2, \ldots, N-1. \end{aligned} \tag{13.105}$$

While the choice of the opening transform as a sieve automatically fulfills some requirements concerning the type of operator, the selection of the gradually increasing masks (structuring elements) H_i needs to also meet certain theoretical conditions [70], [13]; naturally, it should follow size dependence suitable for a given problem. The usually used square, line, or circular structuring elements all meet the needed theoretical properties.

The individual images of the sequence $\{g_i\}$ may be further analyzed to determine the total number of particles on the image, the total area covered by objects, or the difference describing the loss of the object area between successive steps of sieving. This way, *granulometric curves*, i.e., plots of the mentioned quantities vs. mask size, are provided. This method is particularly suitable for microscopic images of, e.g., blood cells and other textured tissues or materials.

14

Medical Image Processing Environment

In the previous chapters, the concepts of the principal methods used in medical image processing were introduced. The present chapter will point to related areas of the field that are beyond the main scope of the book, but deserve mentioning in the wider context of medical image processing and analysis. Different from the previous chapters, the purpose is not an in-depth treatment, but rather only discussion of relationships to the working environment and to perspective applications. The remarks will concern some features of the used hardware and software and will also briefly comment on the current trends in analyzing, archiving, and communicating the medical images. In this frame, the important aspect — the image compression methods — closely connected with the concepts forming the basic content of the book and not included in the previous chapters, will be dealt with briefly.

Of the references to Part III, the main sources for this chapter are [41], [4], [57], [6], [34], [35], [14], [48], suggested also for further reading together with the numerous references therein; useful additions may also be found in more generic books, e.g., [26], [22], [12], [71], [3].

14.1 HARDWARE AND SOFTWARE FEATURES

14.1.1 Hardware Features

The enormous development in availability of distributed computing power has led to an approach completely different from the concepts applied some 20 years ago. Instead of using (and maintaining) huge mainframe computers, possibly connected to terminals of a local network, the main workload has been shifted to the physically small but powerful local working stations that can do practically all image processing and analysis tasks in a reasonable time. Nowadays, even the power of personal computers may be sufficient for this purpose, providing that the internal RAM is large enough (in the range of units of gigabytes); the processors working within the gigahertz range may even fulfill the requirements of relatively complicated processing. Whether the archiving possibilities are also partly distributed or completely concentrated on the centralized servers depends on the information system architecture; however, with today's easily available hard disc capacities in hundreds of gigabytes, at least a temporary store of, e.g., daily work can easily be realized.

Naturally, the conversion of raw measurement data to images is accomplished mostly locally by computers embedded in the imaging systems; this part of processing is accessible to users usually only via adjustment of some imaging and processing parameters — otherwise (as for the methods and algorithms applied) only occasionally, mostly in research centers. In the dedicated embedded computers, the use of specialized hardware — matrix computers or pipelined structures, signal processors, etc. — may be expected. However, the large area of image restoration, enhancement, and low-level analysis is in principle available to the end user and his local computing possibilities. The character of the work with the image data differs substantially when some experimenting or research is being done, and when the routine serial image evaluation by a radiologist should be performed; this difference should be reflected in the required response times. While a delay of seconds or even minutes may be acceptable when experimenting, any such delay longer than that appearing when working with classical films would lead to distraction of the radiologist and, under the pressure of the serial evaluation tasks, consequently to mental stress and fatigue. This should be considered when the needed local computing power (and also the communication capacity of the connecting network) is estimated.

Particular care should be devoted to image display. The classical hard-copy film attached to the lightbox remains the golden

standard of the quality of presentation and possibilities of inspecting. The soft-copy presentation on monitors should provide a comparable environment. This includes a possibility for inspecting multiple images (either on a large high-resolution monitor or on a couple of monitors). Though the possibility of showing images sequentially on a single monitor, offered by the electronic presentation, may be advantageous when a series of correlated images is presented (as, e.g., a series of computed tomography (CT) or magnetic resonance imaging (MRI) slices), it should not be considered an equivalent to the tiled presentation on at least two monitors, as required for simultaneous evaluation of different projection images.

Two types of monitors are currently used: cathode-ray tube (CRT) and liquid-crystal display (LCD). We will not discuss the difference in the physics of displaying principles; let us only point to the difference in the source of light: the CRT screen is itself the source of light, so that aging of CRT, which is the most expensive part of the monitor, leads (besides other deterioration) to lessening of brightness. The LCD monitor, on the other hand, is backlit by a fluorescent lamp that is relatively cheap and can easily be replaced; this promises a substantially longer service life of LCD monitors. Other differences important from the image quality point of view will be mentioned below where appropriate.

The important issue with monitors is their spatial *resolution*. The present standard image monitors offer resolution in the range from about 1 megapixel (1280×1024 pixels) to about 5 megapixels (2560×2048), occasionally more. Naturally, a higher resolution is desirable, but economical limits lead to compromises. As the monitors can only be economical when mass-produced (namely, modern LCD monitors), only the above range may be considered reasonably available, of which only the highest end corresponds to the resolution of film on the lightbox when viewed from about 0.5 m distance. To achieve the film resolution when viewed more closely, a possibility to look closer at the image displayed on a monitor via zooming is required — either locally ("magnifying glass") or on the total display area. In the latter case, easy navigation via an alternate view of the complete image with the indicated position of the zoomed area is needed.

As for the *brightness* of the display, the film on the lightbox is mostly still remarkably (up to 10 times*) brighter than most electronic

*However, the human sight sensitivity is approximately logarithmic, so that the brightness sensation is not that different.

monitors, of either CRT or LCD type. This requires well-designed illumination of a (probably darkened) room, so that the disturbance by scattered and reflected light is eliminated or strongly limited, and the sight of the image user well accommodated. Nevertheless, higher brightness is desirable to enable working in the range of the logarithmic sight sensitivity that enables good appreciation of contrast. CRT monitors may reach higher brightness at the cost of service life shortening. On the other hand, LCD monitors (namely when the color filter, unnecessary for gray-scale representation, is removed) may, without their service life being substantially influenced, in the future enable brightness comparable to lightboxes. It seems that radiologists prefer higher brightness to a high resolution when a compromise has to be made.

Of even greater importance is the *contrast range* (dynamics) of the presented image. While film copies can have a ratio of maximum and minimum brightness up to 800, the CRT monitor, though theoretically having a maximum contrast about one decimal order higher, practically achieves only about 200 due to light scatter from bright parts of the displayed image. The modern LCD monitors commonly achieve the maximum contrast ratio 400 (up to over 1000) and are thus better comparable with the film–lightbox combination; when the color filter is removed, this parameter would be improved as well.

A good *gray-scale* (contrast) reproduction means that a sufficient number of shades of gray can be well recognized on the screen. This is not too difficult to fulfill as the human sight has a relatively low resolution in this respect (only several tens of shades, about 50 under favorable conditions). However, to guarantee the same impression of the same image presented on different monitors, the usually nonlinear characteristic (brightness vs. input signal) of each monitor must be identified and linearized with respect to human perception — equal changes in the input data value should lead to equal changes in the perceived brightness. This should be done via software processing of the display data; the result of the nonlinear brightness conversion is implemented via a digital-to-analogue (D/A) converter providing the control signal for the display. Though 8-bit precision is obviously sufficient to express more than a recognizable number of gray levels, the linearization requires that the used D/A converters be of a higher-input resolution, at least 10 bits, to allow finer scaling on the digital side.

The *geometrical fidelity* of the displayed image is obviously of high importance in medical imaging; it guarantees that the shape of organs and consequently the measurements would not be distorted.

While CRT monitors require a complex sweeping circuitry to achieve a reasonable linearity, better than is usual in TV equipment, LCD monitors have a physically fixed grid of pixels that are geometrically precisely positioned; the geometrical linearity thus depends only on the production mask precision and may be considered ideal. This is an important advantage of LCD monitors.

Another important monitor property is the *screen uniformity*, in the sense of both the contrast transfer and basic brightness, and the imaging properties (local point-spread function (PSF)). While the center of the image on a CRT may be of a high quality, the imaging properties always deteriorate more or less toward margins, namely due to defocusing of the electron beams and to worse convergence of R, G, and B beams. The uniformity of phosphors is now usually provided as very good. The LCD monitors do not suffer with imaging deterioration toward the borders; the absolute pixel accuracy and focusing are maintained equally perfect during the whole service life, as given by the principle. However, LCD monitors still suffer with a low viewing angle: they provide good contrast only when they are looked at perpendicularly. A deviation from this viewing angle leads at least to darkening of the image and change of contrast, which may be visible on greater LCD monitors at margins even when they are looked at perpendicularly, but from a close position. A different viewing angle may cause the image to change the contrast completely, perhaps reversing it, as well as a complete color distortion. This is particularly disturbing when more than a single person should evaluate an image at a time. This property is improving gradually, and the best monitors can now be viewed well from a relatively wide angle (say ±45° from the perpendicular direction). In this respect, CRT monitors are better; their screen is practically a cosine radiator, and thus is visible under each angle with the same brightness, contrast, and colors.

