8 3-D Imaging

8.1 Basics

In this chapter we discuss various imaging techniques that can retrieve the depth coordinate which is lost by the projection of the object onto an image plane. These techniques fall into two categories. They can either retrieve only the depth of a surface in 3-D space or allow for a full reconstruction of volumetric objects. Often *depth imaging* and *volumetric imaging* are both called *3-D imaging*. This causes a lot of confusion.

Even more confusing is the wide variety of both depth and volumetric imaging techniques. Therefore this chapter will not detail all available techniques. It rather focuses on the basic principles. Surprisingly or not, there are only a few principles on which the wide variety of 3-D imaging techniques is based. If you know them, it is easy to understand how they work and what accuracy you can expect.

We start with the discussion of the basic limitation of projective imaging for 3-D vision in Section 8.1.1 and then give a brief summary of the basic principles of depth imaging (Section 8.1.2) and volumetric imaging (Section 8.1.3). Then one section is devoted to each of the basic principles of 3-D imaging: depth from triangulation (Section 8.2), depth from time-of-flight (Section 8.3), depth from phase (interferometry) (Section 8.4), shape from shading and photogrammetric stereo (Section 8.5), and tomography (Section 8.6).

8.1.1 Basic Limitation of Projective Imaging

As we have discussed in detail in Sections 7.6.1 and 7.6.2, a projective optical system is a linear shift-invariant system that can be described by a point spread function (PSF) and optical transfer function (OTF).

The 3-D OTF for geometrical optics shows the limitations of a projective imaging system best (see Section 7.6.2):

$$\hat{h}(q, k_3) = \frac{2I_0}{\pi |q \tan \alpha|} \left(1 - \frac{k_3^2}{q^2 \tan^2 \alpha} \right)^{1/2} \Pi \left(\frac{k_3}{2q \tan \alpha} \right). \qquad (8.1)$$

The symbols q and k_3 denote the radial and axial components of the wave number vector, respectively. Two severe limitations of 3-D imaging immediately follow from the shape of the 3-D OTF.

Complete loss in wide wave number range. As shown in Fig. 7.13b, the 3-D OTF is rotationally symmetric around the k_3 axis (z direction) and nonzero only inside an angle cone of $\pm \alpha$ around the xy plane. Structures with a wide range of wave numbers especially around the z axis are completely lost. We can "see" only structures in those directions from which the optics collect rays.

Loss of contrast at high wave numbers. According to Eq. (8.1), the OTF is inversely proportional to the radial wave number q. Consequently, the contrast of a periodic structure is attenuated in proportion to its wave number. As this property of the OTF is valid for all optical imaging — including the human visual system — the question arises why can we see fine structures at all?

The answer lies in a closer examination of the geometric structure of the objects observed. Most objects in the natural environment are opaque. Thus, we see only the surfaces, i. e., we do not observe real 3-D objects but only 2-D surface structures. If we image a 2-D surface onto a 2-D image plane, the 3-D PSF also reduces to a 2-D function. Mathematically, this means a multiplication of the PSF with a δ plane parallel to the observed surface. Consequently, the unsharpness disk corresponding to the distance of the surface from the lens now gives the 2-D PSF. The restriction to 2-D surfaces thus preserves the intensity of all structures with wavelengths larger than the disk. We can see them with the same contrast.

We arrive at the same conclusion in Fourier space. Multiplication of the 3-D PSF with a δ plane in the x space corresponds to a convolution of the 3-D OTF with a δ line along the optical axis, i. e., an integration in the corresponding direction. If we integrate the 3-D OTF along the k coordinate, we actually get a constant independent of the radial wave number q:

$$\frac{2I_0}{\pi} \int_{-q\tan\alpha}^{q\tan\alpha} \frac{1}{|q\tan\alpha|} \left[1 - \left(\frac{z'}{q\tan\alpha} \right)^2 \right]^{1/2} \mathrm{d}z' = I_0. \qquad (8.2)$$

To solve the integral, we substitute $z'' = z'/(q\tan\alpha)$ which yields an integral over a unit semicircle.

In conclusion, there is a significant difference between surface imaging (and thus *depth imaging*) and *volumetric imaging*. The OTF for surface structures is independent of the wave number. However, for volumetric structures, we still have the problem of the decrease of the OTF with the radial wave number. When observing such structures by eye or with a camera, we will not be able to observe fine details. Projective imaging systems are not designed to image true 3-D objects. Consequently, volumetric imaging requires different techniques.

