1

Fundamentals of Biomedical Image Processing

Thomas M. Deserno

Summary. This chapter gives an introduction to the methods of biomedical image processing. After some fundamental preliminary remarks to the terminology used, medical imaging modalities are introduced (Sect. 1.2). Sections 1.3 and 1.4 deal with low-level image processing and visualization, respectively, as far as necessary to understand the following chapters. Subsequently, the core steps of image analysis, namely: feature extraction, segmentation, classification, quantitative measurements, and interpretation are presented in separate sections. On account of its high relevance, the focus is on segmentation of biomedical images. Special segmentation methods and techniques have been developed in the medical application domain. Section 1.9 provides a brief summary of image communication. The electronic transmission and exchange of medical images will become more important in future for multimedia applications such as electronic patient records in health telematics and integrated care. Section 1.10 completes this chapter with an overview of past, present, and future challenges to biomedical image processing.

1.1 Introduction

By the increasing use of direct digital imaging systems for medical diagnostics, digital image processing becomes more and more important in health care. In addition to originally digital methods, such as Computed Tomography (CT) or Magnetic Resonance Imaging (MRI), initially analogue imaging modalities such as endoscopy or radiography are nowadays equipped with digital sensors. Digital images are composed of individual pixels (this acronym is formed from the words "picture" and "element"), to which discrete brightness or color values are assigned. They can be efficiently processed, objectively evaluated, and made available at many places at the same time by means of appropriate communication networks and protocols, such as Picture Archiving and Communication Systems (PACS) and the Digital Imaging and Communications in Medicine (DICOM) protocol, respectively. Based on digital imaging techniques, the entire spectrum of digital image processing is now applicable in medicine.

Fig. 1.1. *Modules of image processing.* In general, image processing covers four main areas: image formation, visualization, analysis, and management. The algorithms of image enhancement can be assigned as pre- and post-processing in all areas
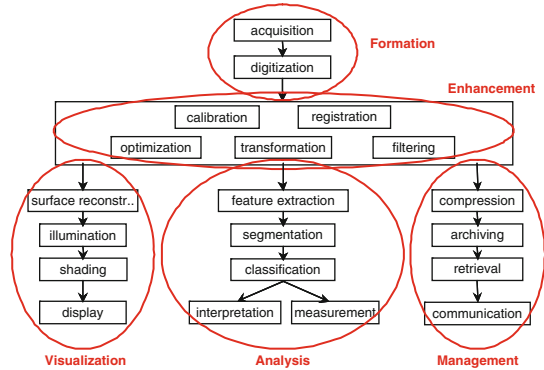
1.1.1 Steps of Image Processing

The commonly used term "biomedical image processing" means the provision of digital image processing for biomedical sciences. In general, digital image processing covers four major areas (Fig. 1.1):

1. *Image formation* includes all the steps from capturing the image to forming a digital image matrix.
2. *Image visualization* refers to all types of manipulation of this matrix, resulting in an optimized output of the image.
3. *Image analysis* includes all the steps of processing, which are used for quantitative measurements as well as abstract interpretations of biomedical images. These steps require a priori knowledge on the nature and content of the images, which must be integrated into the algorithms on a high level of abstraction. Thus, the process of image analysis is very specific, and developed algorithms can be transferred rarely directly into other application domains.
4. *Image management* sums up all techniques that provide the efficient storage, communication, transmission, archiving, and access (retrieval) of image data. Thus, the methods of telemedicine are also a part of the image management.

In contrast to image analysis, which is often also referred to as high-level image processing, low-level processing denotes manual or automatic techniques, which can be realized without a priori knowledge on the specific content of images. This type of algorithms has similar effects regardless of the content of the images. For example, histogram stretching of a radiograph improves the contrast as it does on any holiday photograph. Therefore, low-level processing methods are usually available with programs for image enhancement.
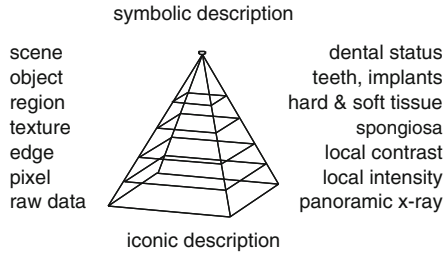
Fig. 1.2. *Levels of abstraction.* The general terms (*left*) are exemplified for a panoramic radiograph of upper and lower jaws (*right*). At the pyramid's top, the dental status corresponds to an abstract scene analysis, which only contains standardized information (existence and condition) on the tooth positions

1.1.2 Remarks on Terminology

The complexity of an algorithm, the difficulty of its implementation, or the computation time required for image processing plays a secondary role for the distinction between low-level and high-level processing methods. Rather, the degree of abstraction of the a priori knowledge is important for this meaning. Although the following definitions are not standardized in the literature, they are used consistently within this book (Fig. 1.2):

- The *raw data level* records an image as a whole. Therefore, the totality of all raw data pixels is regarded on this level.
- The *pixel level* refers to discrete individual pixels.
- The *edge level* represents the One-dimensional (1D) structures, which are composed of at least two neighbored pixels.
- The *texture level* refers to Two-Dimensional (2D) or Three-Dimensional (3D) structures. On this level however, the delineation of the area's contour (in three dimensions: the surface of the volume) may be unknown.
- The *region level* describes 2D or 3D structures with a well-defined boundary or surface.
- The *object level* associates textures or regions with a certain meaning or name, i.e., semantics is introduces on this level.
- The *scene level* considers the ensemble of image objects in spatial and/or temporal terms. If 3D structures are imaged over the time, also Four-Dimensional (4D) data is acquired.

From an iconic (concrete) to a symbolic (abstract) description of images, information is gradually reduced. Methods of low-level image processing operate on the raw data as well as on pixel, edge, or texture levels, and thus at a minimally level of abstraction. Methods of high-level image processing include the texture, region, object, and scene levels. The required abstraction can be achieved by increased modeling of a priori knowledge.

1.1.3 Biomedical Image Processing

With these definitions, a particular problem in high-level processing of biomedical images is inherently apparent: resulting from its complex nature, it is difficult to formulate medical a priori knowledge such that it can be integrated directly and easily into automatic algorithms of image processing. In the literature, this is referred to as the *semantic gap*, which means the discrepancy between the cognitive interpretation of a diagnostic image by the physician (high level) and the simple structure of discrete pixels, which is used in computer programs to represent an image (low level). In the medical domain, there are three main aspects hindering bridging this gap:

1. *Heterogeneity of images*: Medical images display living tissue, organs, or body parts. Even if captured with the same modality and following a standardized acquisition protocol, shape, size, and internal structures of these objects may vary remarkably not only from patient to patient (inter-subject variation) but also among different views of a patient and similar views of the same patients at different times (intra-subject variation). In other words, biological structures are subject to both inter- and intra-individual alterability. Thus, universal formulation of a priori knowledge is impossible.
2. *Unknown delineation of objects*: Frequently, biological structures cannot be separated from the background because the diagnostically or therapeutically relevant object is represented by the entire image. Even if definable objects are observed in biomedical images, their segmentation is problematic because the shape or borderline itself is represented fuzzily or only partly. Hence, medically related items often can be abstracted at most on the texture level.
3. *Robustness of algorithms*: In addition to these inherent properties of medical images, which complicate their high-level processing, special requirements of reliability and robustness of medical procedures and, when applied in routine, image processing algorithms are also demanded in the medical area. As a rule, automatic analysis of images in medicine should not provide wrong measurements. That means that images, which cannot be processed correctly, must be automatically classified as such, rejected and withdrawn from further processing. Consequently, all images that have not been rejected must be evaluated correctly. Furthermore, the number of rejected images is not allowed to become large, since most medical imaging procedures are harmful and cannot be repeated just because of image processing errors.

1.2 Medical Image Formation

Since the discovery of X-rays by Wilhelm Conrad Röntgen in 1895, medical images have become a major component of diagnostics, treatment planning and procedures, and follow-up studies. Furthermore, medical images are used
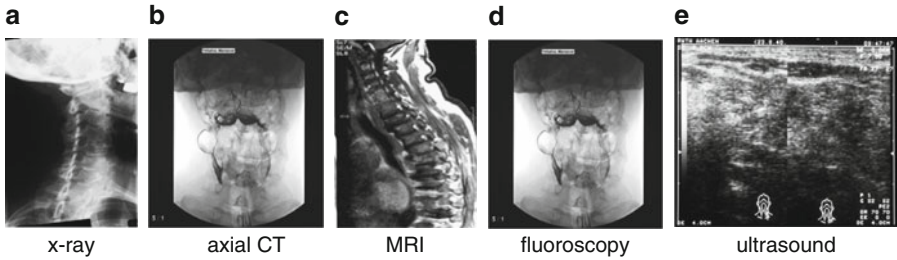
a   **b**   **c**   **d**   **e**

x-ray   axial CT   MRI   fluoroscopy   ultrasound

Fig. 1.3. *Medical imaging modalities.* The body region (here: cervical vertebra) appears completely different when altering the imaging modality

Name	Symbol	Mass	Charge
Proton	p	1 u	+1 e
Neutron	n	1 u	0 e
Alpha particle	α	4 u	+2 e
Electron	β	0 u	−1 e
Positron	β^+	0 u	+1 e
Photon	γ	0 u	0 e

Table 1.1. *Atomic particles.* The given values for mass and charge are only rough estimates. The atomic mass unit $1\,\mathrm{u} = 1.660538782 \cdot 10^{-27}$ kg. The elementary charge $1\,\mathrm{e} = 1.602176487 \cdot 10^{-19}$ C

for education, documentation, and research describing morphology as well as physical and biological functions in 1D, 2D, 3D, and even 4D image data (e.g., cardiac MRI, where up to eight volumes are acquired during a single heart cycle). Today, a large variety of imaging modalities have been established, which are based on transmission, reflection or refraction of light, radiation, temperature, sound, or spin. Figure 1.3 emphasizes the differences in image characteristic with respect to the imaging modality. Obviously, an algorithm for delineation of an individual vertebra shape that works with one imaging modality will not be applicable directly to another modality.

1.2.1 Basic Physics

To understand the different nature of medical images and imaging modalities, we need to recall some basic physics of matter. Roughly, all matter is build from atoms, where a nucleus composed of protons and neutrons is surrounded by a electron shell. Table 1.1 lists charge and mass of nuclear particles.

The number of protons determines the element number. In the equilibrium state, the number of electrons equals the number of protons and there is no external Coulomb field. However, the positions of the particles are not constant. In particular, the electrons orbit the nucleus. According to the Maxwell laws, accelerated (continuously changing its direction) charge induces electromagnetic radiation: the electron would lose energy gradually spiraling inwards and collapsing into the nucleus.

Within the Bohr model of the atom, there are certain shells where an electron can orbit its nucleus without releasing electromagnetic radiation. These

Fig. 1.4. *Bohr model of the atom.* The shells where an electron can orbit the nucleus without releasing electromagnetic radiation are numbered, and there is a maximal number of electrons for each shell. Sometimes, the shells are also referred to by letters k, l, m, etc. The difference of energy between shells is released as radiation when an electron changes its position
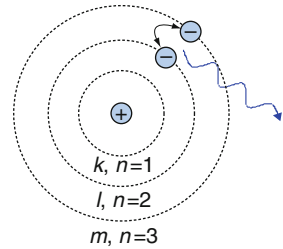
shells are numbered n (Fig. 1.4) and allow for $2 \cdot n^2$ electrons. The energy of an electron $E_n = (-13.6 \text{ eV}) \frac{1}{n^2}$ depends on the orbit number n, where inner shells are energetically preferred and ionizing needs higher energy if an electron of an inner shell is removed. The unit Electron Volt (eV) refers to the kinetic energy of an electron after passing the acceleration voltage of 1.0 V.

1.2.2 Imaging Modalities

From the plenty of medical imaging modalities, we will focus on X-ray imaging, CT, MRI, and ultrasound. However, optical modalities such as endoscopy, microscopy, or photography are not less important.

X-Ray Imaging

According to the Bohr model, *X-radiation* – the term was initially introduced by Röntgen – can be generated, for instance, if an electron from a higher shell jumps over into a free position of an inner shell (Fig. 1.4). The discrete difference of energy ΔE is released as a photon (γ particle). ΔE is characteristic to the numbers of shells and the element.

Technically, free positions in inner shells are produced from shooting electrons to the atom. Figure 1.5 schematically shows an X-ray tube. The high voltage between cathode and anode accelerates the electrons that are released from a filament. Passing the acceleration voltage, these electrons are loaded with kinetic energy. Hitting the target material, usually tungsten for skeletal imaging and molybdenum for mammography, two types of interactions may occur, i.e., the accelerated electron interacts with the:

- *Nucleus*, where the electron is slowed down by the Coulomb field of the protons, and a photon is released with an energy equal to the loss of kinetic energy of the electron (*Bremsstrahlung*).
- *Shell*, where the *characteristic radiation* is released as described above.

When X-radiation passes through matter, e.g., the human body we would like to image, the X-ray photons again may interact with the nucleus or the shell resulting in:
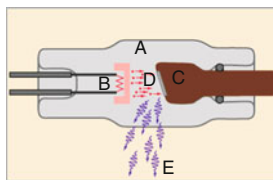
Fig. 1.5. *X-ray tube.* The vacuum tube (**A**) houses cathode (**B**) and anode (**C**). A current heats up the filament, releasing electrons (**D**), which are accelerated towards the anode. Interacting with either the nucleus or the shell of the target material, Bremsstrahlung and characteristic radiation are released (**E**), respectively

- *Absorption*: The photon is completely vanished giving all its energy to the absorbing material. This effect is harmful and causes damage to living cells, but it is required to obtain a contrasted image.
- *Scattering*: A secondary photon is produced, that might be coherent (Thomson effect) or incoherent (Compton effect). Both effects lower the Signal to Noise Ratio (SNR), since the secondary photon usually travels in another direction and contributes to the image at a wrong location, and scatter rasters from lead are used to filter the scattered radiation.

The absorption coefficient μ uniquely describes the material, and is mapped to the gray scale for image display. In plain radiography, high-attenuating material (e.g., bone) is displayed in white (see Fig. 1.3a) while in fluoroscopy, the scale is inverted (see Fig. 1.3d), and the high-absorbing contrast agent is displayed in black.

However, the absorption sums up along the path through the matter. In particular, the absorption is described by an exponential function. In a first approximation, the intensity I of radiation depends on the thickness d of the imaged material $I = I_0 e^{-\mu d}$. However, a human body is not made from constant material $\mu \sim \mu(d)$, and furthermore, the absorption depends on the photon's energy $\mu \sim \mu(E)$. Since X-radiation cannot be obtained mono-energetic (Bremsstrahlung), the absorption equation yields

$$I = \int I_0(E) e^{-\int \mu(z,E)\mathrm{d}z} \mathrm{d}E \qquad (1.1)$$

The dependence of the absorption on the energy of the photon is obvious. Photons with low energy are more likely absorbed or scattered than high-energetic photons. Consequently, the spectrum of X-radiation, which is released from the X-ray tube, hardens when passing matter. This effect is called beam hardening.