The monitors of either type can present only two-dimensional images, though time, as a further dimension of data, can be represented via video sequences. The presentation of the third spatial dimension is possible either by means of the already mentioned sequences of slices or via three-dimensional rendering — displaying the three-dimensional objects represented by three-dimensional image data as a kind of two-dimensional views or projections. This is a task for suitable software (see below). However, a special type of three-dimensional imaging — *stereo vision* — requires a hardware component, which should provide individual images for each eye of the observer. This can be done by a special head-mounted display

(helmet) containing two small independent CRTs or LCDs, which is a rather uncomfortable and expensive solution, used commonly only in virtual reality applications (see below). An alternative is to display both images of the stereo pair alternatively or simultaneously on a single display while allowing only the proper image to enter the respective eye, using some special selective means. When the images of the stereo couple are alternating on the display sufficiently fast (i.e., with the frame rate, some 30 to 50 complete couples per second), the continuous stereo perception can be obtained when wearing special glasses opening, in each instant, the optical path only to the proper eye via electronically steered polarization filters. Switching of the glasses must be obviously synchronized with alternating of the images on the display; the glasses thus form a special periphery of the working station. Still another alternative is simpler, but only suitable for gray-scale images: wearing simple optical glasses with disjoint band-pass filters (e.g., red and cyan) and displaying each of the images in the corresponding color. The observer then perceives subjectively the gray-scale stereo image.

When specifying the hardware to be used for medical image processing and evaluation, the general ergonomics should be considered as well with respect to the possibility that a concrete type of work will be performed for a long time. Thus, the entire user interface should be understandable, possibly simple, reliable, and well within the operator's reach. The same requirement, and perhaps even with a greater weight, concerns the software design (see below).

14.1.1.1 Software Features

The software side of image processing begins with the representation of image data. So far, we only considered the digital images in the most comfortable form of matrices of pixel values, possibly converted to vectors by scanning along columns. Such a form is not suitable for archiving or communication for two reasons. Primarily, any possibility of posteriorly identifying the image properties (patient data, date of imaging, imaging details, including positioning of the patient and possibly medication or contrast agent used, comments, etc.) is missing; this is obviously unacceptable for both practical and legislative reasons in the medical environment. Second, the files containing the matrices are very large: e.g., a gray-scale image with 12-bit-intensity representation and a medium resolution of 2000×2000 pixels requires some 8 Mbytes of memory.

Though this is not much (by today's scale) in the case of a few images, it becomes overwhelming when thousands of images are to be quickly communicated over the local (or even international) network, and tens or hundreds of thousands images must be safely stored for a long (tens of years) time. It is intuitively clear that there is much redundancy in the images, so that the useful information should be expressible by substantially smaller files.

The practical *image file formats* thus try to solve both mentioned problems. Each image file therefore contains a certain file header that carries the external information on the image as mentioned above and perhaps also the data necessary to decode the image content when it is needed. The second part of the image file contains the internal image data. In the simplest form, this may be the pixel matrix scanned along rows or columns; then we speak about bitmap formats. However, most modern image formats keep the image data in a compressed form; i.e., the image data are transformed so that they have a smaller extent (sometimes substantially), though preserving all or at least a substantial part of the image content. Naturally, both the compression and the image reconstruction (decoding) from the compressed data require relatively complex computations. Thus, the decision on whether to use the compression, and if so, what should be the degree of compression, is a compromise between the computational load and amount of memory/communication channel requirements. The compression principles and main approaches will be briefly described in the Section 14.2.

Historically, many different image formats arose that differ in header organization, compression methods used, and ordering of bits in the resulting bitstreams to be communicated or saved; their detailed description can be found in respective definitions and manuals. For an image processing user, it is important to realize that conversion programs are now available that enable conversion of the image from any standard format to any other format; naturally, with the quality that is enabled by the input and output formats—one of them may limit the resulting quality substantially. Unsurprisingly, no gain in informational content is possible when converting from a low-quality format to a higher-quality one. Most general-purpose image processing packages now contain these conversion routines ("filters"), thus enabling automatic conversion of common input formats to the internal image representation, and also export of the result of processing in any of the common image formats as chosen by the user.

Of particular interest in medical applications are the DICOM* standard image formats, determined for exchanging image data and associated information in medical environments. The DICOM standard uses open compression systems, primarily JPEG and run-length coding (RLC) as its compression standards. The newer JPEG 2000 compression method is supposed to be accepted by the DICOM standard as well. The DICOM format can be roughly described as a uniform encapsulation of other formats that enables easy image data communication among systems produced by different manufacturers and using different internal formats; it thus enables an effective connection of different components of an imaging department.

14.1.1.2 Some Features of Image Processing Software

The *image processing packages* consist of several basic parts. The already mentioned input routines enable acceptance of image files in different formats and conversion of the image data into the internal format of the particular program (unfortunately, again program specific). The main block of the package provides for organization of the internal data, enables, e.g., multiple matrices to be adjoined to a single image, thus defining layers (i.e., images that after joining together form a fused image, though usually only internally formed, and may be used, e.g., for textual description of, or drawing on, the image without damaging the image content). So-called channels have a similar form — they are gray-scale images of identical size used for individual color components (e.g., R, G, B) of a color image. Still another type of auxiliary image is represented by the masks — binary images determining the regions to which a certain operation should or should not be applied. The programs allow the defining of different areas (e.g., regions of interest or selections) by means of different tools; these might also include the segmentation of the image into meaningful parts according to a given criterion.

The core of each such *experimental program* is formed by procedures performing the algorithms according to the methods described in the previous chapters. Depending on the expertise of the expected user, the packages classify the operations, offered via a menu or a selection table, either according to their mathematical

*DICOM is the abbreviation for the Digital Imaging and Communications in Medicine committee formed in 1983 by the American College of Radiology (ACR) and the National Equipment Manufacturers Association (NEMA).

background, like in this book, or according to the appearance of the result. These procedures include, on one hand, simple operations such as primitive geometrical transforms (rotation, magnification, etc.), contrast transforms (e.g., via different precomputed lookup tables or simple functions like the gamma correction), or local operations such as sharpening or smoothing. On the other hand, complex restoration procedures may be available as preprogrammed sequences of operations, including identification of image properties as needed by the respective methods. In this case, more images (e.g., of the distortion PSF, noise realizations, etc., plus the distorted image itself) may be needed as inputs to the procedure to finally provide the resulting restored image. In the middle of the scale of complexity are the operations that require substantial intervention of, or cooperation with, the user, as different types of higher-level morphological transforms or segmentation by advanced methods. In many cases, the procedures need to point manually to specific features in the image or to adjust some parameters of the algorithms, both actions requiring a deeper knowledge of the principles of the used methods. Applying these procedures is perhaps the area of immediate practical use of the knowledge offered by the previous chapters. Naturally, the provided theoretical insight is also useful and usually necessary for the active user, when designing his or her own original approaches. Some of the packages offer comfortable possibilities of performing such new procedures, primarily as freely programmed sequences of the elementary operations, or alternatively as attached (linked) procedures in a low- or middle-level programming language. Though designing and programming such methods is already demanding work, still the packages may save a substantial part of the work that would otherwise be necessary to arrange the user interface, display routines, input and output drivers, memory management, etc. Often, the new procedures can utilize with advantage many of the offered internal operations and subroutines as their elements.

Though the book has been written with this type of experimenting reader in mind, *routine medical image processing* as practiced by radiologists should not be forgotten. Then the programs of the above type, or at least the described kind of the rich user interface, are definitely unsatisfactory. The primary requirements in any routine use are speed and simplicity. Only the most necessary image manipulation and enhancement should be available, as needed in daily diagnostics. It may be expected that the most required would be a possibility of zooming (magnification), together

with an easy roam possibility enabling positioning of the magnified window anywhere in the original image and, conversely, identification or confirmation of its position in the complete view. This corresponds to the close inspection of a detail when evaluating the classical film image. A simple contrast-and-brightness adjustment, possibly spatially limited to a previously encircled region of interest, is definitely the other necessary operation, also corresponding to a former possibility — viewing the classical film under different angles to reduce or increase the local illumination. An alternative may be a set of *a priori* designed contrast-transforming lookup tables suitable for different types of organs or tissues, among which the radiologist can switch easily. Still another possibility that is more sophisticated is a kind of semiautomatic adaptive contrast adjustment controlled by one or at most two adjustable parameters. Depending on the type of processed images and purpose of the diagnostics, simple geometrical measurements may be enabled, including the possibility of recording the drawn lines or curves and borders of the delimited areas. The ability to accept hand drawings on a separate layer of the image, as information for further processing or reevaluation by other specialists, may also be considered a useful function.

