

Daniel L. Rubin, Hayit Greenspan,  
and James F. Brinkley

After reading this chapter, you should know the answers to these questions:

- What makes images a challenging type of data to be processed by computers when compared to non-image clinical data?
- Why are there many different imaging modalities, and by what major two characteristics do they differ?
- How are visual and knowledge content in images represented computationally? How are these techniques similar to representation of non-image biomedical data?
- What sort of applications can be developed to make use of the semantic image content made accessible using the Annotation and Image Markup model?
- What are four different types of image processing methods? Why are such methods assembled into a pipeline when creating imaging applications?
- What is an imaging modality with high spatial resolution? What is a modality that provides functional information? Why are most imaging modalities not capable of providing both?
- What is the goal in performing segmentation in image analysis? Why is there more than one segmentation method?
- What are two types of quantitative information in images? What are two types of semantic information in images? How might this information be used in medical applications?
- What is the difference between image registration and image fusion? What are examples of each?

---

D.L. Rubin, MD, MS (✉)  
Departments of Radiology and Medicine, Stanford University, 1201 Welch Road, P285, Stanford 94305, CA, USA  
e-mail: dlrubin@yahoo.com

H. Greenspan, PhD  
Department of Biomedical Engineering, Faculty of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel  
e-mail: hayit@eng.tau.ac.il

J.F. Brinkley, MD, PhD  
Departments of Biological Structure, Biomedical Education and Medical Education, Computer Science and Engineering, University of Washington, 357420, Seattle 98195, WA, USA  
e-mail: brinkley@u.washington.edu

---

## 9.1 Introduction

Imaging plays a central role in the health care process. The field is crucial not only to health care, but also to medical communication and education, as well as in research. In fact much of our recent progress, particularly in diagnosis, can be traced to the availability of increasingly sophisticated imaging techniques that not only show the structure of the body in incredible detail, but also show the function of the tissues within the body.

---

This chapter is adapted from an earlier version in the third edition authored by James F. Brinkley and Robert A. Greenes.

Although there are many types (or modalities) of imaging equipment, the images the modalities produce are nearly always acquired in or converted to digital form. The evolution of imaging from analog, film-based acquisition to digital format has been driven by the necessities of cost reduction, efficient throughput, and workflow in managing and viewing an increasing proliferation in the number of images produced per imaging procedure (currently hundreds or even thousands of images). At the same time, having images in digital format makes them amenable to image processing methodologies for enhancement, analysis, display, storage, and even enhanced interpretation.

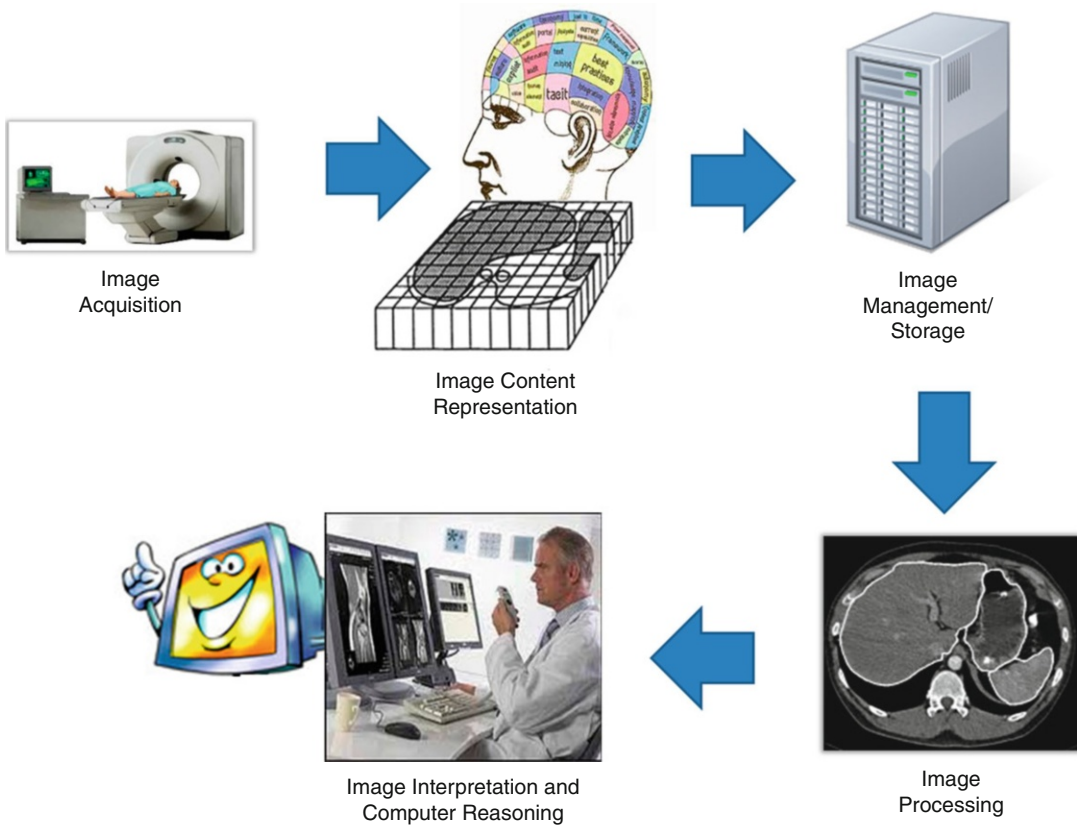
Because of the ubiquity of images in biomedicine, the increasing availability of images in digital form, the rise of high-powered computer hardware and networks, and the commonality of image processing solutions, digital images have become a core data type that must be considered in many biomedical informatics applications. Therefore, this chapter is devoted to a basic understanding of the unique aspects of images as a core data type and the unique aspects of imaging from an informatics perspective. Chapter 20, on the other hand, describes the use of images and image processing in various applications, particularly those in radiology since that field places the greatest demands on imaging methods.

The topics covered by this chapter and Chap. 20 comprise the growing discipline of biomedical imaging informatics (Kulikowski 1997), a subfield of biomedical informatics (see Chap. 1) that has arisen in recognition of the common issues that pertain to all image modalities and applications once the images are converted to digital form. Biomedical imaging informatics is a dynamic field, recently evolving from focusing purely on image processing to broader informatics topics such as representing and processing the semantic contents (Rubin and Napel 2010). At the same time, imaging informatics shares common methodologies and challenges with other domains in biomedical informatics. By trying to understand these common issues, we can develop general solutions that can be applied to all images, regardless of the source.

The major topics in biomedical imaging informatics include image acquisition, image content representation, management/storage of images, image processing, and image interpretation/computer reasoning (Fig. 9.1). **Image acquisition** is the process of generating images from the modality and converting them to digital form if they are not intrinsically digital. **Image content representation** makes the information in images accessible to machines for processing. **Image management/storage** includes methods for storing, transmitting, displaying, retrieving, and organizing images. **Image processing** comprises methods to enhance, segment, visualize, fuse, or analyze the images. **Image interpretation/computer reasoning** is the process by which the individual viewing the image renders an impression of the medical significance of the results of imaging study, potentially aided by computer methods. Chapter 20 is primarily concerned with information systems for image management and storage, whereas this chapter concentrates on these other core topics in biomedical imaging informatics.

An important concept when thinking about imaging from an informatics perspective is that images are an *unstructured data type*; as such, while machines can readily manage the raw image data in terms of storage/retrieval, they cannot easily access image contents (recognize the type of image, annotations made on the image, or anatomy or abnormalities within the image). In this regard, biomedical imaging informatics shares much in common with natural language processing (NLP; Chap. 8). In fact, as the methods of computationally representing and processing images is presented in this chapter, parallels to NLP should be considered, since there is synergy from an informatics perspective.

As in NLP, a major purpose of the methods of imaging informatics is to extract particular information; in biomedical informatics the goal is often to extract information about the structure of the body and to collect features that will be useful for characterizing abnormalities based on morphological alterations. In fact, imaging provides detailed and diverse information very useful for characterizing disease, providing an “imaging



**Fig. 9.1** The major topics in biomedical imaging informatics follow a workflow of activities and tasks commencing with include image acquisition, followed by

image content representation, management/storage of images, image processing, and image interpretation/computer reasoning

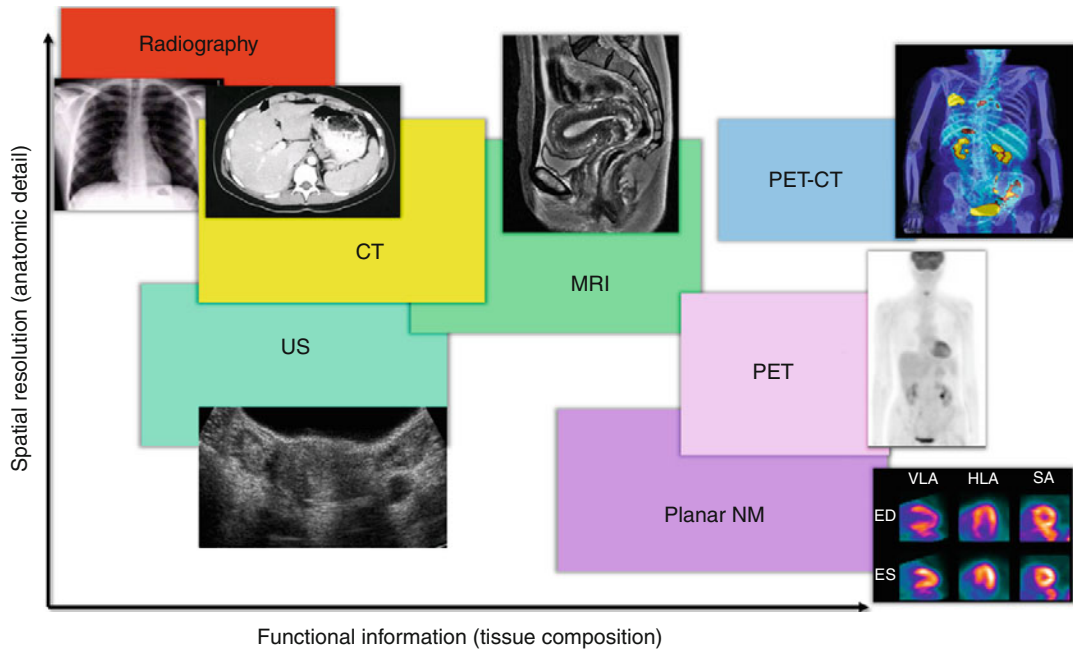
phenotype” useful for characterizing disease, since “a picture is worth a thousand words.”<sup>1</sup> However, to overcome the challenges posed by the unstructured image data type, recent work is applying semantic methods from biomedical informatics to images to make their content explicit for machine processing (Rubin and Napel 2010). Many of the topics in this chapter therefore involve how to represent, extract and characterize the information that is present in images, such as anatomy and abnormalities. Once that task is completed, useful applications that process the image contents can be developed, such as image search and decision support to assist with image interpretation.

While we seek generality in discussing biomedical imaging informatics, many examples in this chapter are taken from a few selected domains such as brain imaging, which is part of the growing field of **neuroinformatics** (Koslow and Huerta 1997). Though our examples are specific, we attempt to describe the topics in generic terms so that the reader can recognize parallels to other imaging domains and applications.

## 9.2 Image Acquisition

In general, there are two different strategies in imaging the body: (1) delineate *anatomic structure* (anatomic/structural imaging), and (2) determine *tissue composition or function* (functional imaging) (Fig. 9.2). In reality, one does not

<sup>1</sup>Frederick Barnard, “One look is worth a thousand words,” *Printers’ Ink*, December, 1921.



**Fig. 9.2** The various radiology imaging methods differ according to two major axes of information of images, spatial resolution (anatomic detail) and functional infor-

mation depicted (which represents the tissue composition—e.g., normal or abnormal). A sample of the more common imaging modalities is shown

choose between anatomic and functional imaging; many modalities provide information about both morphology and function. However, in general, each imaging modality is characterized primarily as being able to render high-resolution images with good contrast resolution (anatomic imaging) or to render images that depict tissue function (functional imaging).

### 9.2.1 Anatomic (Structural) Imaging

Imaging the structure of the body has been and continues to be the major application of medical imaging, although, as described in Sect. 9.2.2, functional imaging is a very active area of research. The goal of anatomic imaging is to accurately depict the structure of the body—the size and shape of organs—and to visualize abnormalities clearly. Since the goal in anatomic imaging is to depict and understand the structure of anatomic entities accurately, high spatial resolution is an important requirement of the imaging method (Fig. 9.2). On the other hand, in anatomic

imaging, recognizing tissue function (e.g., tissue ischemia, neoplasm, inflammation, etc.) is not the goal, though this is crucial to functional imaging and to patient diagnosis. In most cases, imaging will be done using a combination of methods or modalities to derive both structural/anatomic information as well as functional information.

### 9.2.2 Functional Imaging

Many imaging techniques not only show the structure of the body, but also the function, where for imaging purposes function can be inferred by observing changes of structure over time. In recent years this ability to image function has greatly accelerated. For example, ultrasound and angiography are widely used to show the functioning of the heart by depicting wall motion, and ultrasound Doppler can image both normal and disturbed blood flow (Mehta et al. 2000). Molecular imaging (Sect. 9.2.3) is increasingly able to depict the expression of particular genes

superimposed on structural images, and thus can also be seen as a form of functional imaging.

A particularly important application of functional imaging is for understanding the cognitive activity in the brain. It is now routinely possible to put a normal subject in a scanner, to give the person a cognitive task, such as counting or object recognition, and to observe which parts of the brain light up. This unprecedented ability to observe the functioning of the living brain opens up entirely new avenues for exploring how the brain works.

Functional brain imaging modalities can be classified as *image-based* or *non-image based*. In both cases it is taken as axiomatic that the functional data must be mapped to the individual subject's anatomy, where the anatomy is extracted from structural images using techniques described in the previous sections. Once mapped to anatomy, the functional data can be integrated with other functional data from the same subject, and with functional data from other subjects whose anatomy has been related to a template or probabilistic atlas. Techniques for generating, mapping and integrating functional data are part of the field of Functional Brain Mapping, which has become very active in the past few years, with several conferences (Organization for Human Brain Mapping 2001) and journals (Fox 2001; Toga et al. 2001) devoted to the subject.

### 9.2.2.1 Image-Based Functional Brain Imaging

Image-based functional data generally come from scanners that generate relatively low-resolution volume arrays depicting spatially-localized activation. For example, **positron emission tomography** (PET) (Heiss and Phelps 1983; Aine 1995; Alberini et al. 2011) and **magnetic resonance spectroscopy** (MRS) (Ross and Bluml 2001) reveal the uptake of various metabolic products by the functioning brain; and **functional magnetic resonance imaging** (fMRI) reveals changes in blood oxygenation that occur following neural activity (Aine 1995). The raw intensity values generated by these techniques must be processed by sophisticated statistical algorithms to sort out how much of the observed

intensity is due to cognitive activity and how much is due to background noise.

As an example, one approach to fMRI imaging is language mapping (Corina et al. 2000). The subject is placed in the **magnetic resonance imaging** (MRI) scanner and told to silently name objects shown at 3-s intervals on a head-mounted display. The actual objects ("on" state) are alternated with nonsense objects ("off" state), and the fMRI signal is measured during both the on and the off states. Essentially the **voxel** values at the off (or control) state are subtracted from those at the on state. The difference values are tested for significant difference from non-activated areas, then expressed as t-values. The voxel array of t-values can be displayed as an image.

A large number of alternative methods have been and are being developed for acquiring and analyzing functional data (Frackowiak et al. 1997). The output of most of these techniques is a low-resolution 3-D image volume in which each voxel value is a measure of the amount of activation for a given task. The low-resolution volume is then mapped to anatomy guided by a high-resolution structural MR dataset, using one of the linear registration techniques described in Sect. 9.4.7.

Many of these and other techniques are implemented in the SPM program (Friston et al. 1995), the AFNI program (Cox 1996), the Lyngby toolkit (Hansen et al. 1999), and several commercial programs such as Medex (Sensor Systems Inc. 2001) and Brain Voyager (Brain Innovation B.V. 2001). The FisWidgets project at the University of Pittsburgh is developing an approach that allows customized creation of graphical user interfaces in an integrated desktop environment (Cohen 2001). A similar effort (VoxBox) is underway at the University of Pennsylvania (Kimborg and Aguirre 2002).

The ultimate goal of functional neuroimaging is to observe the actual electrical activity of the neurons as they perform various cognitive tasks. fMRI, MRS and PET do not directly record electrical activity. Rather, they record the results of electrical activity, such as (in the case of fMRI) the oxygenation of blood supplying the active neurons. Thus, there is a delay from the time of

activity to the measured response. In other words these techniques have relatively poor temporal resolution (Sect. 9.2.4). **Electroencephalography** (EEG) or **magnetoencephalography** (MEG), on the other hand, are more direct measures of electrical activity since they measure the electromagnetic fields generated by the electrical activity of the neurons. Current EEG and MEG methods involve the use of large arrays of scalp sensors, the output of which are processed in a similar way to CT in order to localize the source of the electrical activity inside the brain. In general this “source localization problem” is under-constrained, so information about brain anatomy obtained from MRI is used to provide further constraints (George et al. 1995).

### 9.2.3 Imaging Modalities

There are many different approaches that have been developed to acquire images of the body. A proliferation in imaging modalities reflects the fact that there is no single perfect imaging modality; no single imaging technique satisfies all the desiderata for depicting the broad variety of types of pathology, some of which are better seen on some modalities than on others. The primary difference among the imaging modalities is the energy source used to generate the images. In radiology, nearly every type of energy in the electromagnetic spectrum has been used, in addition to other physical phenomena such as sound and heat. We describe the more common methods according to the type of energy used to create the image.

#### 9.2.3.1 Light

The earliest medical images used visible light to create photographs, either of gross anatomic structures or, if a microscope was used, of histological specimens. Light is still an important source for creation of images, and in fact optical imaging has seen a resurgence of interest and application for areas such as molecular imaging (Weissleder and Mahmood 2001; Ray 2011) and imaging of brain activity on the exposed surface of the cerebral cortex (Pouratian et al. 2003).

Visible light is the basis for an emerging modality called “optical imaging” and has promising applications such as cancer imaging (Solomon, Liu et al. 2011). Visible light, however, does not allow us to see more than a short distance beneath the surface of the body; thus other modalities are used for imaging structures deep inside the body.

#### 9.2.3.2 X-Rays

X-rays were first discovered in 1895 by Wilhelm Conrad Roentgen, who was awarded the 1901 Nobel Prize in Physics for this achievement. The discovery caused worldwide excitement, especially in the field of medicine; by 1900, there already were several medical radiological societies. Thus, the foundation was laid for a new branch of medicine devoted to imaging the structure and function of the body (Kevles 1997).