Computed Tomography (CT)

X-ray imaging produces summation images, where all attenuation coefficients along the path are integrated. From a single image, one cannot determine the

order of overlapping objects. This is different with CT, where the absorption is determined in 3D for each volume element (voxel). For imaging, a volume is acquired slice by slice, and each slice is reconstructed from several measures in different angulation.

Although back projection or filtered back projection in k-space (spatial frequency domain after Fourier transform) are nowadays used for image reconstruction, we will explain the principle based on the arithmetic reconstruction technique, which in fact was applied to the first CT machines. Suppose a slice being divided into four pixels, and two parallel rays passing it in two different directions (Fig. 1.6). This results in four independent equation (Fig. 1.6B) allowing to compute the four absorption coefficients μ_{ij} (Fig. 1.6E). To obtain more rows and columns, the number of parallel rays and the number of angles must be increased accordingly. Today, fan beam gantries are build (Fig. 1.7), enabling continuous rotation of the imaging fan and continuous longitudinal movement of the patient (spiral CT). Further speedup of acquisition is obtained from scanners, where up to 64 lines of X-ray receptors are mounted. From 1972, the acquisition time per slice has decreased about 10^5 (Table 1.2).

Fig. 1.6. *Arithmetic CT reconstruction.* Two parallel X-rays pass the slice in 90° and the measures are recorded. (**A**) logarithm allowing assignment of absorption coefficients μ; (**B**) four linear equations are obtained; (**C**) iterative solution; (**D**) assignment; (**E**) inverse logarithm
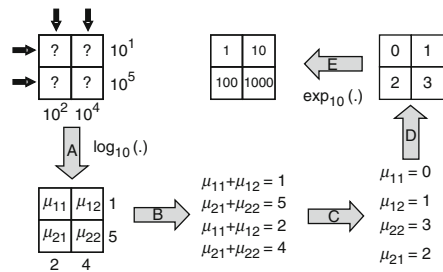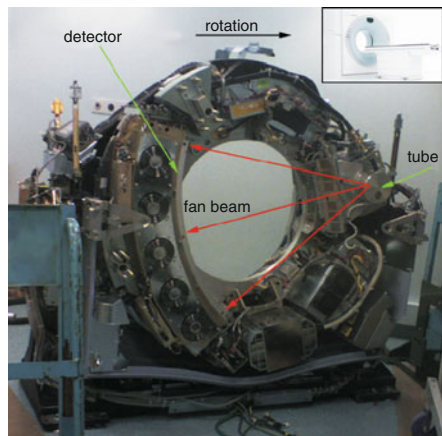

Fig. 1.7. *CT gantry.* Detaching the housing from the CT exposes the X-ray tube and the detector fan (http://wikepedia.org)

Year	Resolution	Gray scales	Thickness	Time
1974	80×80	64 (6 bit)	10 mm	300 s
1984	256×256	256 (8 bit)	5 mm	10 s
1994	512×512	512 (9 bit)	0.8 mm	0.5 s
2004	1024×1024	1024 (10 bit)	0.4 mm	0.005 s

Table 1.2. *CT slice parameters.* Today, a 64 line scanner rotates about three times a second yielding up to 192 slices per second
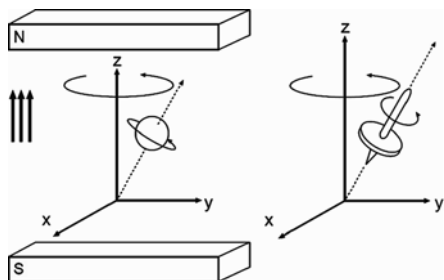

Fig. 1.8. *Precession [1].* A spinning proton in a magnetic field (*left*) moves like a gyroscope in the mass gravitation field of the earth (*right*). According to the Larmor theorem, the precession frequency is determined by the strength of the external magnetic field

Magnetic Resonance Imaging (MRI)

Almost simultaneously to CT, MRI has been introduced to medicine. It is based on electromagnetic effects of the nucleus. Since the human body consists of about 70 % water, we focus on hydrogen. Its nucleus is composed of only one proton. As mentioned before, the particles forming the nucleus are continuously moving. For hydrogen, this movement is a self rotation (spin), which has a magnetic moment. As shown in Fig. 1.8, the magnetic moment is aligned to an outer magnetic field, and precession ν is started.

To understand MRI, we need to regard a probation of tissue, which is composed of billions of hydrogen atoms. In other words, we move from a microscopic to a macroscopic view, where the spins sum up to a macroscopic magnetic moment M. Suppose the external magnetic field B_z is directed along the z-axis, the magnetic moments can align parallel or anti-parallel, where the latter occurs slightly minor (six less in a million). Therefore, $M_z > 0$. In addition, all precession is dephased ($M_{xy} = 0$).

The next component of MRI is a so called Radio Frequency (RF) impulse. Such an impulse can excite the system of spinning protons if the electromagnetic frequency equals the precession frequency. Depending on the amplitude and time of excitement, M can be arbitrarily directed. In Fig. 1.9b, for example, M has been turned into the (x, y)-plane; the corresponding RF impulse is referred to as 90° impulse.

RF excitement is followed by exponential relaxation, where the system is restoring its equilibrium state. The stored energy is released as signal (i.e., the Free Induction Decay (FID) when measured in Fourier domain), which can be detected and transformed to an image. However, the relaxation process is complex, since two independent effects superimpose:
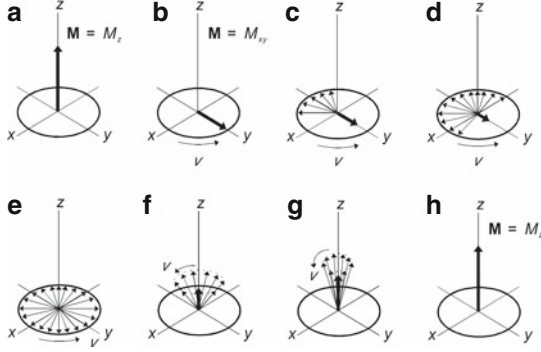
Fig. 1.9. *Excitement and relaxation [1].* Without excitement, $M = M_z$ (**a**); a 90°
RF impulse turns $M = M_{xy}$ (**b**); dephasing (spin–spin relaxation) starts first,
continuously decreasing M_{xy} (**c,d**); spin–spin relaxation completed (**e**); dealign-
ment (spin–lattice relaxation) starts, steadily increasing M_z (**f,g**); relaxation
completed (**h**)

- *spin–spin relaxation* with relaxation time T_2 affects the phase of the spins.
  For water-based and fat-based tissues, T_2 is in the $40 - 200\,\mathrm{ms}$ and $10 -$
  $100\,\mathrm{ms}$ range, respectively.
- *spin–lattice relaxation* with relaxation time T_1 affects the parallel vs. anti-
  parallel alignment of spins. For water-based and fat-based tissues, T_1 is in
  the $0.4 - 1.2\,\mathrm{s}$ and $0.10 - 0.15\,\mathrm{s}$ range, respectively.

Therefore, spin–spin relaxation is almost completed before spin–lattice
relaxation is detectable. Relaxation is visualized in Fig. 1.9. After 90° impulse,
$M = M_{xy}$ and $M_z = 0$. Note that M_{xy} rotates with precession frequency ν
in the (x, y)-plane. When T_2-relaxation is completed (Fig. 1.9e), $M_{xy} = 0$ and
$M_z = 0$. The T_1-relaxation is visualized in Fig. 1.9f–h. In Fig. 1.9h, the spins
gave back the energy they obtained from the RF pulse to the surrounding
lattice.

To obtain a high-quality relaxation signal, spin-echo sequences are applied,
where different RF impulses are induced, and readout of T_1 and T_2 is per-
formed in between. Therefore, spin-echo sequences are characterized by the:

- *echo time* T_E determining half of the delay between a 90° and a 180° RF
  impulse, and the
- *repetition time* T_R denoting the rate of re-applying a 90°/180° sequence.

Figure 1.10 emphasizes the differences in contrast and appearance depend-
ing on the echo sequence. In particular, a M_0-, T_1-, or T_2-weighted MRI is
obtained if ($T_E \ll T_2$ and $T_R \gg T_1$), ($T_E \ll T_2$ and $T_R \approx T_1$), or ($T_E \approx T_2$
and $T_R \gg T_1$), respectively.

However, the theory we discovered so far does not allow us to obtain
such images because we do not have any spatial alignment with the signal
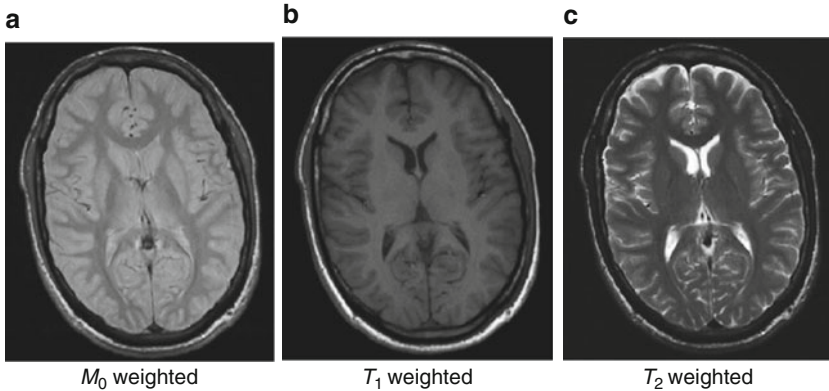yet. This is obtained using gradient fields, which are superimposed to the

a   **b**   **c**

M_0 weighted T_1 weighted T_2 weighted

Fig. 1.10. *Forming MRI with spin-echo sequences.* (Courtesy: Tony Stöcker, FZ Jülich)

constant external field B. For instance, let us superimpose a gradient field in x-direction: $B = B_0 + B_G(x)$, $B_G(x_1) < B_G(x_2)$ $\forall x_1 < x_2$. Now, the Larmor frequency of precession $\nu \sim \nu(x)$ is slightly shifted along the x-axis. Since the induced RF impulse covers all frequencies, excitement results in the entire probation, and from the frequency of the FID signal, the according slice can be located. Another gradient, for instance in y-direction, allows for addressing a line rather than a plane. As we have seen with CT reconstruction, capturing signals from different lines through the volume finally allows voxel assignment. Advantageous to CT, gradient fields in MRI can be generated with gradient coils, where the current is adopted, and no mechanical rotation is required. In fact, steering the gradient fields produces the noise of MRI devices, since strong currents need to be turned on and off quickly.

Ultrasound

In contrast to CT and MRI, ultrasound is a medical imaging modality that is based on reflection of sound waves. Depending on the transducer, 1D to 4D data is obtained. We start from the 1D case (signal), where a longitudinal sound wave is traveling through the tissue of the human body. At transitions between different matter (e.g., muscle and fat), the sound wave is partly reflected and transmitted (refracted if the surface is not hit perpendicular). In other words, the echo runtime indicates the distance between transducer and tissue border while the echo strength is related to material properties. More precisely, these sound-relevant properties of matter are described by the speed of sound c_s and the density ρ, yielding the acoustic impedance $Z = c_s \cdot \rho$. Interfacing two materials $Z_1 = Z_0$ and $Z_2 = Z_0 + \Delta Z$, the reflection ratio r and transmission ratio t are given by

Table 1.3. *Speed of sound in matter.* The acoustic impedance $Z = c_s\rho$ computes from density ρ and speed of sound c_s. All numbers refer to body temperature of $37°$ centigrade

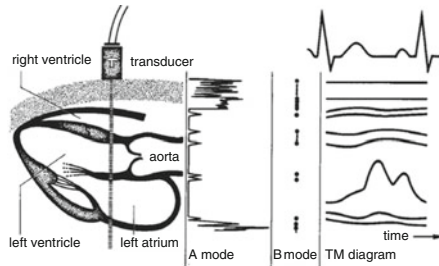| Material	c_s in m/s	ρ in kg/m^3	Z in kg/m^{2}s
Bone	3,600	$1.70 \cdot 10^3$	$6.12 \cdot 10^6$
Marrow	1,700	$0.97 \cdot 10^3$	$1.65 \cdot 10^6$
Blood	1,570	$1.02 \cdot 10^3$	$1.61 \cdot 10^6$
Muscle	1,568	$1.04 \cdot 10^3$	$1.63 \cdot 10^6$
Water	1,540	$0.99 \cdot 10^3$	$1.53 \cdot 10^6$
Fat	1,400	$0.97 \cdot 10^3$	$1.36 \cdot 10^6$
Air	340	1.20	$4.08 \cdot 10^2$

Fig. 1.11. *Ultrasound visualization modes [2].* A section through the heart is drawn schematically (*left*). From a simple sampling line, reflections of sound may be plotted (**i**) according to their amplitude (A-mode), (**ii**) coded as dots with a gray scale mapped to the amplitude (B-mode), which supports 2D images if a array of Piezo-electric crystals is applied, and (**iii**) in their position over the time, to visualize motion (TM diagram)

$$r = \sqrt{\frac{I_0}{I_A}} = \frac{Z_2 - Z_1}{Z_2 + Z_1} = \frac{\Delta Z}{2Z_0 + \Delta Z} \quad \text{and} \quad t = 1 - r \approx \begin{cases} 1, \text{ if } \Delta Z \ll Z_0 \\ 0, \text{ if } \Delta Z \gg Z_0 \end{cases} \tag{1.2}$$

where I_0 and I_R denote the intensity of the initial and reflected wave, respectively. As we can see from Table 1.3, $t \approx 0$ from air to water and soft tissue to bone, while $t \approx 1$ within the soft tissue. Therefore, a sonographic view behind bony structures or through organs filled with air is almost impossible. Furthermore, water-based gel must be used for air-free coupling the transducer to the human body.

Furthermore, the sound intensity is attenuated from expansion. The attenuation increases linear with the sound frequency but spatial resolution requires high frequency. Therefore, typical diagnostic scanners operate in the frequency range of $2 - 18$ MHz trading-off spatial resolution and imaging depth.