A particularly complex task is to present three-dimensional image data under simple and comfortable user control. This might be done as sequences of slice images with a user-controlled pace of alternation, possibly also reversing the order of images. Another possibility is to display the three-dimensional objects in axonometric or perspective projection, which requires fast and reliable rendering via computer graphics methods, basically based either on precomputed surfaces or on ray tracing. The first method finds its use when high-quality data enable the reliable segmentation and consequential determination of organ borders (as in CT or MRI), while noisy data, as in three-dimensional ultrasonography, are better presented by ray tracing with user-defined light-influencing properties of tissues. Whatever the rendering method, the calculation should provide the response quickly enough to enable the objects to be smoothly rotated, possibly under the user control along two or three axes, thus mediating a good grasp of the spatial information contained in the data.

In any case, the user interface for routine use should be kept as simple as possible in order to prevent any intrusion of the radiologist by the computer-related details that might lead to a loss of concentration on the diagnostic task, possibly delay, stress, and consequently

fatigue, exhausting the radiologist's energy and endangering his long-term performance. The peripheral devices mediating the user response and requirements (like mouse, joysticks, graphical tablet, or even pedal handles) should be comfortable, easy to handle, and ergonomic with respect to whole-day and long-term use. This concerns not only the physical properties of hardware but, perhaps to an even greater degree, the hardware drivers as well. An important condition for the interface to be felt as user-friendly under such conditions is the instant response. With this in mind, not only the workstation power and programming environment, but also the capacity of the channels connecting the workstation to the image database or other sources and destinations of image data, should be designed.

14.2 PRINCIPLES OF IMAGE COMPRESSION FOR ARCHIVING AND COMMUNICATION

While processing an image of several megabytes in the memory of today's workstation is not a problem, the same cannot be said about communicating the image to a remote destination via a channel with an always limited capacity, and also about archiving large numbers of images, as in a hospital information system.

14.2.1 Philosophy of Image Compression

The image data compression utilizes primarily the strong correlations inside every natural image (intraframe correlations — neighboring pixels are mostly similar in their intensities, and there may be mutual relations even among distant pixels). Moreover, images in a time series (animation) or in a spatial series (set of slices) are also mutually correlated (interframe correlation). The correlation means that it is possible, knowing a pixel value, to predict the values of other pixels, at least in a neighborhood, with a high probability of a good estimate. It would then obviously suffice to transmit only the usually small differences from the estimates, instead of full intensities, which requires less bits per pixel. The correlation therefore means that there is redundancy in a fully preserved image matrix or a sequence of matrices. This is particularly well seen on the interframe correlation: the following image of a series usually differs only slightly from the previous one, and only some small amount of correction information is thus needed to update the older image — the estimate — to obtain the new frame. Another example of obvious spatial redundancy is the

transmission of a chain of identical pixel values: instead of this, just the value and the number of repetitions suffice for exact reconstruction, so that the amount of transmission bytes is minimized. The *spatial* or *temporal redundancy* due to correlation is the first source of the compression possibilities.

Another source of possible compression is the *redundancy with respect to human sight properties*: both its spatial and time resolutions are limited, so that communicating or archiving any features beyond the resolution limits is wasting of resources (memory or channel capacity). This way, we arrive at a different definition of image fidelity: not the mathematical differences between the original and the reconstructed image, but the differences between the perception of the original and the *resulting perception* of the reconstructed image are decisive. This is the current* basic idea behind transmission of all multimedia data, including audio and video. Obviously, the decompressed signal would differ from the original, and there are no means to recover the precise form of the original image; the procedure is thus called a *lossy compression*.

The third source of possible compression is the informational redundancy in the bitstream transferring the pixel intensities (or any other values, here called *symbols*), when coded without respect to probability of individual values. Obviously, the total code length is shorter when the frequent symbols are coded by shorter bitchains, and only the rare values by longer codes, than if all the symbols were coded by a uniform-length code. The optimal code from this point of view is provided by *entropy coding*. The symbol (value) chain coded efficiently via entropic coding is fully recoverable when the code is properly designed; this type of compression thus belongs to the *lossless compression* methods.

As this book deals exclusively with still images, we shall restrict ourselves to explanation of principles of still-image data compression (from now on, we will use the shorter term *image compression*, understanding that not the image but the respective data are compressed).

14.2.2 Generic Still-Image Compression System

The complete chain of compression–transfer (or archiving)–decompression is schematically depicted in Figure 14.1. The digital image is first transformed to a shorter form utilizing the spatial correlations

*And novel, in comparison with the LMSE and similar methods.

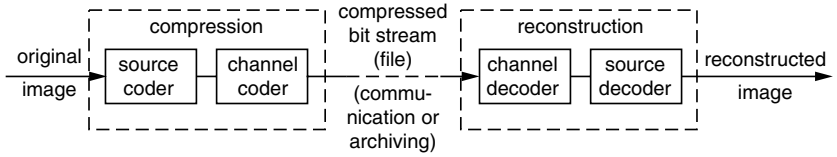


Figure 14.1 Standard chain of image compression and decompression.

and perceptual limitations. This transform must conform to the criterion of perceptual fidelity to the required extent; whether there is full recoverability or not depends on the type of transform that is denoted as *source coding*. Usually, the source coding is a lossy transform. The reduced amount of data obtained by source coding is further submitted to a usually lossless *channel (entropy) coding*. The cascade of the source coder and entropic (channel) coder forms the *encoder* or *compressor*. The resulting compressed series of binary codes — sc., *bitstream* — is then communicated via a suitable digital channel or saved in a digital archive. On the receive (or retrieval) end of the chain, the bitstream has to be first decoded by the channel decoder, dual to the entropic encoder on the source side. The obtained source code is subsequently decoded by the source decoder, approximately complementary to the source coder; the cascade of both partial decoders is the *decoder or decompressor*. As a result, we obtain the reconstructed image that should be perceived as a close approximation of the original when a lossy compression method is used, or is identical with the original when the compression is lossless (when disregarding the rounding errors).

The compression–decompression procedure can be characterized by the following properties or parameters:

- *Compression ratio*: The rate of original vs. compressed file lengths; this is the basic parameter evaluating the compression efficiency.
- *Quality* of image after decompression (theoretically absolute for lossless compression), and the types of distortion (some types are rather annoying, while others, though perhaps with larger differences from the original, are barely visible); the ultimate criterion is whether the distortion is acceptable for a particular task.
- *Complexity* of coding and decoding algorithms and their memory requirements; this influences the response time in duplex transmission or on retrieval.

- *Error robustness:* Describes how false bits in the bitstream due to transmission noise influence the quality of the decompressed image.
- *Scalability:* Evaluates whether the compression method enables the transfer or decoding of the image gradually, progressively improving the spatial and dynamic resolution; this way, only partial information is transmitted or decoded when it suffices.
- *Repeatability*: Determines how the quality deteriorates when compression–decompression is applied repeatedly.
- *Compatibility* with other compression systems: How the decompressed image behaves when submitted to another compression method; how an image compressed and decompressed by another system would be processed by the present system.

All these aspects should be taken into account when selecting the compression method for a particular purpose.

14.2.3 Principles of Lossless Compression

The methods of lossless compression are known and utilized a long time in the generic coding area. We shall therefore limit the explanation to the main ideas, without going into details. This application is also why the vocabulary is tied to the communication field: the image is considered to be a *message* (a word) of a finite length, consisting of *symbols* (often corresponding to pixel values) that are elements of a finite set called *alphabet* (e.g., all forms of a byte, thus 256 different symbols, possibly meaning the gray shades).

The simplest (and probably historically oldest) heuristic method of this kind is the *run-length coding*: instead of transmitting a chain of n identical symbols (a run), the symbol and the count n are sent (when n is limited to $n < 256$ and each symbol is expressed by a single byte; instead of n bytes only two are transmitted). Thus, when the average run length is (substantially) larger than 2, a (substantial) shortening of the finite code is achieved. Though it is not optimal, this method is still often used in coding of linearly scanned images thanks to its implementation simplicity.