Radiography is the primary modality used in radiology departments today, both to record a static image (Fig. 9.3) as well as to produce a real-time view of the patient (fluoroscopy) or a movie (cine). Both film and fluoroscopic screens were used initially for recording X-ray images, but the fluoroscopic images were too faint to be used clinically. By the 1940s, however, television and image-intensifier technologies were developed to produce clear real-time fluorescent



**Fig. 9.3** A radiograph of the chest (Chest X-ray) taken in the frontal projection. The image is shown as if the patient is facing the viewer. This patient has abnormal density in the left lower lobe

images. Today, a standard procedure for many types of examinations is to combine real-time television monitoring of X-ray images with the creation of selected higher resolution film images. Until the early 1970s, film and fluoroscopy were the only X-ray modalities available. Recently, nearly all radiology departments have shifted away from acquiring radiographic images on film (analog images) to using digital radiography (Korner et al. 2007) to acquire digital images.

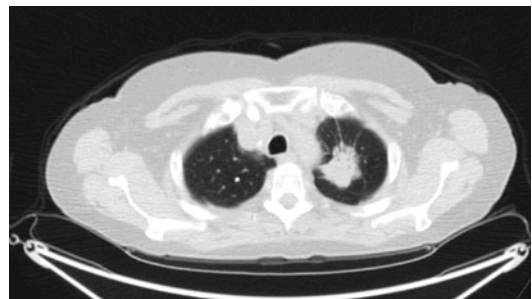
X-ray imaging is a projection technique; an X-ray beam—one form of ionizing radiation—is projected from an X-ray source through a patient’s body (or other object) onto an X-ray array detector (a specially coated cassette that is scanned by a computer to capture the image in digital form), or film (to produce a non-digital image). Because an X-ray beam is differentially absorbed by the various body tissues based on the thickness and atomic number of the tissues, the X-rays produce varying degrees of brightness and darkness on the radiographic image. The differential amounts of brightness and darkness on the image are referred to as “image contrast;” differential contrast among structures on the image is the basis for recognizing anatomic structures. Since the image in radiography is a projection, radiographs show a superposition of all the structures traversed by the X-ray beam.

**Computed radiography (CR)** is an imaging technique that directly creates digital radiographs from the imaging procedure. Storage phosphor replaces film by substituting a reusable phosphor plate in a standard film cassette. The exposed plate is processed by a reader system that scans the image into digital form, erases the plate, and packages the cassette for reuse. An important advantage of CR systems is that the cassettes are of standard size, so they can be used in any equipment that holds film-based cassettes (Horii 1996). More recently, **Digital Radiography (DR)** uses charge-coupled device (CCD) arrays to capture the image directly.

Radiographic images have very high spatial resolution because a high photon flux is used to produce the images, and a high resolution detector (film or digital image array) that captures many line pairs per unit area is used. On the

other hand, since the contrast in images is due to differences in tissue density and atomic number, the amount of functional information that can be derived from radiographic images is limited (Fig. 9.2). Radiography is also limited by relatively poor contrast resolution (compared with other modalities such as **computed tomography (CT)** or MRI), their use of ionizing radiation, the challenge of spatial localization due to projection ambiguity, and their limited ability to depict physiological function. As described below, newer imaging modalities have been developed to increase contrast resolution, to eliminate the need for X-rays, and to improve spatial localization. A benefit of radiographic images is that they can be generated in real time (fluoroscopy) and can be produced using portable devices.

Computed Tomography (CT) is an important imaging method that uses X-ray imaging to produce cross sectional and volumetric images of the body (Lee 2006). Similar to radiography, X-rays are projected through the body onto an array of detectors; however, the beam and detectors rotate around the patient, making numerous views at different angles of rotation. Using computer reconstruction algorithms, an estimate of absolute density at each point (volume element or **voxel**) in the body is computed. Thus, the CT image is a computed image (Fig. 9.4); CT did not become practical for generating high quality images until the advent of powerful computers



**Fig. 9.4** A CT image of the upper chest. CT images are slices of a body plane; in this case, a cross sectional (axial) image of the chest. Axial images are viewed from below the patient, so that the patient’s left is on viewer’s right. This image shows a cancer mass in the left upper lobe of the lung

and development of computer-based reconstruction techniques, which represent one of the most spectacular applications of computers in all of medicine (Buxton 2009). The spatial resolution of images is not as high in CT as it is in radiography, however, due to the computed nature of the images, the contrast resolution and ability to derive functional information of tissues in the body is superior for CT than for radiography (Fig. 9.2).

### 9.2.3.3 Ultrasound

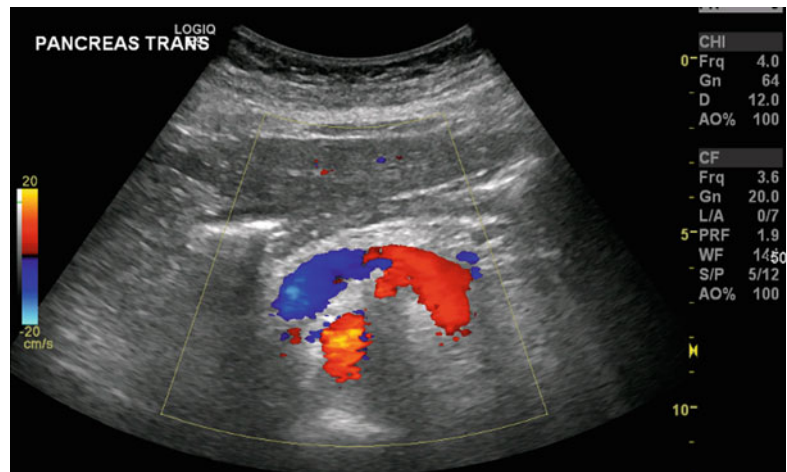
A common energy source used to produce images is **ultrasound**, which developed from research performed by the Navy during World War II in which sonar was used to locate objects of interest in the ocean. Ultrasonography uses pulses of high-frequency sound waves rather than ionizing radiation to image body structures (Kremkau 2006). The basis of image generation is due to a property of all objects called acoustical impedance. As sound waves encounter different types of tissues in a patient's body (particularly interfaces where there is a change in acoustical impedance), a portion of the wave is reflected and a portion of the sound beam (which is now attenuated) continues to traverse into deeper tissues. The time required for the echo to return is proportional to the distance into the body at which it is reflected; the amplitude (intensity) of a returning echo depends on the acoustical properties of the tissues encoun-

tered and is represented in the image as brightness (more echoes returning to the source is shown as image brightness). The system constructs two-dimensional images (B-scans) by displaying the echoes from pulses of multiple adjacent one-dimensional paths (A-scans). Ultrasound images are acquired as digital images from the outset, and saved on computer disks. They may also be recorded as frames in rapid succession (cine loops) for real-time imaging. In addition, Doppler methods in ultrasound are used to measure and characterize the blood flow in blood vessels in the body (Fig. 9.5).

Since the image contrast in ultrasound is based on differences in the acoustic impedance of tissue, ultrasound provides functional information (e.g., tissue composition and blood flow). On the other hand, the flux of sound waves is not as dense as the photon flux used to produce images in radiography; thus ultrasound images are generally lower resolution images than other imaging modalities (Fig. 9.2).

Current ultrasound machines are essentially specialized computers with attached peripherals, with active development of three-dimensional imaging. The ultrasound transducer now often sweeps out a 3-D volume rather than a 2-D plane, and the data are written directly into a three-dimensional array memory, which is displayed using volume or surface-based rendering techniques (Ritchie et al. 1996).

**Fig. 9.5** An ultrasound image of abdomen. Like CT and MRI, ultrasound images are slices of a body, but because a user creates the images by holding a probe, any arbitrary plane can be imaged (so long as the probe can be oriented to produce that plane). This image shows an axial slice through the pancreas, and flow in nearby blood vessels (*in color*) is seen due to Doppler effects incorporated into the imaging method



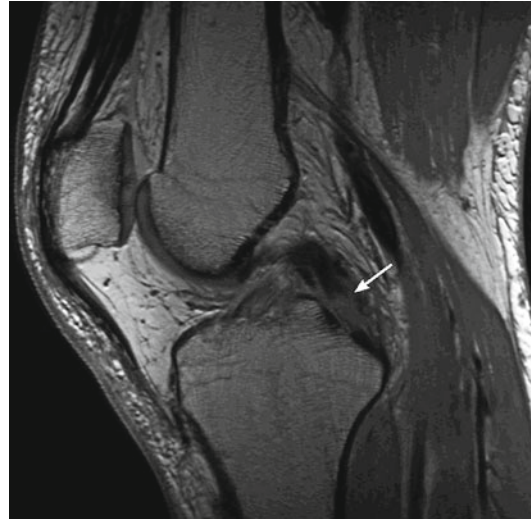


### 9.2.3.4 Magnetic Resonance Imaging (MRI)

Creation of images from the resonance phenomena of unpaired spinning charges in a magnetic field grew out of **nuclear magnetic resonance (NMR) spectroscopy**, a technique that has long been used in chemistry to characterize chemical compounds. Many atomic nuclei within the body have a net magnetic moment, so they act like tiny magnets. When a small chemical sample is placed in an intense, uniform magnetic field, these nuclei line up in the direction of the field, spinning around the axis of the field with a frequency dependent on the type of nucleus, on the surrounding environment, and on the strength of the magnetic field.

If a radio pulse of a particular frequency is then applied at right angles to the stationary magnetic field, those nuclei with rotation frequency equal to that of the radiofrequency pulse resonate with the pulse and absorb energy. The higher energy state causes the nuclei to change their orientation with respect to the fixed magnetic field. When the radiofrequency pulse is removed, the nuclei return to their original aligned state (a process called “relaxation”), emitting a detectable radiofrequency signal as they do so. Characteristic parameters of this signal—such as intensity, duration, and frequency shift away from the original pulse—are dependent on the density and environment of the nuclei. In the case of traditional NMR spectroscopy, different molecular environments cause different frequency shifts (called chemical shifts), which we can use to identify the particular compounds in a sample. In the original NMR method, however, the signal is not localized to a specific region of the sample, so it is not possible to create an image.

Creation of medical images from NMR signals, known as Magnetic Resonance Imaging (MRI), had to await the development of computer-based reconstruction techniques, similar to CT. The basis of image formation in MRI is based on proton relaxation (referred to as T1 and T2 relaxation); differences in T1 and T2 are inherent properties of tissue and they vary among tissues. Thus, MRI provides detailed functional information about tissue and can be valuable in clinical



**Fig. 9.6** An MRI image of the knee. Like CT, MRI images are slices of a body. This image is in the sagittal plane through the mid knee, showing in a tear in the posterior cruciate ligament (*arrow*)

diagnosis (Fig. 9.6). At the same time, the flux of radiofrequency waves used to produce the images is high, and MRI thus has high spatial resolution (Fig. 9.2).

Many new modalities are being developed based on magnetic resonance. For example, magnetic resonance arteriography (MRA) and venography (MRV) are used to image blood flow (Lee 2003) and diffusion tensor imaging (DTI) is increasingly being used to image white matter fiber tracts in the brain (Le Bihan et al. 2001; Hasan et al. 2010; de Figueiredo et al. 2011; Gerstner and Sorensen 2011).

### 9.2.3.5 Nuclear Medicine Imaging

In **nuclear medicine imaging**, the imaging approach is a reverse of the radiographic imaging: instead of the imaging beam being outside the subject and projecting into the subject, the imaging source is inside the subject and projects out. Specifically, a radioactive isotope is chemically attached to a biologically active compound (such as an analogue of glucose) and then is injected into the patient’s peripheral circulation. The compound collects in the specific body compartments or organs (such as metabolically-active tissues), where it is stored or processed by the

body. The isotope emits radiation locally, and the radiation is measured using a special detector. The resultant nuclear-medicine image depicts the level of radioactivity that was measured at each spatial location of the patient. Because the counts are inherently quantized, digital images are produced. Multiple images also can be processed to obtain temporal dynamic information, such as the rate of arrival or of disappearance of isotope at particular body sites.

Nuclear medicine images, like radiographic images, are usually acquired as projections—a large planar detector is positioned outside the patient and it collects a projected image of all the radioactivity emitted from the patient. The images are similar in appearance to radiographic projection images. However, since the photon flux is extremely low (to minimize the radiation dose to the patient), the spatial resolution of nuclear medicine images is low. On the other hand, since the only places where radioisotope accumulates will be places in the body that are targeted by the injected agent, nearly all the information in nuclear medicine images is functional information; thus nuclear imaging methods have high functional information and low spatial resolution (Fig. 9.2). Nuclear medicine techniques have recently attracted much attention because of an explosion in novel imaging probes and targeting mechanisms to localize the imaging agent.

In addition to projection images, a computed tomography-like method called **single-photon emission computed tomography** (SPECT) (Alberini et al. 2011) has been developed. A camera rotates around the patient similar to CT, producing a computed volumetric image that may be viewed and navigated in multiple planes. A technique called Positron Emission Tomography (PET) uses a special type of radioactive isotope that emits positrons, which, upon encountering an electron, produces an annihilation event that sends out two gamma rays in opposite directions that are simultaneously detected on an annular detector array and used to compute a cross sectional slice through the patient, similar to CT and SPECT (Fig. 9.7). These volumetric nuclear medicine imaging methods, like the projection

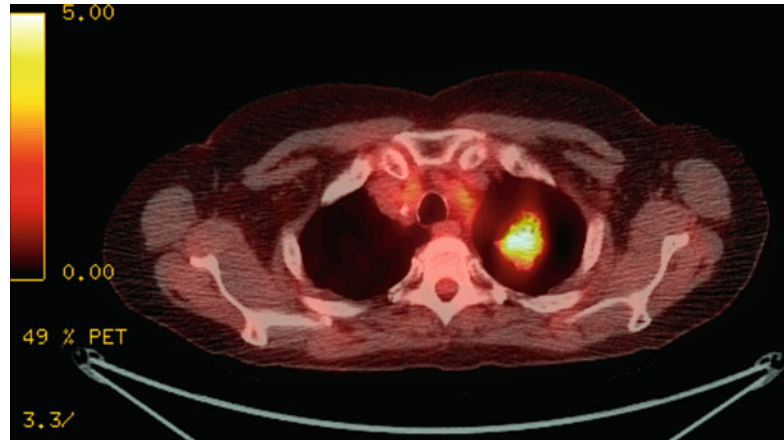


**Fig. 9.7** A PET image of the body in a patient with cancer in the left lung (same patient as in Fig. 9.4). This is a projection image taken in the frontal plane after injection of a radioactive isotope that accumulates in cancers. A *small black spot* in the left upper lobe is abnormal and indicates the cancer mass in the upper lobe of the left lung

methods, have high functional information and low spatial resolution. However, recently a newer modality called PET/CT has been developed that integrates a PET scanner and CT with image fusion (discussed below) to get the best of both worlds—functional information about lesions in the PET image plus spatial localization of the abnormality on the CT image (Figs. 9.8 and 9.2).

A subdomain of nuclear imaging called **molecular imaging** has emerged that embodies this work on molecularly-targeted imaging (and therapeutic) agents (Weissleder and Mahmood 2001; Massoud and Gambhir 2003; Biswal et al. 2007; Hoffman and Gambhir 2007; Margolis et al. 2007; Ray and Gambhir 2007; Willmann et al. 2008; Pysz et al. 2010). Molecularly-tagged molecules are increasingly being introduced into the living organism, and imaged with optical, radioactive, or magnetic energy sources, often using reconstruction techniques and often in 3-D. It is becoming possible to combine gene sequence

**Fig. 9.8** A PET/CT fused image. The axial slice from the PET study (Fig. 9.7) and the corresponding axial slice from the CT study (Fig. 9.4) are combined into a single image that has both good spatial resolution and functional information, showing that the lung mass has abnormal uptake of isotope, indicating it is metabolically active



information, gene expression array data, and molecular imaging to determine not only which genes are expressed, but where they are expressed in the organism (Kang and Chung 2008; Min and Gambhir 2008; Singh et al. 2008; Lexe et al. 2009; Smith et al. 2009; Harney and Meade 2010). These capabilities will become increasingly important in the post-genomic era for determining exactly how genes generate both the structure and function of the organism.

## 9.2.4 Image Quality

### 9.2.4.1 Characteristics of Image Quality

The imaging modalities described above are complex devices with many parameters that need to be specified in generating the image, and most of the parameters can have substantial impact on the following key characteristics of the final image appearance: spatial resolution, contrast resolution, and temporal resolution, all of which have substantial impact on image quality and diagnostic value of the image. These characteristics provide an objective means for comparing images formed by digital imaging modalities.

- **Spatial resolution** is related to the sharpness of the image; it is a measure of how well the imaging modality can distinguish points on the object that are close together. For a digital image, spatial resolution is generally related to the number of pixels per image area.

- **Contrast resolution** is a measure of the ability to distinguish small differences in intensity in different regions of the image, which in turn are related to differences in measurable parameters, such as X-ray attenuation. For digital images, the number of bits per pixel is related to the contrast resolution of an image.
- **Temporal resolution** is a measure of the time needed to create an image. We consider an imaging procedure to be a real-time application if it can generate images concurrent with the physical process it is imaging. At a rate of at least 30 images per second, it is possible to produce unblurred images of the beating heart.

Other parameters that are specifically relevant to medical imaging are the degree of invasiveness, the dosage of ionizing radiation, the degree of patient discomfort, the size (portability) of the instrument, the ability to depict physiologic function as well as anatomic structure, and the availability and cost of the procedure at a specific location.

A perfect imaging modality would produce images with high spatial, contrast, and temporal resolution; it would be available, low in cost, portable, free of risk, painless, and noninvasive; it would use nonionizing radiation; and it would depict physiological function as well as anatomic structure. As seen above, the different modalities differ in these characteristics and none is uniformly strong across all the parameters (Fig. 9.2).

### 9.2.4.2 Contrast Agents

One of the major motivators for development of new imaging modalities is the desire to increase contrast resolution. A contrast agent is a substance introduced into the body to enhance the imaging contrast of structures or fluids in medical imaging. Contrast agents can be introduced in various ways, such as by injection, inspiration, ingestion, or enema. The chemical composition of contrast agents vary with modality so as to be optimally visible based on the physical basis of image formation. For example, iodinated contrast agents are used in radiography and CT because iodine has high atomic number, greatly attenuating X-rays, and thus greatly enhancing image contrast in any tissues that accumulate the contrast agent. Contrast agents for radiography are referred to as “radiopaque” since they absorb X-rays and obscure the beam. Contrast agents in radiography are used to highlight the anatomic structures of interest (e.g., stomach, colon, urinary tract). In an imaging technique called angiography, a contrast agent is injected into the blood vessels to opacify them on the images. In pathology, histological staining agents such as haematoxylin and eosin (H&E) have been used for years to enhance contrast in tissue sections, and magnetic contrast agents such as gadolinium have been introduced to enhance contrast in MR images.