Technically, a piezoelectric crystal is used to convert an electrical signal into a mechanical movement, and the deformation of the crystal is coupled into the body. Then, the same transducer is used to detect the echos. There are several options to form an image from this pulse-echo signal (Fig. 1.11):

- *A-mode*: In amplitude mode, the echo intensity is plotted on the screen as a function of depth;
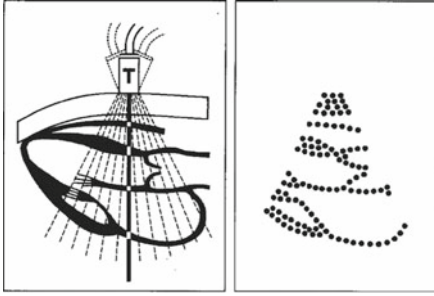
Fig. 1.12. *B-mode sector scan [2].* An array of transducers is used to scan a sector. In B-mode, reflections at tissue borders are displayed within a fan-shaped aperture, which is typically for medical ultrasound. Turning the transducer array perpendicularly allows for imaging cone-shaped volumes

- *B-mode*: In brightness mode, the echo intensity is coded with gray scales. This allows composing an array of transducers simultaneously scanning a plane through the body (Fig. 1.12). Parallel and sector scanners are available;
- *TM-mode*: Time motion diagrams visualize movements of sound-reflecting tissue borders. This mode offers functional rather than morphological inspection;
- *M-mode*: In motion mode, a sequence of rapidly acquired B-mode scans is displayed as moving picture. This is the most common mode in clinical ultrasound;
- *D-mode*: The doppler mode makes use of the doppler effect (i.e., a shift in frequency that occurs if the source of sound, the receptor, or the reflector is moved) in measuring and visualizing blood flow. Several visualization modes are used:
  - *Color Doppler*: The velocity information is presented as a color-coded overlay on top of a B-mode image;
  - *Continuous Doppler*: Doppler information is sampled along a line through the body, and all velocities detected at each point in time are presented (on a time line);
  - *PW Doppler*: Pulsed-wave Doppler information is sampled from only a small sample volume (defined in the 2D B-mode image), and presented on a time line;
  - *Duplex*: Color and (usually) PW Doppler are simultaneously displayed.

1.2.3 Digitalization

Digital image processing implies a discrete nature of the images. Regardless whether a film-based radiograph is digitized secondarily with a scanner, or the device primarily delivers a digital pixel (voxel) matrix, digitization effects alter the image. Digitization applies to both the definition (sampling) and the value range (quantization).
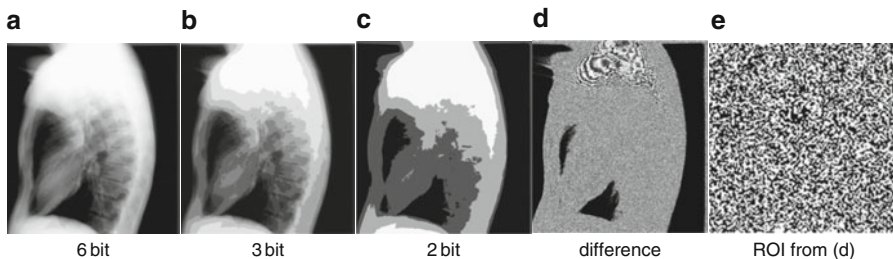
Fig. 1.13. *Quantization.* A lateral chest radiograph is shown with 6 bit, 3 bit, and 2 bit, which equals 64, 8, and 4 different gray scales in panel (**a**), (**b**), and (**c**), respectively. Disregarding saturation effects occurring within bone and air, the difference between 8 bit and 7 bit representation results in white noise (**d**,**e**). Panel (**d**) shows the histogram-optimized difference image, and (**e**) a centered Region of Interest (ROI) of 100×100 pixel

Quantization

Quantization refers to the digitization of the value range. We need to determine the maximal number of gray scales for every image. Usually, 8 bit and 24 bit are chosen for gray scale and full color images, respectively, allowing 256 different values in each band. In medicine, radiography or CT usually delivers 12 bit = 4,096 different values. If we assume a continuous brightness, quantization always worsen the image quality. The alteration can be modeled as additive noise, and the SNR of our digital image is improved by an increased number of gray scales.

Quantization noise is visualized in Fig. 1.13. With printing, we do not see any differences between 8 bit, 7 bit, or 6 bit quantization. If the number of gray scales becomes small, artefacts are apparent (Fig. 1.13b,c). Subtracting the 7 bit representation from the original 8 bit image illustrates the quantization noise (Fig. 1.13d,e).

Sampling

Sampling refers to the digitization of the definition range. According to the linear system theory, an analogue signal can be unambiguously represented with a discrete set of samples if the sampling rate exceeds two times the highest frequency occurring in the image (Nyquist theorem). Shannon's popular version of the sampling theorem states [3]:

> If a function $x(t)$ contains no frequencies higher than f_b Hz, it is completely determined by giving its ordinates at a series of points spaced $t = \frac{1}{2f_b}$ seconds apart.

Once the sampling theorem is satisfied, we cannot improve the image quality adding more pixels, which is contrarily to the effects of quantization. In
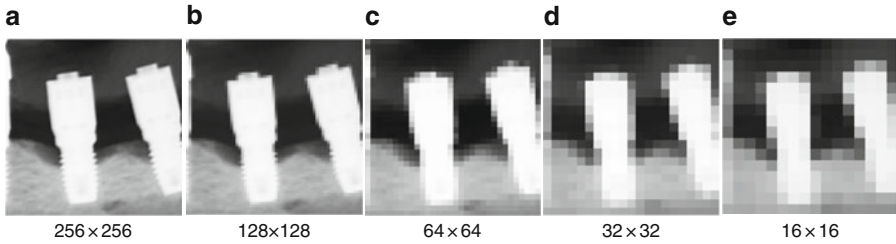
a 256×256   **b** 128×128   **c** 64×64   **d** 32×32   **e** 16×16

Fig. 1.14. *Sampling.* Artefacts from sub-sampling are visible if the printer resolution is smaller than the pixel scale. In panel (**b**), the first partial pixel effects should be notable
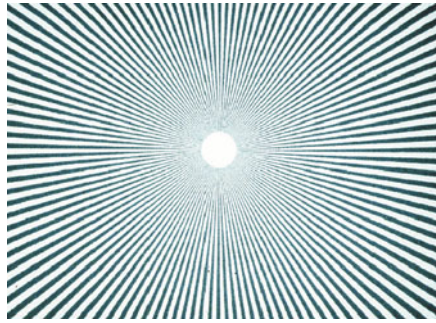

Fig. 1.15. *Moiré pattern.* This pattern is obtained from radiographing a lead grid. The spatial resolution of the entire X-ray imaging chain, disregarding whether it ends analogously or digitally, is measured by the distance from the center to that radius where individual lines can be differentiated. Furthermore, a four leaf clover can be seen in the center although it is neither with the lead lamella nor the squared pixel grid

spatial discretization, increasing the number of samples beyond the Nyquist rate only increases the file size of raw data, but not the information coded in it.

Figure 1.14 emphasizes the loss of information that results from applying an insufficient number of pixels for image acquisition. For instance, the spongious structure of the jaw bone disappears (Fig. 1.14c–e). Furthermore, gray scales are obtained misleadingly indicating a different material. For instance at the border of the implants, pixel values with μ of bone are obtained. In CT imaging, this partial pixel effect is also known as partial volume effect, see Sect. 1.8.1.

Similar to insufficient quantization, subsampling suppresses information in the digital image representation. In contrast to quantization, information is falsely added to an image if the sampling theorem is not fulfilled. Partial effects are one example. More important are aliasing effects. In 2D imaging, Moiré patterns are obtained whenever a regular structure mismatches the grid (Fig. 1.15).

1.3 Image Enhancement

Low-level methods of imaging processing, i.e., procedures and algorithms that are performed without a priori knowledge about the specific content of an image, are mostly applied to pre- or post-processing of medical images (Fig. 1.1). Therefore, the basic methods of histogram transforms, convolution and (morphological) filtering are mostly disregarded unless required for further understanding of this text (see the list of related textbooks on page 49). As a special preprocessing method for medical images, techniques for calibration and registration are briefly introduced.

1.3.1 Histogram Transforms

Point operations (pixel transforms) are based on the histogram of the image. Modifying the pixel values, all pixels are transformed independently from their positions in the image and their immediate neighborhood. Therefore, these type of transform is also referred to as point operation.

Histogram

The histogram shows the frequency distribution of pixel values (e.g., gray scales) disregarding the certain positions where the gray scales occur in the image. Simple pixel transforms can be defined using a histogram. For example, through the stretching of gray scales, the contrast of an image is improved (Fig. 1.16). After determining the histogram, upper and lower bounds are

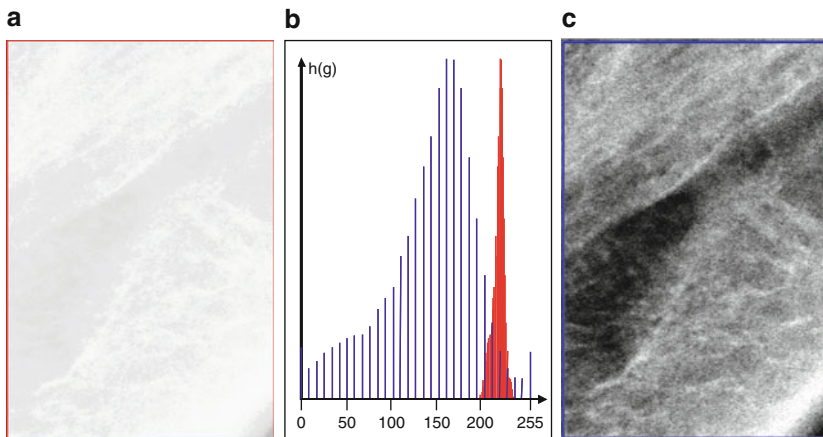**Fig. 1.16.** *Histogram stretching.* A ROI is taken in the area of the temporomandibular joint from an intra-oral radiograph (**a**). Resulting from under-exposure, the spongy bone structure is displayed quite poorly. The associated histogram (**b**) is only narrow occupied (*red*). By stretching the histogram, the columns are linearly pulled apart (*blue*) and the contrast of the transformed radiograph is increased (**c**)

Old	New pixel value			Old	New pixel value		
Gray	Red	Green	Blue	Gray	Red	Green	Blue
0	0	0	0	...	...	...	...
1	1	0	2	246	254	244	239
2	2	0	3	247	255	245	243
3	3	0	5	248	255	245	244
4	4	1	7	249	255	246	246
5	5	1	9	250	255	247	248
6	5	2	12	251	255	249	250
7	5	2	14	252	255	251	252
8	5	3	16	253	255	251	253
9	5	4	18	254	255	253	254
...	...	...	...	255	255	255	255

Table 1.4. *Look-up table for pseudo coloring.* For each value in the range of the input image, the lookup table holds a value from the range of the output image. The color palette shown here is used for pseudo coloring keeping the original brightness progression of the input image [4]
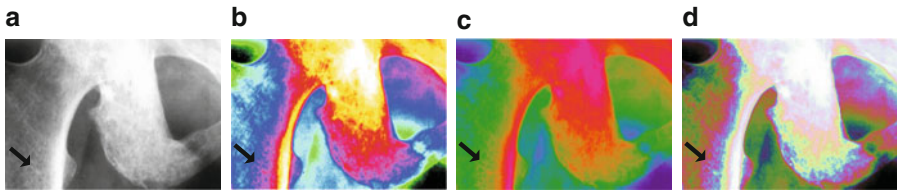

Fig. 1.17. *Pseudo coloring [4].* X-ray image of a pelvic bone metastasis after radiotherapy (**a**); pseudo-colored image of (**a**) using the corner colors of the RGB cube (**b**), colors of constant brightness (**c**), and colors with continuous brightness progression obtained from a spiral around the gray diagonal of the RGB cube (**d**). The arrow indicates local contrast enhancement

located, and a linear transform is applied that maps the lower bound to zero and the upper bound to the maximal gray scale (i.e., 255 for 8 bit images). If the histogram of the initial image does not contain all possible gray scales, the gray scale distance between neighbored pixels is enlarged, which results in an enhanced contrast.

Look-Up Table (LUT)

Technically, computation of histogram transforms is based on a Look-Up Table (LUT). For all pixel values, the lookup table contains a new value, which can also originate from another range of values. The example in Table 1.4 assigns each gray scale with a triple for Red, Green, and Blue (RGB). This transform is called pseudo coloring, and it is frequently used in the biomedical domain to enhance local contrast (Fig. 1.17). Computer graphic boards may limit the number of gray scales to 256 (8 bit), but offer $256^3 = 16,777,216$ colors. Special algorithms are recommended for the pseudo coloring in the medical context. In other words, pseudo coloring allows presentation of data, where the range of values exceeds the length of the RGB cube's edges without reducing the information as it would result from windowing.

a			**b**			**c**			**d**			**e**			**f**		
1	1	1	1	2	1	−1	−2	−1	0	−1	0	−1	−2	−1	−2	−1	0
1	1	1	2	4	2	−2	12	−2	−1	5	−1	0	0	0	−1	0	1
1	1	1	1	2	1	−1	−2	−1	0	−1	0	1	2	1	0	1	2

Fig. 1.18. *Convolution templates.* The sliding average (**a**) and the binomial low-pass filter (**b**) cause a smoothing of the image. The binomial high-pass filter (**c**), however, increases contrast and edges, but also the noise in the image. The templates (**a**) to (**c**) must be normalized to make sure that the range domain of values is not exceeded. The contrast filter (**d**) is based on integer pixel values. The convolution with (**d**) is therefore easy to calculate. The anisotropic templates (**e**) and (**f**) belong to the family of Sobel operators. Eight Sobel masks can be generated by rotation and mirroring for direction-selective edge filtering (see Fig. 1.24)

1.3.2 Convolution

In contrast to point operations (histogram transforms), the considered pixels are combined with the values of their neighborhood when discrete filtering is applied. The underlying mathematical operation, i.e., convolution, can be characterized with the help of so-called templates (Fig. 1.18). A *template* is a mostly small, squared mask of usually odd lateral length. This template is mirrored along two axes (hence, the name "convolution" is commonly used) and positioned in one corner of the input image. The image pixels under the mask are named *kernel*[1]. Each pair of corresponding pixel values of template and kernel are multiplied and then summed up. The result is registered at the position of the mask's center pixel in the output image. Then, the template is shifted row by row and column by column to the next positions on the input image, until all the positions have been visited, and thus, the output image has been calculated completely.

The pixel values of the template determine the effect of the filter. If only positive values are used in the template, basically a (weighted) averaging is calculated in the local neighborhood of each pixel (Fig. 1.18a,b). The resulting image is smoothed and appears with reduced noise. However, the sharpness of edges is also reduced. If the template is composed of positive and negative coefficients, the contrast in the image is intensified, and the edges are highlighted (Fig. 1.18c–f). Anisotropic (i.e., not rotationally symmetric) templates also have a preferred direction (Fig. 1.18e,f). Hereby, the contrasts can be direction-selectively strengthened.