All the modern lossless compression methods are based on *entropy coding*, the principle of which is as follows. Let the probability distribution among the symbols of the alphabet $\mathbf{s} = \{s_1, s_2, \cdots, s_N\}$ be $\mathbf{p} = \{p_1, p_2, \cdots, p_N\}, \sum_{i=1}^{N} p_i = 1$. The digital representation of an image will be considered a word in this alphabet. If there is no

important context among the symbols of the alphabet, the amount of information carried by a single symbol of the alphabet is

$$I(s_i) = \log_2 \frac{1}{p_i} \quad \text{[bits per symbol]}. \tag{14.1}$$

This Shannon's measure of information has been suggested based on the idea that the information obtained by receiving a symbol is higher the lower is its probability (if only one symbol was possible, $p_k = 1$, the respective information is clearly 0). The mean information per symbol of the alphabet **s** is obviously

$$H(\mathbf{s}) = \sum_{i=1}^{N} p_i \log_2 \frac{1}{p_i} = -\sum_{i=1}^{N} p_i \log_2 p_i \quad \text{[bits per symbol]}; \tag{14.2}$$

this quantity is called the *entropy* of the alphabet. It can easily be shown that the entropy is maximal when the probability distribution is uniform, while it becomes zero if any individual probability is 1.

One of the fundamental results of the information theory is the *coding theorem* (here presented without a proof): when a message of the length L in the alphabet **s** with the distribution **p** is to be losslessly coded in a binary alphabet, the minimum length of the binary code (the number of needed bits) is

$$LH(\mathbf{s}) \quad \text{[bits]}. \tag{14.3}$$

In other words, one symbol of the message (in **s**) needs on average $H(s)$ bits.

This has immediate consequences for efficient coding: the shortest code for the word of a fixed length (in **s**) will be maximal when the distribution of the symbol probability of **s** is uniform. The more uneven (compact) is the distribution, the shorter will be the code. Therefore, if it is possible to find a transform that would provide a new (informationally identical) message of the same length in an alphabet with lower entropy, the minimal code length for the transformed word will be shorter. All the formalized methods of lossless image compression thus aim at finding coding systems that utilize a description of the image in a low-entropy alphabet.

14.2.3.1 Predictive Coding

The commonly used transform of this kind is *predictive coding*, which expresses the image not by the pixel intensities, but instead by the corrections to values predicted using some extrapolation formula based on the already coded pixels. The corrections can be shown to have substantially more compact distribution* than the original pixel values of natural images. It can be immediately seen in the example of the elementary *delta-coding* when the scanned image samples are expressed by differences between neighboring pixels (i.e., the predicted value of a newly coded pixel is simply the intensity of the previous pixel). As it can be intuitively estimated, the probabilities of low correction values are much higher (due to similarity of neighboring pixels) than those of higher values; the probability thus becomes highly uneven and the entropy is lowered. Other, more advanced types of predictive coding can be found in most still-image compression schemes; mostly they use linear prediction,

$$x_{i,k} = \sum_m \sum_n w_{i-m,k-n} X_{i-m,k-n}, \tag{14.4}$$

based on a certain neighboring range of previously coded pixels X. The well-known JPEG standard uses, in its lossless part, a simple prediction based on the causal neighborhood formed by the three closest already coded pixels: left ($X_{i,k-1}$), upper ($X_{i-1,k}$), and left-upper ($X_{i-1,k-1}$) neighbor. The prediction $x_{i,k}$ of a real value of a pixel $X_{i,k}$ may be selected by the user out of seven possibilities:

$$\begin{aligned}
&x_{i,k} = X_{i,k-1}, \quad x_{i,k} = X_{i-1,k}, \quad x_{i,k} = X_{i-1,k-1}, \\
&x_{i,k} = X_{i,k-1} + X_{i-1,k} - X_{i-1,k-1}, \quad x_{i,k} = X_{i,k-1} + (X_{i-1,k} - X_{i-1,k-1})/2, \\
&x_{i,k} = X_{i-1,k} + (X_{i,k-1} - X_{i-1,k-1})/2, \quad x_{i,k} = (X_{i,k-1} + X_{i-1,k})/2.
\end{aligned} \tag{14.5}$$

The chosen formula that is then used for the complete image is indicated in the header of the compressed file. The value to be coded is the prediction error, $e_{i,k} = X_{i,k} - x_{i,k}$. Obviously, the first predictor corresponds to the simple delta-coding; the more complicated formulae may provide slightly better compression (by ~10%).

*The distribution is approximately Laplacian, with zero mean, regardless of the original pixel intensity distribution.

It is then required to find a binary coding of the new alphabet such that the shortest codes correspond to the most probable symbols and the code length increases with the lowering probability; such a method is called *entropy coding*. The overall length of the binary code is thus minimized (or approaches the minimum given by Equation 14.3). The most commonly used systems of nearly optimum entropy coding are *Hufmann coding* and *arithmetic coding*, or their combinations and modifications. Alternatively, *Rice–Golomb coding* may be used. Though simple, the descriptions of these generic coding methods, which are not specific for images, will be omitted; they can be found in any book on communication, coding, or information theory (see also [35]).

Let us conclude that entropy coding can generally compress the data quite remarkably, while the message remains completely recoverable, as it is a lossless compression. Similar methods are used to compress any data files, thus obtaining the compression ratio in the range of about 1.5 to 3.

14.2.4 Principles of Lossy Compression

Lossy compression of still images utilizes the redundancy of the image data with respect to the human sight limitations in both the spatial and dynamic (e.g., with respect to brightness and color) resolutions.

The theoretical considerations concerning lossy compression may be based on the distortion D objectively defined as the mean quadratic error of pixel intensities,

$$D = \frac{1}{M}\sum_{i=1}^{M}(X_i - \hat{X}_i)^2, \tag{14.6}$$

where M is the number of pixels in the image, X_i the real value, and $\hat{X}_i$ the reconstructed (decompressed) value of a pixel intensity. It is then possible to theoretically find the relation between the compression ratio R and the distortion D, sc., *rate-distortion function* $R(D)$, when a model of spatial two-dimensional (or three-dimensional) correlation in the image is known (e.g., via analysis of typical images). The basic *theorem of lossy compression* says that there is a certain maximum achievable compression ratio $R_{\max}$ for a given distortion D and vice versa; it is theoretically possible to realize a compression system with this $R_{\max}$. The inspection of this function reveals in general what could be expected intuitively: the compression is more

effective when the method considers two-dimensional (or even three-dimensional) correlation, instead of only one-dimensional correlation. The other conclusion is that the attainable $R_{\max}$ depends positively on the extent of the main lobe of the correlation function and on the allowed D. The function $R(D)$ thus formulates a certain golden standard with which the real system can be compared.

However, the criterion in Equation 14.6 is not decisive for the *subjective quality* of the decompressed image. Because the subjective perception has been declared as the final criterion, the above theory has only a limited application and may be considered just auxiliary. The practical evaluation of a particular lossy compression should be based on statistically assessed experiments with groups of human observers, independently evaluating suitably selected images reconstructed via the tested compression method in comparison with the originals. The images should belong to the area to which the method is intended. The design of new compression methods is in turn also based on the psychophysiological findings determined by similar experiments and formulated, e.g., as the frequency transfer function of the sight, its dynamic characteristic (dependence of the sensation vs. the intensity of the light stimulus), sensitivity to particular types of artifact, etc.

In the following brief explanation, we shall only describe some of the principles of lossy compression approaches; the details, as well as evaluations of the individual systems with respect to different types of images, can be found in the indicated literature. The common compression methods may be classified as pixel-oriented, block- (or region-) oriented, and global approaches, the latter being applied to the image as a whole.

14.2.4.1 Pixel-Oriented Methods

The *pixel-oriented methods* are based on the same scheme using prediction, as the lossless predictive coding, with the important difference that the prediction differences are quantized before they are coded by the channel encoder (Figure 14.2). The quantization of the residual (error) $e_{i,k} = f_{i,k} - p_{i,k}$ is the lossy step that, on one hand, introduces distortion via the nonlinear quantizing function Q, providing the quantized residual $q_{i,k} = Q(e_{i,k})$, but, on the other hand, also enables shortening of the source alphabet by limiting the number of possible values of the coded differences. The number of quantization steps and their distribution determine both the distortion and the compression ratio. The predictor, working according to Equation 14.4, contains a memory

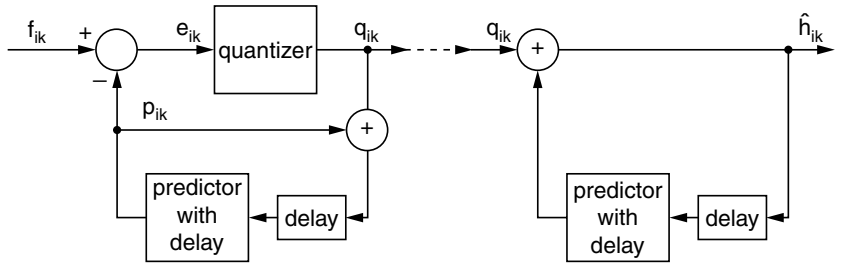


Figure 14.2 Scheme of pixel-oriented lossy compression and reconstruction.