Although these methods have been very successful, they generally are somewhat non-specific. In recent years, advances in molecular biology have led to the ability to design contrast agents that are highly specific for individual molecules. In addition to radioactively tagged molecules used in nuclear medicine, molecules are tagged for imaging by magnetic resonance and optical energy sources. Tagged molecules are imaged in 2-D or 3-D, often by application of reconstruction techniques developed for clinical imaging. Tagged molecules have been used for several years *in vitro* by such techniques as immunocytochemistry (binding of tagged antibodies to antigen) (Van Noorden 2002) and *in situ* hybridization (binding of tagged nucleotide sequences to DNA or RNA) (King et al. 2000). More recently, methods have been developed to image these molecules in the living organism, thereby opening up entirely new

avenues for understanding the functioning of the body at the molecular level (Biswal et al. 2007; Hoffman and Gambhir 2007; Margolis et al. 2007; Ray and Gambhir 2007; Willmann et al. 2008; Pysz et al. 2010).

## 9.2.5 Imaging Methods in Other Medical Domains

Though radiology is a core domain and driver of many clinical problems and applications of medical imaging, several other medical domains are increasingly relying on imaging to provide key information for biomedical discovery and clinical insight. The methods of biomedical informatics presented in this chapter, while focusing on radiology in our examples, are generalizable and applicable to these other domains. We briefly highlight these other domains and the role of imaging in them.

### 9.2.5.1 Microscopic/Cellular Imaging

At the microscopic level, there is a rapid growth in **cellular imaging** (Larabell and Nugent 2010; Toomre and Bewersdorf 2010; Wessels et al. 2010), including use of computational methods to evaluate the features in cells (Carpenter et al. 2006). The confocal microscope uses electronic focusing to move a two-dimensional slice plane through a three-dimensional tissue slice placed in a microscope. The result is a three-dimensional voxel array of a microscopic, or even submicroscopic, specimen (Wilson 1990; Paddock 1994). At the electron microscopic level electron tomography generates 3-D images from thick electron-microscopic sections using techniques similar to those used in CT (Perkins et al. 1997).

### 9.2.5.2 Pathology/Tissue Imaging

The radiology department was revolutionized by the introduction of digital imaging and **Picture Archiving and Communication Systems (PACS)**. Pathology has likewise begun to shift from an analog to a digital workflow (Leong and Leong 2003; Gombas et al. 2004). Pathology informatics is a rapidly emerging field (Becich 2000; Gabriel and Yousef 2010), with goals and research problems similar to those in radiology, such as managing

huge images, improving efficiency of workflow, learning new knowledge by mining historical cases, identifying novel imaging features through correlative quantitative imaging analysis, and decision support. A particularly promising area is deriving novel quantitative image features from pathology images to improve characterization and clinical decision making (Giger and MacMahon 1996; Nielsen et al. 2008; Armstrong 2010). Given that pathology and radiology produce images that characterize phenotype of disease, there is tremendous opportunity for information integration and linkage among pathology, radiology, and molecular data for discovery.

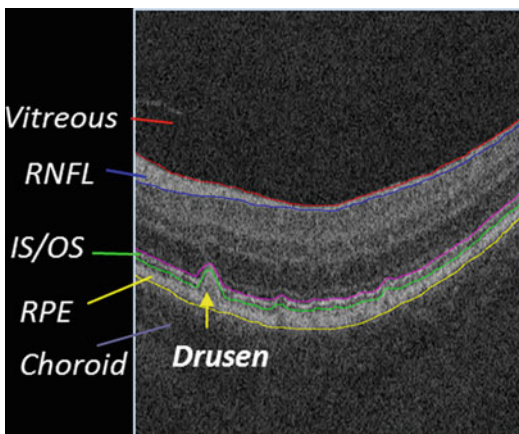
### 9.2.5.3 Ophthalmologic Imaging

Visualization of the retina is a core task of ophthalmology to diagnose disease and to monitor treatment response (Bennett and Barry 2009). Imaging modalities include retinal photography, autofluorescence, and fluorescein angiography. Recently, tomographic-based imaging has been introduced through a technique called **optical coherence tomography** (OCT; Fig. 9.9) (Figurska et al. 2010). This modality is showing great progress in evaluating a variety of retinal diseases (Freton and Finger 2012; Schimmel et al. 2011; Sohrab et al. 2011). As with radiological

imaging, a number of quantitative and automated segmentation methods are being created to evaluate disease objectively (Cabrera Fernandez et al. 2005; Baumann et al. 2010; Hu et al. 2010a, b). Likewise, image processing methods for image visualization and fusion are being developed, similar to those used in radiology.

### 9.2.5.4 Dermatologic Imaging

Imaging is becoming an important component of dermatology in the management of patients with skin lesions. Dermatologists frequently take photographs of patients with skin abnormalities, and while initially this was done for clinical documentation, increasingly this is done to leverage imaging informatics methods for training, to improve clinical care, for consultation, for monitoring progression or change in skin disease, and for image retrieval (Bittorf et al. 1997; Diepgen and Eysenbach 1998; Eysenbach et al. 1998; Lowe et al. 1998; Ribaric et al. 2001). Like radiology and pathology, recent work is being done to extract quantitative features from the images to enable decision support (Seidenari et al. 2003).



**Fig. 9.9** An OCT image of the retina. Like ultrasound, OCT produces an image slice at any arbitrary angle (depending on how the light beam can be oriented), but it is limited to visualizing superficial structures due to poor penetration by light. In this image, the layered structure of the retina can be seen, as well as abnormalities (drusen)

## 9.3 Image Content Representation

The image contents comprise two components of information, the *visual content* and the *knowledge content*. The visual content is the raw values of the image itself, the information that a computer can access in a digital image directly. The knowledge content arises as the observer, who has biomedical knowledge about the image content, views the visual information in the image. For example, a radiologist viewing a CT image of the upper abdomen immediately recognizes that the image contains the liver, spleen, and stomach (anatomic entities), as well as image abnormalities such as a mass in the liver with rim enhancement (imaging observation entities). Unlike the visual content, the knowledge content of images is not directly accessible to computers from the image itself. However, semantic methods are being developed to make this content machine-accessible (Sect. 9.3.2). In this section we describe

imaging informatics methods for representing the visual and knowledge content of images.

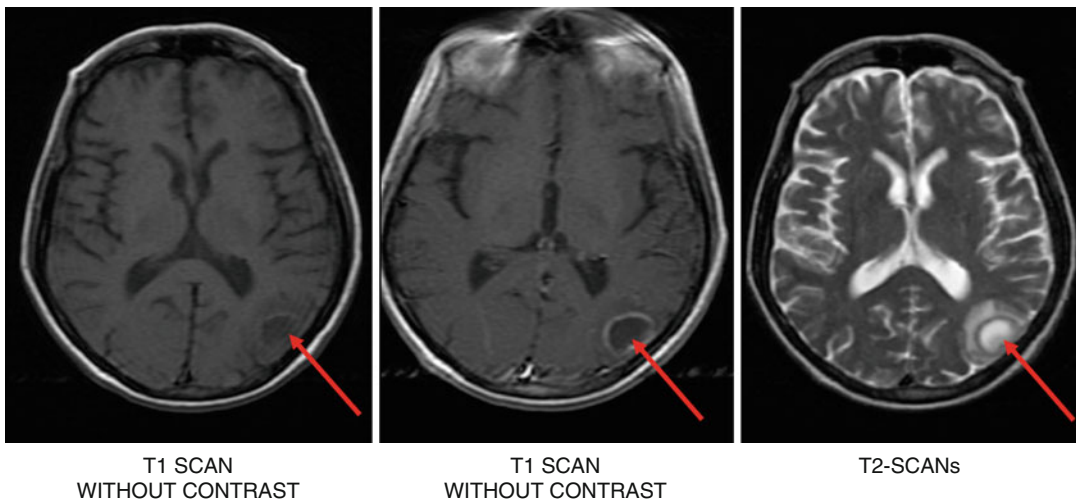
### 9.3.1 Representing Visual Content in Digital Images

The visual content of digital images typically is represented in a computer by a two-dimensional array of numbers (a bit map). Each element of the array represents the intensity of a small square area of the picture, called a picture element (or pixel). Each **pixel** element corresponds to a volume element (or **voxel**) in the imaged subject that produced the pixel. If we consider the image of a volume, then a three-dimensional array of numbers is required. Another way of thinking of a volume is that it is a stack of two-dimensional images. However, it is also important to be aware of the voxel dimensions that correspond to the pixels when doing this. In many 2-D imaging applications, the in-plane resolution (the size of the voxels in the  $x, y$  plane) is higher than the resolution in the  $z$ -axis (i.e., the slice thickness). This creates a problem when re-sampling the volume data to create other projections, such as coronal or sagittal from primary axial image data. If the dimensions

of the voxels (and pixels) are uniform in all dimensions, they are referred to as *isotropic*.

We can store any image in a computer as a matrix of integers (or real-valued numbers), either by converting it from an analog to a digital representation or by generating it directly in digital form. Once an image is in digital form, it can be handled just like all other data. It can be transmitted over communications networks, stored compactly in databases on magnetic or optical media, and displayed on graphics monitors. In addition, the use of computers has created an entirely new realm of capabilities for image generation and analysis: images can be computed rather than measured directly. Furthermore, digital images can be manipulated for display or analysis in ways not possible with film-based images.

In addition to the 2D (slice) and 3D (volume) representation for image data, there can be additional dimensions to representing the visual content of images. It is often the case that multi-modality data are required for the diagnosis; this can be a combination of varying modalities, (e.g., CT and PET, CT and MRI) and can be a combination of imaging sequences *within* a modality (e.g., T1, T2, or other sequences in MRI) (Fig. 9.10). Pixel (or voxel) content, from



**Fig. 9.10** Multi-modality imaging. Images of the brain from three modalities (T1 without contrast, T1 with contrast, and T2) are shown. The patient has a lesion in the left occipital lobe that has distinctive image features on

each of these modalities, and the combination of these different features on different modalities establishes characteristic patterns useful in diagnosis

each of the respective acquisition modalities, are combined in what is known as a “feature-vector” in the multi-dimensional space. For example, a 3-dimensional intensity-based feature vector, based on three MRI pulse sequences, can be defined as a set of three values for each pixel in the image, where the intensity of each pixel in each of the three MRI images is extracted and recorded (e.g., [Intensity(Sequence 1), Intensity(Sequence 2), Intensity(Sequence 3)]. Any imaging performed over time (e.g., cardiac echo videos) can be represented by the set of values at each time point, thus the time is added as an additional dimension to the representation.

Finally, in addition to representing the visual content, medical images also need to represent certain information about that visual content (referred to as **image metadata**). Image metadata include such things as the name of the patient, date the image was acquired, the slice thickness, the modality that was used to acquire the image, etc. All image metadata are usually stored in the header of the image file. Given that there are many different types of equipment and software that produce and consume images, standards are crucial. For images, the Digital Imaging and Communications in Medicine (DICOM) standard is for distributing and viewing any kind of medical image regardless of the origin (Bidgood and Horii 1992). DICOM has become pervasive throughout radiology and is becoming a standard in other domains such as pathology, ophthalmology, and dermatology. In addition to specifying a standard file syntax and metadata structure, DICOM specifies a standard protocol for communicating images among imaging devices.

### 9.3.2 Representing Knowledge Content in Digital Images

As noted above, the knowledge content related to images is not directly encoded in the images, but it is recognized by the observer of the images. This knowledge includes recognition of the anatomic entities in the image, imaging observations and characteristics of the observations (sometimes called “findings”), and interpretations

(probable diseases). Representing this knowledge in the imaging domain is similar to knowledge representation in other domains of biomedical informatics (see Chap. 22). Specifically, for representing the entities in the domain of discourse, we adopt terminologies or ontologies. To make specific statements about individuals (images), we use information models that reference ontological entities as necessary.

#### 9.3.2.1 Knowledge Representation of Anatomy

Given segmented anatomical structures, whether at the macroscopic or microscopic level, and whether represented as 3-D surface meshes or extracted 3-D regions, it is often desirable to attach labels (names) to the structures. If the names are drawn from a controlled terminology or ontology, they can be used as an index into a database of segmented structures, thereby providing a qualitative means for comparing structures from multiple subjects.

If the terms in the vocabulary are organized so as to assert relationships true of all individuals (“ontologies”), they can support systems that manipulate and retrieve image contents in “intelligent” ways. If anatomical ontologies are linked to other ontologies of physiology and pathology they can provide increasingly sophisticated knowledge about the *meaning* of the various images and other data that are increasingly becoming available in online databases. This kind of knowledge (by the computer, as opposed to the scientist) will be required in order to achieve the seamless integration of all forms of imaging and non-imaging data.

At the most fundamental level, *Nomina Anatomica* (International Anatomical Nomenclature Committee 1989) and its successor, *Terminologia Anatomica* (Federative Committee on Anatomical Terminology 1998) provide a classification of officially sanctioned terms that are associated with macroscopic and microscopic anatomical structures. This canonical term list, however, has been substantially expanded by synonyms that are current in various fields, and has also been augmented by a large number of new terms that designate structures omitted from

Terminologia Anatomica. Many of these additions are present in various controlled terminologies (e.g., MeSH (National Library of Medicine 1999), SNOMED (Spackman et al. 1997), Read Codes (Schultz et al. 1997), GALEN (Rector et al. 1993)). Unlike Terminologia these vocabularies are entirely computer-based, and therefore lend themselves for incorporation in computer-based applications.

The most complete primate *neuroanatomical* terminology is NeuroNames, developed by Bowden and Martin at the University of Washington (Bowden and Martin 1995). NeuroNames, which is included as a knowledge source in the National Library of Medicine's Unified Medical Language System (UMLS) (Lindberg et al. 1993), is primarily organized as a part-of hierarchy of nested structures, with links to a large set of ancillary terms that do not fit into the strict part-of hierarchy. Other neuroanatomical terminologies have also been developed (Paxinos and Watson 1986; Swanson 1992; Bloom and Young 1993; Franklin and Paxinos 1997; Bug et al. 2008). A challenge for biomedical informatics is either to come up with a single consensus terminology or to develop Internet tools that allow transparent integration of distributed but commonly agreed-on terminology, with local modifications.

Classification and ontology projects to date have focused primarily on arranging the terms of a particular domain in hierarchies. As noted with respect to the evaluation of Terminologia Anatomica (Rosse 2000), insufficient attention has been paid to the relationships among these terms. These relationships are named (e.g., "*is-a*" and "*part-of*") to indicate how the entities connected by them are related (e.g., Left Lobe of Liver *part-of* Liver). Linking entities with relations encodes knowledge and is used by computer reasoning applications in making inferences. Terminologia, as well as anatomy sections of the controlled medical terminologies, mix *-is a-* and *-part of-* relationships in the anatomy segments of their hierarchies. Although such heterogeneity does not interfere with using these term lists for keyword-based retrieval, these programs will fail to support higher level

knowledge (reasoning) required for knowledge-based applications. To meet this gap, the Foundational Model of Anatomy (FMA) was developed to define a comprehensive symbolic description of the structural organization of the body, including anatomical concepts, their preferred names and synonyms, definitions, attributes and relationships (Rosse et al. 1998a, b; Rosse and Mejino 2003) (Fig. 9.11).

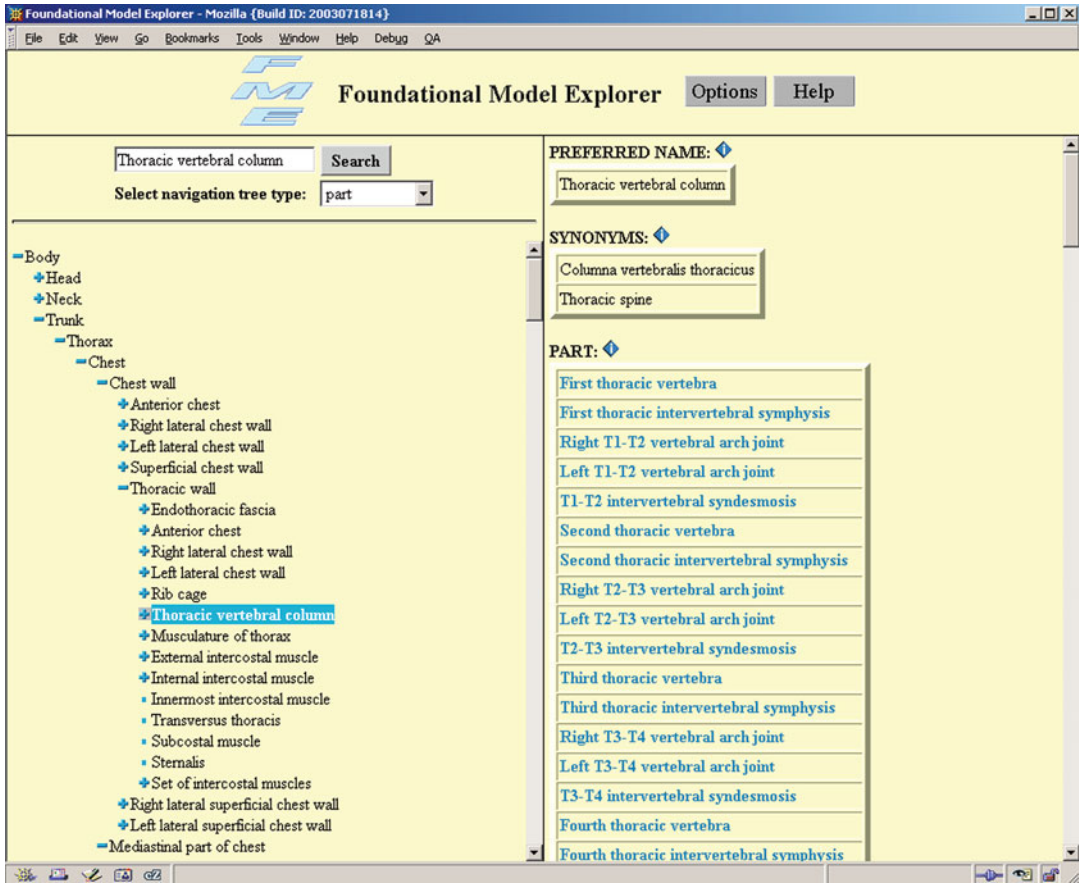
In the FMA, anatomical entities are arranged in class-subclass hierarchies, with inheritance of defining attributes along the *is-a* link, and other relationships (e.g., parts, branches, spatial adjacencies) represented as additional descriptors associated with the concept. The FMA currently consists of over 75,000 concepts, represented by about 120,000 terms, and arranged in over 2.1 million links using 168 types of relationships. These concepts represent structures at all levels: macroscopic (to 1 mm resolution), cellular and macromolecular. Brain structures have been added by integrating NeuroNames with the FMA as a Foundational Model of Neuroanatomy (FMNA) (Martin et al. 2001).