1.3.3 Mathematical Morphology

Another approach to filtering is adapted from the mathematical morphology. Although morphologic operators can also be defined for gray scale images,

---

[1] In the literature, "mask", "kernel", and "template" frequently are used as synonyms.
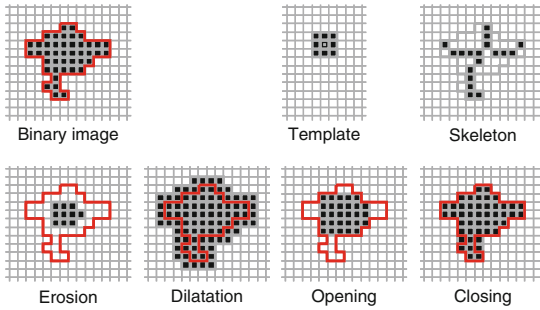
Fig. 1.19. *Binary morphology.* The binary pattern is outlined in red. The small circle marks the center pixel in disk-shaped template with radius 1

morphologic filtering is principally performed on binary input images, i.e., each pixel is assigned either TRUE or FALSE. According to a general convention, the white pixels in the binary image indicate relevant segments and the black pixels indicate the background. For printing, however, this assignment may be inverted. The binary template, which is also referred to as *structural element* (structuring element, structel, strel), is associated to the binary image using logical operations, in particular:

- *erosion* (based on logical AND of structel and binary image),
- *dilation* or dilatation (based on logical OR of structel and binary image),
- *opening* (erosion followed by dilatation using the same structel),
- *closing* (dilation followed by erosion using the same structel), and
- *skeleton* (e.g., by erosion with various structels).

As it can be seen in Fig. 1.19, the erosion reduces the size of a segment, and the dilation leads to its enlargement. The opening removes small details on the outline of segments or the background, without affecting the total size of relevant regions. The closing is able to remove holes in the interior of a region and smooth its contour. Here, the size of the segment is roughly maintained, too. The skeleton is a path with thickness of one pixel, which is located in the middle of the segment.

Binary morphology is applied frequently in medical image processing, for instance to clean up shapes after pixel-based segmentation (see Sect. 1.6.1). Gray scale morphology is simply a generalization from 1 bit (binary) images to images with multiple bits per pixel, where MIN and MAX operations replace the AND and OR operations of binary morphology, respectively.

1.3.4 Calibration

If the physician intents to take quantitative measurements from an image, a careful calibration of the imaging modality is required. Both, geometry (spatial domain) and brightness or color intensity (value domain) must be adapted to the modality. Calibration is device-specific but disregards the biological content captured, and thus, it is part of low-level processing methods. While reading a radiograph, calibration is made unconsciously by the radiologist.
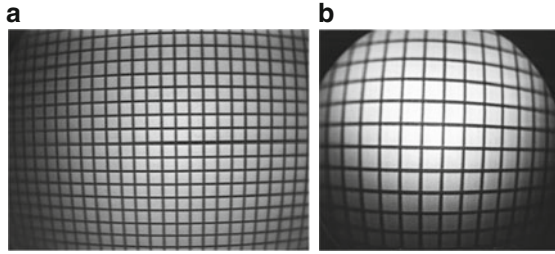
Fig. 1.20. *Geometric distortion and brightness variation [5].* By endoscopic examinations, barrel distortions are often generated, which must be corrected before the image can be analyzed quantitatively. In addition, the boundary areas in the video appear darker and blurred. Image (**a**) is generated with a rigid laryngoscope, which is used for the examination of the larynx. Image (**b**) is taken with a flexible endoscope for nasal laryngoscopy. Both endoscopes are used in clinical routine. Microscopy and other optical methods may produce similar artifacts

However, it must be explicitly implemented for computerized image analysis and measurements.

Geometric aberrations (distortions) have the consequence, that relevant structures of the same size are displayed depending on the position within the image. In the biomedical sciences, the positioning of the imaging device must not affect any measurements. For example in endoscopy, resulting from the optical devices in use, so called barrel distortions are originated (Fig. 1.20). Even in simple planar radiography, the objects, which are far away from the image plane, appear larger than those, which are located close to the imaging device. This must be kept in mind whenever geometric measurements in digital X-rays are taken and displayed to the physicians: point distances in digital images can be converted into length measurements only if a fixed scale is assumed, which is often not fulfilled.

In the same way, the absolute assignment of the pixel values to physical measurements usually is problematic. For example in X-ray imaging, the linear correspondence of brightness values to the accumulated absorption coefficient of the imaged structure is possible, if an aluminum (step) wedge with known X-ray absorption properties is placed beside the object. In digital video recording, white balancing must be performed such that the color values corresponds with reality. However, different illumination of the same scene may still alter the captured colors.

1.3.5 Registration

Often, an absolute calibration of examination procedures is not possible or only limitedly feasible. Then, registration can be used to achieve an approximation of two or more images such that at least a change in measured dimensions can be quantified. For example, an acute inflammation turns tissue
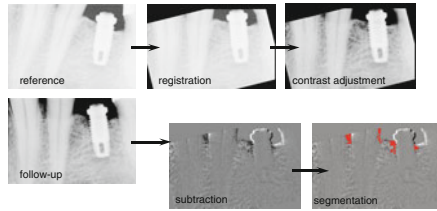
Fig. 1.21. *Unimodal Registration.* In dental implantology, reference and follow-up images are taken at various points of time. Geometric registration with subsequent contrast adjustment enables pixel-by-pixel subtraction. In the subtraction image, bone destruction is clearly emphasized and can be segmented easily on the pixel level of features (*red*)

into a reddish color. Under treatment, the absolute redness of the tissue is less interesting than its relative change as compared to the findings of previous recordings.

Unimodal Registration

This term refers to the relative calibration of images that have been acquired with the same modality. For instance, images that have been taken from the same patient but at different points of time are adjusted in order to quantify the course of the disease. As in the field of calibration, we differ between geometric registration and color or contrast adjustment, if the registration is performed in the spatial domain or the value range, respectively. Figure 1.21 illustrates the diagnostic potential of registration in dental implantology. After registration, the appraisal of the status of peri-implant bone is significantly simplified by the subtraction of recall and follow-up recordings.

Multi-Modal Registration

The images to be compared are captured with different modalities. For example, a 3D rigid registration is illustrated as the movement of the hat on the head. Especially in neurology, these methods have a crucial meaning. Since tumor resection in the brain must be executed very carefully, in order to avoid damage of neighbored brain areas, functional and morphological brain images are registered to plan the procedure. While morphology can be adequately represented in MRI or CT data, function of brain areas is frequently localized using Positron Emission Tomography (PET) or Single Photon Emission Computed Tomography (SPECT). Thus, multi-modal registration of functional and morphological data provides valuable additional information for diagnosis and therapy (Fig. 1.22).

Fig. 1.22. *Multi-modal registration and fusion [6]. 1. Row*: T1-weighted MRI of a 66 year old subject with right parietal glioblastoma; *2. Row*: Corresponding PET layers after multi-modal registration; *3. Row*: Fusion of registered layers to support intervention planning; *4. Row*: The fusion of MRI with PET of the sensorimotor-activated cortex area proves that the relevant area is out of focus
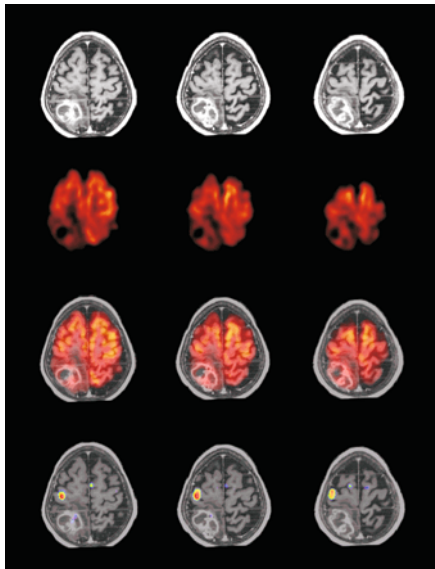

Table 1.5. *Taxonomy of 3D visualization methods.* Triangulation for surface-based rendering is described in textbooks on computer graphics. The marching cube approach is described in the text. As a simple example of surface-based direct volume rendering methods, depth shading visualizes the length of rays passing through the volume until they hit the surface. Integral shading codes the sum of voxel values along the ray as gray scale. It is therefore frequently used to obtain radiograph-like images based on CT data

Concept	Surface-oriented method	Volume-oriented method
Surface reconstruction and rendering	Triangulation	Cuberille approach
		Marching cube
Direct volume rendering	Depth shading	Integral shading
	Depth gradient shading	Transparent shading
	Gray gradient shading	Maximum projection

1.4 Image Data Visualization

Under the concept of image visualization, we had summarized all the transforms which serve the optimized output of the image. In medicine, this includes particularly the realistic visualization of 3D data. Such techniques have found broad applications in medical research, diagnostics, treatment planning and therapy. In contrast to problems from the general area of computer graphics, the displayed objects in medical applications are not given implicitly by formal, mathematical expressions, but as an explicit set of voxel. Consequently, specific methods have been established for medical visualization. These methods are based either on a surface reconstruction or on a direct volume visualization, and lighting and shading are also regarded (Table 1.5).

1.4.1 Marching Cube Algorithm

The marching cube algorithm was specifically developed for surface reconstruction from medical volumes. Here, the voxel is no longer interpreted as a cube of finite edge length but as a point. It is equivalent to a point grid for visualizing volumes. In this volume, a cube is considered with four corners in each of the two adjacent layers. Utilizing symmetry, the complex problem of surface production is reduced to only 15 different topologies, which can be calculated most efficiently since the polygon descriptions that belong to the basic topologies can be stored in a lookup table. Similar to the process of spatial convolution, the cube is positioned successively at all points in the volume dataset (marching). After completion of the marching cube algorithm, a segmented volume is transformed into a triangulated surface. However, the surface is build from a very large number of triangles, which may be reduced significantly by heuristic procedures without any discernible loss of quality. Reducing the number of elements to be visualized supports real-time visualization of the volume.

1.4.2 Surface Rendering

To generate photo-realistic presentations of the volume surface, the lighting is simulated analog to natural scenes. According to the lighting model by Phong, ambient light is created through overlapping of multiple reflections, diffuse scattering on non-shiny surfaces, and direct mirroring on shiny surfaces. While the intensity of the ambient light remains constant in the scene for all surface segments, the intensities of diffuse and speckle reflections depend on the orientation and characteristics of surfaces as well as their distances and directions to the light source and the observing point of viewing.

Without shading, one can recognize the initial triangles. This is a nasty artifact in computer graphics. Therefore, various strategies for shading have been developed to improve significantly the visual impression. For instance, the Gouraud shading results in smooth blunt surfaces, and the Phong shading also provides realistic reflections. In newer applications, transparencies are also modeled to glance at encapsulated objects. Moreover, textures or other bitmaps on the surfaces can be projected to reach a more realistic impression of the scene.

1.4.3 Volume Rendering

Direct volume visualization is abstained from preliminary calculation of the object surface. The visualization is based directly on the voxel data and, therefore, possible without any segmentation. This strategy allows visualization of medical 3D and 4D data by radiologists for interactive localization of pathological areas. The volume is processed either along the data layers (back-to-front or front-to-back) or along an imaginary light ray. Based on

the observer position, rays will be pursued through the volume (ray-tracing). Hereby, the recursive follow-up of secondarily reflected rays is also possible (ray-casting). Although quite realistic visualizations can be provided to the observer, problems arising from the discrete nature of pixel topology (see above) have led to a multitude of algorithmic variants.

In general, parameters are extracted from voxel intensity along the rays and applied as gray or color value at the corresponding position in the viewing plane. This procedure is also referred to as shading. By the methods of the surface-based shading, light source and image plane are placed on the same side of the object, while the volume-oriented procedures radiograph the entire object according to X-ray imaging, i.e., the object is located between light sources and the observation (Table 1.5). Combining direct volume with surface-based approaches, amazingly realistic scenes can be created (Fig. 1.23).
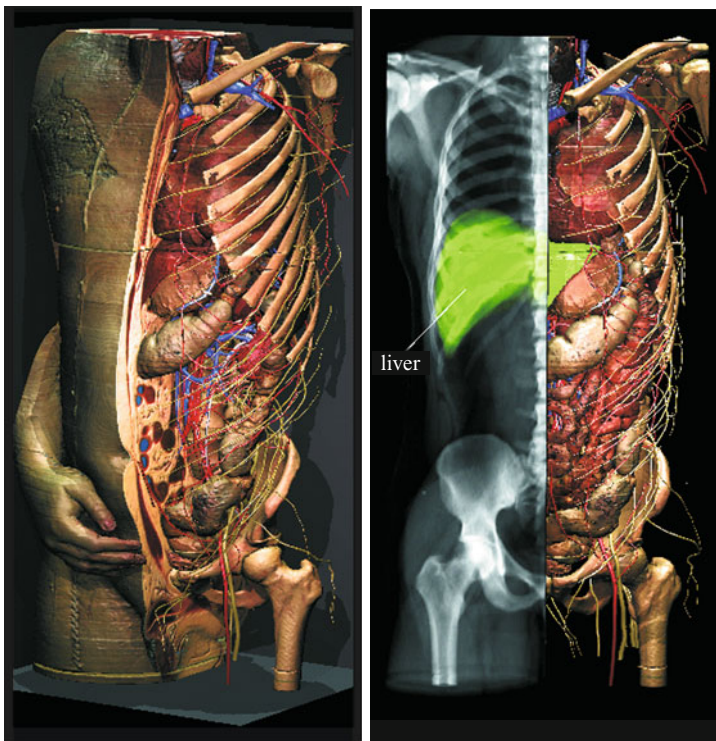

Fig. 1.23. *3D-visualization with Voxel–Man [7].* This 3D model of the internal organs is based on the Visible Human data. The Voxel–Man 3D-Navigator provides unprecedented details and numerous interactive possibilities (*left*). Direct volume rendering and surface-based visualization of segmented objects are combined with integral shading (*right*)

1.5 Visual Feature Extraction

In Fig. 1.1, feature extraction is defined as the first stage of intelligent (high level) image analysis. It is followed by segmentation and classification, which often do not occur in the image itself, i.e., the data or pixel level, but are performed on higher abstraction levels (Fig. 1.2). Therefore, the task of feature extraction is to emphasize image information on the particular level, where subsequent algorithms operate. Consequently, information provided on other levels must be suppressed. Thus, a data reduction to obtain the characteristic properties is executed. The schema in Fig. 1.1 is greatly simplified because many connections between the modules were left out on behalf of readability. So for example, cascades of feature extraction and segmentation at various levels of abstraction can be realized gradually, before classification is eventually performed at a high level of abstraction. Just before classification, a step of feature extraction that is based on the region level is often performed as well.

1.5.1 Data Level

Data-based features depend on the joint information of all pixels. Therefore, all transforms manipulating the whole matrix of an image at once can be regarded for data feature extraction. The most famous example of a data feature transform is the Fourier transform, which describes a 2D image in terms of frequencies, according to their amplitude and phase. Furthermore, the Hough, wavelet or Karhunen-Loève transforms provide possibilities of data feature extraction (see list of textbooks on image processing on page 49). These methods are not in the focus of research in biomedical image processing. In fact, these procedures are rather adapted from technical areas into medical applications.

1.5.2 Pixel Level

Since pixel-based features depend on the values of individual pixels, all point operations that have been defined in Sect. 1.3 can be regarded as feature extraction on the pixel level. Another example was already presented in Fig. 1.21, namely, the arithmetic combination of two images. The subtraction of reference and recall images after appropriate registration in both spatial and value ranges enforce local changes in the images as characteristic pixels.