(internal delay) of the older pixel values; the new *reconstructed* pixel value $\hat{h}_{i,k} = p_{i,k} + q_{i,k}$ can enter the predictor only in the next step; therefore, the delay must be inserted. Note that the reconstructed, not the original, values $\hat{h}_{i,k}$ are used in the encoder prediction; this corresponds to the only possible situation at the receiving end, where only reconstructed values $\hat{h}_{i,k} = p_{i,k} + q_{i,k}$ are available.

14.2.4.2 Block-Oriented Methods

The *block-oriented methods* divide the image mechanically into usually rectangular (possibly square) blocks, each of which might be coded separately as a stand-alone image without any regard to the neighboring blocks. Theoretically (and even in practice) the two-dimensional block organization can better utilize the redundancy than the pixel-oriented approach; the efficacy theoretically improves with the block size. Unfortunately, this approach brings an annoying artifact, *blocking effect*: the individual blocks of the reconstructed image do not continue smoothly into the neighboring blocks. Human sight is very sensitive to such regularly spaced discontinuities. Therefore, the block size is mostly chosen as small as $N \times N = 8 \times 8$ bits; then the artifacts usually remain under the recognition threshold when the compression is not too aggressive.

The *original (spatial) domain block-oriented compression* is realized via *vector quantization*. The ordered set of $N \times N$ pixel values is considered a vector in the vector space R^{N^2}. Based on statistical analysis of the distribution of the vectors in the space, the space is subdivided into areas such that each area contains similar vectors. A vector that represents reasonably well all the vectors belonging to a certain area is chosen as its representative; finding the suitable representatives and the respective area borders

(the codebook design) is a complex step. Once the codebook is defined, the vector quantization consists of replacing any input vector by the representative of the area to which the vector belongs. This procedure aims at obtaining a shorter transformed alphabet (of representatives), instead of a much more extensive alphabet of all possible N^2-sized vectors. Clearly, the representatives do not describe the elementary images exactly, being only their approximations; the chosen number and positioning of representatives influence the mean error as well as the compression ratio. However, it seems that block-oriented compression via frequency domain is easier and more effective.

The *block-oriented compression in the frequency domain* is presently the most successful image compression method in terms of applicability; among other systems, it also forms the basis of the lossy part of the JPEG standard. The basic idea is to transform each block image into the spectral domain by a unitary transform that provides the best possible compaction of energy into as small a number of spectral coefficients as possible. It implies that nonzero values of some spectral coefficients are more probable than others, meaning decreasing of entropy; also, the spectral coefficients should be uncorrelated, so that the redundancy is minimized. An optimum transform from this point of view is the Karhunen–Loeve (KL) transform; regrettably, it is an image-specific transform, and thus not a unified algorithm, and it lacks a fast version. Fortunately, there is a good approximation to the KL transform for most natural images: the discrete cosine transform (DCT) with compaction properties almost as good, being separable, real-valued, and having the fast algorithm. This is the reason why DCT is the most frequently met transform in the signal and image compression fields.

The principle of the block-oriented frequency-domain compression is described in the following steps:

- Divide the image matrix into blocks (usually 8×8 sized).
- Each partial block is DCT transformed, thus obtaining the partial DCT spectrum.
- The spectral coefficients are quantized, depending on the individual statistics of every spectral position; this is the lossy step, where the degree of compression can be adjusted by choosing the number of levels and character of quantization for each coefficient.
- The quantized spectrum, without zero-valued elements, is losslessly entropy coded, thus yielding the compressed binary chain (compressed file).

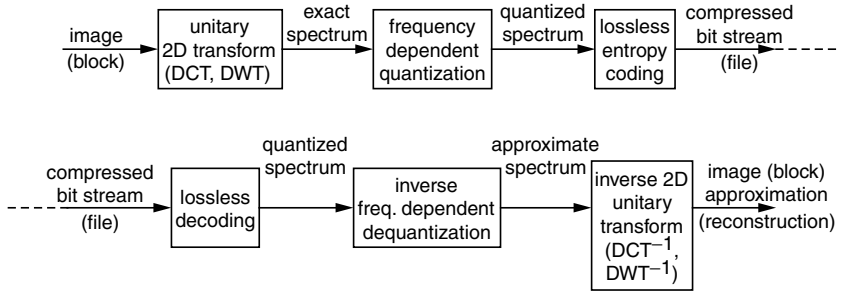


Figure 14.3 Frequency-domain lossy compression and reconstruction.

The decompression has the reverse order:

- Lossless entropy decoding of each block code.
- Approximate restoration of the spectral coefficients, providing the approximate DCT spectrum.
- Inverse (8×8) DCT.
- The whole image is reconstructed by placing the blocks into proper positions.

The steps repeated for every block are depicted schematically in Figure 14.3. Note that the block size is irrelevant at this level; we shall return to this figure when discussing global frequency-domain-based compression.

Obviously, the crucial point of the method is the determination of suitable quantization. It is, on one hand, a result of statistical analysis of the probability distributions of individual spectral coefficients for experimental sets of natural images (possibly images of a particular type). This yields the range of values for each spectral coefficient; obviously, coefficients, of which the values are always small, need a lower number of bits, even when losslessly coded. On the other hand, complex psychophysiological experiments yield the information on the relevance of individual spectral components for the subjective quality. Consequently, the less important components need only to be roughly described (i.e., only a coarse quantization is needed). The quantization is made in a simple way: the number $z_{i,k}$ representing a particular spectral coefficient $F_{i,k}$ to be coded and transferred is obtained as

$$z_{i,k} = \mathrm{int}\left(\frac{F_{i,k}}{q_{i,k}} + 0.5\right) \tag{14.7}$$

i.e., by rounding the ratio of the exact spectral value and the frequency-dependent factor $q_{i,k}$ resulting from the above statistical and psychophysiological considerations. The more important a spectral coefficient, the lower the corresponding q value; when the exact DCT coefficients are normalized and rounded into an integer range (e.g., ±1023), obviously no further quantization error would be introduced when $q = 1$. Conversely, the higher is q, the higher the allowed relative error due to rounding (but also the higher the compression). The q factors form the matrix (8×8), which may look like, e.g.,

$$\mathbf{q} = \begin{bmatrix} 17 & 11 & 10 & 14 & \cdots \\ 11 & 12 & 16 & \cdots & 80 \\ 10 & 16 & \ddots & 110 & 100 \\ 14 & \vdots & 110 & 120 & 101 \\ \vdots & 80 & 100 & 101 & 105 \end{bmatrix}; \tag{14.8}$$

the low-frequency end of the spectrum needs the highest precision, while the higher frequencies are only roughly described. Many of the z values become zero by rounding and need not be transferred. The z-matrix is scanned for encoding in a zigzag manner starting from the low-frequency end. This way, most of the (higher-frequency) zero coefficients would be at the end of the chain and may be omitted by simply stopping the chain with a special symbol; the zero sequences among nonzero values may be substituted by run-length data. Both of these replacements save the code length substantially. The compression ratio may obviously be controlled by compression adjustment via the choice of q values (or rather of a complete matrix $\mathbf{q}$); the higher the q values, the coarser the quantization and the higher the obtained compression ratio (and distortion after decompression).

At the decompressing end of the chain, the estimates of the spectral coefficients are to be provided, evidently by

$$\widehat{F}_{i,k} = z_{i,k} q_{i,k} \tag{14.9}$$

thus obtaining the approximate spectrum. The factor matrix $\mathbf{q}$ is to either be settled by convention or be a part of the compressed file (or of a set of equally compressed files).