8.1.2 Basic Principles of Depth Imaging

Depth imaging of a single opaque surface requires one additional piece of information besides the brightness at each pixel of the image in order to produce a depth image or range image. We can distinguish four basic principles of depth imaging known as *depth from paradigms*. In addition, depth can be inferred from the slope of surfaces by a paradigm known as *shape from shading*.

Depth from triangulation. If we observe an object from two different points of view separated by a base line b, the object will be seen under a different angle to the base line from both positions. This technique is known as *triangulation* and constitutes one of the basic techniques in *geodesy* and *cartography*.

The triangulation technique is at the heart of a surprisingly wide variety of techniques. At first glance these techniques appear so different that it is difficult to believe that they are based on the same principle.

Depth from time-of-flight. This is another straightforward principle of distance measurement. A signal is sent out, propagates with a characteristic speed to the object, is reflected and travels back to the camera. The travel time is directly proportional to the sum of the distances between the sender and the object and the object and the receiver.

Depth from phase: interferometry. *Interferometry* can be regarded as a special form of time-of-flight distance measurement. This technique measures distances of a fraction of the wavelength of the radiation by measuring not only the amplitude (energy) of the radiation but also its phase. Phase measurements are possible by superimposition of coherent radiation (Section 6.3.3) leading to high intensities when the two superimposing wave fronts are in phase (constructive interference) and to low intensities when they show a phase shift of 180° (π, destructive interference). Light has wavelengths between 400 and 700 nm (Section 6.3.1 and Fig. 6.6). Consequently interferometric distance measurements with light resolve distances in the nanometer range (10^{-9} m) — a small fraction of the wavelength.

Depth from coherency. Another inherent property of radiation is its *coherency length* (Section 6.3.3), i.e., the maximum path difference at which coherent superimposition is still possible. The coherency length can easily be measured by the ability to generate interference patterns. Coherency lengths can be as short as a few wavelengths. Depth from coherency techniques fill in the gap in the distance range that can be measured between interferometric techniques and time-of-flight techniques.

Shape from shading.   The shape of surfaces can also be determined from the local orientation of the surface elements. This is expressed mathematically by the surface normal. Then, of course, the absolute depth of surface is lost, but the depth profile can be computed by integrating the surface inclination. The surface normal can be inferred from the shading because the radiance of a surface depends on the angle of incidence of the illumination source.

8.1.3   Basic Principles of Volumetric Imaging

Any depth from technique that can measure multiple depths simultaneously is also useful for volumetric imaging. The capability to measure multiple depths is thus another important characteristic of a depth imaging technique. In addition to the depth imaging techniques, there are two new basic principles for volumetric images:

Illumination slicing.   In projective imaging, we do not know from which depth the irradiance collected at the image plane originates. It could be from any position of the projection ray (see Section 7.3.1 and Fig. 7.3). However, the illumination can be arranged in such a way that only a certain depth range receives light. Then we know from which depth the irradiance at the image plane originates. When we scan the illumination depth, a volumetric image can be taken.

Depth from multiple projections: tomography.   A single projection contains only partial information from a volumetric object. The question therefore is, whether it is possible to take multiple projections from different directions and to combine the different pieces of partial information to a complete 3-D image. Such *depth from multiple projections* techniques are known as *tomography*.

8.1.4   Characterization of 3-D Imaging Techniques

Depth imaging is characterized by two basic quantities, the *depth resolution* σ_z and the *depth range* Δz. The depth resolution denotes the statistical error of the depth measurement and thus the minimal resolvable depth difference. Note that the systematic error of the depth measurement can be much larger (see discussion in Section 3.1). How the resolution depends on the distance z is an important characteristic of a depth imaging technique. It makes a big difference, for example, whether the resolution is uniform, i. e., independent of the depth, or decreasing with the distance z.