The FMA can be useful for symbolically organizing and integrating biomedical information, particularly that obtained from images. But in order to answer non-trivial queries in neuroscience and other basic science areas, and to develop "smart tools" that rely on deep knowledge, additional ontologies must also be developed (e.g., for physiological functions mediated by neurotransmitters, and pathological processes and their clinical manifestations, as well for the radiological appearances with which they correlate). The relationships that exist among these concepts and anatomical parts of the body must also be explicitly modeled. Next-generation informatics efforts that link the FMA and other anatomical ontologies with separately developed functional ontologies will be needed in order to accomplish this type of integration.

### 9.3.2.2 Knowledge Representation of Radiology Imaging Features

While FMA provides a comprehensive knowledge representation for anatomy, it does not cover other portions of the radiology domain. As is





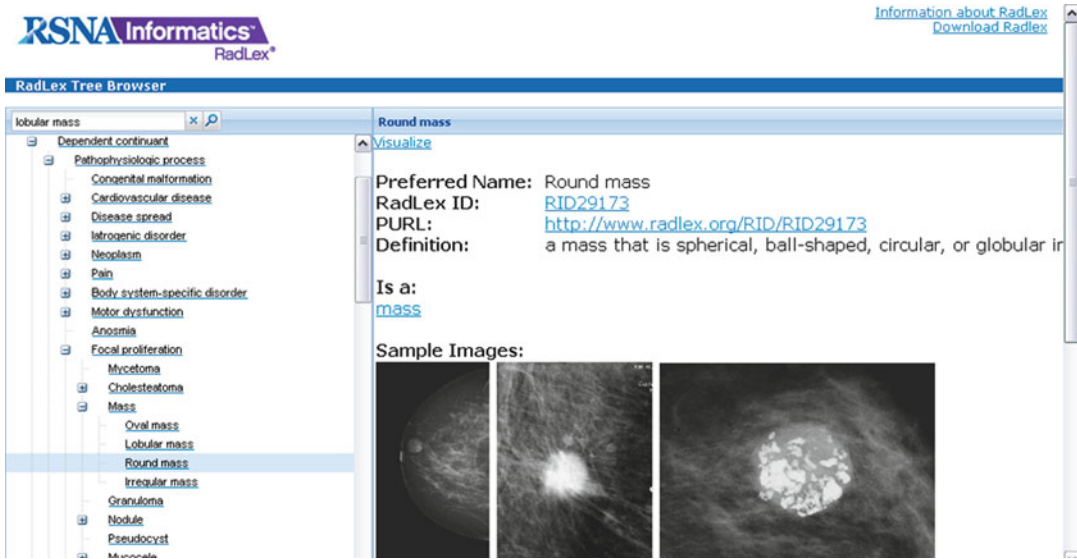
**Fig. 9.11** The Foundational Model Explorer, a Web viewer for the frame-based University of Washington Foundational Model of Anatomy (FMA). The left panel shows a hierarchical view along the part of link. Hierarchies along other links, such as is-a, branch-of, tributary-of, can also be viewed in this panel. The right

hand panel shows the detailed local and inherited attributes (slots) associated with a selected structure, in this case the thoracic vertebral column (Photograph courtesy of the Structural Informatics Group, University of Washington)

discussed in Chap. 7, there are controlled terminologies in other domains, such as MeSH, SNOMED, and related terminologies in the UMLS (Cimino 1996; Bodenreider 2008); however, these lack terminology specific to radiology for describing the features seen in imaging. The Radiological Society of North America (RSNA) recently developed RadLex, a controlled terminology for radiology (Langlotz 2006; Rubin 2008). The primary goal of RadLex is to provide a means for radiologists to communicate clear, concise, and orderly descriptions of imaging findings in understandable, unambiguous language. Another goal is to promote an orderly

thought process and logical assessments and recommendations based on observed imaging features based on terminology-based description of radiology images and to enable decision support (Baker et al. 1995; Burnside et al. 2009). Another goal of RadLex is to enable radiology research; data mining is facilitated by the use of standard terms to code large collections of reports and images (Channin et al. 2009a, b).

RadLex includes thousands of descriptors of visual observations and characteristics for describing imaging abnormalities, as well as terms for naming anatomic structures, radiology imaging procedures, and diseases (Fig. 9.12).



**Fig. 9.12** RadLex controlled terminology (<http://radlex.org>). RadLex includes term hierarchies for describing anatomy (“anatomical entity”), imaging observations (“imaging observation”) and characteristics (“imaging observation characteristic”), imaging procedures and procedure steps (“procedure step”), diseases (“pathophysio-

logic process”), treatments (“treatment”), and components of radiology reports (“report”). Each term includes definitions, preferred name, image exemplars, and other term metadata and relationships such as subsumption (Figure reprinted with permission from Rubin 2011)

Each term in RadLex contains a unique identifier as well as a variety of attributes such as definition, synonyms, and foreign language equivalents. In addition to a lexicon of standard terms, the RadLex ontology includes term relationships—links between terms to relate them in various ways to encode radiological knowledge. For example, the *is-a* relationship records subsumption. Other relationships include part-of, connectivity, and blood supply. These relationships are enabling computer-reasoning applications to process image-related data annotated with RadLex.

RadLex has been used in several imaging informatics applications, such as to improve search for radiology information. RadLex-based indexing of radiology journal figure captions achieved very high precision and recall, and significantly improved image retrieval over keyword-based search (Kahn and Rubin 2009). RadLex has been used to index radiology reports (Marwede et al. 2008). Work is underway to introduce RadLex controlled terms into radiology reports to reduce radiologist variation in

use of terms for describing images (Kahn et al. 2009). Tools are beginning to appear enabling radiologists to annotate and query image databases using RadLex and other controlled terminologies (Rubin et al. 2008; Channin et al. 2009a, b).

In addition to RadLex, there are other important controlled terminologies for radiology. The Breast Imaging Reporting and Data System (BI-RADS) is a lexicon of descriptors and a reporting structure comprising assessment categories and management recommendations created by the American College of Radiology (D’Orsi and Newell 2007). Terminologies are also being created in other radiology imaging domains, including the Fleischner Society Glossary of terms for thoracic imaging (Hansell et al. 2008), the Nomenclature of Lumbar Disc Pathology (Appel 2001), terminologies for image guided tumor ablation (Goldberg et al. 2009) and transcatheter therapy for hepatic malignancy (Brown et al. 2009), and the CT Colonography Reporting and Data System (Zalis et al. 2005).

### 9.3.2.3 Semantic Representation of Image Contents

While ontologies and controlled terminologies are useful for representing knowledge related to images, they do not provide a means to directly encode assertions for recording the semantic content in images. For example, we may wish to record the fact that “there is a mass 4×5 cm in size in the right lobe of the liver.” The representation of this semantic image content certainly will use ontologies and terminologies to record the entities to which such assertions refer; however, an **information model** is required to provide the required grammar and syntax for recording such assertions. There are two approaches to recording these assertions, no formal information model (narrative text) and a formal information model.

#### Narrative Text

In the current workflow, nearly all semantic image content is recorded in narrative text (radiology reports). The advantage of text reports is that they are simple, quick to produce (the radiologist speaks freely into a microphone), and they can be expressive, capturing the subtle nuances (and ambiguities) that the English language provides. There are several downsides, however. First, text reports are unstructured; there is no adherence to controlled terminology and not consistent structure that would permit reliable information extraction. Second, the reports may be incomplete, vague, or contradictory. Further, free text is challenging for computers (see Chap. 8), which makes it difficult to leverage free text in applications. Finally, radiology images and the corresponding radiologist report are currently disconnected; e.g., the report may describe a mass in an organ, and the image may contain a **region of interest (ROI)** measuring the lesion, but there is no information directly linking the description of the lesion in the report with the ROI in the image. Such linkage could enable applications such as content-based image retrieval, as described below.

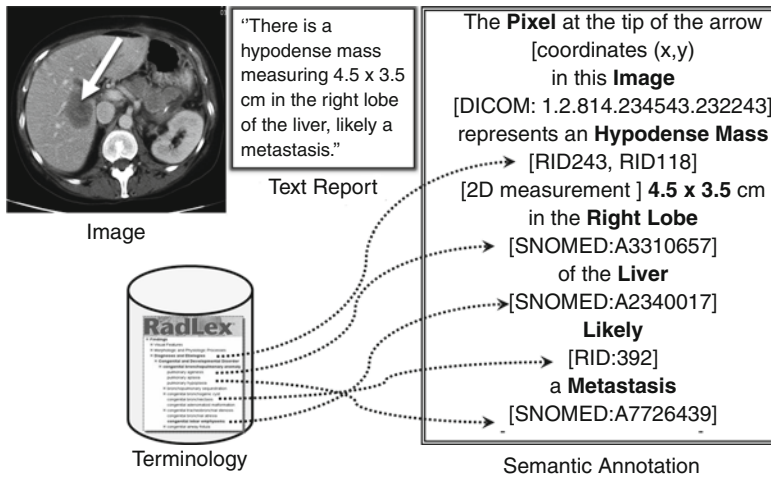
#### Information Model

An information model provides an explicit specification of the types of data to be collected and the syntax by which it will be saved. So-called

“semantic annotation” methods are being developed to adapt the semantic content about images that would have been put into narrative text so that it can instead be put in structured annotations compliant with the information model. The information model conveys the pertinent image information explicitly and in human-readable and machine accessible format. For example, a semantic annotation might record the coordinates of the tip of an arrow and indicate the organ (anatomic location) and imaging observations (e.g., mass) in that organ. These annotations can be recorded in a standard, searchable format, such as the Annotation and Image Markup (AIM) schema, recently developed by the National Cancer Institute’s Cancer Biomedical Informatics Grid (caBIG) initiative (Channin et al. 2009a, b; Rubin et al. 2009b). AIM captures a variety of information about image annotations, e.g., ROIs, lesion identification, location, measurements, method of measurement, and other qualitative and quantitative features (Channin et al. 2009a, b).

The AIM information model includes use of controlled terms as semantic descriptors of lesions (e.g., RadLex). It also provides a syntax associating an ROI in an image with the aforementioned information, enabling raw image data to be linked with semantic information, and thus bridges the current disconnect between semantic terms and the lesions in images being described. In conjunction with RadLex, the AIM information model provides a standard syntax (in XML schema) to create a structured representation of the semantic contents of images (Fig. 9.13). Once the semantic contents are recorded in AIM (as XML instances of the AIM XML schema), applications can be developed for image query and analysis. Tools for creating semantic annotation of images as part of the routine clinical and research workflow is underway (Rubin et al. 2008). Automated semantic image annotation methods are also being pursued (Carneiro, Chan et al. 2007; Mechouche et al. 2008; Yu and Ip 2008) that will ultimately make the process of generating this structured information efficient.

In addition, tools to facilitate creating semantic annotations on images as part of the image viewing workflow are being developed. One such



**Fig. 9.13** Semantic annotation of images. The radiologist's image annotation (left) and interpretation (middle) associated with the annotation are not represented in a form such that the detailed content is directly accessible. The same information can be put into a structured representa-

tion as a semantic annotation (right), comprising terms from controlled terminologies (Systematized Nomenclature of Medicine (SNOMED) and RadLex) as well as numeric values (coordinates and measurements) (Figure reprinted with permission from (Rubin and Napel 2010)

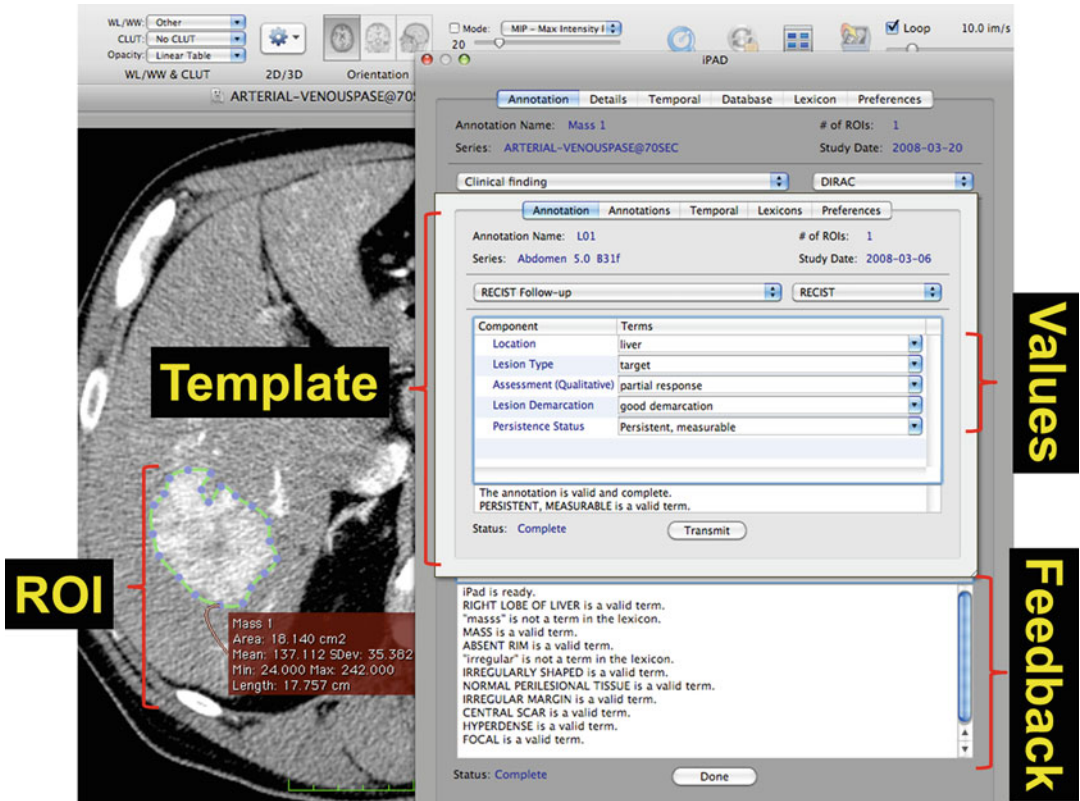
tool, the electronic Imaging Physician Annotation Device (ePAD, formerly called iPAD; (Rubin et al. 2008)), a plug in to the Osirix image viewing program, is freely available. The ePAD tool permits the user to draw image annotations in a manner in which they are accustomed while viewing images, while simultaneously collecting semantic information about the image and the image region directly from the image itself as well as from the user using a structured reporting template (Fig. 9.14). The tool also features a panel to provide feedback so as to ensure complete and valid annotations. Image annotations are saved in the AIM XML format.

By making the semantic content of images explicit and machine-accessible, these structured annotations of images will help radiologists analyze data in large databases of images. For example, cancer patients often have many serial imaging studies in which a set of lesions is evaluated at each time point. Automated methods will be able to use semantic image annotations to identify the measurable lesions at each time point and produce a summary of, and automatically reason about, the total tumor burden over time, helping physicians to determine how well patients are responding to treatment (Levy and Rubin 2008).

### 9.3.2.4 Atlases

Spatial representations of anatomy, in the form of segmented regions on 2-D or 3-D images, or 3-D surfaces extracted from image volumes, are often combined with symbolic representations to form digital atlases. A digital atlas (which for this chapter refers to an atlas created from 3-D image data taken from real subjects, as opposed to artists' illustrations) is generally created from a single individual, which therefore serves as a "canonical" instance of the species. Traditionally, atlases have been primarily used for education, and most digital atlases are used the same way.

As an example in 2-D, the Digital Anatomist Interactive Atlases (Sundsten et al. 2000) were created by outlining ROIs on 2-D images (many of which are snapshots of 3-D scenes generated by reconstruction from serial sections) and labeling the regions with terminology from the FMA. The atlases, which are available on the web, permit interactive browsing, where the names of structures are given in response to mouse clicks; dynamic creation of "pin diagrams", in which selected labels are attached to regions on the images; and dynamically-generated quizzes, in which the user is asked to point to structures on the image (Brinkley et al. 1997).



**Fig. 9.14** The electronic Imaging Physician Annotation Device (ePAD). This tool creates structured semantic annotations on images using a graphical interface to minimize impact on image viewing workflow. The user views the image in and draws a region of interest (left). ePAD incorporates ontologies so that users can specify con-

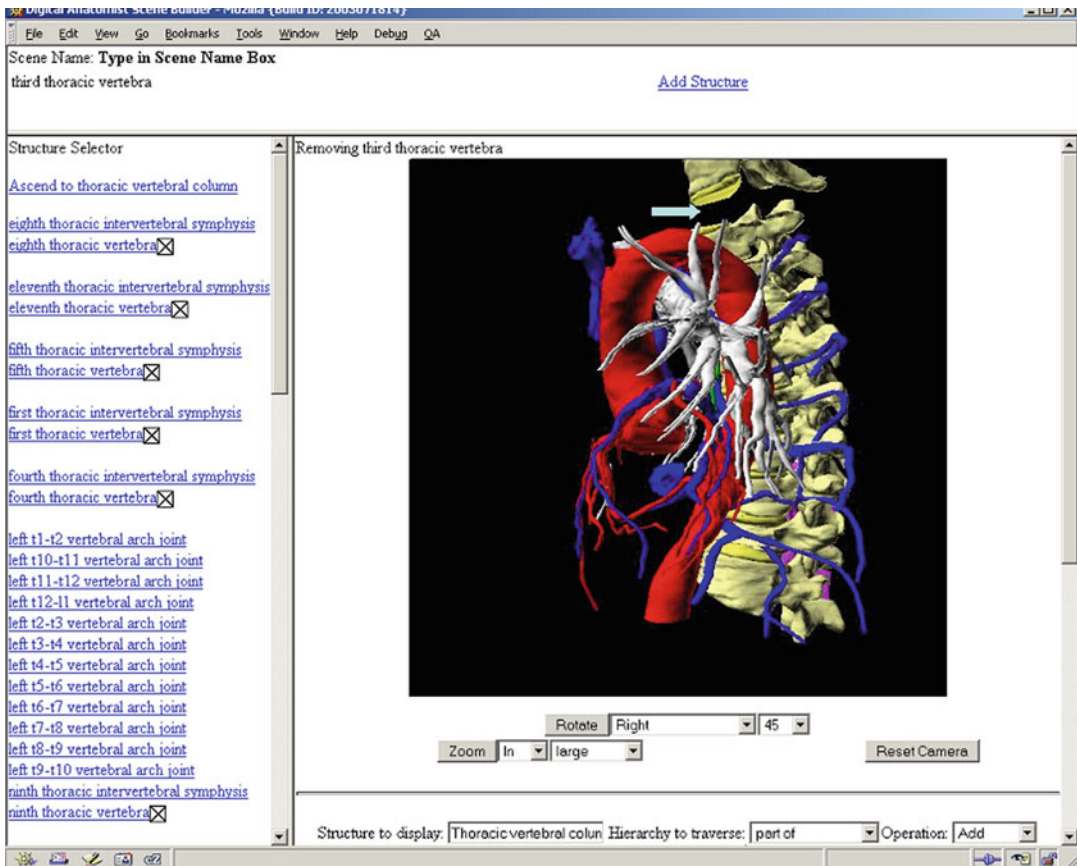
trolled terms as values in making their annotations (*pull down panel on right*). As they make their annotation, they receive feedback to ensure data entries are complete and that there are no violations of pre-specified annotation logic (*panel on lower right*). The ePAD tool saves image annotations in the AIM information model XML format

As an example 3-D, the Digital Anatomist Dynamic Scene Generator (DSG, Fig. 9.15) creates interactive 3-D atlases “on-the-fly” for viewing and manipulation over the web (Brinkley et al. 1999; Wong et al. 1999). In this case the 3-D scenes generated by reconstruction from serial sections are broken down into 3-D “primitive” meshes, each of which corresponds to an individual part in the FMA. In response to commands such as “Display the branches of the coronary arteries” the DSG looks up the branches in the FMA, retrieves the 3-D model primitives associated with those branches, determines the color for each primitive based on its type in the FMA is-a hierarchy, renders the assembled scene as a 2-D snapshot, then sends it to a web-browser, where the user may change

the camera parameters, add new structures, or select and highlight structures.