Region-oriented compression (ROC) is another method of compressing the image by parts. In a sense, it is a modern form of a

two-dimensional lossy compression based on a similar principle as the lossy run-length coding (coding of runs of not identical, but similar values). It consists of:

- Segmenting the image into regions, each with a (almost) uniform gray shade, color, or texture
- For each of the regions:
  - Describing the segmented region geometrically by its closed border (perimeter), expressed by a suitable method, e.g., by a chain code or parametric curves (e.g., cubic splines)
  - Expressing the content of the region (shade, color, or texture) by a vector of a few parameters, based on a texture model
- Entropy coding just the border description and the parameter vector of the region content, instead of pixel values, for all regions covering completely the image area.

The decoding consists of approximately the inverse procedure:

- Entropy decoding
- For each of the regions:
  - Border reconstruction
  - Content (shade, color, or texture) reconstruction, possibly based on a conventional texture model utilizing the encoded parameters
  - Filling the regions with the artificial content

Obviously, the image content is substantially changed this way; however, the subjective impression may be close to the original when the regions are small enough and their contents expressed well. Though suitable for unambitious multimedia applications, this method is in principle naturally far from fidelity and might be used for medical images only very cautiously, perhaps for segmentation results or some other approximate regional description.

14.2.4.3 Global Compression Methods

The *global compression methods* process the complete image at once to obtain the shortened code. The primary advantage of the global methods is the absence of any blocking artifacts (or unnatural appearance of artificially formed regions in ROC). Also, other desirable features, like scalability (selectable resolution and quality), can be better implemented.

14.2.4.3.1 Pyramidal Decomposition

As for *original-domain global compression*, *the pyramidal decomposition*, as explained in Section 13.2.2 on segmentation via region splitting, may be considered an example. This approach is obviously well adapted to the requirement of selectable resolution — when a lower resolution suffices, the more detailed lower levels of the hierarchy may be omitted, this way saving on transmission. However, pyramidal decomposition itself does not offer any data compression (on the contrary, the amount of data for the pyramid is higher), and some kind of prediction difference scheme (sc., Laplacian pyramid) has to be introduced to enable more effective coding. A hierarchical approach, in a sense similar, is nowadays used in the modern wavelet-based compression schemes (see below), though formulated in the scale-space domain.

14.2.4.3.2 Subband Coding

Subband coding, as a *frequency-domain global compression* method, is one of the older approaches, which regains attention in the modern form of wavelet decomposition. The basic principle is as follows:

- Using a bank of two-dimensional filters, with ideally non-overlapping two-dimensional frequency bands completely covering the frequency extent of the image, provides the narrowband-filtered (subband) versions of the image.
- Thanks to the reduced bandwidths, the subband images may be subsampled without a danger of aliasing; this way, the total sample number remains identical to the original one (no compression achieved so far).
- The individual subband images may be coded separately; this way, a suitable quantization and coding/decoding strategy may be adjusted individually to the particular properties of each subband.

Again, the decompression phases are complementary to the compression steps; the last phase consists of summing the reconstructed (approximate) subband images together to obtain the image reconstruction. The filters are usually hierarchically arranged into pairs; a pair on a level subdivides the input frequency band into a high-pass subband and a low-pass subband, this way enabling use of quadrature mirror filters that compensate for overlaps in the frequency responses of the realizable filters. When the number of

subbands equals the number of blocks in the block-oriented frequency-domain compression, it can easily be seen that there is a close relation between the sets of frequency coefficients in both approaches.

14.2.4.3.3 Wavelet-Based Compression

Wavelet transform (WT) provides the decomposition of the image into subband images, ideally into two-dimensional octave (dyadic) bands, as described and depicted in Section 2.3.4. As the spectral domain of WT is the scale-space domain,* the decimation applies not only to frequency (scale) coordinates, but also to the spatial coordinates. The WT spectral representation therefore also has the features of the pyramidal decomposition, as visible from Figure 2.27b. Thanks to subsampling enabled by narrowing the bandwidth via subband coding, the lower-frequency components can be accommodated in smaller matrices; at each pyramidal level, just one quarter of the matrix is needed for the higher-level (less detailed) representation. Thus, the complete pyramid can be accommodated in a matrix of the original size. The wavelet transform provides for decreasing entropy in the subband images: as visible in Figure 2.27a, three high-frequency (detail) components of the spectrum, consisting of only small details or lines, are to a large extent decorrelated and their gray-scale histograms are usually much narrower, thus allowing for efficient coding. The fourth quadrant is the low-pass (approximation) version of the image, with a high degree of spatial correlation, and may be further decomposed. Mostly only partial (incomplete) wavelet decomposition on several levels is performed, as in Figure 2.27b; the pyramidal highest-level approximation image may serve as a thumbnail, e.g., in a database search.

No compression is obtained so far, but the same principle of quantizing the spectral coefficients as used in block-oriented compression may now be applied, based on psychophysiological findings concerning the importance of different wavelet components. Naturally, the matrix **q** of the quantizing dividers is of the same size as the image, in contrast to the usual 8×8 size of block-based systems; otherwise, the compression concepts (Figure 14.3) are identical. Thanks to the properties of the wavelet base, the components of

*The scale corresponds roughly to the inverse frequency, see Section 2.3.4.

which are well suited to the description of image features, the global approach to compression is well acceptable even with only a small number of reconstruction components.

The modern global WT-based compression schemes, also forming the basis for the lossy part of the new JPEG 2000 standard, have several important advantages over the older block-based schemes:

- No blocking effects
- Scalability as to multiple-resolution concerns
- Scalability of the reconstruction quality (with respect to signal-to-noise ratio (SNR))
- Region-of-interest delimiting

The first feature has the obvious explanation of the compression being global. The last possibility enables reconstruction of only a part of the image with a high-quality resolution, thus saving on the rest of the image, which serves only as the orientation environment.

Multiresolution possibilities can be understood as the result of the dyadic decomposition embedded in the dyadic WT (DWT). It is well visible in Figure 2.27: if sufficient, only a lower-resolution part of the two-dimensional dyadic spectrum (a quarter, of possibly a quarter, etc., of the complete data) needs to be transferred and decoded. Similarly, the representation quality of spectral values (and consequently of the resulting SNR) depends primarily on the number of bits preserved by quantization and also on how many of them are utilized. When the (quantized) spectral matrix is expressed in a direct binary code, it may be decomposed into binary planes expressing the individual bits of the codes. The coarsest spectrum description is obtained, when only the most significant bit-plane is utilized (thus saving the decoding of all other bit-planes); the quality naturally improves monotonously with an increase of the number of the used bit-planes.

The *scalability* in both mentioned directions allows for *progressive decoding*: while preserving the maximum information in the complete encoded bitstream, only the necessary parts of it need to be transferred and decoded according to the requirements and preferences of a concrete application. When, e.g., the bit-planes are transferred gradually, starting with the most significant plane, the reconstruction can be terminated after receiving any of the planes, thus limiting the transfer length and also the amount of reconstruction calculations (naturally at the cost of

limited reconstruction quality). Similarly, when the lowest-resolution parts of the WT spectrum are transferred first, with the larger spectral submatrices added consecutively, the transfer may be terminated after receiving enough data to reconstruct the image with the required resolution.

The wavelet transform–based compression is presently considered the most promising lossy image compression method; the achieved compression ratios are slightly higher than those offered by the classical block-based methods at similar SNR. The most visible advantage is the lack of annoying blocking artifacts; the whole-image artifacts of the WT-based compression are generally much less noticeable. On the other hand, this may be a disadvantage for applications where the objective fidelity is critical, like for medical diagnostic images. While the disturbing blocking artifacts, to which human sight is rather sensitive, may warn the user of the insufficient SNR, the hardly visible smooth artifacts due to WT compression may remain undetected, which might cause the use of misleadingly distorted images.

14.3 PRESENT TRENDS IN MEDICAL IMAGE PROCESSING

Medical imaging became a very interesting field concerning both the physical principles of the imaging modalities and the aspects of image data processing. In both directions, substantial methodological achievements have been obtained in the past few decades, substantially supported (or even at all enabled) by enormous technological development. This advancement is most probably not at its end; the imaging principles and technology, as well as the methods of image data processing and analysis, are likely to continue in their development and refining. The increasing computational power available, together with the improved data processing methodology, may allow consideration of *new imaging principles* so far regarded less promising or infeasible. To name a few, electrical impedance tomography, passive infrared imaging, transmission ultrasonic tomography, or microwave imaging might show certain clinical potential.

Concerning the standard modalities, there seems to be a strong trend in *implementing a priori medical knowledge* into the image processing procedures. A typical example of such a *model-based approach* is utilizing the spatial anatomical knowledge during the

segmentation phase of image analysis. Models based on physiological or biophysical knowledge may contribute to tissue characterization in images or imaged three-dimensional structures. The tendency to automate mechanical operations, namely by applying the knowledge-based approaches, is continuing; on the other hand, it can hardly be expected that the human factor of medical expert supervision would be eliminated from the process of image processing and evaluation.