The depth range Δz is the difference between the minimum and maximum depth that can be measured by a depth imaging technique. Consequently, the ratio of the depth range and depth resolution, $\Delta z/\sigma_z$, denotes the *dynamic range* of depth imaging.
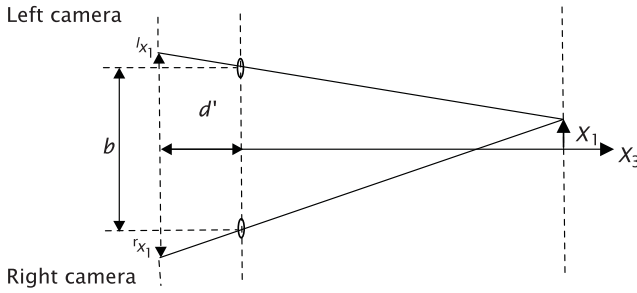
Figure 8.1: A stereo camera setup.

8.2 *Depth from Triangulation*

Looking at the same object from different points of view separated by a base vector *b* results in different viewing angles. In one way or the other, this difference in viewing angle results in a shift on the image plane, known as *disparity*, from which the depth of the object can be inferred.

Triangulation-based depth measurements include a wide variety of different techniques that — at first glance — have not much in common, but are still based on the same principle. In this section we will discuss stereoscopy (Section 8.2.1), active triangulation, where one of the two cameras is replaced by a light source (Section 8.2.2), depth from focus (Section 8.2.3), and confocal microscopy (Section 8.2.4). In the section about stereoscopy, we also discuss the basic geometry of triangulation.

8.2.1 *Stereoscopy*

Observation of a scene from two different points of view allows the distance of objects to be determined. A setup with two imaging sensors is called a *stereo system.* Many biological visual systems perform depth perception in this way. Figure 8.1 illustrates how depth can be determined from a stereo camera setup. Two cameras are placed close to each other with parallel optical axes. The distance vector *b* between the two optical axes is called the *stereoscopic basis*.

An object will be projected onto different positions of the image plane because it is viewed from slightly different angles. The difference in the position is denoted as the *disparity* or *parallax*, *p*. It is easily calculated from Fig. 8.1:

$$p = {}^{r}x_1 - {}^{l}x_1 = d'\frac{X_1 + b/2}{X_3} - d'\frac{X_1 - b/2}{X_3} = b\frac{d'}{X_3}. \tag{8.3}$$

The parallax is inversely proportional to the distance X_3 of the object (zero for an object at infinity) and is directly proportional to the stereoscopic basis and the focal length of the cameras ($d' \approx f$ for distant objects). Thus the distance estimate becomes more difficult with increasing distance. This can be seen more clearly by using the law of error propagation (Section 3.3.3) to compute

the error of X_3 from :

$$X_3 = \frac{bd'}{p} \quad \rightsquigarrow \quad \sigma_{X_3} = \frac{bd'}{p^2} \sigma_p = \frac{X_3^2}{bd'} \sigma_p. \tag{8.4}$$

Therefore, the absolute sensitivity for a depth estimate decreases with the distance squared. As an example, we take a stereo system with a stereoscopic basis of 200 mm and lenses with a focal length of 100 mm. Then, at a distance of 10 m the change in parallax is about 200 μm/m (about 20 pixel/m), while it is only 2 μm/m (0.2 pixel/m) at a distance of 100 m.

Parallax is a vector quantity and parallel to the stereoscopic basis b. This has the advantage that if the two cameras are exactly oriented we know the direction of the parallax beforehand. On the other hand, we cannot calculate the parallax in all cases. If an image sector does not show gray value changes in the direction of the stereo basis, then we cannot determine the parallax. This problem is a special case of the so-called *aperture problem* which occurs also in motion determination and will be discussed in detail in Section 14.2.2.

The depth information contained in stereo images can be perceived directly with a number of different methods. First, the left and right stereo image can be represented in one image, if one is shown in red and the other in green. The viewer uses spectacles with a red filter for the right and a green filter for the left eye. In this way, the right eye observes only the green and the left eye only the red image. This method — called the *anaglyph method* — has the disadvantage that no color images can be used. However, this method needs no special hardware and can be projected, shown on any RGB monitor, or printed out with standard printers.

Vertical stereoscopy also allows for the viewing of color stereo images [114]. The two component images are arranged one over the other. When viewed with prism spectacles that refract the upper image to the right eye and the lower image to the left eye, both images fuse into a 3-D image.

Other stereoscopic imagers use dedicated hardware. A common principle is to show the left and right stereo image in fast alternation on a monitor and switch the polarization direction of the screen synchronously. The viewer wears polarizing spectacles that filter the correct images out for the left and right eye.