1.5.3 Edge Level

Edge-based features are defined as local contrast, i.e., a strong difference of (gray scale or color) values of adjacent pixels. Thus, the discrete convolution introduced in Sect. 1.3 can be used with appropriate templates for edge

Fig. 1.24. *Edge extraction using the Sobel operator.* The X-ray image (*center*) was convolved with the eight direction-selective Sobel templates. The strong contrasts on the edges of metallic implants are further strengthened by binarization of the edge images. An isotropic edge image is obtained if, e.g., the maximum at each pixel position is chosen from the eight direction-selective sub-images
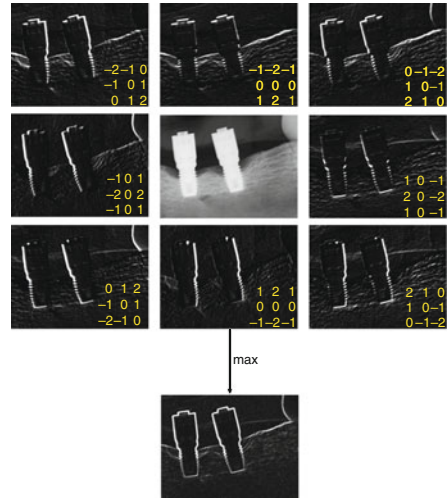
extraction. All masks for high-pass filtering amplify edges in an image. The templates of the so called Sobel operator (Fig. 1.18) are particularly suited for edge extraction. Figure 1.24 exemplarily presents the result of the orientation-selective Sobel masks when applied to a dental radiograph. The edges of the metallic implants are clearly highlighted. An isotropic Sobel-based edge image is achieved, e.g., by a linear or maximum combination of the eight sub-images.

1.5.4 Texture Level

Textural features have been used in medicine for a long time. In textbooks on pathology one can read many metaphors to describe texture, such as a cobblestone-shaped mucosal relief, onion-like stratification of subintima, or honeycomb-structured lung tissue. As intuitive as these metaphors are for people, as difficult is their computational texture processing, and a variety of procedures and approaches have been developed.

Texture analysis attempts to quantify objectively the homogeneity in a heterogeneous but at least subjectively periodic structure (see the spongious bone structure in Fig. 1.18c as an example). In general, we can distinguish:

- *structural approaches* that are based on texture primitives (textone, texture element, texel) and their rules of combinations and
- *statistical approaches* that describe texture by a set of empirical parameters.

1.5.5 Region Level

Regional features are used primarily for object classification and identification. They are normally calculated for each segment after the segmentation process.

The most important parameters to be mentioned here are:

- *localization-descriptive* measurements such as size, position, and orientation of the major axis and
- *delineation-descriptive* measures such as shape, convexity, and length of the border.

Since the degree of abstraction on the region level is rather high as compared to the previous levels, a priori knowledge has already been largely integrated into the image processing chain. Therefore, universal examples cannot be specified. In fact, the definition of regional feature extraction is strongly dependent on the respective application (see Sects. 1.5.5 and 1.6.3).

1.6 Segmentation

Segmentation generally means dividing an image into connected regions. With this definition, the production of regions is emphasized as the pre-stage of classification. Other definitions accentuate the various diagnostically or therapeutically relevant image areas and, thus, focus the most common application of medical imaging, namely, the discrimination between healthy anatomical structures and pathological tissue. By definition, the result of segmentation is always on the regional level of abstraction (cf., Fig. 1.2). Depending on the level of feature extraction as an input to the segmentation, we can methodically classify pixel-, edge-, and texture- or region-oriented procedures. In addition, there are hybrid approaches, which result from combination of single procedures.

1.6.1 Pixel-Based Segmentation

Pixel-based procedures of segmentation only consider the gray scale or color value of current pixels disregarding its surroundings. It should be noted that pixel-based approaches are not segmentation procedures in the strict sense of our definition. Since each pixel is considered only isolated from its neighborhood, it cannot be ensured that actually only connected segments are obtained. For this reason, post-processing is required, e.g., by morphologic filtering (see Sect. 1.3.3). Most pixel-based procedures use thresholds in the histogram of an image and employ more or less complex methods to determine this threshold. Furthermore, statistical methods for pixel clustering are used.

Static Thresholding

If the assignment of pixel intensities is well known and constant for a certain type of tissue, static thresholds are applicable. A static threshold is independent of the individual instance in a set of similar images. For example, bone or soft tissue windows in the CT can be realized (Fig. 1.25) with static thresholds on the Hounsfield Unit (HU).

Fig. 1.25. *Static thresholding [5].*
Pixel-based segmentation in CT relies
on Hounsfield Units (HU), which
allow the definition of windows for
different types of tissue: bone
$[200\ldots3{,}000]$, water $[-200\ldots200]$,
fat $[-500\cdots-200]$, or air
$[-1{,}000\cdots-500]$

● air   ● fat   ● water   ○ bone
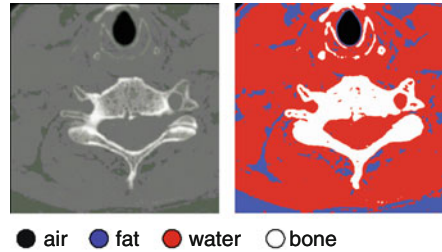
Adaptive Thresholding

Globally adaptive thresholds result from analyzing each individual image
entirely. They are exclusively used in this image. The well-known method
of Otsu is based on a simple object vs. background model. The threshold
in the histogram is determined such that the two resulting classes minimize
the intra-class variance of gray scale values, while the inter-class variance is
maximized. For example in skeletal radiography, bone, soft tissue and back-
ground can be seen, but the actual mean gray scale of this tissue classes may
vary with respect to illumination and exposure parameters. By adopting the
threshold to the image, the Otsu segmentation is able to balance this variation
in imaging.

Using *locally adaptive* thresholds, the threshold is computed not only for
each image individually, but also for each region within an image. In the
extreme case, an individual threshold is determined for every pixel posi-
tion (i.e., pixel-adaptive). This is particularly necessary if the simple object
to background assumption is globally invalid because of continuous bright-
ness gradients. For example, due to the irregularity of optical illumination,
the background in microscopy imaging of cell cultures (Fig. 1.26a) runs from
light shades of gray (top right) to dark shades of gray (bottom left), where
also the gray scale values of the cells are located. A global threshold deter-
mined with the dynamic procedure of Otsu (Fig. 1.26b) does not separate
the cells from backgrounds, although the global threshold had been deter-
mined image-individually. The locally adaptive segmentation (Fig. 1.26c) leads
to a significantly improved result, but isolated block artifacts appear. These
artifacts can be avoided only by pixel-adaptive thresholding (Fig. 1.26d).

Clustering

Pixel clustering is another way of pixel-based segmentation. This statistical
method is particularly suitable if more than one value is assigned to each pixel
and regarded in the segmentation process (e.g., color images). Figure 1.27
illustrates the iso-data clustering algorithm (also referred to as k-means clus-
tering) in a simple 2D case. All pixel values are registered as data points in
the 2D feature space. Initialized by the number of segments to be obtained,

Fig. 1.26. *Dynamic thresholding in microscopy [8].* The microscopy of a cell culture (**a**) was segmented using a global threshold (**b**), locally adaptive (**c**) and pixel-adaptive (**d**). According to morphological post-processing for noise reduction and a connected components analysis, the final segmentation is shown in (**e**)
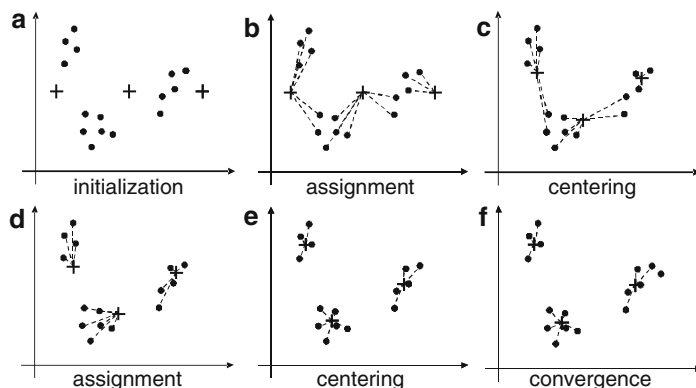

Fig. 1.27. *Iso-data pixel clustering.* The iterative iso-data algorithm for pixel clustering is exemplified in a 2D feature space. The number of clusters is given a priori. After arbitrary initialization, the data points are assigned to the nearest cluster center. Then, the positions of the centers are recalculated and the assignment is updated until the process finally converges. The final location of cluster centers is not affected by their initial position. This may only have impact to the number of iterations

the initial cluster centers are arbitrarily placed by the algorithm. Then, the following two steps are repeated iteratively until the process converges:

1. Each data point is aligned to the closest cluster center.
2. Based on the current assignment, the cluster centers are recalculated.

It can be proven mathematically that the resulting cluster centers are independent of initial positions, which may only impact the number of iterations and hence, the calculation time. However, either a fixed distance metrics (e.g., Euclidean (geometric) distance) or a data-adaptive metrics (e.g., Mahalanobis distance) must be selected, which certainly impacts the clustering result. Also, the predefined number of cluster centers is an important parameter. If the application domain does not allow to determine the number of segments a priori, pixel clustering can be performed for a different number

of centers and the residual error of the computed model can be analyzed to determine the appropriate number of centers.

Post-Processing

Segments obtained from pixel-based analysis usually are incoherent and highly noisy (see Fig. 1.25 or Fig. 1.26b). Therefore, post-processing is required. Noisy structures can be effectively reduced with methods of mathematical morphology. While a morphologic opening removes spread parts from the segments, holes are closed by morphologic closing (see Sect. 1.3.3). The connected components algorithm provides each separated segment with a unique reference number. In the segmentation of the cell image (Fig. 1.26a); clustering provides a rough cluster of "cells", which is separated from the "background", although many individual cells are shown separately in Panel 1.26(d). After morphological post-processing and connected components analysis, cells are separated and colored (labeled) differently according to their segment number. Now, they can be further processed as independent objects (Fig. 1.26e).

1.6.2 Edge-Based Segmentation

This type of segmentation is based on the abstract level of edges and tries to capture the objects due to their closed outline in the image. Hence, edge-based segmentation procedures are only used for such problems, where objects are represented as clearly defined boundaries. As described in Sect. 1.1.3, this occurs rather seldom when biological tissue is imaged. One of these special cases is a metallic implant, which is displayed in a radiograph.

In general, the image processing chain for edge-based segmentation is composed of edge extraction and edge completion. Edge extraction is usually obtained by edge-based feature extraction, as described in Sect. 1.5.3, such as generated with the Sobel filter (see Fig. 1.24). The next steps of processing are binarization, to obtain only edge pixels and non-edge pixels, morphological filtering to reduce noise and artifacts, and, finally, a skeleton of the edge is computed. Tracing and closing of binary contours are the main tasks of the edge-based segmentation. Almost exclusively, heuristic methods are used. For example, one can search along differently directed rays to find connecting pieces of a contour. This procedure aims at bridging local gaps on the edge profile.

Livewire Segmentation

In practice, edge-based segmentation is often realized semi-automatically. By the interactive livewire segmentation, the user clicks onto or near by the edge of the Object of Interest (OOI), and the computer determines the exact edge location based on local gradients. Then, the computer calculates a cost function, which again is based on local gradients. For all paths (wire) to the current

position of the cursor, the path with the lowest cost is displayed in real time (live) as the cursor is moved manually. Therefore, the metaphor "livewire" is commonly used to refer to this interactive method of segmentation. If the cursor moves far from the object, the contour is lost but if the cursor is placed near to the contour again, the cost function ensures that the wire snaps back to the desired object. Finally, the user must provide only a few supporting points by hand and can directly verify the correctness of segmentation (Fig. 1.28). Application of such procedures can be found at computer-assisted (semi-automatic) segmentations in layers of CT data, e.g., to produce a model for surgical intervention planning. Guided by the cost function, the segmentation result (delineation) is independent of the user placing the supporting point (localization).

1.6.3 Region-Based Segmentation

As an advantage of region-based segmentation, only connected segments are produced, and morphological post-processing is avoided. There are agglomerative (bottom-up) and divisive (top-down) approaches. All approaches are based on a certain distance or similarity measure to guide the assignment of neighbored pixels or regions. Here, plenty of methods are used. Easiest, one can compare the mean gray value but complex texture measures (see Sect. 1.5.4) are often used, too.
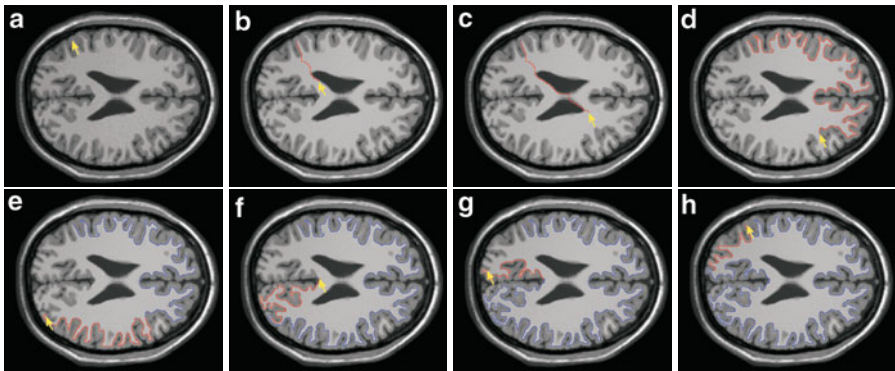

Fig. 1.28. *Edge-based interactive livewire segmentation [9].* The user marks a starting point with the cursor (*yellow*) on the border between white and gray matter (**a**). The connection to the current cursor position is denoted with red, cf. (**b**) to (**e**). Depending on the cursor position, the contour can also jump between very different courses (**d**, **e**). So, the user can interactively place an appropriate fix point. The fixed curve segment is shown in blue, cf. (**e**) to (**g**). In this example, only five points are manually marked to achieve a complete segmentation (**h**)

Agglomerative Algorithm

Region growing, in 3D also referred to as volume growing, is a well known example of an agglomerative procedure. Starting from seed points, which may be placed either automatically or manually, neighbored pixels are iteratively associated to the growing areas if the distance measure is below a certain threshold. This process is iterated until no more merges can be carried out. From this qualitative description, the variety and sensitivity of the parameters of such procedures are already clear. Special influence on the result of agglomerative segmentation has:

- the number and position of seed points,
- the order in which the pixels or voxels are iteratively processed,
- the distance or similarity measure applied, and
- the threshold used to guide merging.

Therefore, agglomerative algorithms for segmentation often are affected by small shifts or rotations of the input image. For instance, if x- and y-axis of the image matrix are transposed, the result of segmentation is different regarding size and shape of OOI, which is an unwanted effect in medical image processing.