Another expressed present trend is in *fusion of information* from two or more imaging modalitites, perhaps also combining the fused images with nonpictorial information. The functional and perfusion imaging is an excellent example of a recent qualitative breakthrough enabled by image fusion, which in turn influences the physiology and biophysics by bringing so far unknown information on correlation among different phenomena.

Expansion of *three-dimensional imaging* into areas traditionally only two-dimensional, as, e.g., ultrasonography, or replacing the classical two-dimensional modalities by their three-dimensional counterparts, which provides images that better describe spatial structures and their functions and also more closely correspond to the anatomical knowledge of medical staff, seems to be another steady trend. Conversely, but again with the aim of offering a known type view, a two-dimensional display can be derived from three-dimensional data, as in virtual (computed) endoscopy. Wherever feasible, four-dimensional imaging — i.e., three-dimensional time-dependent image sequences (cine loops, movies) — can be expected to expand to so far empty niches.

Sharing and communicating images is another strong present trend. Digital imaging enabled paperless or filmless imaging, thus not only simplifying image acquisition and manipulation (and consequently, perhaps, also the evaluation), but also transporting images easily to multiple places, thus providing direct access to different users. It seems obvious that in comparison with the usual short verbal description, as provided by a radiologist, the image carries much more information that may also be used by other specialists, or even the family physician, to support and complement the radiologist's conclusions[*]. It is technically possible to communicate medical images in a reasonable quality during an acceptable time and for tolerable costs to remote places, when utilizing modern image data compression methods. The data can be sent inside a fast

[*]It should be mentioned that not all radiological clinics accept this view.

intranet, namely in the frame of a *hospital information system*, or via the public Internet or a switched data line network, or even, using modems, via a fixed-line or cellular (mobile) telephone network — possibly even via satellites.

However, there are several serious problems connected with communicating medical images. Primarily, the medical data, including images, are sensitive *private personal data* that must be secured against unauthorized approaches; this implies complicated problems of managing the transfer and access authorization, as well as encrypting the images with the use of modern cryptographic means. This itself requires careful management of cipher distribution and update. The other side of the same problem is the authorization of images by electronic signature of the sending radiologist or clinic, perhaps also in the form of an advanced watermark — an image processing problem. It is also necessary to use standardized image formats, including the attached textual and numeric information, so that any confusion of data among patients, or of image interpretation (e.g., left–right orientation), is excluded. The current development leads to establishing such methodology and standards, on one hand reliable and robust, while on the other user-friendly.

Real or simulated medical images may also become part of *virtual reality* forming a training environment, either on a medical school preparatory level or for simulation and preparation of complicated surgical operations or for planning of treatment procedures. This is closely connected with using endoscopic imaging in minimally invasive operations, possibly also in connection with virtual reality elements. The already mentioned virtual endoscopy is a simple example of providing a vivid virtual view based on real medical three-dimensional image data. The virtual reality-based systems may also be used for ergonomics studies and as a training system for patient rehabilitation, or also as a supporting environment for the disabled. Obviously, intensive image processing is a substantial part of every virtual reality arrangement.

Telemedicine, which can be roughly defined as providing medical care remotely, is crucially dependent on image communication and processing. Primarily, the already mentioned *teleradiology* — sending medical images of a patient to an expert at a distant place for evaluation — is one of the earliest components of telemedicine, besides the naturally easier remote evaluation of ECG and other signals. Sending video sequences, namely in real time, is essential in *telesurgery*, the most demanding field of telemedicine, where the surgeon

acts indirectly, via a suitable interface (data gloves, etc.), a telecommunication link, and mechanical actuators, while tracking his action via a real-time video, possibly in high-resolution stereo vision, mediated by a reverse communication link. This may be denoted as *telepresence* of the surgeon. Obviously, both the physical distance and duplex communication channel complexity influence the signal delay, which limits the maximum range of such a service. Again, the effective image processing, namely, the video data compression, is crucial for the success of the telepresence system. Image communication, with a possibility of experimental processing and analysis, also forms one of the pillars of *teleeducation* (distant electronic study) of some medical skills, which is also considered an important part of telemedicine.

The above-mentioned are just a few examples of relatively new or emerging applications of image data processing in medicine. Should this book (with the previous chapters) contribute to a deeper understanding of the underlying principles of this highly topical technology and, by means of this, also to the intellectual satisfaction of the reader, its intended mission will have been fulfilled.

REFERENCES for Part III

[1] Bates, R.H.T. and McDonnell, M.J. *Image Restoration and Reconstruction*. Clarendon Press, Oxford, 1986.

[2] Bhaskaran, V. and Konstantinides, K. *Image and Video Compression Standards: Algorithms and Architectures*, 2nd ed. Kluwer Academic, Dordrecht, Netherlands, 1997.

[3] Bronzino, J.D. (Ed.). *Biomedical Engineering Handbook*, 2nd ed. CRC Press/IEEE Press, Boca Raton, FL, 2000.

[4] Brown, M.S. and McNitt-Gray, M.F. Medical image interpretation. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M. and Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[5] Dawant, B.M. and Zijdenbos, A.P. Image segmentation. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[6] Dhawan, A.P. *Medical Image Analysis*. John Wiley & Sons/IEEE Press, New York, 2003.

[7] Dougherty, E.R. (Ed.). *Digital Image Processing Methods*. Marcel Dekker, New York, 1994.

[8] Fessler, J.A. Statistical image reconstruction methods for transmission tomography. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[9] Fitzpatrick, J.M., Hill, D.L.G., and Maurer, C.R., Jr. Image registration. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[10] Gimel'farb, G. Stereo terrain reconstruction by dynamic programming. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[11] Gold, B. and Rader, C.M. *Digital Processing of Signals*. McGraw-Hill, New York, 1969.

[12] Gonzalez, R.C. and Woods, R.E. *Digital Image Processing*. Addison-Wesley, Reading, MA, 1992.

[13] Goutsias, J. and Batman, S. Morphological methods for biomedical image analysis. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[14] Greenleaf, W. Piantanida: medical applications of virtual reality technology. In *Biomedical Engineering Handbook*, 2nd ed., Bronzino, J.D. (Ed.). CRC Press/IEEE Press, Boca Raton, FL, 2000.

[15] Haindl, M. Texture Synthesis: *CWI Quarterly*, 4, 305–331, 1991.

[16] Hajnal, J.V., Hill, D.L.G., and Hawkes, D.J. (Eds.). *Medical Image Registration*. CRC Press, Boca Raton, FL, 2001.

[17] Haralick, R.M. and Shapiro, L.G. *Computer and Robot Vision*. Addison-Wesley, Reading, MA, 1992.

[18] Haussecker, H. and Spies, H. Motion. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[19] Haykin, S. *Neural Networks*. Prentice Hall, Englewood Cliffs, NJ, 1994.

[20] Herman, G.T. Algorithms for computed tomography. In *The Digital Signal Processing Handbook*, Madisetti, V.K., Williams, D.B. (Eds.). CRC Press/IEEE Press, Boca Raton, FL, 1998.

[21] Chrastek, R., Skokan, M., Kubecka, L., Wolf, M., Donath, K., Jan, J., Michelson, G., Niemann, H. Multimodal retinal image registration for optic disc segmentation. *Methods of Information in Medicine*, 4, 336–345, 2004.

[22] Jahne, B. and Haussecker, H., Geissler, P. (Eds.). *Handbook of Computer Vision and Applications*, Vol. 2. Academic Press, New York, 1999.

[23] Jahne, B. *Image Processing for Scientific Applications*. CRC Press, Boca Raton, FL, 1997.

[24] Jahne, B. Interpolation and image warping. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[25] Jahne, B. Local structure. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[26] Jain, A.K. *Fundamentals of Digital Image Processing*. Prentice Hall, Englewood Cliffs, NJ, 1989.

[27] Jan, J. and Janova, D. Complex approach to surface reconstruction of microscopic samples from bimodal image stereo data. *Mach. Graphics Vision*, 10, 261–288, 2001 (special issue on stereogrammetry and related topics).

[28] Jan, J. and Kylian, P. Modified Wiener Approach to Restoration of Ultrasonic Scans via Frequency Domain. Paper presented at Proceedings of 9th Scandinavian IAPR Conference on Image Analysis, Uppsala, Sweden, 1995, pp. 1173–1180.