However, the anaglyph method has the largest potential for most applications, as it can be used with almost any image processing workstation, the only additional piece of hardware needed being red/green spectacles. A stimulating overview of scientific and technical applications of stereo images is given by Lorenz [127].

8.2.2 *Depth from Active Triangulation*

Instead of a stereo camera setup, one camera can be replaced by a light source. For a depth recovery it is then necessary to identify at each pixel from which direction the illumination is coming. This knowledge is equivalent to knowledge of the disparity. Thus an active triangulation technique shares all basic features with the stereo system that we discussed in the previous section.

Sophisticated techniques have been developed in recent years to code the light rays in a unique way. Most commonly, light projectors are used that project
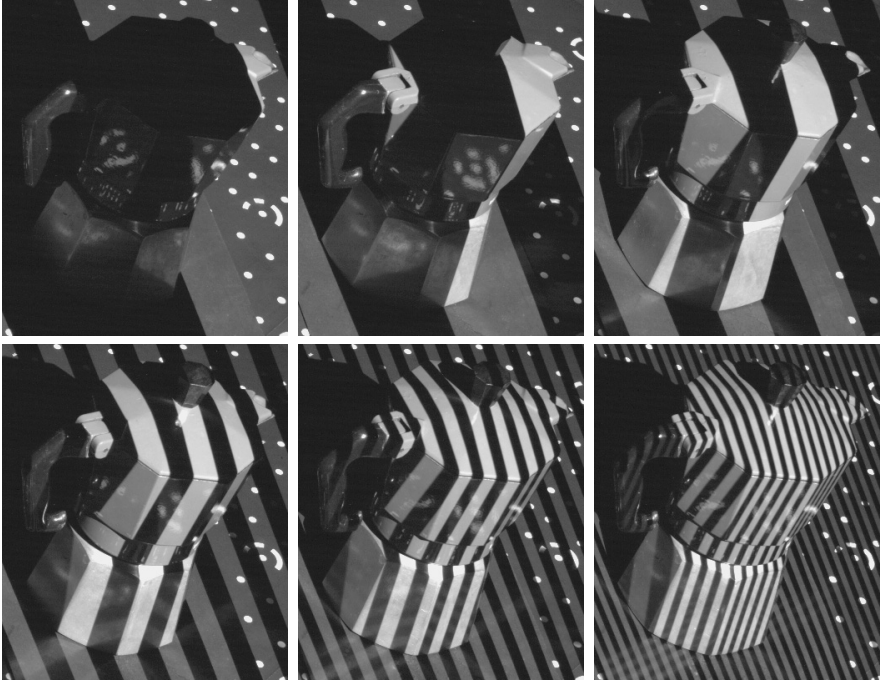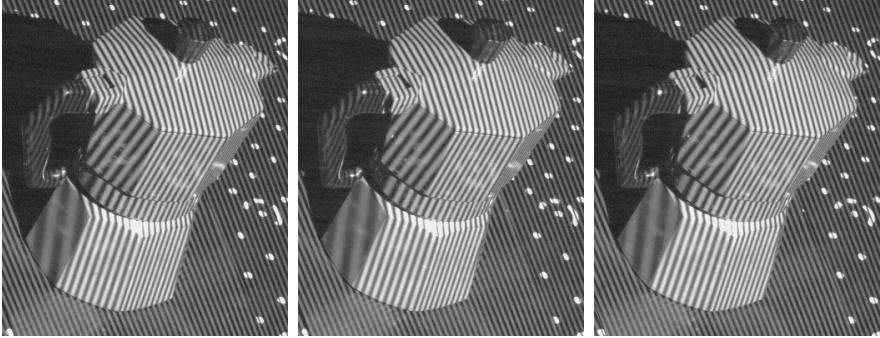
Figure 8.2: *Active triangulation by projection of a series of fringe patterns with different wavelengths for binary coding of the horizontal position; from Wiora [218].*

fringe patterns with stripes perpendicular to the triangulation base line onto the scene. A single pattern is not sufficient to identify the position of the pattern on the image plane in a unique way, but with a sequence of fringe patterns with different wavelengths, each horizontal position at the image plane of the light projector can be identified by a unique sequence of dark and bright stripes. A partial series of six such patterns is shown in Fig. 8.2.