Divisive Algorithm

The divisive approach somehow inverts the agglomerative strategy. By splitting, the regions are iteratively subdivided until they are considered sufficiently homogeneous in terms of the chosen similarity measure. As an advantage, seed points are not required anymore, because the first split is performed throughout the whole image. As a drawback, the dividing lines are usually drawn horizontally or vertically, and this arbitrary separation may separate the image objects. Therefore, split is unusually performed as a self standing segmentation procedure, but rather combined with a subsequent merging step (split and merge). Another drawback of divisive segmentation procedures is the resulting wedge-formed boundary of objects, which may require post-processing such as contour smoothing.

1.6.4 Over- and Under-Segmentation

A fundamental problem of pixel- and region-based segmentation is the dualism between over- and under-segmentation. For a definition of these terms, we rely on the general model of the image processing chain (see Fig. 1.1). Here, segmentation is regarded as a pre-stage for classification, in which the extracted image segments are assigned to their semantic meaning. This can take the form of automatically assigning concrete terms for the segments (for example, the organ "heart" or the object "TPS implant screws" or, more abstract, a "defect" or an "artifact").

In any case, the segment should be related directly to an object, if an automatic classification is desired. In this context, under-segmentation occurs if resulting segments are composed from parts of several objects. Analogously, over-segmentation is obtained if a particular object is disintegrated into several segments or parts of segments. The big problem with segmentation of medical images is that over- and under-segmentation usually occur simultaneously.

Hierarchical Algorithm

Hierarchical procedures are one of the concepts to deal with the dualism between over- and under segmentation. Starting on a lower resolution of the image, where it is represented with a small number of pixles only, the chance of splitting objects into more than one segment is decreased. Then, the exact outline of each segment is reconstructed on higher resolutions, where more details are contained (Fig. 1.29).

Hybrid Algorithm

In the practice of medical image processing, hybrid approaches of segmentation have come to the greatest importance. Here, one is trying to combine the advantages of individual (usually edge- and region-based) algorithms without maintaining their disadvantages.

For example, the watershed transform extends an agglomerative, regional segmentation procedure with edge-based aspects of segmentation. Indeed, it is based on the very intuitive analogy of the image with a topographic surface: the gray levels of the pixels correspond to the altitude of the relief. In
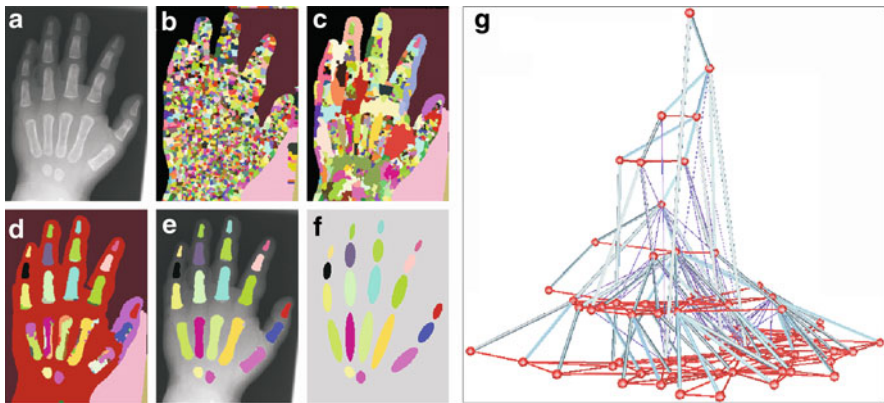

Fig. 1.29. *Hierarchical region merging.* The skeletal radiograph of the hand (**a**) has been segmented at various levels of resolution, cf. (**b**) to (**d**). The initial step is obtained with the watershed transform (see Sect. 1.6.4). Depending on the size of the objects, they can be localized in the appropriate level (**e**), approximated by ellipses (**f**), or visualized as nodes in a graph (**g**)

hydrology, a catchment basin relative to a body of water (e.g., a river or an ocean) is defined as a region where water from rain drains downhill into this reference body, whereas the watershed lines (also known as water parting or divide) separate different drainage basins. Similarly, a catchment basin surrounded with watershed lines can be defined for every regional minimum in the image. In general, the image gradient is taken as the topographic surface, so as the catchment basins to correspond to connected and homogeneous regions (structures of interest), and watershed lines to lie on higher gradient values.

The so-called classical watershed transform takes into account all regional minima of the image to compute a primitive catchment basin for each one. As natural images contain many regional minima, in general, too many basins are created. The image is over-segmented (see, for example, Fig. 1.29b). However, over-segmentation can be reduced by filtering the image and, therefore, decreasing the number of minima.

On the other hand, when applied to segmentation of medical images, the watershed transform especially has the following advantages:

- From the region-based idea of the flooding process, contiguous segments are determined inherently.
- From the edge-based approach of the watersheds, the objects are exactly delineated.
- The problem of under-segmentation is avoided, since the merging of smaller pools is prevented by the watersheds.

1.6.5 Model-Based Segmentation

State of the art methods for model- or knowledge-based segmentation involve active contour models and deformable templates as well as active shape and active appearance models.

Active Contour Model

Active contour models apply edge-based segmentation considering region-based aspects and an object-based model of a priori knowledge. In the medical application domain, so called snake and balloon approaches are applied for segmentation of 2D and 3D image data and the tracing of contours in 2D image and 3D image sequences, i.e., 3D and 4D data, respectively. The contour of the objects, which is usually closely modeled, is presented by individual nodes, which are – in the simplest case – piecewise connected with straight lines forming a closed polygon. For the nodes, a scalar quality measure (e.g., energy) is calculated and optimized in the local environment of the nodes. Alternatively, adjusted forces are determined that directly move the nodes. The iterative segmentation process completes at minimal energy or if an optimum balance of forces was found. Thus, the potential of this approach is kept in the choice of capable quality criteria (e.g., energy) or forces.

Snake

In 1988, Kass et al. have introduced classical snake approach [10]. It models an internal and an external quality criterion, both as undirected energy. The internal energy results from a predefined elasticity and stiffness of the contour, which is high in places of strong bends or on buckling. The external energy is calculated from an edge-filtered image. The external energy is small, if the contour runs along edges. The idea behind this approach is an edge-based segmentation combined with the a priori knowledge that biological objects rarely have sharp-bending boundaries. With an optimal weighting of energy terms, the contour course is primarily determined by the information of edges in the image. However, if the object's contour is partially covered or incompletely captured, the internal energy ensures an appropriate interpolation of the region's shape.

So simple this approach has been formulated verbally, so difficult it is to implement. During the iteration, the number of nodes must be constantly adjusted to the current size of the contour. Furthermore, crossovers and entanglements of the moving contour must be avoided. The classical snake approach also requires an already precisely positioned starting contour, which often must be defined interactively. Then, the two steps of segmentation, i.e., localization and delineation are performed again by man and machine, respectively. This concept was also applied in the first publications of this segmentation method. For a contour tracking of moving objects in image sequences, the segmentation of image at time t serves as initial contour of iteration in image $t + 1$. After a single initialization for the image $t = 0$, the procedure runs automatically. Hence, fluoroscopy and endoscopy are suitable modalities for the application of the snake approach to track the shape of moving objects.

Balloon

Balloons are based on forces rather than energies. Besides the internal and external force, an inner pressure or suction is modeled, which lets the contour continuously expand or shrink. Figure 1.30 shows the inflation movement of a balloon to segment the cell membrane, which is visualized by the synaptic boutons of contacting dendrites in a microscopy of a motoneuron. Although segmentation is done without an accurate initial contour, in the course of iteration the balloon nestles onto the real contour of cell membrane. Another advantage of the balloon model is that this concept is directly transferable into higher dimensions (Fig. 1.31).

Other Variants

In recent developments of active contour models, it is attempted to incorporate further a priori knowledge, e.g., in the form of anatomical models. Prototypes of the expected object shapes are integrated into the algorithm: In each iteration, the distance of the current object shape to a suitable selected
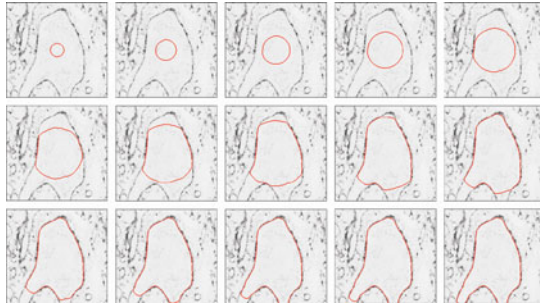
Fig. 1.30. *Balloon segmentation of motoneuron cell membrane [11].* The frames show the balloon at different iterations. By touching the cell membrane, the strong image forces prevent further movement of the active contour. In this application, the internal forces correspond physically to a membrane. This is clearly recognizable at the "adhesion border" of the balloons reaching the dendrites (*bottom left*)
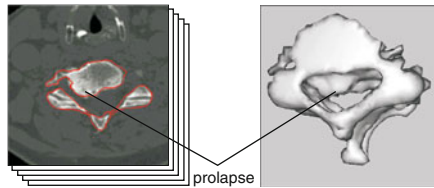

Fig. 1.31. *Segmentation with a 3D balloon model [12].* The CT of a spine (*left*) was segmented with a 3D balloon. In the surface-based rendering after automatic segmentation, the prolapse is clearly visible (*right*). The visualization is based on Phong shading (see Sect. 1.4.2)

prototype is modeled as an additional force on the node. With those extensions, a "break out" of the active contour model is prevented also for long passages of the local object boundary without sufficient edge information.

The complex and time-consuming parameterization of an active contour model for a specific application can be based on manual and also automatic reference segmentations. For the latter approach, different combinations of parameters are determined and the segmentation is performed for all cases. All resulting segmented contours are compared with the appropriate reference contour, a priori defined as the ground truth of the training data. Then, that set of parameters with the best approximation of the reference contour is selected automatically.

Active Shape Model

In the biomedical sciences, OOIs such as bones or organs often have a similar form or projected shape that may vary between individuals or different

points of time. Therefore, a probabilistic model may be applied to explain the shape variation. Segmenting an image imposes constraints using this model as a prior. Usually, such a task involves:

1. registration of the training examples to a common pose,
2. probabilistic representation of the variation of the registered samples, and
3. statistical inference between the model and the image.

Introduced by Cootes et al. in 1995, active shapes aim at matching the model to a new image [13]: for probabilistic representation (Step 2), the shapes are constrained by the Point Distribution Model (PDM) allowing variation only in ways that have been seen in the training set of labeled examples. For statistical inference (Step 3), a local neighborhood in the image around each model point is analyzed for a better position. Alternating, the model parameters are updated to best match to these newly determined positions, until convergence is reached.

Similar to active contour models, each training shape is represented by a set of points, where each point corresponds to a certain landmark. To form a feature vector x_i, all landmark coordinates are concatenated. A mean shape $\bar{x}$ and its covariance matrix S from N training sets is obtained by

$$\bar{x} = \frac{1}{N} \sum_{i=0}^{N-1} x_i \quad \text{and} \quad S = \frac{1}{N} \sum_{i=0}^{N-1} (x_i - \bar{x})(x_i - \bar{x})^T \tag{1.3}$$

The Principle Component Analysis (PCA) is applied for dimension reduction computing normalized eigenvectors and eigenvalues of S across all training shapes. The base Φ of eigenvectors ϕ represents the principle modes of variation, and the eigenvalues λ indicate the variance per mode. The prior model is generated from the t largest eigenvalues. Now, any shape x may be approximated by $x \approx \bar{x} + \Phi \nu$, where the weighting vector ν is determined minimizing a distance measure in the image, e.g., the Mahalanobis distance.

Figure 1.32 shows an application of the active shape method for bone age assessment. The BoneXpert[®] method[2] robustly detects carpal bones and phalanges as well as epiphysis using active shapes.

1.7 Classification

According to the general processing chain (see Fig. 1.1), the task of the classification step is to assign all connected regions, which are obtained from the segmentation, to particularly specified classes of objects. Usually, region-based features that sufficiently abstract the characteristics of the objects are used to guide the classification process. In this case, another feature extraction step is performed between segmentation and classification, which is not visualized

---

[2] Visiana Ltd, Holte, Denmark, `http://www.bonexpert.com`

Fig. 1.32. *Active shape segmentation of hand radiopraph.* BoneXpert® detects relevant bones and measures distances, geometry and sizes to compute skeletal maturity. Although in this example the first metacarpal bone is misaligned, the automatically suggested bone age, which is computed over several regions, lies within the range of human inter-observer variation

in Fig. 1.1. These features must be sufficiently discriminative and suitably adopted to the application, since they fundamentally impact the resulting quality of the classifier.

For all types of classifiers, we can differ supervised (trained), unsupervised (untrained) and learning classification. For example, pixel clustering, which has been already introduced for pixel-based segmentation, is an unsupervised classification process (see Fig. 1.27). As a goal, individual objects are divided into similar groups. If the classification is used for identification of objects, the general principles or an exemplary reference must be available, from which the ground truth of classification can be created. The features of these samples are then used for parameterization and optimization of the classifier. Through this training, the performance of the classifier can be drastically improved. However, supervised object classification is always problematic, if the patterns that are classified differ remarkably from the trained patterns. In such cases, the training set does not sufficiently reflect the real world. A learning classifier has advantages here, because it changes its parameterization with each performed classification, even after the training phase. In the following, however, we assume a suitable set of features that are sufficiently characteristic and large set of samples.

The classification itself reverts mostly to known numerical (statistical) and non-numerical (syntactic) procedures as well as the newer approaches of Computational Intelligence (CI), such as neural networks, evolutionary algorithms, and fuzzy logic. In general, the individual features, which can be determined by different procedures, are summarized either to numerical feature vectors (also referred to as signature) or abstract strings of symbols. For example, a closed contour object can be described by its Fourier-descriptors as a feature

vector, or by means of basic line items such as "straight", "convex", and "concave" forming a symbol chain.

1.7.1 Statistic Classifiers

Statistical classification regards object identification as a problem of the statistical decision theory. Parametric procedures for classification are based on the assumption of distribution functions for the feature specifications of the objects, and the parameters of the distribution functions are determined from the sample. Non-parametric methods, however, waive such model assumptions, which are sometimes unobtainable in biomedical image processing. A common example of such a non-parametric statistical object classifier is the Nearest Neighbor (NN) classifier. All features span the feature space, and each sample is represented by a point in this feature space. Based on the signature of a segment, which has not been included in the training and now is assigned to its nearest neighbor in feature space, the segment is classified to the associated class of the assigned feature vector. The k-Nearest Neighbor (k-NN) classifier assigns the majority class from the k nearest neighbors in feature space (usually, $k = 3$ or $k = 5$). An example of the k-NN classifier is given in Fig. 1.33.

1.7.2 Syntactic Classifiers

In symbol chains, it is neither useful nor possible to define distance measurements or metrics and to evaluate the similarity between two symbol chains, such as used for feature vectors. An exception of this statement is given with the Levenshtein distance, which is defined as the smallest number of modifications such as exchange, erase, or insert, required to transform a symbol chain into another.