An example of a 3-D brain atlas created from the Visible Human is Voxelman (Hohne et al. 1995), in which each voxel in the Visible Human head is labeled with the name of an anatomic structure in a “generalized voxel model” (Hohne et al. 1990), and highly-detailed 3-D scenes are dynamically generated. Several other brain atlases have also been developed, primarily for educational use (Johnson and Becker 2001; Stensaas and Millhouse 2001).

Atlases have also been developed for integrating functional data from multiple studies (Bloom and Young 1993; Toga et al. 1994, 1995; Swanson 1999; Fougousse et al. 2000; Rosen et al. 2000; Martin and Bowden 2001). In their



**Fig. 9.15** The Digital Anatomist Dynamic Scene Generator. This scene was created by requesting the following structures from the scene generator server: the parts of the aorta, the branches of the ascending aorta, the tributaries of the right atrium, the branches of the tracheobronchial tree, and the parts of the thoracic vertebral column. The server was then requested to rotate the camera 45°, and to provide the name of a structure selected with

the mouse, in this case the third thoracic vertebra. The selected structure was then hidden (note the gap indicated by the arrow). The *left frame* shows a partial view of the hierarchy for the thoracic vertebral column. Checked structures are associated with three-dimensional “primitive” meshes that were loaded into the scene (Photograph courtesy of the Structural Informatics Group, University of Washington)

original published form these atlases permit manual drawing of functional data, such as neurotransmitter distributions, onto hardcopy print-outs of brain sections. Many of these atlases have been or are in the process of being converted to digital form. The Laboratory of Neuroimaging (LONI) at the University of California Los Angeles has been particularly active in the development and analysis of digital atlases (Toga 2001), and the California Institute of Technology Human Brain Project has released a web-accessible 3-D mouse atlas acquired with micro-MR imaging (Dhenain et al. 2001).

The most widely used human brain atlas is the Talairach atlas, based on post mortem sections from a 60-year-old woman (Talairach and Tournoux 1988). This atlas introduced a proportional coordinate system (often called “Talairach space”) which consists of 12 rectangular regions of the target brain that are piecewise affine transformed to corresponding regions in the atlas. Using these transforms (or a simplified single affine transform based on the anterior and posterior commissures) a point in the target brain can be expressed in Talairach coordinates, and thereby related to similarly transformed points

from other brains. Other human brain atlases have also been developed (Schaltenbrand and Warren 1977; Hohne et al. 1992; Caviness et al. 1996; Drury and Van Essen 1997; Van Essen and Drury 1997).

---

## 9.4 Image Processing

**Image processing** is a form of **signal processing** in which computational methods are applied to an input image to produce an output image or a set of characteristics or parameters related to the image. Most image processing techniques involve treating the image as a two-dimensional signal and analyzing it using signal-processing techniques or a variety of other transformations or computations. There are a broad variety of image processing methods, including transformations to enhance visualization, computations to extract features, and systems to automate detection or diagnose abnormalities in the images. The latter two methods, referred to as computer-assisted detection and diagnosis (CAD) is discussed in Sect. 9.5.2. In this section we discuss the former methods, which are more elemental and generic processing methods.

The rapidly increasing number and types of digital images has created many opportunities for image processing, since one of the great advantages of digital images is that they can be manipulated just like any other kind of data. This advantage was evident from the early days of computers, and success in processing satellite and spacecraft images generated considerable interest in biomedical image processing, including automated image analysis to improve radiological interpretation. Beginning in the 1960s, researchers devoted a large amount of work to this end, with the hope that eventually much of radiographic image analysis could be improved. One of the first areas to receive attention was automated interpretation of chest X-ray images, because, previously, most patients admitted to a hospital were subjected to routine chest X-ray examinations. (This practice is no longer considered cost effective except for selected subgroups of patients.) Subsequent research, however, con-

firmed the difficulty of completely automating radiographic image interpretation, and much of the initial enthusiasm stagnated long ago. Currently, there is less emphasis on completely automatic interpretation and more on systems that aid the user of images, except in specialized use cases.

Medical image processing utilizes tools similar to general image processing. But there are unique features to the medical imagery that present different, and often more difficult, challenges from those that exist in general image processing tasks. To begin with, the images analyzed all represent the 3D body; thus, the information extracted (be it in 2D or 3D) is based on a 3D volumetric object. The images themselves are often taken from multi-modalities (CT, MRI, PET), where each modality has its own unique physical characteristics, leading to unique noise, contrast and other issues that need to be addressed. The fusion of information across several modalities is a challenge that needs to be addressed as well.

When analyzing the data, it is often desirable to segment and characterize specific organs. The human body organs, or various tissue of interest within them, cannot be described with simple geometrical rules, as opposed to objects and scenes in non-medical images that usually can be described with such representations. This is mainly because the objects and free-form surfaces in the body cannot easily be decomposed into simple geometric primitives. There is thus very little use of geometric shape models that can be defined from *a-priori* knowledge. Moreover, when trying to model the shape of an organ or a region, one needs to keep in mind that there are large inter-person variations (e.g., in the shape and size of the heart, liver and so on), and, as we are frequently analyzing images of patients, there is a large spectrum of abnormal states that can greatly modify tissue properties or deform structures. Finally, especially in regions of interest that are close to the heart, complex motion patterns need to be accounted for as well. These issues make medical image processing a very challenging domain.

Although completely automated image-analysis systems are still in the future, the

widespread availability of digital images, combined with image management systems such as PACS (Chap. 20) and powerful workstations, has led to many applications of image processing techniques. In general, routine techniques are available on the manufacturer's workstations (e.g., a vendor-provided console for an MR machine or an ultrasound machine), whereas more advanced image-processing algorithms are available as software packages that run on independent workstations.

The primary uses of image processing in the clinical environment are for image enhancement, screening, and quantitation. Software for such image processing is primarily developed for use on independent workstations. Several journals are devoted to medical image processing (e.g., *IEEE Transactions on Medical Imaging*, *Journal of Digital Imaging*, *Neuroimage*), and the number of journal articles is rapidly increasing as digital images become more widely available. Several books are devoted to the spectrum of digital imaging processing methods (Yoo 2004; Gonzalez et al. 2009), and the reader is referred to these for more detailed reading on these topics. We describe a few examples of image-processing techniques in the remainder of this section.

#### 9.4.1 Types of Image-Processing Methods

Image processing methods are applied to representations of image content (Sect. 9.3). One may use the very low-level, pixel representation. The computational effort is minimal in the representation stage, with substantial effort (computational cost) in further analysis stages such as segmentation of the image, matching between images, registration of images, etc. A second option is to use a very high-level image content representation, in which each image is labeled according to its semantic content (medical image categories such as "abdomen vs chest", "healthy vs pathology"). In this scenario, a substantial computational effort is needed in the representation stage, including the use of automated image segmentation methods to recognize ROIs as well

as advanced learning techniques to classify the regions of image content. Further analysis can utilize knowledge resources such as ontologies, linked to the images using category labels. A mid-level representation exists, that balances the above two options, in which a transition is made from pixels to semantic features. Feature vectors are used to represent the spectrum of image content compactly and subsequent analysis is done on the feature vector representation.

Much of the current work uses the mid-level representation. In such work, a transition is made from pixel values to features, including: intensity, color, texture and in some cases also spatial coordinates or relative location features. Several main issues need to be addressed when selecting the feature set and the representation scheme: defining a global image representation (such as a histogram representation) or a more localized region-based representation, selecting a feature set that is robust or flexible to variability across the image archive, invariance issues such as the degree of sensitivity to rotation and scale. Some work has raised the issue of a hierarchical representation, such that images can be compared on the organ level in the categorization stage and on the pathology level in a higher-up stage of processing. In any scheme suggested, the representation needs to be general enough to accommodate multiple modalities and robust enough to handle the large variability of the data.

Image processing is the foundation for creating image-based applications, such as image enhancement to facilitate human viewing, to show views not present in the original images, to flag suspicious areas for closer examination by the clinician, to quantify the size and shape of an organ, and to prepare the images for integration with other information. To create such applications, several types of image processing are generally performed sequentially in an *image processing pipeline*, although some processing steps may feed back to earlier ones, and the specific methods used in a pipeline varies with the application. Most image processing pipelines and applications generalize from two-dimensional to three-dimensional images, though three-dimensional images pose unique image



processing opportunities and challenges. Image processing pipelines are generally built using one or more of the following fundamental image processing methods: **global processing, image enhancement, image rendering/visualization, image quantitation, image segmentation, image registration, and image reasoning** (e.g., classification). In the remainder of this section we describe these methods, except for image reasoning which is discussed in Sect. 9.5.

### 9.4.2 Global Processing

Global processing involves computations on the entire image, without regard to specific regional content. The purpose is generally to enhance an image for human visualization or for further analysis by the computer (“pre-processing”). A simple but important example of global image processing is gray-scale windowing of CT images. The CT scanner generates pixel values (**Hounsfield numbers**, or CT numbers) in the range of  $-1,000$  to  $+3,000$ . Humans, however, cannot distinguish more than about 100 shades of gray. To appreciate the full precision available with a CT image, the operator can adjust the mid-point and range of the displayed CT values. By changing the level and width (i.e., intercept and slope of the mapping between pixel value and displayed gray scale or, roughly, the brightness and contrast) of the display, radiologists enhance their ability to perceive small changes in contrast resolution within a subregion of interest.

Other types of global processing change the pixel values to produce an overall enhancement or desired effect on the image: *histogram equalization*, *convolution*, and *filtering*. In histogram equalization, the pixel values are changed, spreading out the most frequent intensity values to increase the global contrast of the image. It is most effective when the usable data of the image are represented by a narrow range of contrast values. Through this adjustment, the intensities can be better distributed on the histogram, improving image contrast by allowing for areas of lower local contrast to gain a higher contrast. In convolution and filtering, mathematical functions are

applied to the entire image for a variety of purposes, such as de-noising, edge enhancement, and contrast enhancement.

### 9.4.3 Image Enhancement

Image enhancement uses global processing to improve the appearance of the image either for human use or for subsequent processing by computer. All manufacturers’ consoles and independent image-processing workstations provide some form of image enhancement. We have already mentioned CT windowing. Another technique is unsharp masking, in which a blurred, or “unsharp,” positive is created to be used as a “mask” that is combined with the original image, creating the illusion that the resulting image is sharper than the original. The technique increases local contrast and enhances the visibility of fine-detail (high-frequency) structures. Histogram equalization spreads the image gray levels throughout the visible range to maximize the visibility of those gray levels that are used frequently. Temporal subtraction subtracts a reference image from later images that are registered to the first. A common use of temporal subtraction is **digital-subtraction angiography** (DSA) in which a background image is subtracted from an image taken after the injection of contrast material.

### 9.4.4 Image Rendering/ Visualization

Image rendering and visualization refer to a variety of techniques for creating image displays, diagrams, or animations to display images more in a different perspective from the raw images. Image volumes are comprised of a stack of 2-D images. If the voxels in each image are isotropic, then a variety of arbitrary projections can be derived from the volume, such as a sagittal or coronal view, or even curved planes. A technique called maximum intensity projection (MIP) and minimum intensity projection (MinIP) can also be created in which imaginary rays are cast



**Fig. 9.16** Three-dimensional ultrasound image of a fetus, in utero. The ultrasound probe sweeps out a three-dimensional volume rather than the conventional two-dimensional plane. The volume can be rendered directly using volume-rendering techniques, or as in this case, fetal surfaces can be extracted and rendered using surface-rendering techniques (Source: [http://en.wikipedia.org/wiki/File:3dultrasound\\_20\\_weeks.jpg](http://en.wikipedia.org/wiki/File:3dultrasound_20_weeks.jpg))

through the volume, recording the maximum or minimum intensity encountered along the ray path, respectively, and displaying the result as a 2-D image.

In addition to these planar visualizations, the volume can be visualized directly in its entirety using **volume rendering** techniques (Foley et al. 1990; Lichtenbelt et al. 1998) (Fig. 9.16) which project a two-dimensional image directly from a three-dimensional voxel array by casting rays from the eye of the observer through the volume array to the image plane. Because each ray passes through many voxels, some form of segmentation (usually simple thresholding) often is used to remove obscuring structures. As workstation memory and processing power have advanced, volume rendering has become widely used to display all sorts of three-dimensional voxel data—

ranging from cell images produced by confocal microscopy, to three-dimensional ultrasound images, to brain images created from MRI or PET.

Volume images can also be given as input to image-based techniques for warping the image volume of one structure to other. However, more commonly the image volume is processed in order to extract an explicit *spatial* (or quantitative) representation of anatomy (Sect. 9.4.5). Such an explicit representation permits improved visualization, quantitative analysis of structure, comparison of anatomy across a population, and mapping of functional data. It is thus a component of most research involving 3-D image processing.

### 9.4.5 Image Quantitation

Image quantitation is the process of extracting useful numerical parameters or deriving calculations from the image or from ROIs in the image. These values are also referred to as “quantitative imaging features.” These parameters may themselves be informative—for example, the volume of the heart or the size of the fetus. They also may be used as input into an automated classification procedure, which determines the type of object found. For example, small round regions on chest X-ray images might be classified as tumors, depending on such features as intensity, perimeter, and area.

Mathematical models often are used in conjunction with image quantitation. In classic pattern-recognition applications, the mathematical model is a classifier that assigns a label to the image; e.g., to indicate if the image contains an abnormality, or indicates the diagnosis underlying an abnormality.

#### 9.4.5.1 Quantitative Image Features

Quantitation uses global processing and segmentation to characterize regions of interest in the image with numerical values. For example, heart size, shape, and motion are subtle indicators of heart function and of the response of the heart to therapy (Clarysse et al. 1997). Similarly, fetal

head size and femur length, as measured on ultrasound images, are valuable indicators of fetal well-being (Brinkley 1993b).

Image features/descriptors are derived from visual cues contained in an image. Two types of quantitative image features are *photometric* features, which exploit color and texture cues, derived directly from raw pixel intensities, and *geometric* features, which use shape-based cues. While color is one of the visual cues often used for content description (Hersh et al. 2009), most medical images are grayscale. Texture features encode spatial organization of pixel values of an image region. Shape features describe in quantitative terms the contour of a lesion and complement the information captured by color or texture. In addition, the histogram of pixel values within an ROI or transforms on those values is commonly performed to compute quantitative image features.

Quantitative image features are commonly represented by feature-vectors in a N-dimensional space, where each dimension of the feature vector describes an aspect of the individual pixel (e.g., color, texture, etc.) (Haralick and Shapiro 1992). Image analysis tasks that use the quantitative features, such as segmentation and classification are then approached in terms of distance measurements between points (samples) in the chosen N-dimensional feature space.

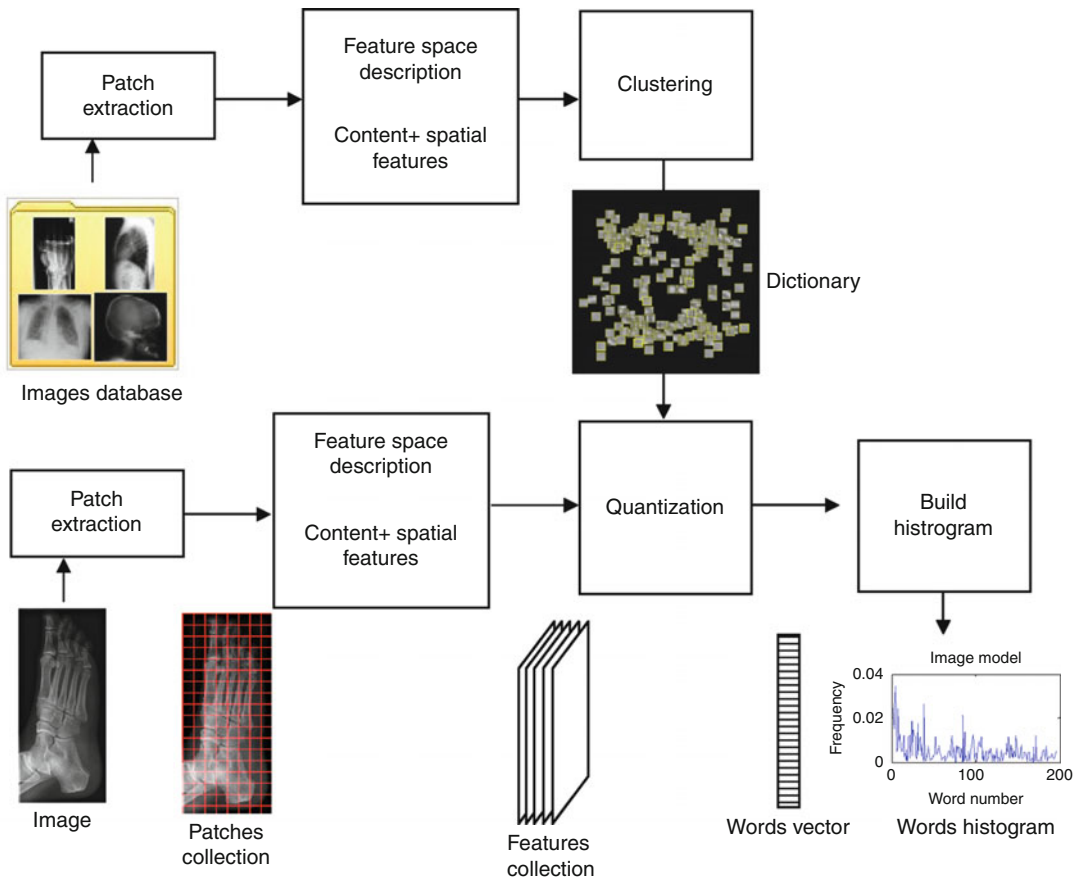
#### 9.4.5.2 Image Patches

In the last several years, “patch-based” representations and “bag-of-features” classification techniques have been proposed and used as an approach to processing image contents (Jurie and Triggs 2005; Nowak et al. 2006; Avni 2009). An overview of the methodology is shown in Fig. 9.17. In these approaches, a shift is made from the pixel as being the atomic entity of computation to a “patch” – a small window centered on the pixel, thus region-based information is included. A very large set of patches is extracted from an image. Each small patch shows a localized “glimpse” at the image content; the collection of thousands and more such patches, randomly selected, have the capability to identify the entire image content (similar to a puzzle being formed from its pieces).

