	4:4:4	4:2:2	4:2:0	4:1:1	4:1:0	4:4:0	4:4:1	4:2:1
horizontal chroma downsampling	1	2	**2**	4	4	1	1	2
vertical chroma downsampling	1	1	**2**	1	2	2	4	4
total chroma downsampling	1	2	**4**	4	8	2	4	8
number of luma bytes in 2 × 2 window	4	4	**4**	4	4	4	4	4
number of chroma bytes in 2 × 2 window	8	4	**2**	2	1	4	2	1

TABLE 2.2 The nomenclature for chroma subsampling is $J{:}a{:}b$, where J is nearly always equal to 4. From top to bottom, the rows are $h = 4/a$, $v = 2a/(a+b)$, $hv = 8/(a+b)$, 4, and $8/hv$; the final two columns show the special case when $v = 4$. Boldface is used to indicate that 4:2:0 is the most common case.

an 8-bit image) and 2 bytes of chroma data* since the chroma data is downsampled by 2 in both directions). To allow for filter overshoot and undershoot, video standards typically do not allow pixels to use all the available values but instead reserve a certain amount of **headroom** and **footroom**, so that black has a value of 16, and white has a value of 235. (While luma ranges from 16 to 235, chroma ranges from 16 to 240.)

A digital image is stored as an array of values in memory or as a sequence of bytes in a file. A large number of file formats exist. Some of the most common include **PNM**, a barebones format for uncompressed images used by researchers that includes **PGM** (for grayscale) and **PPM** (for color); **BMP**, a simple format widely used for its connection to the Windows operating system; **GIF**, an unusual format that supports multiple images for animation but only a limited color palette, making it suitable to simple shapes and logos; **PNG**, an open-source successor to GIF that supports lossless compression; **TIFF**, a flexible format with an extremely wide range of options, making it important for high-end manipulation of photographs but limiting its support in other applications such as Web browsers; **JPEG**, a widely-used format that makes use of lossy compression to reduce the file size; and **JPEG 2000**, a successor to JPEG that never gained widespread acceptance. The **EXIF** file format is increasingly being used, rather than the original **JFIF** format, to store JPEG files in order to allow metadata to be stored with the image, such as when and where the image was captured, the settings of the camera, the color space, and so forth. Another format is **OpenEXR**, which is used in the movie industry for high-dynamic range images using 16- or 32-bit floating point numbers. For video, one option is to store the video as a sequence of JPEG frames, known as **M-JPEG**. More common file formats include the historic **MPEG-1** format, or the more recent **MPEG-2**, **MPEG-4**, **AVI**, and **QuickTime** formats, which come with a dizzying array of choices for the **codec** (compressor-decompressor).† The foundational video compression standard is **H.261**, which is used by MPEG-1 and forms the basis for all later standards. The more recent standards are **H.262** (used by MPEG-2 and DVD discs) and the ubiquitous **H.264** (used by MPEG-4, Blu-ray discs, streaming Internet video, and HDTV broadcasts).

2.4 Other Imaging Modalities

Now that we have spent considerable effort explaining the imaging process for a standard optical camera, in this section we consider several alternate imaging modalities to help appreciate the great diversity of techniques for gathering images, as well as the peculiar properties of each.

* That is, 1 byte for C_B and 1 byte for C_R; see Section 9.5.3 (p. 427).

† Compression and decompression are discussed in more detail in Chapter 8.

2.4.1 Consumer Imaging: Catadioptric, RGBD, and Light-Field

We mentioned earlier that some animals, like lobsters, focus light not by using lenses, but rather by using mirrors. In a similar way, cameras can be made that either focus or bend light using mirrors. A standard optical imaging system using a lens is called **dioptric**, while a camera that uses mirrors is called **catoptric**. Putting these two together, a system that uses both lenses and mirrors is called **catadioptric**. One of the most widely used catadioptric imaging systems is the **omnidirectional** sensor, in which a camera points upward at a hyperbolic (or parabolic) mirror, which allows the camera to see 360 degrees around the scene, as shown in Figure 2.37. Such a camera system has an effective focal point at the focus of the hyperboloidal-shaped (or paraboloid-shaped) mirror, so that it is called a **central panoramic camera**. The resulting image is donut-shaped, with approximately half the pixels wasted.

Another useful sensor is the **RGBD camera** that captures not only the RGB values for each pixel but also the depth of the scene point from the camera plane. Such sensors operate by either time-of-flight, stereo processing after projecting an invisible, infrared texture onto the scene to simplify the correspondence problem, or shape from shading. Currently the most popular sensors are the Microsoft Kinect, Asus Xtion, and Intel RealSense, which are revolutionizing robotics and user interfaces due to their richer capturing modality. An entirely different approach is achieved by a **light-field camera** (also known as a plenoptic camera) that samples the light field using an array of microlenses; by tracing the rays of light using the appropriate computation, the image can be refocused after it has been captured, or it can be viewed as a 3D stereoscopic image whose appearance changes with the viewing angle.