Such a sequence of fringe patterns also has the advantage that — within the limits of the dynamic range of the camera — the detection of the fringe patterns becomes independent of the reflection coefficient of the object and the distance-dependent irradiance of the light projector. The occlusion problem that is evident from the shadow behind the espresso machine in Fig. 8.2 remains.

The binary coding by a sequence of fringe patterns no longer works for fine fringe patterns. For high-resolution position determination, as shown in Fig. 8.3, phase-shifted patterns of the same wavelength work much better and result in a subpixel-accurate position at the image plane of the light projector. Because the phase shift is only unique within a wavelength of the fringe pattern, in practice a hybrid code is often used that determines the coarse position by binary coding and the fine position by phase shifting.

Figure 8.3: *Active triangulation by phase-shifted fringe patterns with the same wavelength. Three of four patterns are shown with phase shifts of 0, 90, and 180 degrees; from Wiora [218].*

8.2.3   *Depth from Focus*

The limited *depth of field* of a real optical system (Section 7.4.3) is another technique for depth estimation. An object is only imaged without blurring if it is within the depth of field. At first glance, this does not look like a depth from triangulation technique. However, it has exactly the same geometry as the triangulation technique. The only difference is that instead of two, multiple rays are involved and the radius of the blurred disk replaces the disparity. The triangulation base corresponds to the diameter of the optics. Thus depth from focus techniques share all the basic properties of a triangulation technique. For given optics, the resolution decreases with the square of the distance (compare Eq. (8.4) with Eq. (7.21)).

The discussion on the limitations of projective imaging in Section 8.1.1 showed that the depth from focus technique does not work for volumetric imaging, because most structures, especially those in the direction of the optical axis, vanish. Depth from focus is, however, a very useful and simple technique for depth determination for opaque surfaces.

Steurer et al. [194] developed a simple method to reconstruct a *depth map* from a light microscopic focus series. A depth map is a two-dimensional function that gives the depth of an object point d — relative to a reference plane — as a function of the image coordinates $[x, y]^T$.

With the given restrictions, only one depth value for each image point needs to be found. We can make use of the fact that the 3-D point spread function of optical imaging discussed in detail in Section 7.6.1 has a distinct maximum in the focal plane because the intensity falls off with the square of the distance from the focal plane. This means that at all points where we get distinct image points such as edges, lines, or local extremes, we will also obtain an extreme in the gray value on the focal plane. Figure 8.4 illustrates that the point spread functions of neighboring image points only marginally influence each other close to the focal plane.
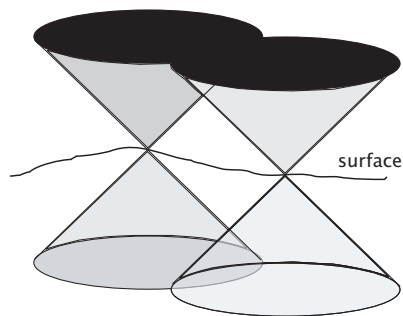
Figure 8.4: *Superposition of the point spread function of two neighboring points on a surface.*