The patch size needs to be larger than a few pixels across, in order to capture higher-level semantics such as edges or corners. At the same time, the patch size should not be too large if it is to serve as a common building block of many images. Patch extraction approaches include using a regular sampling grid, a random selection of points, or the selection of points with high information content using salient point detectors, such as SIFT (Lowe 1999). Once patches are selected, the information content within a patch is extracted. It is possible to take the patch information as a collection of pixel values, or to shift the representation to a different set of features based on the pixels, such as SIFT features. Frequently, the dimensionality of the representation is reduced via dimensionality reduction techniques, such as principal-component analysis (PCA) (Duda et al. 2001). In addition to patch content information represented either by PCA coefficients or SIFT descriptors, it is possible to add the patch center coordinates to the feature vector. This addition introduces spatial information into the image representation, without the need to model explicitly the spatial dependency between patches. Special care needs to be taken when combining features of different units, such as coordinates and PCA coefficients. The relative feature weights are often tuned experimentally on a cross-validation set.

A final step in the process is to learn a dictionary of words over a large collection of patches, extracted from a large set of images. The vector represented patches are converted into “visual words” which form a representative “dictionary”. A visual word can be considered as a representative of several similar patches. A frequently-used method is to perform K-means clustering (Bishop 1995) over the vectors of the initial collection, and then cluster them into K groups in the feature space. The resultant cluster centers serve as a vocabulary of K visual words, with K often in the hundreds and thousands).

Once a global dictionary is learned, each image is represented as a collection of words (also known as a “bag of words”, or “bag of features”), using an indexed histogram over the defined words. Various image processing tasks



**Fig. 9.17** A block diagram of the patch-based image representation. A radiographic image is shown with a set of patches indicated for processing the image data. Subsequent image processing is performed on each patch, and on the entire set of patches, rather than on individual

pixels in the image. A dictionary of visual words is learned from a large set of images, and their respective patches. Further analysis of the image content can then be pursued based on a histogram across the dictionary words (Figure courtesy of Greenspan)

can then be undertaken, ranging from the categorization of the image content, giving the image a “high-level,” more semantic label, the matching between images, or between an image and an image class, using patches for image segmentation and region-of-interest detection within an image. For these various tasks, images are compared using a distance measure between the representative histograms. In categorizing an image as belonging to a certain image class, well-known classifiers, such as the k- nearest neighbor and support-vector machines (SVM) (Vapnik 2000), are used.

In recent years, using patches or bags-of-visual-words (BoW) has successfully been

applied to general scene and object recognition tasks (Fei-Fei and Perona 2005; Varma and Zisserman 2003; Sivic and Zisserman 2003; Nowak et al. 2006; Jiang et al. 2007). These approaches are now gradually emerging in medical tasks as well. For example, in (André et al. 2009) BoW is used as the representation of endoscopic images and achieves high accuracy in the tasks of classifying the images into neoplastic (pathological) and benign. In (Bosch et al. 2006) an application to texture representation for mammography tissue classification and segmentation is presented. The use of BoW techniques for large scale radiograph archive categorization can be found in the ImageCLEF competition, in a

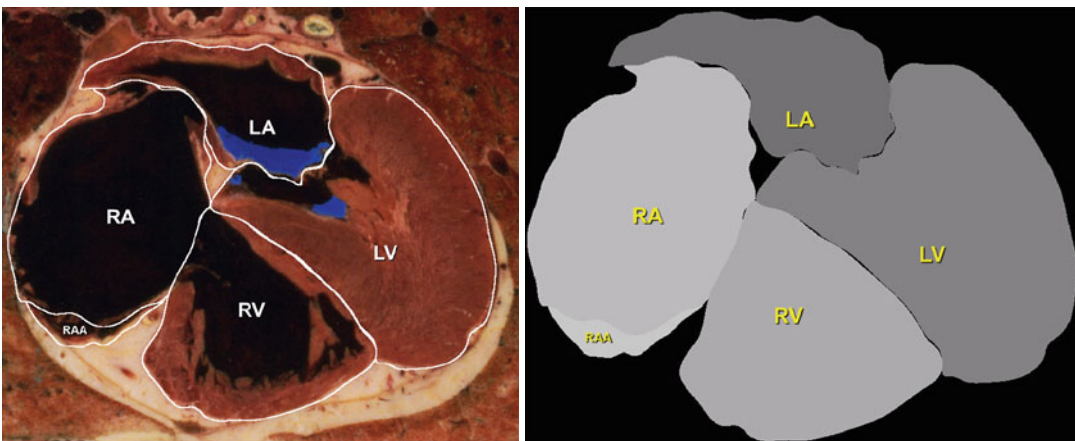
task to classify over 12,000 X-ray images to 196 different (organ-level) categories (Tommasi et al. 2010). This competition provides an important benchmarking tool to assess different feature sets as well as classification schemes on large archives of Radiographs. It is interesting to note that in the last few years, approaches based on local patch representation achieved the highest scores for categorization accuracy (Deselaers et al. 2006; Caputo et al. 2008; Greenspan et al. 2011). Current challenges entail extending from automatic classification of organs in X-ray data, to the identification and labeling of pathologies – achieving automatic healthy vs. pathology diagnostic-level categorization, as well as pathology level discrimination (e.g., work in Chest radiographs (Greenspan et al. 2011)).

#### 9.4.6 Image Segmentation

Segmentation of images involves the extraction of ROIs from the image. The ROIs usually correspond to anatomically meaningful structures, such as organs or parts of organs, or they may be lesions or other types of regions in the image pertinent to the application. The structures may be delineated by their borders, in which case

edge-detection techniques (such as edge-following algorithms) are used, or by their composition in the image, in which case region-detection techniques (such as texture analysis) are used (Haralick and Shapiro 1992). Neither of these techniques has been completely successful as fully automated image segmentation methods; regions often have discontinuous borders or non-distinctive internal composition. Furthermore, contiguous regions often overlap. These and other complications make segmentation the most difficult subtask of the medical image processing problem. Because segmentation is difficult for a computer, it is usually performed either by hand or in a semi-automated manner with assistance by a human through operator-interactive approaches (Fig. 9.18). In both cases, segmentation is time intensive, and it therefore remains a major bottleneck that prevents more widespread application of image processing techniques.

A great deal of progress has been made in automated segmentation in the brain, partially because the anatomic structures tend to be reproducibly positioned across subjects and the contrast delineation among structures is often good. In addition, MRI images of brain tend to be high quality. Several software packages are currently available for automatic segmentation, particularly



**Fig. 9.18** Image segmentation. This figure illustrates the process of segmenting and labeling the chambers of the heart. On the *left*, a cross sectional atlas image of the heart has been segmented by hand and each chamber was labeled (*RAA* right atrial appendage, *RA* right atrium, *LA*

left atrium, *RV* right ventricle, *LV* left ventricle). The boundary of each circumscribed anatomic region can be converted into a digital mask (*right*) which can be used in different applications where labeling anatomic structures in the image is needed

for normal macroscopic brain anatomy in cortical and sub-cortical regions (Collins et al. 1995; Friston et al. 1995; Subramaniam et al. 1997; Dale et al. 1999; MacDonald et al. 2000; Brain Innovation B.V. 2001; FMRIB Image Analysis Group 2001; Van Essen et al. 2001; Hinshaw et al. 2002). The Human Brain Project's Internet Brain Segmentation Repository (Kennedy 2001) has been developing a repository of segmented brain images to use in comparing these different methods.

Popular segmentation techniques include reconstruction from serial sections, region-based methods, edge-based methods, model or knowledge-based methods, and combined methods.

#### 9.4.6.1 Region-Based and Edge-Based Segmentation

In region-based segmentation, voxels are grouped into contiguous regions based on characteristics such as intensity ranges and similarity to neighboring voxels (Shapiro and Stockman 2001). A common initial approach to region-based segmentation is first to classify voxels into a small number of tissue classes. In brain MR images, a common class separation is into: gray matter, white matter, cerebrospinal fluid and background. One then uses these classifications as a basis for further segmentation (Choi et al. 1991; Zijdenbos et al. 1996). Another region-based approach is called region-growing, in which regions are grown from seed voxels manually or automatically placed within candidate regions (Davatzikos and Bryan 1996; Modayur et al. 1997). The regions found by any of these approaches are often further processed by mathematical morphology operators (Haralick 1988) to remove unwanted connections and holes (Sandor and Leahy 1997).

Edge-based segmentation is the complement to region-based segmentation: intensity gradients are used to search for and link organ boundaries. In the 2-D case, contour-following connects adjacent points on the boundary. In the 3-D case, isosurface-following or marching-cubes (Lorensen and Cline 1987) methods connect border voxels in a region into a 3-D surface mesh.

Both region-based and edge-based segmentation are essentially low-level techniques that only look at local regions in the image data.

#### 9.4.6.2 Model- and Knowledge-Based Segmentation

A popular alternative method for medical image segmentation that is popular in brain imaging is the use of deformable models. Based on pioneering work called "Snakes" by Kass, Witkin and Terzopoulos (Kass et al. 1987), deformable models have been developed for both 2-D and 3-D. In the 2-D case the deformable model is a contour, often represented as a simple set of linear segments or a spline, which is initialized to approximate the contour on the image. The contour is then deformed according to a cost function that includes both intrinsic terms limiting how much the contour can distort, and extrinsic terms that reward closeness to image borders. In the 3-D case, a 3-D surface (often a triangular mesh) is deformed in a similar manner. There are several examples of using deformable models for brain segmentation (Davatzikos and Bryan 1996; Dale et al. 1999; MacDonald et al. 2000; Van Essen et al. 2001).

An advantage of deformable models is that the cost function can include knowledge of the expected anatomy of the brain. For example, the cost function employed in the method developed by MacDonald (MacDonald et al. 2000) includes a term for the expected thickness of the brain cortex. Thus, these methods can become somewhat knowledge-based, where knowledge of anatomy is encoded in the cost function.

An alternative knowledge-based approach explicitly records shape information in a geometric constraint network (GCN) (Brinkley 1992), which encodes local shape variation based on a training set. The shape constraints define search regions on the image in which to search for edges. Found edges are then combined with the shape constraints to deform the model and reduce the size of search regions for additional edges (Brinkley 1985, 1993a, b). The advantage of this sort of model over a pure deformable model is that knowledge is explicitly represented in the model, rather than implicitly represented in the cost function.

### 9.4.6.3 Combined Methods

Most brain segmentation packages use a combination of methods in a sequential pipeline. For example, a GCN model has been used to represent the overall cortical “envelope”, excluding the detailed gyri and sulci (Hinshaw et al. 2002). The model is semi-automatically deformed to fit the cortex, then used as a mask to remove non-cortex such as the skull. Isosurface-following is then applied to the masked region to generate the detailed cortical surface. The model is also used on aligned MRA and MRV images to mask out non-cortical veins and arteries prior to isosurface-following. The extracted cortical, vein and artery surfaces are then rendered to produce a composite visualization of the brain as seen at neurosurgery (Fig. 9.9).

MacDonald et al. describe an automatic multi-resolution surface deformation technique called ASP (Anatomic Segmentation using Proximities), in which an inner and outer surface are progressively deformed to fit the image, where the cost function includes image terms, model-based terms, and proximity terms (MacDonald et al. 2000). Dale et al. describe an automated approach that is implemented in the FreeSurfer program (Dale et al. 1999; Fischl et al. 1999). This method initially finds the gray-white boundary, then fits smooth gray-white (inner) and white-CSF (outer) surfaces using deformable models. Van Essen et al. describe the SureFit program (Van Essen et al. 2001), which finds the cortical surface midway between the gray-white boundary and the gray-CSF boundary. This mid-level surface is created from probabilistic representations of both inner and outer boundaries that are determined using image intensity, intensity gradients, and knowledge of cortical topography. Other software packages also combine various methods for segmentation (Davatzikos and Bryan 1996; Brain Innovation B.V. 2001; FMRIB Image Analysis Group 2001; Sensor Systems Inc. 2001; Wellcome Department of Cognitive Neurology 2001).

### 9.4.6.4 Parametric and Non-Parametric Clustering for Segmentation

The core operation in a segmentation task is the division of the image into a finite set of regions, which are smooth and homogeneous in their

content and their representation. When posed in this way, segmentation can be regarded as a problem of finding clusters in a selected feature space.

The segmentation task can be seen as a combination of two main processes: (a) The generation of an image representation over a selected feature space. This can be termed the modeling stage. The model components are often viewed as groups, or *clusters* in the high-dimensional space. (b) The assignment of pixels to one of the model components or segments. In order to be directly relevant for a segmentation task, the clusters in the model should represent homogeneous regions of the image. In general, the better the image modeling, the better the segmentation produced. Since the number of clusters in the feature space is often unknown, segmentation can be regarded as an *unsupervised clustering* task in the high-dimensional feature space.

There is a large body of work on **clustering algorithms**. We can categorize them into three broad classes: (a) deterministic algorithms, (b) probabilistic model-based algorithms, and (c) graph-theoretic algorithms. The simplest of these are the deterministic algorithms such as k-means (Bishop 1995), mean-shift (Comaniciu and Meer 2002), and agglomerative methods (Duda et al. 2001). For certain data distributions, i.e., distributions of pixel feature vectors in a feature space, such algorithms perform well. For example, k-means provides good results when the data is convex or blob-like and the agglomerative approach succeeds when clusters are dense and there is no noise. These algorithms, however, have a difficult time handling more complex structures in the data. Further, they are sensitive to initialization (e.g., choice of initial cluster centroids). The probabilistic algorithms, on the other hand, model the distribution in the data using parametric models (McLachlan and Peel 2000). Such models include auto-regressive (AR) models, Gaussian mixture models (GMM), Markov random fields (MRF), conditional random fields, etc. Efficient ways of estimating these models are available using maximum likelihood algorithms such as the Expectation-maximization (EM) algorithm (Dempster et al. 1977). While probabilistic models offer a principled way to explain the

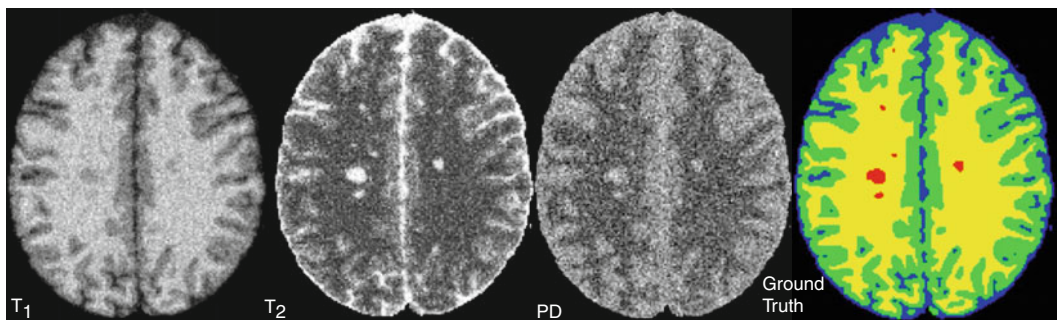
structures present in the data, they could be restrictive when more complex structures are present. Another type of clustering algorithms is non-parametric in that this class imposes no prior shape or structure on the data. Examples of these are graph-theoretic algorithms based on spectral factorization (e.g., (Ng et al. 2001; Shi and Malik 2000)). Here, the image data are modeled as a graph. The entire image data along with a global cost function are used to partition the graph, with each partition now becoming an image segment. In this approach, global considerations determine localized decisions. Moreover, such optimization procedures are often compute-intensive.

Consider an example application in brain image segmentation using parametric modeling and clustering. The tissue and lesion segmentation problem in Brain MRI is a well-studied topic of research. In such images, there is interest in three main tissue types: white matter (WM), gray matter (GM) and cerebro-spinal fluid (CSF). The volumetric analysis of such tissue types in various part of the brain is useful in assessing the progress or remission of various diseases, such as Alzheimer's disease, epilepsy, sclerosis and schizophrenia. A segmentation example is shown in Fig. 9.19. In this example, images from three MRI imaging sequences are input to the system, and the output is a segmentation map, with different colors representing three different normal brain tissues, as well as a separate color to indicate regions of abnormality (multiple-sclerosis lesions).

Various approaches to the segmentation task are reviewed in (Pham et al. 2000). Among the approaches used are pixel-level intensity based clustering, such as K-means and Mixture of Gaussians modeling (e.g., (Kapur et al. 1996)). In this approach, the intensity feature is modeled by a mixture of Gaussians, where each Gaussian is assigned a semantic meaning, such as one of the tissue regions (or lesion). Using pattern recognition methods and learning, the Gaussians can be automatically extracted from the data, and once defined, the image can be segmented into the respective regions.

Algorithms for tissue segmentation using pixel-level intensity-based classification often exhibit high sensitivity to various noise artifacts, such as intra-tissue noise, inter-tissue intensity contrast reduction, partial-volume effects and others. Due to the artifacts present, classical voxel-wise intensity-based classification methods, including the K-means modeling and Mixture of Gaussians modeling, often give unrealistic results, with tissue class regions appearing granular, fragmented, or violating anatomical constraints. Specific works can be found addressing various aspects of these concerns (e.g., partial-volume effect quantification (Dugas-Phocion et al. 2004)).

One way to address the smoothness issue is to add spatial constraints. This is often done during a pre-processing phase by using a statistical atlas, or as a post-processing step via Markov Random Field models. A statistical atlas provides the prior probability for each pixel to originate from a particular



**Fig. 9.19** Brain MRI segmentation example. Brain slice from multiple acquisition sequences (with 9 % noise) was taken from BrainWEB (<http://www.bic.mni.mcgill.ca/brainweb/>). From left to right: T1-, T2-, and proton den-

sity (PD)-weighted image. Segmentation of the images is shown on the right: Blue: CSF, Green: Gray matter (GM), Yellow: white matter (WM), Red: Multiple-sclerosis lesions (MSL) (Friefeld et al. 2009)



tissue class (e.g., (Van Leemput et al. 1999; Marroquin et al. 2002; Prastawa et al. 2004)).