2.4.2 Medical Imaging: CAT, PET, MRI, and Sonar

Medical applications use a variety of imaging technologies. One of the most well-known is **X-ray radiography**, which produces images by transmission rather than reflection. A generator emits X-rays toward an object of interest (such as part of a person's body), and a detector measures the photons that make it to the other side, rather than being absorbed by the object. The high amount of calcium in bones, for example, along with their high

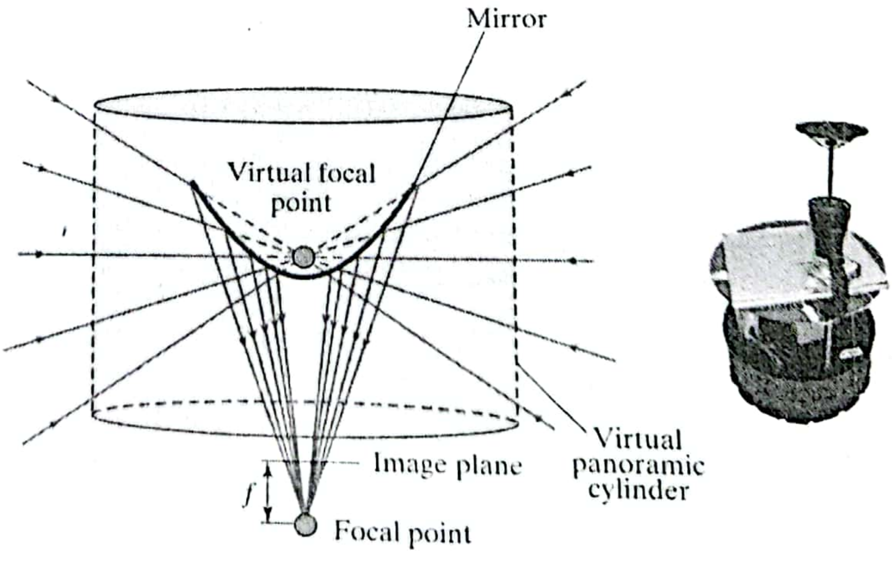

Figure 2.37 An omnidirectional camera can be achieved by attaching a hyperbolic or parabolic mirror to an upward-facing camera. Based on Valdir Grassi Junior and Jun Okamoto Junior, Development of an omnidirectional vision system, J. Braz. Soc. Mech. Sci. & Eng. vol. 28 no. 1 Rio de Janeiro Jan./Mar. 2006, http://www.scielo.br/scielo.php?script=sci_arttext&pid=S1678-58782006000100007

density, causes them to absorb X-rays, which explains why bone structure is revealed so prominently in an X-ray. X-ray technology can also be used to capture 3D structure as an array of slices. The word *tomography* refers to imaging by slices, so this approach is known as **computed tomography (CT)**. The term *axial* refers to the horizontal plane through the human body when standing upright, and since the slices are parallel to this plane, the approach is also known as **computed axial tomography (CAT)**, so that CT scan and CAT scan are essentially synonymous. A patient is enclosed in a ring of scintillation detectors, and the X-ray emitting tube is rotated around the patient, collecting an image for each slice. Reconstruction of the patient's body is then obtained using algorithms such as filtered back projection, or iterative reconstruction.

Another common technique is **magnetic resonance imaging (MRI)**, which is safer than X-ray because it does not use ionizing radiation. Soft tissue in the human body contains water, and MRI uses powerful magnets to align the hydrogen nuclei (that is, protons) in these water molecules. A radio frequency (RF) signal pulse at the resonance frequency is emitted that systematically alters the alignment of the nuclei by flipping the spin of the protons. As the nuclei return to their original state, their motion generates an RF signal which is detected by receiver coils. MRI is widely used for medical diagnosis, and its extension called **functional MRI (fMRI)** uses MRI technology to detect change in magnetization between oxygen-rich and oxygen-poor blood to measure brain activity.

Positron emission tomography (PET) is another technique for nuclear imaging. A patient is injected with a radioactive isotope which, as it undergoes positron emission decay, emits a positron. The positron travels a short distance, decelerates, and then interacts with an electron. Both the electron and positron are annihilated, emitting a pair of gamma photons in opposite directions in the process, which are detected by a scintillator. Unlike CT or MRI, PET can detect details at the level of molecular biology.