a

b
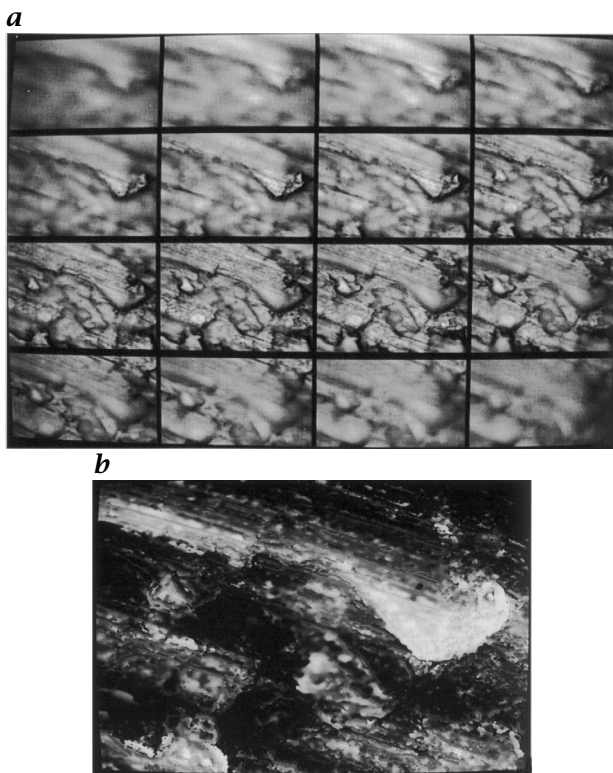
Figure 8.5: *a Focus series with 16 images of a metallic surface taken with depth distances of 2 µm; the focal plane becomes deeper from left to right and from top to bottom. b Depth map computed from the focus series. Depth is coded by intensity. Objects closer to the observer are shown brighter. From Steurer et al. [194].*
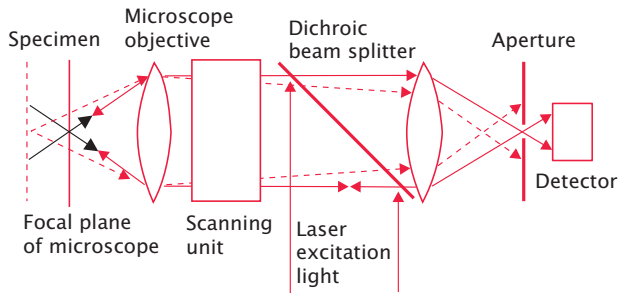
Figure 8.6: *Principle of confocal laser scanning microscopy.*

Steurer's method makes use of the fact that a distinct maximum of the point spread function exists in the focal plane. His algorithm includes the following four steps:

1. Take a focus series with constant depth steps.
2. Apply a suitable filter such as the *variance operator* (Section 15.2.2) to emphasize small structures. The highpass-filtered images are segmented to obtain a mask for the regions with significant gray value changes.
3. In the masked regions, search for the maximum magnitude of the difference in all the images of the focus series. The image in which the maximum occurs gives a depth value for the depth map. By interpolation of the values the depth position of the maximum can be determined more exactly than with the depth resolution of the image series [178].
4. As the depth map will not be dense, interpolation is required. Steurer used a region-growing method followed by an adaptive lowpass filtering which is applied only to the interpolated regions in order not to corrupt the directly computed depth values. However, other valid techniques, such as normalized convolution (Section 11.6.2) or any of the techniques described in Section 17.2, are acceptable.

This method was successfully used to determine the surface structure of worked metal pieces. Figure 8.5 shows that good results were achieved. A filing can be seen that projects from the surface. Moreover, the surface shows clear traces of the grinding process.

This technique works only if the surface shows fine details. If this is not the case, the confocal illumination technique of Scheuermann et al. [178] can be applied that projects statistical patterns into the focal plane (compare Section 1.2.2 and Fig. 1.3).

8.2.4   *Confocal Microscopy*

Volumetric microscopic imaging is of utmost importance for material and life sciences. Therefore the question arises, whether it is possible to change the image formation process — and thus the point spread function — so that the optical transfer function no longer vanishes, especially in the z direction.
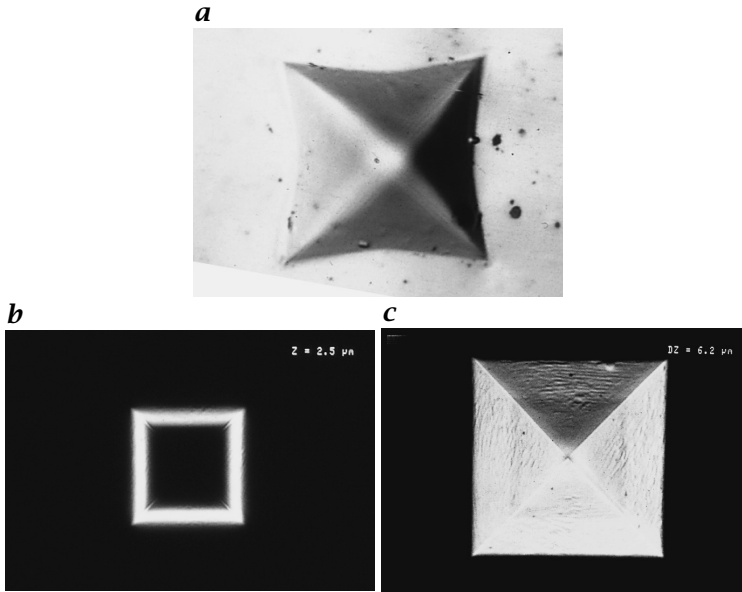
Figure 8.7: *Demonstration of confocal laser scanning microscopy (CLSM). **a** A square pyramid-shaped crystal imaged with standard microscopy focused on the base of the pyramid. **b** Similar object imaged with CLSM: only a narrow height contour range, 2.5 µm above the base of the square pyramid, is visible. **c** Image composed of a 6.2 µm depth range scan of CLSM images. Images courtesy of Carl Zeiss Jena GmbH, Germany.*