Algorithms exist that use the maximum-a-posteriori (MAP) criterion to augment intensity information with the atlas. However, registration between a given image and the atlas is required, which can be computationally prohibitive (Rohlfing and Maurer 2003). Further, the quality of the registration result is strongly dependent on the physiological variability of the subject and may converge to an erroneous result in the case of a diseased or severely damaged brain. Finally, the registration process is applicable only to complete volumes. A single slice cannot be registered to the atlas. Therefore it cannot be segmented using these state-of-the-art algorithms.

Segmentation can also be improved using a post-processing phase in which smoothness and immunity to noise can be achieved by modeling the interactions among neighboring voxels. Such interactions can be modeled using a Markov Random Field (MRF), and thus this technique has been used to improve segmentation (Held et al. 1997; Van Leemput et al. 1999; Zhang et al. 2001).

Finally, there are algorithms that use deformable models to incorporate tissue boundary information (McInerney and Terzopoulos 1997). They often imply inherent smoothness but require careful initialization and precisely calibrated model parameters in order to provide consistent results in the presence of a noisy environment.

In yet another approach, the image representation is augmented to include spatial information in the feature space, and GMM clustering is utilized to provide coherent clusters in feature space that correspond to coherent spatial localized regions in the image space. In this methodology, the atlas pre-processing step and the smoothing post-processing are not required components. For regions of complex shapes in the image plane, for which a single convex hull is not sufficient (will cover two or more different segments of the image), a plausible approach is to utilize very small spatial supports per Gaussian. This in turn implies the use of a large number of Gaussians, a modeling that was shown to be useful in the brain segmentation task (Greenspan et al. 2006) as well as extended to multiple-sclerosis lesion modeling task (Friefeld et al. 2009).

## 9.4.7 Image Registration

The growing availability of 3-D and higher dimensionality structural and functional images leads to exciting opportunities for realistically observing the structure and function of the body. Nowhere have these opportunities been more widely exploited than in brain imaging. Therefore, this section concentrates on 3-D brain imaging, with the recognition that many of the methods developed for the brain have been or will be applied to other areas as well.

The basic 2-D image processing operations of global processing, segmentation, feature detection, and classification generalize to higher dimensions, and are usually part of any image processing application. However, 3-D and higher dimensionality images give rise to additional informatics issues, which include image *registration* (which also occurs to a lesser extent in 2-D), *spatial* representation of anatomy, *symbolic* representation of anatomy, integration of spatial and symbolic anatomic representations in *atlases*, *anatomical variation*, and *characterization* of anatomy. All but the first of these issues deal primarily with anatomical structure, and therefore could be considered part of the field of structural informatics. They could also be thought of as being part of imaging informatics and **neuroinformatics**.

As noted previously, 3-D image volume data are represented in the computer by a 3-D volume array, in which each voxel represents the image intensity in a small volume of space. In order to depict anatomy accurately, the voxels must be accurately registered (or located) in the 3-D volume (*voxel registration*), and separately acquired image volumes from the same subject must be registered with each other (*volume registration*).

### 9.4.7.1 Voxel Registration

Imaging modalities such as CT, MRI, and confocal microscopy (Sects. 9.2.3 and 9.2.5) are inherently 3-D: the scanner generally outputs a series of image slices that can easily be reformatted as a 3-D volume array, often following alignment algorithms that compensate for any patient motion during the scanning procedure. For this

reason, almost all CT and MR manufacturers' consoles contain some form of three-dimensional reconstruction and visualization capabilities.

As noted in Sect. 9.4.4, two-dimensional images can be converted to 3-D volumes if they are closely spaced parallel sections through a tissue or whole specimen and contain isotropic voxels. In this case the problem is how to align the sections with each other. For whole sections (either frozen or fixed), the standard method is to embed a set of thin rods or strings in the tissue prior to sectioning, to manually indicate the location of these **fiducials** on each section, then to linearly transform each slice so that the corresponding fiducials line up in 3-D (Prothero and Prothero 1986). A popular current example of this technique is the Visible Human, in which a series of transverse slices were acquired, then reconstructed to give a full 3-D volume (Spitzer and Whitlock 1998) (Chap. 20).

It is difficult to embed fiducial markers at the microscopic level, so intrinsic tissue landmarks are often used as fiducials, but the basic principle is similar. However, in this case tissue distortion may be a problem, so non-linear transformations may be required. For example Fiala and Harris (Fiala and Harris 2001) have developed an interface that allows the user to indicate, on electron microscopy sections, corresponding centers of small organelles such as mitochondria. A non-linear transformation (warp) is then computed to bring the landmarks into registration.

An approach being pursued (among other approaches) by the National Center for Microscopy and Imaging Research<sup>2</sup> combines reconstruction from thick serial sections with electron tomography (Soto et al. 1994). In this case the tomographic technique is applied to each thick section to generate a 3-D digital slab, after which the slabs are aligned with each other to generate a 3-D volume. The advantages of this approach over the standard serial section method are that the sections do not need to be as thin, and fewer of them need be acquired.

An alternative approach to 3-D voxel registration from 2-D images is stereo-matching, a tech-

nique developed in computer vision that acquires multiple 2-D images from known angles, finds corresponding points on the images, and uses the correspondences and known camera angles to compute 3-D coordinates of pixels in the matched images. The technique is being applied to the reconstruction of synapses from electron micrographs by a Human Brain Project collaboration between computer scientists and biologists at the University of Maryland (Agrawal et al. 2000).

#### 9.4.7.2 Volume Registration

A related problem to that of aligning individual sections is the problem of aligning separate image volumes from the same subject, that is, *intra-subject* alignment. Because different image modalities provide complementary information, it is common to acquire more than one kind of image volume on the same individual. This approach has been particularly useful for brain imaging because each modality provides different information. For example, PET (Sect. 9.2.3) provides useful information about function, but does not provide good localization with respect to the anatomy. Similarly, MRV and MRA (Sect. 9.2.3) show blood flow but do not provide the detailed anatomy visible with standard MRI. By combining images from these modalities with MRI, it is possible to show functional images in terms of the underlying anatomy, thereby providing a common neuroanatomic framework.

The primary problem to solve in multimodality image fusion is volume registration—that is, the alignment of separately acquired image volumes. In the simplest case, separate image volumes are acquired during a single sitting. The patient's head may be immobilized, and the information in the image headers may be used to rotate and resample the image volumes until all the voxels correspond. However, if the patient moves, or if examinations are acquired at different times, other registration methods are needed. When intensity values are similar across modalities, registration can be performed automatically by intensity-based optimization methods (Woods et al. 1992; Collins et al. 1994). When intensity values are not similar (as is the case with MRA, MRV and MRI), images can be aligned to templates of the same modalities that are already

<sup>2</sup><http://ncmir.ucsd.edu/> (accessed 4/26/13).

aligned (Woods et al. 1993; Ashburner and Friston 1997). Alternatively, landmark-based methods can be used. The landmark-based methods are similar to those used to align serial sections (see earlier discussion of voxel registration in this section), but in this case the landmarks are 3-D points. The Montreal Register Program (MacDonald 1993) is an example of such a program.

## 9.5 Image Interpretation and Computer Reasoning

The preceding sections of this chapter as well as Chap. 20 describe informatics aspects of image generation, storage, manipulation, and display of images. Rendering an interpretation is a crucial final stage in the chain of activities related to imaging. *Image interpretation* is this final stage in which the physician has direct impact on the clinical care process, by rendering a professional opinion as to whether abnormalities are present in the image and the likely significance of those abnormalities. The process of image interpretation requires reasoning—to draw inferences from facts; the facts are the image abnormalities detected and the known clinical history, and the inferred information is the diagnosis and management decision (what to do next, such as another test or surgery, etc.). Such reasoning usually entails uncertainty, and optimally would be carried out using probabilistic approaches (Chap. 3), unless certain classic imaging patterns are recognized. In reality, radiology practice is usually carried out without formal probabilistic models that relate imaging observations to the likelihood of diseases. However, variation in practice is a known problem in image interpretation (Robinson 1997), and methods to improve this process are desirable.

Informatics methods can enhance radiological interpretation of images in two major ways: (1) image retrieval systems and (2) decision support systems. The concept of *image retrieval* is similar to that of information retrieval (see Chap. 21), in which the user retrieves a set of documents pertinent to a question or information need. The information being sought when doing image

retrieval is images with specific content—typically to find images that are similar in some ways to a query image (e.g., to find images in the PACS containing similar-appearing abnormalities to that in an image being interpreted). Finding images containing similar content is referred to as **content based image retrieval** (CBIR). By retrieving similar images and then looking at the diagnosis of those patients, the radiologist can gain greater confidence in interpreting the images from patients whose diagnosis is not yet known.

As with the task of medical diagnosis (Chap. 22), radiological diagnosis can be enhanced using *decision support systems*, which assist the physician through a process called *computer reasoning*. In computer reasoning, the machine takes in the available data (the images and possibly other clinical information), performs a variety of image processing methods (Sect. 9.4), and uses one or more types of knowledge resources and/or mathematical models to render an output comprising either a decision or a ranked list of possible choices (e.g., diagnoses or locations on the image suspected of being abnormal).

There are two types of decision support in radiology, *computer-assisted detection* (CAD) and *computer-assisted diagnosis* (CADx). In the former, the computer locates ROIs in the image where abnormalities are suspected and the radiologist must evaluate their medical significance. In CADx, the computer is given an ROI corresponding to a suspected abnormality (possible with associated clinical information) and it outputs the likely diagnoses and possibly management recommendations (ideally with some sort of confidence rating as well as explanation facility).

In this section we describe informatics methods for image retrieval and decision support (computer reasoning with images).

### 9.5.1 Content-Based Image Retrieval

Since a key aspect of radiological interpretation is recognizing characteristic patterns in the imaging features which suggest the diagnosis, searching databases for similar images with known diagnoses

could be an effective strategy to improving diagnostic accuracy. CBIR is the process of performing a match between images using their visual content. A query image can be presented as input to the system (or a combination of a query image and the patient's clinical record), and the system searches for similar cases in large archive settings (such as PACS) and returns a ranked list of such similar data (images). This task requires an informative representation for the image data, along with similarity measures across image data. CBIR methods are already useful in non-medical applications such as consumer imaging and on the Web (Wang et al. 1997; Smeulders et al. 2000; Datta et al. 2008).

There has also been ongoing work to develop CBIR methods in radiology. The approach generally is based on deriving quantitative characteristics from the images (e.g., pixel statistics, spatial frequency content, etc.; Sect. 9.4.5), followed by application of similarity metrics to search databases for similar images (Lehmann et al. 2004; Muller et al. 2004; Greenspan and Pinhas 2007; Datta et al. 2008; Deserno et al. 2009). The focus of the current work is on entire images, describing them with sets of numerical features, with the goal of retrieving similar images from medical collections (Hersh et al. 2009) that provide benchmarks for image retrieval. However, in many cases only a particular region of the image is of interest when seeking similar images (e.g., finding images containing similar-appearing lesions to those in the query image). More recently, "localized" CBIR methods are being developed in which a part of the image containing a region of interest is analyzed (Deselaers et al. 2007; Rahmani et al. 2008; Napel et al. 2010).

There are several unsolved challenges in CBIR. First, CBIR has been largely focused on query based on single 2-D images; methods need to be developed for 3D retrieval in which a volume is the query "image." A second challenge is the need to integrate images with non-image clinical data to permit retrieval based on entire patient cases and not single images (e.g., the CBIR method should take into consideration the clinical history in addition to the image appearance in retrieving a similar "case").

Another limitation of current CBIR is that image semantics is not routinely included. The

information reported by the radiologist ("semantic features"), is complementary to the quantitative data contained in image pixels. One approach to capturing image semantics is analyzing and processing "visual words" in images, captured as image patches or codebooks (Sect. 9.4.5). These techniques have been shown to perform well in CBIR applications (Qiu 2002). Another approach to capture image semantics is to use the radiologist's imaging observations as image features. Several studies have found that combining the semantic information obtained from radiologists' imaging reports or annotations with the pixel-level features can enhance performance of CBIR systems (Ruiz 2006; Zhenyu et al. 2009; Napel et al. 2010). The knowledge representation methods described in Sects. 9.3.2 and 9.4.5 make it possible to combine these types of information.

### 9.5.2 Computer-Based Inference with Images and Knowledge

Though image retrieval described above (and information retrieval in general) can be helpful to a practitioner interpreting images, it does not directly answer a specific question at hand, such as, "what is the diagnosis in this patient" or "what imaging test should I order next?" Answering such questions requires reasoning, either by the physician with all the available data, or by a computer, using physician inputs and the images. As the use of imaging proliferates and the number of images being produced by imaging modalities explodes, it is becoming a major challenge for practicing radiologists to integrate the multitude of imaging data, clinical data, and soon molecular data, to formulate an accurate diagnosis and management plan for the patient. Computer-based inference systems (decision support systems) can help radiologists understand the biomedical import of this information and to provide guidance (Hudson and Cohen 2009).

There are two major approaches to computerized reasoning systems for imaging decision support, quantitative imaging-based methods (CAD/CADx) and knowledge-based computer reasoning systems.

### 9.5.2.1 Quantitative Imaging Computer Reasoning Systems (CAD/CADx)

The process of deriving quantitative image features was described in Sect. 9.4.5. Quantitative imaging applications such as CAD and CADx use these quantifiable features extracted from medical images for a variety of decision support applications, such as the assessment of an abnormality to suggest a diagnosis, or to evaluate the severity, degree of change, or status of a disease, injury, or chronic condition. In general, the quantitative imaging computer reasoning systems apply a mathematical model (e.g., a *classifier*) or other machine learning methods to obtain a decision output based on the imaging inputs.

#### CAD

In CAD applications, the goal is to scan the image and identify suspicious regions that may represent regions of disease in the patient. A common use for CAD is *screening*, the task of reviewing many images and identifying those that are suspicious and require closer scrutiny by a radiologist (e.g., mammography interpretation). Most CAD applications comprise an image processing pipeline (Sect. 9.4) that uses global processing, segmentation, image quantitation with feature extraction, and classification to determine whether an image should be flagged for careful review by a radiologist or pathologist. In CAD and in screening in general, the goal is to detect disease; thus, the tradeoff favors having false positive instead of missing false negatives. Thus CAD systems tend to flag a reasonable number of normal images (false positives) and they miss very few abnormal images (false negatives). If the number of flagged images is small compared with the total number of images, then automated screening procedures can be economically viable. On the other hand, too many false positives are time-consuming to review and lessens user confidence in the CAD system; thus for CAD to be viable, they must minimize the number of false positives as well as false negatives.

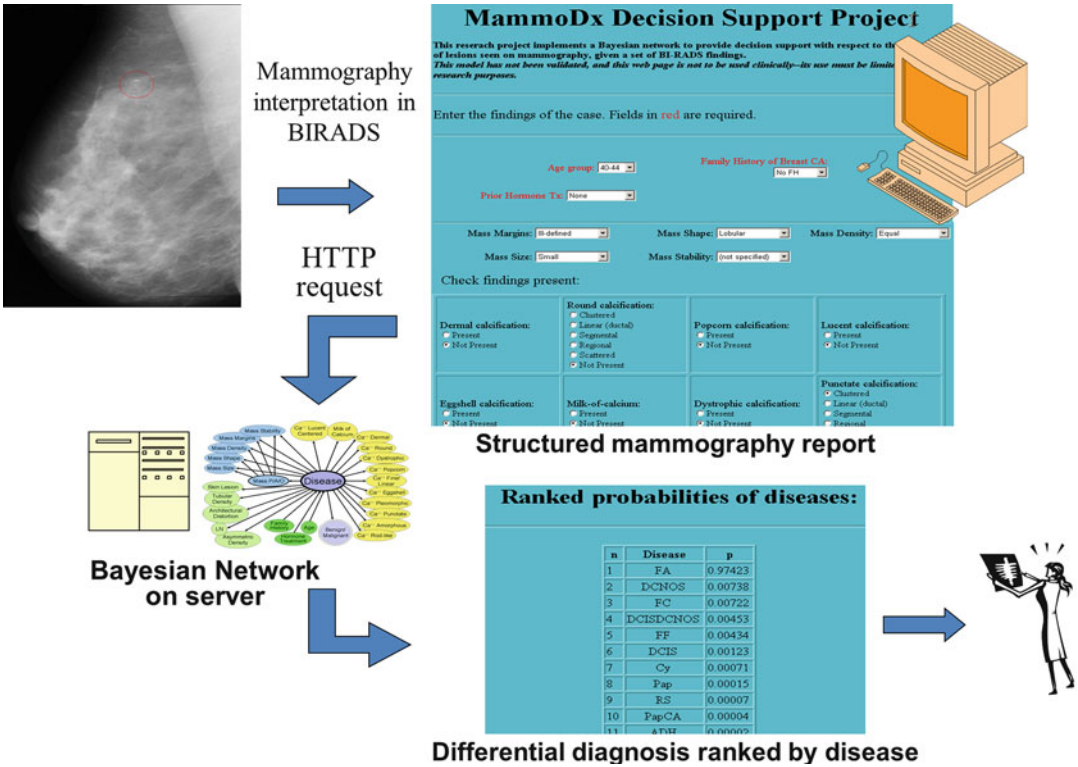
CAD techniques for screening have been applied successfully to many different types of images (Doi 2007), including mammography

images for identifying mass lesions and clusters of microcalcifications, chest X-rays and CT of the chest to detect small cancerous nodules, and volumetric CT images of the colon (“virtual colonoscopy”) to detect polyps. In addition, CAD methods have been applied to Papanicolaou (Pap) smears for cancerous or precancerous cells (Giger and MacMahon 1996), as well as to many other types of non-radiologic images.

#### CADx

In CADx applications, a suspicious region in the image has already been identified (by the radiologist of a CAD application), and the goal is to evaluate it to render a diagnosis or differential diagnosis. CADx systems usually need to be provided an ROI, or they need to segment the image to locate specific organs and lesions in order to perform analysis of quantitative image features that are extracted from the ROI and use that to render a diagnosis. In general, a mathematical model is created to relate the quantitative (or semantic) features to the likely diagnoses. Probabilistic models have been particularly effective (Burnside et al. 2000, 2004a,b, 2006, 2007; Lee et al. 2009; Liu et al. 2009, 2011), because the image features are generated based on the underlying disease, so there is probabilistic dependence on the disease and the quantitative and perceived imaging features. In fact, it can be argued that radiological interpretation is fundamentally a Bayesian task (Lusted 1960; Ledley and Lusted 1991; Donovan and Manning 2007) (see Chaps. 3 and 22), and thus decision-support strategies based on Bayesian models may be quite effective.