Fluoroscopy is a way of obtaining real-time images of a patient using an X-ray image intensifier to convert the X-rays on the sensor to visible light for viewing by a radiologist. A popular fluoroscopy technique is **digital subtraction angiography**, in which a contrast medium has been injected into a structure; by subtracting the precontrast image, an enhanced image is obtained which enables a physician to more easily see the blood vessels for catheters and vascular imaging. Another technique is **fluorescence in situ hybridization (FISH)** which is used to detect DNA sequences on chromosomes using fluorescent probes that bind to certain parts of chromosomes.

Finally, **ultrasound** does not use electromagnetic radiation at all but rather sound waves, which are longitudinal and require a medium for transmission. These broadband sound waves are reflected by the tissue, allowing real-time imaging of moving structures with no ionizing radiation. Ultrasound imaging is widely used for observing babies in the womb, as well as elastography, which is measuring the elastic properties of soft tissue.

2.4.3 Remote Sensing: SAR and Multispectral

Cameras are often attached to aircraft or satellites for remote sensing of the earth for applications in meteorology, agriculture, surveillance, and geology. To enable detailed sensing of the terrain in all weather conditions, these cameras typically sense multiple frequencies simultaneously. A **multispectral** sensor senses a small number of frequencies, typically 5 to 7, while a **hyperspectral** sensor senses a much larger range of frequencies. Due to the larger number of frequencies, it is often not possible to build a 2D array that yields an image directly. Instead, either a **whiskbroom** sensor is used, in which a rotating mirror scans one pixel at a time, or a **pushbroom** sensor, which is a 1D linear array perpendicular to the direction of travel. Comparing the two alternatives, a pushbroom sensor is smaller, lighter, consumes less power, and has high reliability because it has no mechanical parts.

The **Landsat program** is the longest-running program to gather satellite imagery of the earth's surface, beginning in the 1970s and continuing to the present day. The latest version of Landsat uses a whiskbroom multispectral scanner with 8 spectral bands and an opto-mechanical sensor to collect information about earth from space, with calibration used to convert raw sensed values to absolute units of radiance. Another satellite imaging program is **SPOT (Système Pour l'Observation de la Terre)**, which uses a pushbroom camera consisting of a linear array of CCDs to collect 5 spectral bands. The SPOT sensor is able to collect more photons than Landsat, so it has a higher signal-to-noise ratio. The **AVIRIS (Airborne Visible InfraRed Imaging Spectrometer)** instrument, which uses a hyperspectral sensor capable of collecting radiance in 224 contiguous spectral bands from 400 to 2500 nm, is a more recent sensor mounted on aircraft for measuring the Earth's surface atmosphere. In remote sensing, it is common to call the raw digital values from an uncalibrated sensor **digital numbers (DNs)**, to distinguish them from physically meaningful quantities such as radiance or reflectance.

Synthetic aperture radar (SAR) illuminates the scene with radio waves whose wavelength ranges from one meter to millimeters. The received echo waveforms are detected and processed to form an image. SAR is usually mounted on a moving platform with a single beam-forming antenna attached to an aircraft or spacecraft. SAR is an advanced form of side-looking airborne radar (SLAR), which is essentially a virtual phased array. Related to SAR is **ultra-wideband radar**, whose signals are defined as having a bandwidth exceeding 500 MHz or 20% the center frequency of radiation and are sometimes used for through-the-wall imaging.

2.4.4 Scientific Imaging: Microscopy

A **micrograph** is an image obtained by connecting a camera to a microscope or similar device to obtain a magnified image. An optical microscope, also known as a **light microscope**, uses visible light and a system of lenses to focus the image. Some forms of light microscopy are **bright field microscopy**, in which the light shines below the sample, yielding a dark sample on a bright background; **phase contrast microscopy**, which exploits phase shifts that occur when light passes through media, thus avoiding the need to stain the specimen and allowing for *in vivo* imaging; and **fluorescent microscopy**, which illuminates the specimen with a nearly monochromatic light to excite fluorescent stains or proteins. Most fluorescent microscopes use **epifluorescence**, in which reflected light from the specimen combines with the emitted light, yielding a high signal-to-noise ratio. To reduce the out-of-focus light and improve the contrast, the recent approach of **light sheet microscopy** has been gaining in popularity. Another advanced approach is that of a **confocal microscope**, which uses point illumination and a beam splitter to allow 2D or 3D imaging of the object with increased contrast and resolution. Further improvements in resolution are achieved using electron microscopes such as a **scanning electron microscope (SEM)**, which scans the surface using beams of electrons, or a **scanning tunneling microscope (STM)**, which uses quantum tunneling.

2.5 A Detailed Look at Electromagnetic Radiation

You may know that there are three ways to transfer heat energy. If you pick up a pan from the stove, it will feel hot to the touch because of **conduction**. If you sit in front of a rotating fan, the fan will cool your skin due to the movement of the air, known as **convection**. Both of these methods require the source responsible for heat transfer to be nearby. In contrast, if you stand outside on a sunny day, you will feel warmth from the sun, even though the