The answer to this question is *confocal laser scanning microscopy*. Its basic principle is to illuminate only the points in the focal plane. This is achieved by scanning a laser beam over the image plane that is focused by the optics of the microscope onto the focal plane (Fig. 8.6). As the same optics are used for imaging and illumination, the intensity distribution in the object space is given approximately by the point spread function of the microscope. (Slight differences occur, as the laser light is coherent.) Only a thin slice close to the focal plane receives a strong illumination. Outside this slice, the illumination falls off with the distance squared from the focal plane. In this way contributions from defocused objects outside the focal plane are strongly suppressed and the distortions decrease. However, can we achieve a completely distortion-free reconstruction? We will use two independent trains of thought to answer this question.

Let us first imagine a periodic structure in the z direction. In conventional microscopy, this structure is lost because all depths are illuminated with equal radiance. In confocal microscopy, however, we can still observe a periodic variation in the z direction because of the strong decrease of the illumination intensity provided that the wavelength in the z direction is not too small.

The same fact can be illustrated using the PSF. The PSF of confocal microscopy is given as the product of spatial intensity distribution and the PSF of the optical

imaging. As both functions fall off with z^{-2}, the PSF of the confocal microscope falls off with z^{-4}. This much sharper localization of the PSF in the z direction results in a nonzero OTF in the z direction up to the z resolution limit.

The superior 3-D imaging of confocal laser scanning microscopy is demonstrated in Fig. 8.7. An image taken with standard microscopy shows a crystal in the shape of a square pyramid which is sharp only at the base of the pyramid (Fig. 8.7a). Towards the top of the pyramid, the edges become more blurred. In contrast, a single image taken with a confocal laser scanning microscopy images only a narrow height range at all (Fig. 8.7b). An image composed of a 6.2 μm depth scan by adding up all images shows a sharp image for the whole depth range (Fig. 8.7c). Many fine details can be observed that are not visible in the image taken with the conventional microscope. The laser-scanning microscope has found widespread application in medical and biological sciences and materials research.

8.3 *Depth from Time-of-Flight*

Time-of-flight techniques measure the delay caused by the time for a signal to travel a certain distance. If the signal is sent out from the position of the camera, it has to travel twice the distance between the camera and the object reflecting the signal. Therefore the delay τ is given by

$$\tau = \frac{2z}{c},$$ (8.5)

where c is the travel speed of the signal. From Eq. (8.5) it is evident that the statistical error of the depth measurement is independent of the distance to the object. It only depends on the accuracy of the delay measurement:

$$z = \frac{c\tau}{2} \quad \rightsquigarrow \quad \sigma_z = \frac{c}{2}\sigma_\tau.$$ (8.6)

This is a significant advantage over triangulation techniques (Eq. (8.4)).

With time-of-flight techniques one immediately thinks of *pulse modulation*, i.e., measuring the time of flight by the delay between sending and receiving a short pulse. The maximum measurable distance depends on the frequency with which the pulses are sent to the object. With electromagnetic waves, delay measurements are very demanding. Because the light speed c is $3 \cdot 10^8$ m/s, the delay is only 6.7 ns per meter.

Pulse modulation is only one of many techniques to modulate the signal for time-of-flight measurements. Another powerful technique is the *continuous-wave modulation* (CW modulation). With this technique the signal is modulated periodically and the delay is measured as a phase shift between the outgoing and ingoing signal:

$$z = \frac{c}{4\pi\nu}\phi \quad \rightsquigarrow \quad \sigma_z = \frac{c}{4\pi\nu}\sigma_\phi,$$ (8.7)

where ν is the frequency of the modulation. The depth range is given by the fact that the phase can be measured uniquely only in a range of $\pm\pi$:

$$\Delta z = \frac{c}{2\nu} = \frac{cT}{2}.$$ (8.8)

One of the most significant disadvantages of periodic modulation is thus the limited depth range. This problem is overcome by *pseudo-noise modulation* where the signal amplitude is randomly modulated. This technique combines the high resolution of CW modulation with the large distance range of pulse modulation.