[29] Jan, J., Sonka, M., Provaznik, I. (Guest Eds.). Special issue on modality-oriented medical image processing. *EURASIP J. Applied Signal Processing* (Hindawi), 5, 2003.

[30] Jan, J. *Digital Signal Filtering, Analysis and Restoration*. IEE, London, 2000.

[31] Jan, J. Two-Dimensional Non-Linear Matched Filters. Paper presented at Proceedings of 2nd International Conference COFAX '96, Bratislava, Slovakia, 1996, pp. 193–198.

[32] Janova, D. Reliable surface reconstruction from stereo pairs of images provided by scanning electron microscope. *Cz. J. Phys.*, 44, 255–260, 1994.

[33] Jirik, R., Taxt, T.,and Jan, J. Ultrasound attenuation imaging. *J. Electrical Engineering*, 55(7–8), 180–187, 2004.

[34] Jones, P.W. and Rabbani, M. Digital image compression. In *Digital Image Processing Methods*, Dougherty, E.R. (Ed.). Marcel Dekker, New York, 1994.

[35] Jones, P.W. and Rabbani, M. JPEG compression in medical imaging. In *Handbook of Medical Imaging*, Vol. 3, Kim, Y., Horii, S.C. (Eds.). SPIE Press, Bellingham, WA, 2000.

[36] Judy, P.F. Reconstruction principles in CT. In *The Biomedical Engineering Handbook*, Bronzino, J.D. (Ed.). CRC Press, Boca Raton, FL, 1995.

[37] Kak, A.C. and Slaney, M. Principles of Computerized Tomographic Imaging. Paper presented at SIAM Society for Industrial and Applied Mathematics, Philadelphia, 2001.

[38] Kamen, E.W. *Introduction to Signals and Systems*, 2nd ed. Macmillan Publishing Company, New York, 1990.

[39] Katsaggelos, A.K. (Ed.). *Digital Image Restoration*. Springer-Verlag, Heidelberg, 1991.

[40] Katsaggelos, A.K. Iterative image restoration algorithms. In *The Digital Signal Processing Handbook*, Madisetti, V.K., Williams, D.B. (Eds.). CRC Press/IEEE Press, Boca Raton, FL, 1998.

[41] Kim, Y. and Horii, S.C. (Eds.). *Handbook of Medical Imaging*, Vol. 3. SPIE Press, Bellingham, WA, 2000.

[42] Kosko, B. (Ed.). *Neural Networks for Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[43] Kosko, B. *Neural Networks and Fuzzy Systems*. Prentice Hall, Englewood Cliffs, NJ, 1992.

[44] Kreyszig, E. *Advanced Engineering Mathematics*, 4th ed. John Wiley & Sons, New York, 1979.

[45] Kubecka, L., Skokan, M., and Jan, J. Optimization methods for registration of multimodal images of retina. In *Proceedings of 25th Annual International Conference of IEEE-EMBS*, Cancún, Mexico, 2003, pp. 599–601.

[46] Lagendijk, R.L., Franich, R.E.H., and Hendriks, E.A. Stereoscopic image processing. In *The Digital Signal Processing Handbook*, Madisetti, V.K., Williams, D.B. (Eds.). CRC Press/IEEE Press, Boca Raton, FL, 1998.

[47] Lau, C. (Ed.). *Neural Networks, Theoretical Foundations and Analysis*. IEEE Press, New York, 1992.

[48] Lau, Ch., Cabral, J.E., Jr., Haynor, D.R., and Kim, Y. Telemedicine. In *Handbook of Medical Imaging*, Vol. 3, Kim, Y., Horii, S.C. (Eds.). SPIE Press, Bellingham, WA, 2000.

[49] Lindeberg, T. Principles for automatic scale selection. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[50] Loew, M.H. Feature extraction. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[51] Madisetti, V.K. and Williams, D.B. (Eds.). *The Digital Signal Processing Handbook*. CRC Press/IEEE Press, Boca Raton, FL, 1998.

[52] Maes, F. Segmentation and Registration of Multimodal Medical Images. Ph.D. dissertation, Katholieke Universiteit Leuven, Belgium, 1998.

[53] Mallot, H.A. Stereopsis: geometrical and global aspects. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[54] MATLAB Image Processing Toolbox version 2 (manual). The Math-Works Inc., Natick, MA, 1997.

[55] MATLAB version 5.1 (manual). The Math-Works Inc., Natick, MA, 1997.

[56] MATLAB wavelet toolbox (manual). The Math-Works Inc., Natick, MA, 1996.

[57] Niemann, H. Knowledge-based interpretation of images. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[58] Nuyts, J. Quantification of SPECT Images: Simulation, Scatter Correction, Reconstruction and Automated Analysis. Ph.D. thesis, Catholic University Leuven, Belgium, 1991.

[59] Pratt, W.K. *Digital Image Processing*, 3rd ed. John Wiley & Sons, New York, 2001.

[60] Proakis, J.G., Rader, C.M., Ling, F., and Nikias, C.L. *Advanced Digital Signal Processing*. Maxwell Macmillan International, New York, 1992.

[61] Rabiner, L.R. and Gold, B. *Theory and Application of Digital Signal Processing*. Prentice Hall, Englewood Cliffs, NJ, 1975.

[62] Rektorys, K. *Applied Mathematics*, 6th ed. Prometheus, Prague, 1995.

[63] Rosenfeld, A. and Kak, A.C. *Digital Picture Processing*, 2nd ed. Academic Press, New York, 1982.

[64] Rosenfeld, A. and Kak, A.C. *Digital Picture Processing*. Academic Press, New York, 1976.

[65] Rueckert, D. Nonrigid registration: concepts, algorithms and applications. In *Medical Image Registration*, Hajnal, J.V., Hill, D.L.G, Hawkes, D.J. (Eds.). CRC Press, Boca Raton, FL, 2001.

[66] Russ, J.C. *The Image Processing Handbook*, 4th ed. CRC Press, Boca Raton, FL, 2002.

[67] Sangwine, S.J. and Horne, R.E.N. (Eds.). *The Color Image Processing Handbook*. Chapman & Hall, New York, 1998.

[68] Skokan, M., Skoupy, A., and Jan, J. Registration of multimodal images of retina. In *Proceedings of 24th Annual International Conference of IEEE–EMBS*, Houston, TX, 2002, pp. 1094–1096.

[69] Skrzypek, J. and Karplus, W. (Eds.). *Neural Networks in Vision and Pattern Recognition*. World Scientific, Singapore, 1992.

[70] Soille, P. Morphological operators. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., Geissler, P. (Eds.). Academic Press, New York, 1999.

[71] Sonka, M. and Fitzpatrick, J.M. (Eds.). *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*. SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.

[72] Sonka, M., Hlavac, V., and Boyle, R.D. *Image Processing, Analysis and Machine Vision*, 2nd ed. PWS, Boston, 1998.

[73] Sroubek, F., and Flusser, J. Shift-invariant multichannel blind restorations. In *Proc. 3rd Intl. Symp. Image and Signal Processing and Analysis,* Rome, September, 2003, pp. 332–337.

[74] Strang, G. and Nguyen, T. *Wavelets and Filter Banks*. Wellesley/Cambridge Press, 1996.

[75] Taxt, T. and Jirik, R. Superresolution of ultrasound images using the 1st and 2nd harmonic, *IEEE Trans. Ultrason. Ferroelec. Freq. Cont.*, 51(2), 163–175, 2004.

[76] Tekalp, A.M. Image and video restoration. In *The Digital Signal Processing Handbook*, Madisetti, V.K., Williams, D.B. (Eds.). CRC Press/IEEE Press, Boca Raton, FL, 1998.

[77] Vaidyanathan, P.P. *Multirate Systems and Filter Banks*. Prentice Hall PTR, Englewood Cliffs, NJ, 1993.

[78] Vandermeulen, D. Methods for Registration, Interpolation and Interpretation of Three-Dimensional Medical Image Data for Use in Three-Dimensional Display, Three-Dimensional Modelling and Therapy Planning. Ph.D. dissertation, Katholieke Universiteit Leuven, Belgium, 1991.

[79] Wagner, T. Texture analysis. In *Handbook of Computer Vision and Applications*, Vol. 2, Jahne, B., Haussecker, H., and Geissler, P. (Eds.). Academic Press, New York, 1999.

[80] Xu, C., Pham, D.L., and Prince, J.L. Image segmentation using deformable models. In *Handbook of Medical Imaging*, Vol. 2, *Medical Image Processing and Analysis*, Sonka, M., Fitzpatrick, J.M. (Eds.). SPIE, International Society for Optical Engineering, Bellingham, WA, 2000.