CADx can be very effective in practice, reducing variation and improving positive predictive value of radiologists (Burnside et al. 2006). Deploying CADx systems, however, can be challenging. Since the inputs to CADx generally need to be structured (semantic features from the radiologist and/or quantitative features from the image), a means of capturing the structured image information as part of the routine clinical workflow is required. A promising approach is to combine structured reporting with CADx (Fig. 9.20); the radiologist records the imaging



**Fig. 9.20** Bayesian network-based system for decision support in mammography CADx. The radiologist interpreting the image enters the radiology observations and clinical information (patient history) in a structured reporting Web-based data capture form to render the report. This form is sent to a server which inputs the observations into

the Bayesian network to calculate posterior probabilities of disease. A list of diseases, ranked by the probability of each disease, is return to the user who can make a decision based on a threshold of probability of malignancy, or based on shared decision making with the patient (Figure reprinted with permission from (Rubin 2011))

observations with a data capture form, which provides the structured image content required to the CADx system. Ideally the output would be presented immediately to the radiologist as the report is generated so that the output of decision support can be incorporated into the radiology report. Such implementations will be greatly facilitated by informatics methods to extract and record the image information in structured and standard formats and with controlled terminologies (Sect. 9.3.2).

**9.5.2.2 Knowledge-Based Reasoning with Images**

The CAD and CADx systems do not require processing radiological knowledge (e.g., anatomic knowledge) in order to carry out their tasks; they are based on quantitative modeling

of relationships of images features to diagnoses. However, not all image-based reasoning problems are amenable to this approach. In particular, knowledge-based tasks such as reasoning about anatomy, physiology, and pathology—tasks that entail symbolic manipulations of biomedical knowledge and application of logic—are best handled using different methods, such as ontologies and logical inference (see Chap. 22).

Knowledge-based computer reasoning applications use knowledge representations, generally ontologies, in conjunction with rules of logic to deduce information from asserted facts (e.g., from observations in the image). For example, an anatomy ontology may express the knowledge that “if a segment of a coronary artery is severed, then branches distal to the

severed branch will not receive blood,” and “the anterior and lateral portions of the right ventricle are supplied by branches of the right coronary artery, with little or no collateral supply from the left coronary artery.” Using this knowledge, and recognition via image processing that the right coronary artery is severed in an injury, a computer reasoning application could deduce that the anterior and lateral portions of the right ventricle will become ischemic (among other regions; Fig. 9.21). In performing this reasoning task, the application uses the knowledge to draw correct conclusions by manipulating the anatomical concepts and relationships using the rules of logical inference during the reasoning process.

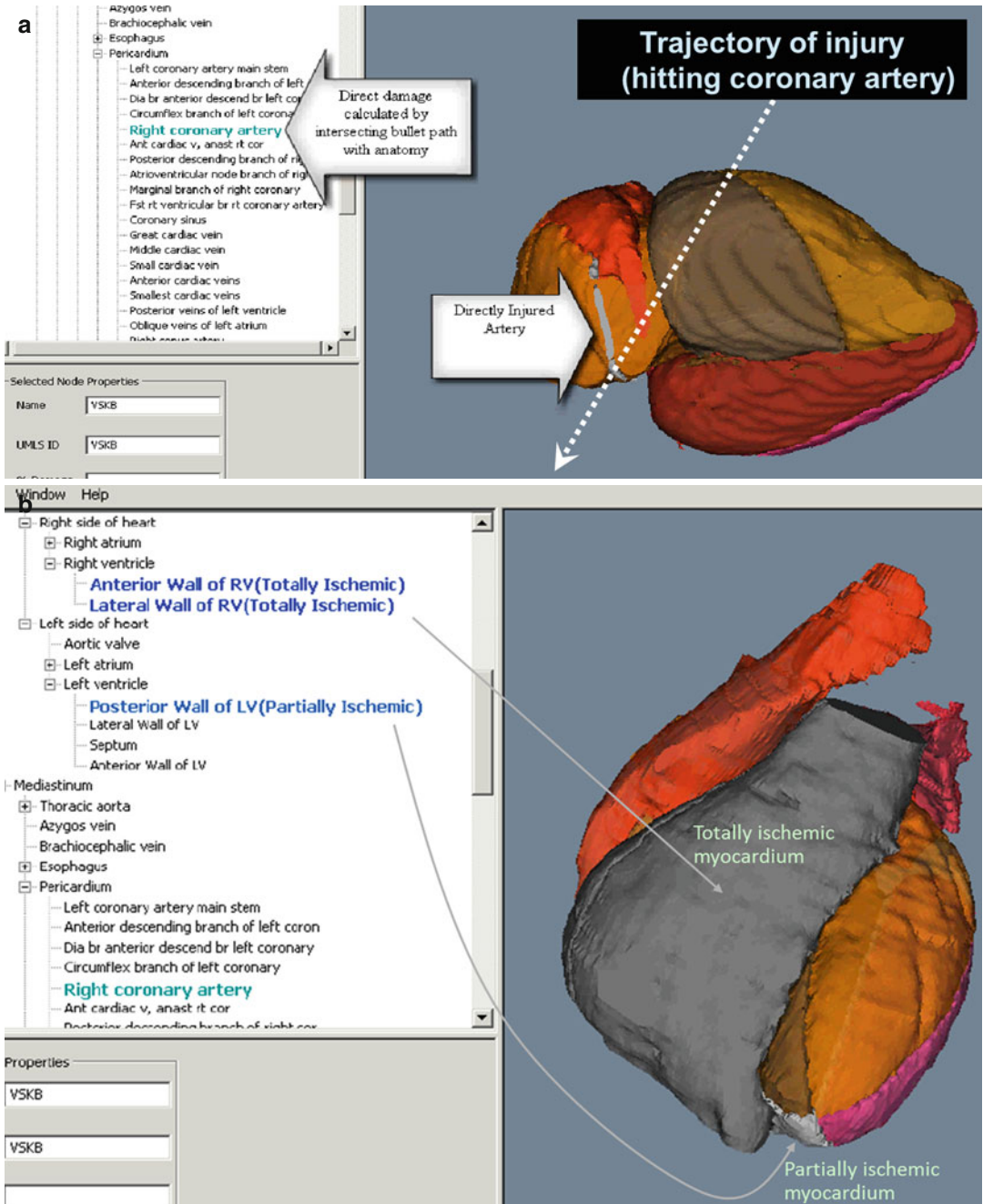
Computer reasoning with ontologies is performed by one of two methods: (1) *reasoning by ontology query* and (2) *reasoning by logical inference*. In reasoning by ontology query, the application traverse relationships that link particular entities in the ontology to directly answer particular questions about how those entities relate to each other. For example, by traversing the *part-of* relationship in an anatomy ontology, a reasoning application can infer that the left ventricle and right ventricle are part-of the chest (given that the ontology asserts they are each part of the heart, and that the heart is part of the chest), without our needing to specify this fact explicitly in the ontology.

In reasoning by logical inference, ontologies that encode sufficient information (“explicit semantics”) to apply generic reasoning engines are used. The Web Ontology Language (OWL) (Bechhofer et al. 2004; Smith et al. 2004; Motik et al. 2008) is an ontology language recommended by the World Wide Web Consortium (W3C) as a standard language for the Semantic Web (World Wide Web Consortium W3C Recommendation 10 Feb 2004). OWL is similar to other ontology languages in that it can capture knowledge by representing the entities (“classes”) and their attributes (“properties”). In addition, OWL provides the capability of defining “formal semantics” or meaning of the entities in the ontology. Entities are defined using logic statements that provide assertions about

entities (“class axioms”) using **description logics (DL)** (Grau et al. 2008). DLs provide a formalism enabling developers to define precise semantics of knowledge in ontologies and to perform automated deductive reasoning (Baader et al. 2003). For example, an anatomy ontology in OWL could provide precise semantics for “hemopericardium,” by defining it as a pericardial cavity that contains blood.

Highly optimized computer reasoning engines (“reasoners”) have been developed for OWL, helping developers to incorporate reasoning efficiently and effectively in their applications (Tsarkov and Horrocks 2006; Motik et al. 2009). These reasoners work with OWL ontologies by evaluating the asserted logical statements about classes and their properties in the original ontology (the “asserted ontology”), and they create a new ontology structure that is deduced from the asserted knowledge (the “inferred ontology”). This reasoning process is referred to as “automatic classification.” The inferences obtained from the reasoning process are obtained by querying the inferred ontology and looking for classes (or individuals) that have been assigned to classes of interest in the ontology. For example, an application was created to infer the consequences of cardiac injury in this manner (Fig. 9.21).

Several knowledge-based image reasoning systems have been developed that use ontologies as the knowledge source to process the image content and derive inferences from them. These include: (1) reasoning about the anatomic consequences of penetrating injury, (2) inferring and simulating the physiological changes that will occur given anatomic abnormalities seen in images, (3) automated disease grading/staging to infer the grade and/or stage of disease based on imaging features of disease in the body (4) surgical planning by deducing the functional significance of disruption of white matter tracts in the brain, (5) inferring the types of information users seek based on analyzing query logs of image searches, and (6) inferring the response of disease in patients to treatment based on analysis of serial imaging studies. We briefly describe these applications.



**Fig. 9.21** Knowledge-based reasoning with images in a task to predict the portions of the heart that will become ischemic after a penetrating injury that injures particular anatomic structures. The application allows the user to draw a trajectory of penetrating injury on the image, a 3-D rendering of the heart obtained from segmented CT images. The reasoning application automatically carries out two tasks. (a) The application first deduces the anatomic structures that will be injured consequent to the trajectory (arrow, right) by interrogating semantic

annotations on the image based on the trajectory of injury (injured anatomic structures shown in bold in the left panel). (b) The anatomic structures that are predicted to be initially injured are displayed in the volume rendering (dark gray=total ischemia; light gray=partial ischemia). In this example, the right coronary artery was injured, and the reasoning application correctly inferred there will be total ischemia of the anterior and lateral wall of the right ventricle and partial ischemia of the posterior wall of the left ventricle



### Reasoning About Anatomic Consequences of Penetrating Injury

In this system, images were segmented and semantic annotations applied to identify cardiac structures. An ontology of cardiac anatomy in OWL was used to encode knowledge about anatomic structures and the portions of them that are supplied by different arterial branches. Using knowledge about part-of relationships and connectivity, the application uses the anatomy ontology to infer the anatomic consequences of injury that are recognized on the input images (Fig. 9.21) (Rubin et al. 2004, 2005, 2006b).

### Inferring and Simulating the Physiological Changes

Morphological changes in anatomy have physiological consequences. For example, if a hole appears in the septum dividing the atria or ventricles of the heart (a septal defect), then blood will flow abnormally between the heart chambers and will produce abnormal physiological blood flow. The simulation community has created mathematical models to predict the physiological signals, such as time-varying pressure and flow, given particular parameters in the model such as capacitance, resistance, etc. The knowledge in these mathematical models can be represented ontologically, in which the entities correspond to nodes in the simulation model; the advantage is that a graphical representation of the ontology, corresponding to a graphical representation of the mathematical model, can be created. Morphological alterations seen in images can be directly translated into alterations in the ontological representation of the anatomic structures, and simultaneously can update the simulation model appropriately to simulate the physiological consequences of the morphological anatomic alteration (Rubin et al. 2006a). Such knowledge-based image reasoning methods could greatly enable functional evaluation of the static abnormalities seen in medical imaging.

### Automated Disease Grading/Staging

A great deal of image-based knowledge is encoded in the literature and not readily available to clinicians needing to apply it. A good example

of this is the criteria used to grade and stage disease based on imaging criteria. For example, there are detailed criteria specified for staging tumors and grading the severity of disease. This knowledge has been encoded in OWL ontologies and used to automate grading of brain gliomas (Marquet et al. 2007) and staging of cancer (Dameron et al. 2006) based on the imaging features detected by radiologists. This ontology-based paradigm could provide a good model for delivering current biomedical knowledge to practitioners “just-in-time” to help them grade and stage disease as they view images and record their observations.

### Surgical Planning

Understanding complex anatomic relationships and their functional significance in the patient is crucial in surgical planning, particular for brain surgery, since there are many surgical approaches possible, and some will have less severe consequences to patients than others. It can be challenging to be aware of all these relationships and functional dependencies; thus, surgical planning is an opportune area to develop knowledge-based image reasoning systems. The anatomic and functional knowledge can be encoded in an ontology and used by an application to plan the optimal surgical approach. In recent work, such an ontological model was developed to assess the functional sequelae of disruptions of motor pathways in the brain, which could be used in the future to guide surgical interventions (Talos et al. 2008; Rubin et al. 2009a)

### Inferring Types of Information Users Seek from Images

Knowledge-based reasoning approaches have been used to evaluate image search logs on Web sites that host image databases to ascertain the types of queries users submit. RadLex (Sect. 9.3.2) was used as the ontology, and by mapping the queries to leaf classes in RadLex and then traversing the subsumption relations, the types of queries could be deduced by interrogating the higher-level classes in RadLex (such as “visual observation” and “anatomic entity”) (Rubin et al. 2011).

## Inferring the Response of Disease Treatment

As mentioned above, the complex knowledge required to grade and stage disease can be represented using an ontology. Similarly, the criteria used to assess the response of patients to treatment is also complex, evolving, and dependent on numerous aspects of image information. The knowledge needed to apply criteria of disease response assessment have been encoded ontologically, specifically in OWL, and used to determine automatically the degree of cancer response to treatment in patients (Levy et al. 2009; Levy and Rubin 2011). The inputs to the computerized reasoning method are the quantitative information about lesions seen in the images, recorded as semantic annotations using the AIM information model (Sect. 9.3.2). This application demonstrates the potential for a streamlined workflow of radiology image interpretation and lesion measurement automatically feeding into decision support to guide patient care.

## 9.6 Conclusions

This chapter focuses on methods for computational representation and for processing images in biomedicine, with an emphasis on radiological imaging and the extraction and characterization of anatomical structure and abnormalities. It has been emphasized that the content of images is complex—comprising both quantitative and semantic information. Methods of making that content explicit and computationally-accessible have been described, and they are crucial to enable computer applications to access the “biomedical meaning” in images; presently, the vast archives of images are poorly utilized because the image content is not explicit and accessible. As the methods to extract quantitative and semantic image information become more widespread, image databases will be as useful to the discovery process as the biological databases (they will even likely become linked), and an era

of “data-driven” and “high-throughput imaging” will be enabled, analogous to modern “high-throughput” biology. In addition, the computational imaging methods will lead to applications that leverage the image content, such as CAD/CADx and knowledge-based image reasoning that use image content to improve physicians’ capability to care for patients.

Though this chapter has focused on radiology, we stress that the biomedical imaging informatics methods presented are generalizable and either have been or will be applied to other domains in which visualization and imaging are becoming increasingly important, such as microscopy, pathology, ophthalmology, and dermatology. As new imaging modalities increasingly become available for imaging other and more detailed body regions, the techniques presented in this chapter will increasingly be applied in all areas of biomedicine. For example, the development of molecular imaging methods is analogous to functional brain imaging, since functional data, in this case from gene expression rather than cognitive activity, can be mapped to an anatomical substrate.

Thus, the general biomedical imaging informatics methods described here will increasingly be applied to diverse areas of biomedicine. As they are applied, and as imaging modalities continue to proliferate, a growing demand will be placed on leveraging the content in these images to characterize the clinical phenotype of disease and relate it to genotype and clinical data from patients to enhance research and clinical care.

## Suggested Readings

- Brinkley, J. F. (1991). Structural informatics and its applications in medicine and biology. *Academic Medicine*, 66(10), 589–591. Short introduction to the field.
- Brinkley, J. F., & Rosse, C. (2002). Imaging and the human brain project: A review. *Methods of Information in Medicine*, 41, 245–260. Review of image processing work related to the brain. Much of the brain-related material for this chapter was adapted from this article.

- Deserno, T. M. (2011). *Biomedical image processing*. Berlin: Springer. Edited book of current approaches to variety of biomedical image processing tasks.
- Foley, D. D., Van Dam, A., Feiner, S. K., Hughes, J. F. (1990). *Computer Graphics: Principles and Practice*. Reading, MA: Addison-Wesley.
- Gonzalez, R. C., Woods, R. E., et al. (2009). *Digital image processing using MATLAB*. Knoxville: S.I. Gatesmark Publishing. A comprehensive overview of image processing methods focusing on MATLAB examples.
- Horii, S. C. (1996). *Image acquisition: Sites, technologies and approaches*. In Greenes, R. A. and Bauman, R. A. (eds.) *Imaging and information management: computer systems for a changing health care environment*. The Radiology Clinics of North America, 34(3):469–494.
- Pham, D. L., Xu, C., & Prince, J. L. (2000). Current methods in medical image segmentation. *Annual Review of Biomedical Engineering*, 2, 315–337. Overview of medical image segmentation.
- Potchen, E. J. (2000). Prospects for progress in diagnostic imaging. *Journal of Internal Medicine*, 247(4), 411–424. Nontechnical description of newer imaging methods such as cardiac MRI, diffusion tensor imaging, fMRI, and molecular imaging. Current and potential use of these methods for diagnosis.
- Robb, R. A. (2000). *Biomedical imaging, visualization, and analysis*. New York: Wiley. Overview of biomedical imaging modalities and processing techniques.
- Shapiro, L. G., & Stockman, G. C. (2001). *Computer vision*. Upper Saddle River: Prentice-Hall. Detailed description of many of the representations and methods used in image processing. Not specific to medicine, but most of the methods are applicable to medical imaging.
- Sonka, M., & Fitzpatrick, J. M. (2000). *Handbook of medical imaging* (Medical image processing and analysis, Vol. 2). Bellingham: SPIE Press. Overview of biomedical imaging modalities and processing techniques.
- Yoo, T. S. (2004). *Insight into images: Principles and practice for segmentation, registration, and image analysis*. Wellesley: A K Peters. Comprehensive overview of digital image processing methods with examples from the Insight Toolkit (ITK).

in a PAP smear? How would you collect and incorporate semantic features into the program?

2. Why is segmentation so difficult to perform? Give two examples of ways by which current systems avoid the problem of automatic segmentation.
3. How might you build a decision-support system that is based on searching the hospital image archive for similar images and returning the diagnosis associated with the most similar images? How might you make use of the semantic information in images to improve the accuracy of retrieval?
4. Give an example of how knowledge about the problem to be solved (e.g., local anatomy in the image) could be used in future systems to aid in automatic segmentation.
5. Both images and free text share the characteristic that they are unstructured information; image processing methods to make the biomedical content in images explicit are very similar to related problems in natural language processing (NLP; Chap. 8). How are image processing methods and NLP similar in terms of (1) computer representation of the raw content? (2) representation of the semantic content? (3) processing of the content (e.g., what is the NLP equivalent of segmentation, or the image processing equivalent of named entity recognition)?

### Questions for Discussion

1. How might you create an image processing pipeline to build an image-analysis program looking for abnormal cells