The syntactic classification is therefore based on grammars, which can possibly generate an infinite amount of symbol chains with finite symbol formalism. A syntactic classifier can be understood as a knowledge-based classification system (*expert system*), because the classification is based on a formal heuristic, symbolic representation of expert knowledge, which is transferred into image processing systems by means of facts and rules. If the expert system is able to create new rules, a learning classifier is also realizable as a knowledge-based system.

It should be noted that the terms "expert system" or "expert knowledge", however, are not standardized in the literature. Therefore, "primitive" image processing systems, which use simple heuristics as implemented distinction of cases to classification or object identification, are also referred to as "knowledge-based".
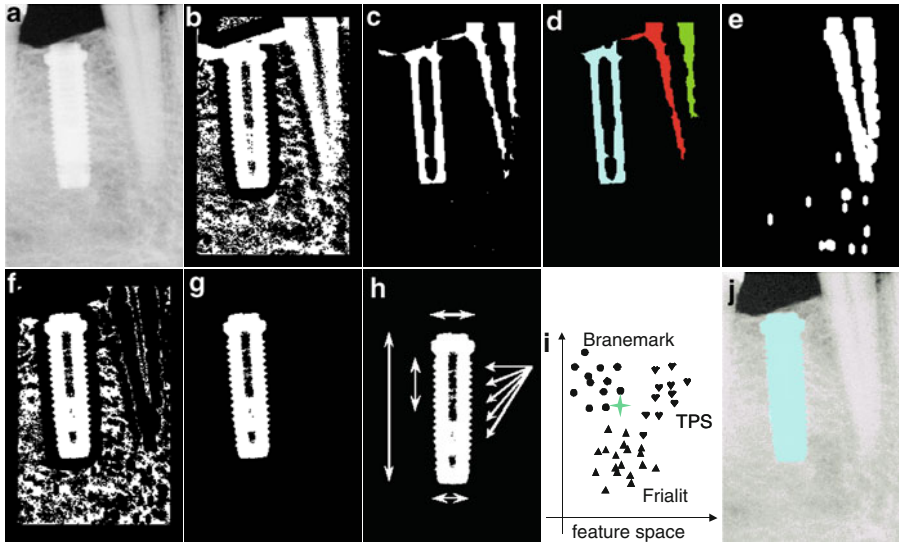
Fig. 1.33. *Identification of dental fixtures [14].* An implant is shown in the intra-oral radiograph of the lower jaw (**a**). For feature extraction, the image is binarized with a local adaptive threshold (**b**). The morphological filtering (erosion) separates individual areas (**c**) and eliminates interference. In this example, three regions were segmented (**d**). Further processing is shown for the blue segment. After its fade-out, the gap of morphological erosion is compensated by a subsequent dilation (**e**), and the result is subtracted from the intermediate image (**b**). Any coordinate of blue segment from (**d**) identifies the corresponding region, which can be extracted now (**g**) and aligned into a normal position using the Karhunen-Loève transform. Geometric dimensions are determined as region-based features and stored in a feature vector (signature). As part of the training, the reference measures of different implant types have been recorded in the feature space. The classification in the feature space is done with the statistical *k*-NN classifier (**i**), which identifies the blue segment reliably as Branemark implant screw (**j**)

1.7.3 Computational Intelligence-Based Classifiers

As part of the artificial intelligence, the methods of CI include neural networks, evolutionary algorithms and fuzzy logic. These methods have their examples in biological information processing. Although they usually require high computational power, they are frequently used in biomedical image processing for classification and object identification. Thereby, all the procedures have a mathematical-based, complex background.

Neural Network

Artificial neural networks simulate the information processing in the human brain. They consist of many simply constructed basic elements (i.e., neurons), which are arranged and linked in several layers. Each neuron calculates

the weighted sum of its input excitations, which is mapped over a nonlinear function (i.e., characteristic curve) to the output. The number of layers, the number of neurons per layer, the network's topology, and the characteristic curve of the neurons are predefined within the network dimensioning step. On the one hand, heuristics are usually applied rather than methodological derivations. On the other hand, the individual weights of the excitements are identified numerically during the training of the network. Then, the network remains unchanged and can be used as a classifier.

Evolutionary Algorithm

Evolutionary algorithms are based on the constant repetition of a cycle of mutation and selection following the Darwinian paradigm of the survival of the fittest. Genetic algorithms work on a number of individuals (the population). The crossing of two randomly selected individuals and afterwards the mutation changes the population. A fitness function evaluates the population in terms of their goodness to problem solution. Although the selections are equipped with a random component, fit individuals are frequently selected for reproduction. Evolutionary algorithms can solve complex optimization problems amazingly well, but for object classification, they are less successfully used than other methods.

Fuzzy Algorithm

The idea of fuzzy logic is to extend the binary (TRUE or FALSE) computer model with some uncertainty or blur, which exists in the real world, too. Many of our sensory impressions are qualitative and imprecise and, therefore, unsuitable for accurate measurements. For example, a pixel is perceived as "dark", "bright" or even "very bright", but not as a pixels with the gray scale value "231". Fuzzy quantities are based mathematically on the fuzzy set theory, in which the belonging of an element to a set of elements is not restricted to the absolute states TRUE (1) or FALSE (0), but continuously defined within the entire interval [0..1].

Beside classification, applications of fuzzy logic in biomedical image processing can be found also for pre-processing (e.g., contrast enhancement), feature extraction (e.g., edge extraction, skeleton), and segmentation.

1.8 Quantitative Measurements and Interpretation

While the visual appraisal by experts is qualitative and sometimes subject to strong inter- as well as intra-individual fluctuations, in principle, a suitable computer-aided analysis of biomedical images can deliver objective and reproducible results. First of all, this requires a precise calibration of the imaging

modality. Furthermore, partial (volume) effects of the imaging system and particularities of the discrete pixel topology must be taken into account and handled accordingly to ensure reliable and reproducible measurements.

Quantitative measurement is focused on automatic detection of objects as well as their properties. Image interpretation steps further towards analyzing the order of individual objects in space and/or time. It may be understood in the sense of analyzing an abstract scene that corresponds to the ambiguous goal of developing a "visual sense for machines", which is as universal and powerful as that of humans.

1.8.1 Partial Volume Effect

The digitalization of the local area or volume of a pixel or voxel, respectively, always yields an averaging of the measured value in the appropriate field. For example in CT, a voxel containing different tissue is assigned a certain Hounsfield value that results from the proportional mean of the individual Hounsfield values of the covered tissue classes. Thus, a voxel containing only bone and air preserves the Hounsfield value of soft tissue and, thus, may distort quantitative measurements. In general, this partial (volume) effect occurs in all modalities and must be accounted appropriately for any automatic measurement (see Fig. 1.14).

1.8.2 Euclidean Paradigm

The common paradigms of the Euclidean geometry do not apply in the discrete pixel domain. For example, the discrete representations of two straight lines may not join in a common pixel although the lines are crossing. Furthermore, different neighborhood concepts of discrete pixel's topology have remarkable impact on the result of automatic image measurements. In particular, the areas identified in region growing may be significantly larger if the 8-neighborhood is applied, i.e., if eight adjacent pixels are analyzed instead of the four direct neighbors (4-neighborhood).

1.8.3 Scene Analysis

The fundamental step of image interpretation is to generate a spatial or temporal scene description on the most abstract level (symbolic image description, see Fig. 1.2). A suitable form of representation is the attributed relational graph (semantic web), which can be analyzed at different hierarchy levels (see Fig. 1.29, right). Therefore, the considered grid matrix of pixels (iconic image description, see Fig. 1.2) so far is inappropriate for image interpretation.

The primitives of the graph (node) and their relationships (edges) must be abstracted from the segmented and identified objects or object parts in the image. So far, only a few algorithms can execute this level of abstraction. Examples for the abstraction of primitives are given by the numerous

approaches to shape reconstruction: Shape-from-shading, -texture, -contour, -stereo, etc. Examples for the abstraction of relationships can be found at the depth reconstruction by trigonometric analysis of the projective perspective. Recently, considerable progress has been achieved in symbolic image analysis in the fields of industrial image processing and robotics. Because of the special peculiarities of the biomedical imagery (see Sect. 1.1.3) the transfer of these approaches into health care applications and medical image processing is only sparingly succeeded so far.

1.8.4 Examples

We will now discuss some examples for image measurements. For instance in Fig. 1.33, geometrical features are used for the automatic classification of implant systems. The feature measures are extracted on the abstract level of regions. Frequently, further measures are extracted after object identification, which use the information of the certain object detected, i.e., they operate on the level of objects. In Fig. 1.33i, we can use the knowledge that the blue segment corresponds to a Branemark implant to parameterize a special morphological filter that is adapted to the geometry of Branemark implants and count the number of windings of the screw.

Another example of object-based image measurements is given in Fig. 1.34. The result of balloon segmentation of a cell membrane (see Fig. 1.30) is labeled automatically with local confidence values based on model assumptions (Fig. 1.34a). These values indicate the contour segment belonging to a cell membrane and thus a classification via fuzzy logic (see Sect. 1.7.3).
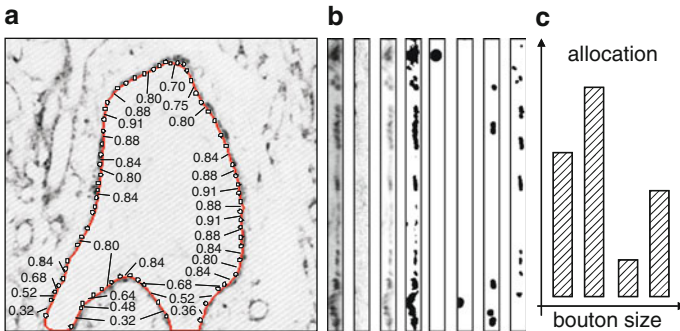

Fig. 1.34. *Quantification of synaptic boutons on a cell membrane [15].* The cell membrane was segmented with a balloon (see Fig. 1.28). Analyzing the impact of internal vs. external forces at a certain vertex, local confidences can be determined to fuzzily classify the affiliation of the contour section to the actual cell membrane (**a**). The cell contour is extracted, linearized, normalized, and binarized before the occupation of the cell membrane with synaptic boutons of different sizes is analyzed by morphological filtering (**b**). The confidence values are considered for averaging the occupation measure along the cell membrane (**c**)
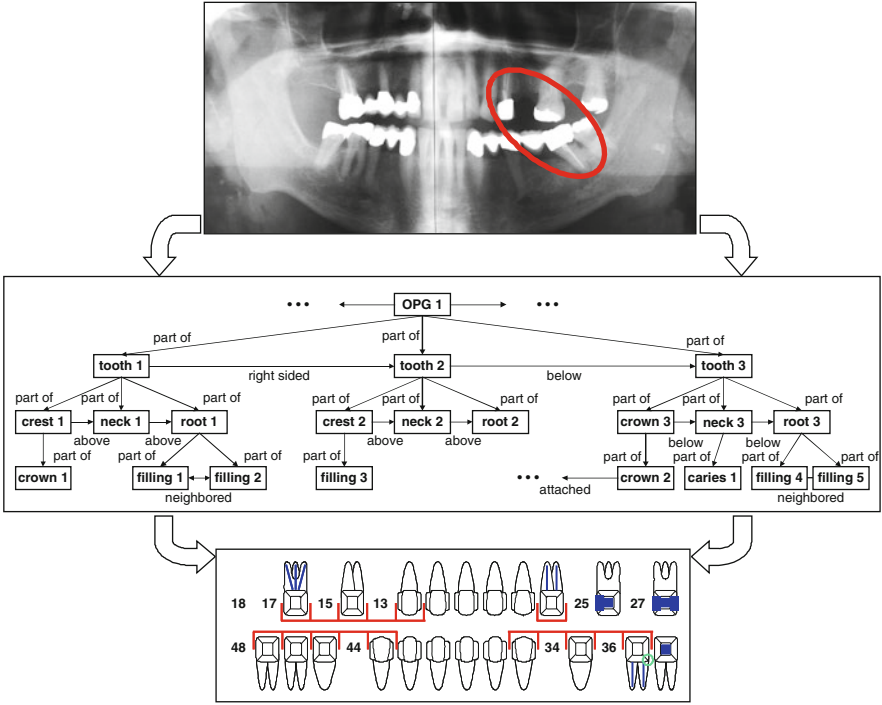
Fig. 1.35. *Scheme of automatic image interpretation.* The panoramic radiograph contains all relevant information of a dental chart. The symbolic description of the scene is obtained with a semantic network. Despite its already considerable complexity, the shown part of the network represents only the marked ROI. In the dental chart, information is coded differently. The teeth are named in accordance with the key of the Fédération Dentaire Internationale (FDI): the leading digit denotes the quadrant clockwise, the second digit refers to the number of the tooth, counting from inside to outside. Existing teeth are represented by templates, in which dental fillings, crowns and bridges are recorded. The green circle at tooth 37 (say: three, seven) indicates a carious process

To increase robustness and reliability of measurements, the confidence values are accounted for an averaging of quantitative measures along the contour, which are extracted, linearized, normalized, and morphologically analyzed (Fig. 1.34b), such that finally a reliable distribution statistics of connecting boutons according to their size is obtained (Fig. 1.34c).

Figure 1.35 displays exemplarily the automatic extraction of a dental chart based on image processing of a panoramic radiograph. It clearly shows the immense difficulties, which have to be faced by the automatic interpretation of biomedical images. Initially, the segmentation and identification of all relevant image objects and object parts must succeed, so that the semantic network can be built. This includes the instances ("tooth 1", "tooth 2", etc.) of the

previously identified objects (e.g., "teeth", "crown", "filling"). The interpretation of the scene based on the network must be carried out in a further, not less difficult step of processing. Thus, all teeth must be named according to their position and shape. Then, crowns, bridges, fillings, and carious processes can be registered in the dental chart. However, the automation of this process, which can be accomplished by dentist in a few minutes, is not yet possible automatically with sufficient robustness.

1.9 Image Management

Introductorily, we have summed with the term "image management" all image manipulation techniques, which serve the effective archiving (short and long term), transmission (communication) and the access (retrieval) of data (see Fig. 1.1). For all three points, the specifics in medical applications and health care environments have led to specific solutions, which are briefly introduced in the following sections.

1.9.1 Archiving

Already in the seventies, the invention of CT and its integration with clinical routine has involved the installation of the first Picture Archiving and Communication System (PACS), which main task is the archiving of image data. The core problem of archiving medical images is the immensely large volume of data. A simple radiography with 40×40 cm (e.g., a chest X-ray) with a resolution of five line pairs per millimeter and 10 bit $= 1,024$ gray levels per pixel already requires a storage capacity of more than 10 MB. Digital mammography, which is captured with high resolution on both breasts in two views results in about 250 MB of raw data for each examination. Ten years ago, radiography, CT, and MRI accumulated in a university hospital to already about 2 TB of image data each year (Table 1.6). This estimate can easily increase tenfold with the resolution-increased novel modalities such as spiral CT and whole-body MRI. For instance in Germany, according to relevant legislations, the data must be kept at least for 30 years. Therefore, efficient storage, retrieval, and communication of medical images have required effective compression techniques and high speed networks. Due to noise in biomedical images, lossless compression usually has a limited effect of compression rates of two or three. Only in recent years, feasible hybrid storage concepts have become available. Storage of and access to medical image data is still of high relevance.

1.9.2 Communication

With increasing digitization of diagnostic imaging, the motto for medical information systems, i.e., to provide "the right information at the right time and

Table 1.6. *Volume of medical image data [5].* The data is taken from the Annual Report 1999 of the University Hospital of RWTH Aachen University, Aachen, Germany (about 1,500 beds). The data is based on the Departments of (i) Diagnostic Radiology, (ii) Neuroradiology, (iii) Nuclear Medicine, and (iv) Dentistry, Oral and Maxillofacial Surgery for a total of 47,199 inpatient and 116,181 outpatient images. Services (such as ultrasound, endoscopic or photographic) from other departments were excluded. For modalities of nuclear medicine, 20 slices per study are assumed. For comparison, the total number of analyses performed in the central laboratory of the Institute for Clinical Chemistry and Pathobiochemistry was estimated with an average of 10 measured values per analysis with highest precision of 64 bit. But still, the annual image data volume is about 10,000 times larger

Modality	Resolution Spatial [pixel]	Range [bit]	Size per image [MB]	Units in year 1999	Total per year [GB]
Chest radiography	4000 × 4000	10	10.73	74,056	775.91
Skeleton radiography	2000 × 2000	10	4.77	82,911	386.09
CT	512 × 512	12	0.38	816,706	299.09
MRI	512 × 512	12	0.38	540,066	197.78
Other radiography	1000 × 1000	10	1.19	69,011	80.34
Panoramic and skull	2000 × 1000	10	2.38	7,599	17.69
Ultrasound	256 × 256	6	0.05	229,528	10.11
Dental radiography	600 × 400	8	0.23	7,542	1.69
PET	128 × 128	12	0.02	65,640	1.50
SPECT	128 × 128	12	0.02	34,720	0.79
Σ					1,770.99
For comparison					
Laboratory tests	1 × 10	64	0.00	4,898,387	0.36

the right place," is projected to the field of medical image processing. Hence, image communication is the core of today's PACS. Image data is not only transferred electronically within a department of radiology or the hospital, but also between widely separated institutions. For this task, simple bitmap formats such as the Tagged Image File Format (TIFF) or the Graphics Interchange Format (GIF) are inadequate, because beside the images, which might have been captured in different dimensions, medical meta information on patients (e.g., Identifier (ID), name, date of birth, ...), the modality (e.g., device, parameters, ...) and organization (e.g., investigation, study, ...) must also be transferred in a standardized way.

Since 1995, the communication is based on the Digital Imaging and Communications in Medicine (DICOM) standard. In its current version, DICOM includes:

- structural information about the contents of the data ("object classes"),
- commands on what should happen to the data ("service classes"), and
- protocols for data transmission.

DICOM is based on the client-server paradigm and allows the coupling of PACS in Radiology Information System (RIS) or Hospital Information Systems (HIS). DICOM incorporates existing standards for communication: the

International Organization for Standardization (ISO) Open System Inter-connection (OSI) model, the Transmission Control Protocol (TCP) Internet Protocol (IP), and the Health Level 7 (HL7) standard. Full DICOM com-pliance for imaging devices and image processing applications is achieved with only a few supported object or service classes, since other DICOM objects, which are not relevant for the current device, simply are handed over to the next system in the DICOM network. The synchronization between the client and server is regularized by conformance claims, which are also specified as part of the DICOM standard. However, the details of imple-mentation of individual services are not specified in the standard, and so in practice, vendor-specific DICOM dialects have been developed, which can lead to incompatibilities when building PACS. In recent years, the Inte-grating the Healthcare Enterprises (IHE) initiative became important. IHE aims at guiding the use of DICOM and other standards such that complete inter-operability is achieved.

1.9.3 Retrieval

In today's DICOM archives, images can be retrieved systematically, only if the patient name with date of birth or the internal system ID is known. Still, the retrieval is based on alphanumerical attributes, which are stored along the image data. It is obvious that diagnostic performance of PACS is magnified significantly if images would be directly available from similar content of a given example image. To provide the Query by Example (QBE) paradigm is a major task of future systems for Contend-Based Image Retrieval (CBIR). Again, this field of biomedical research requires conceptually different strategies as it is demanded in commercial CBIR systems for other application areas, because of the diverse and complex structure of diagnostic information that is captured in biomedical images.

Figure 1.36 shows the system architecture of the Image Retrieval in Med-ical Applications (IRMA) framework[3]. This architecture reflects the chain of processing that we have discussed in this chapter, i.e., registration, feature extraction, segmentation, classification of image objects towards the tip of the pyramid (see Fig. 1.2), which is the symbolic interpretation respective scene analysis. In IRMA, the image information that is relevant for retrieval is gradually condensed and abstracted. The image bitmap is symbolically represented by a semantic network (hierarchical tree structure). The nodes contain characteristic information to the represented areas (segments) of the image. Its topology describes the spatial and/or temporal condition of each object. With this technology, radiologists and doctors are supported similarly in patient care, research, and teaching.
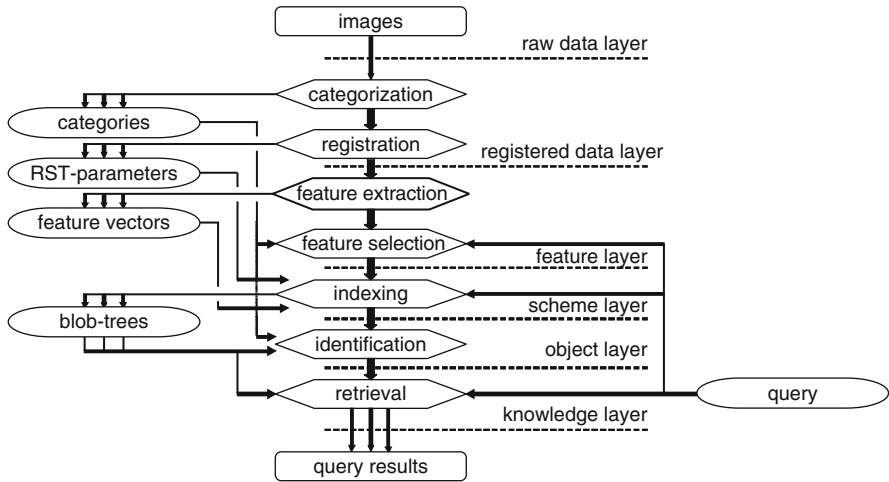
---

[3] http://irma-project.org

Fig. 1.36. *System architecture of the IRMA framework [16].* The processing steps in IRMA are shown in the middle column. Categorization is based on global features and classifies images in terms of imaging modality, view direction, anatomic region, and body system. According to its category, the image geometry and contrast are registered to a reference. The abstraction relies on local features, which are selected specifically to context and query. The retrieval itself is performed efficiently on abstracted and thus information-reduced levels. This architecture follows the paradigm of image analysis (cf. Fig. 1.1). The in-between-representations as presented on the left describe the image increasingly abstract. The levels of abstraction (cf. Fig. 1.20) are named on the right side

1.10 Conclusion and Outlook

The past, present, and future paradigms of medical image processing are composed in Fig. 1.37. Initially (until approx. 1985), the pragmatic issues of image generation, processing, presentation, and archiving stood in the focus of research in biomedical image processing, because available computers at that time had by far not the necessary capacity to hold and modify large image data in memory. The former computation speed of image processing allowed only offline calculations. Until today, the automatic interpretation of biomedical images still is a major goal. Segmentation, classification, and measurements of biomedical images is continuously improved and validated more accurately, since validation is based on larger studies with high volumes of data. Hence, we focused this chapter on image analysis and the processing steps associated with it.

The future development is seen in the increasing integration of algorithms and applications in the medical routine. Procedures in support of diagnosis, treatment planning, and therapy must be easily usable for physicians and,
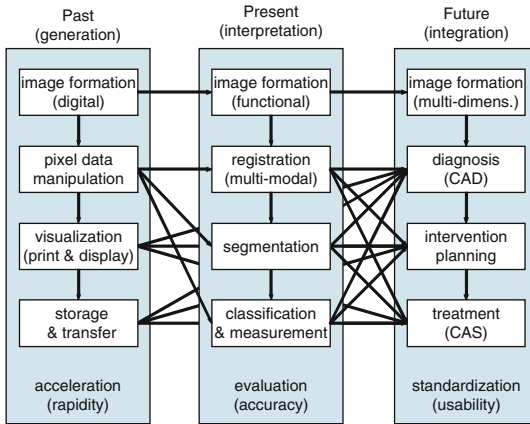
Fig. 1.37. *Changing paradigms in medical image processing [17].* Until now, formation, enhancement, visualization, and management of biomedical images have been in the focus of research. In future, integration, standardization, and validation are seen as major challenges for routine applications in diagnostics, intervention planning, and therapy

therefore, further standardized in order to ensure the necessary interoperability for a clinical use.

References

Related Textbooks

- Bankman IN (ed). Handbook of Medical Image Processing and Analysis. 2nd ed. New York: Academic Press; 2008. ISBN: 0-123-73904-7
- Beolchi L, Kuhn MH (ed). Medical Imaging: Analysis of Multimodality 2D/3D Images. IOS Press; 1995. ISBN: 9-051-99210-6
- Beutel J, Kundel HL, van Metter RL (eds). Handbook of Medical Imaging. Vol. 1: Physics and Psychophysics. Bellingham: SPIE Press; 2000. ISBN 0-819-43621-6
- Bryan RN (ed). Introduction to the Science of Medical Imaging. Cambridge: Cambridge University Press; 2010. ISBN: 978-0-521-74762-2
- Dougherty D. Digital Image Processing for Medical Applications. Cambridge: Cambridge University Press; 2009. ISBN: 0-521-86085-7
- Dougherty ER (ed). Digital Image Processing Methods. New York: CRC Press; 1994. ISBN: 978-0-824-78927-5
- Guy C, Ffytche D. Introduction to the Principles of Medical Imaging. London: Imperial College Press; 2005. ISBN: 1-860-94502-3
- Jan J. Medical Image Processing, Reconstruction and Restoration: Concepts and Methods. Boca Raton: CRC Press; 2005. ISBN: 0-824-75849-8
- Kim Y, Horii SC (eds). Handbook of Medical Imaging. Vol. 3: Display and PACS. Bellingham: SPIE Press; 2000. ISBN: 0-819-43623-2
- Meyer-Baese A. Pattern Recognition in Medical Imaging. San Diego: Academic Press; 2003. ISBN: 0-124-93290-8
- Preim B, Bartz D. Visualization in Medicine. Theory, Algorithms, and Applications. Amsterdam: Morgan Kaufmann; 2007. ISBN: 978-0-123-70596-9

- Rangayyan RM. Biomedical Image Analysis. New York: CRC Press; 2005. ISBN: 0-849-39695-6
- Sonka L, Fitzpatrik JM (eds). Handbook of Medical Imaging. Vol. 2: Medical Image Processing and Analysis. Bellingham: SPIE Press; 2000. ISBN: 0-819-43622-4
- Suetens P. Fundamentals of Medical Imaging. Cambridge: Cambridge University Press; 2002. ISBN: 0-521-80362-4
- Tavares JM, Jorge RMN (eds). Advances in Computational Vision and Medical Image Processing: Methods and Applications. Berlin: Springer; 2008. ISBN: 1-402-09085-4
- Umbaugh SE. Computer Imaging: Digital Image Analysis and Processing. Boca Raton: CRC Press; 2005. ISBN: 0-849-32919-1

Citations

1. Lehmann TM, Oberschelp W, Pelikan E, et al. Bildverarbeitung für die Medizin: grundlagen, Methoden, Anwendungen. Springer, Heidelberg; 1997.
2. Morneburg H, editor. Bildgebende Systeme für die medizinische Diagnostik. 3rd ed. German: Siemens AG, Berlin & Publicis MCD Verlag, Erlangen; 1995.
3. Shannon CE. Communication in the presence of noise. Proc Inst Radio Eng. 1949;37(1):10–21; reprinted in Proc IEEE 1998;86(2):447–57.
4. Lehmann TM, Kaser A, Repges R. A simple parametric equation for pseudo-coloring grey scale images keeping their original brightness progression. Image Vis Comput. 1997;15(3):251–7.
5. Lehmann TM, Hiltner J, Handels H. Medizinische Bildverarbeitung. Chapter 10 In: Lehmann TM, editor. Handbuch der Medizinischen Informatik. Hanser, Munich; 2005.
6. Wagenknecht G, Kaiser HJ, Büll U. Multimodale Integration, Korrelation und Fusion von Morphologie und Funktion: Methodik und erste klinische Anwendungen. Rofo. 1999;170(1):417–6.
7. Pommert A, Höhne KH, Pflesser B, et al. Ein realistisches dreidimensionales Modell der inneren Organe auf der Basis des Visible Human. Munich: Hanser; 2005. p. 72–76.
8. Metzler V, Bienert H, Lehmann TM, et al. A novel method for geometrical shape analysis applied to biocompatibility evaluation. ASAIO J. 1999;45(4):264–1.
9. König S, Hesser J. Live-wires using path-graphs. Methods Inf Med. 2004;43(4):371–5.
10. Kass M, Witkin A, Terzopoulos D. Snakes: active contour models. Int J Comput Vis. 1988;1(5):321–31.
11. Metzler V, Bredno J, Lehmann TM, et al. A deformable membrane for the segmentation of cytological samples. Proc SPIE. 1998;3338:1246–7.
12. Bredno J. Höherdimensionale Modelle zur Quantifizierung biologischer Strukturen. in German: PhD Thesis, RWTH Aachen University, Aachen, Germany; 2001.
13. Cootes TF, Taylor CJ, Cooper DH, et al. Active shape models: their training and application. Comput Vis Image Underst. 1995;61(1):38–59.
14. Lehmann TM, Schmitt W, Horn H, et al. IDEFIX: Identification of dental fixtures in intraoral X-rays. Proc SPIE. 1996;2710:584–5.

15. Lehmann TM, Bredno J, Metzler V, et al. Computer-assisted quantification of axosomatic boutons at the cell membrane of motoneurons. IEEE Trans Biomed Eng. 2001;48(6):706–7.
16. Lehmann TM, Güld MO, Thies C, et al. Content-based image retrieval in medical applications. Methods Inf Med. 2004;43(4):354–61.
17. Lehmann TM, Meinzer HP, Tolxdorff T. Advances in biomedical image analysis: past, present and future challenges. Methods Inf Med. 2004;43(4):308–14.