8.4   *Depth from Phase: Interferometry*

Interferometry can be regarded as a special case of continuous-wave modulation. The modulation is given directly by the frequency of the electromagnetic radiation. It is still useful to regard interferometry as a special class of range measurement technique because coherent radiation (Section 6.3.3) is required. Because of the high frequencies of light, the phases of the outgoing and incoming radiation cannot be measured directly but only by the amplitude variation caused by the coherent optical superimposition of the outgoing and incoming light.

The depth error and depth range for interferometric range measurements is simply given by Eqs. (8.7) and (8.8) and the relations $c = \nu\lambda$ (Section 6.3.1):

$$z = \frac{\lambda}{4\pi}\phi, \quad \sigma_z = \frac{\lambda}{4\pi}\sigma_\phi, \quad \Delta z = \frac{\lambda}{2}. \tag{8.9}$$

Because of the small wavelength of light (0.4–0.7 μm), interferometric measurements are extremely sensitive. The limited depth range of only half a wavelength can be overcome by *multiwavelength interferometry*

A second class of interferometric range measuring techniques is possible with radiation that shows a coherence length of only a few wavelengths. Then interference patterns occur only for a short distance of a few wavelengths and can thus be taken as a depth measurement in a scanning system. This type of interferometry is known as *white-light interferometry* or *coherency radar*.

8.5   *Shape from Shading*

Shape from shading techniques do not infer the depth but the normal of surfaces and thus form an entirely new class of surface reconstruction techniques. It is obvious that shape from shading techniques cannot infer absolute distances.

8.5.1   *Shape from Shading for Lambertian Surfaces*

We first apply this technique for diffuse reflecting opaque objects. For the sake of simplicity, we assume that the surface of a Lambertian object is illuminated by parallel light. The radiance L of a Lambertian surface (Section 6.4.3) does not depend on the viewing angle and is given by:

$$L = \frac{\rho(\lambda)}{\pi}E\cos\gamma, \tag{8.10}$$
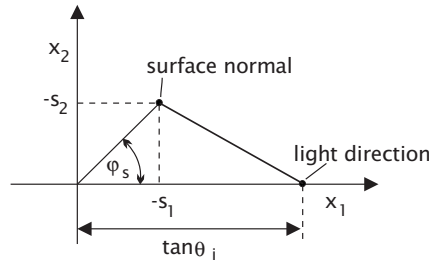
Figure 8.8: *Radiance computation illustrated in the gradient space for a Lambertian surface illuminated by a distant light source with an incidence angle θ_i and an azimuthal angle ϕ_i of zero.*

where E is the irradiance and γ the angle between the surface normal and the illumination direction. The relation between the surface normal and the incident and exitant radiation can most easily be understood in the *gradient space.* This space is spanned by the gradient of the surface height $a(X, Y)$:

$$s = \nabla a = \left[\frac{\partial a}{\partial X}, \frac{\partial a}{\partial Y} \right]^T = [s_1, s_2]^T . \tag{8.11}$$

This gradient is directly related to the surface normal n by

$$n = \left[-\frac{\partial a}{\partial X}, -\frac{\partial a}{\partial Y}, 1 \right]^T = [-s_1, -s_2, 1]^T . \tag{8.12}$$

This equations shows that the gradient space can be understood as a plane parallel to the XY plane at a height $Z = 1$ if we invert the directions of the X and Y axes. The X and Y coordinates where the surface normal vector and other directional vectors intersect this plane are the corresponding coordinates in the gradient space.

The geometry of Lambertian reflection in the gradient space is illustrated in Fig. 8.8. Without loss of generality, we set the direction of the light source as the x direction. Then, the light direction is given by the vector $l = (\tan \theta_i, 0, 1)^T$, and the radiance L of the surface can be expressed as

$$L = \frac{\rho(\lambda)}{\pi} E \frac{n^T l}{|n||l|} = \frac{\rho(\lambda)}{\pi} E \frac{-s_1 \tan \theta_i + 1}{\sqrt{1 + \tan^2 \theta_i} \sqrt{1 + s_1^2 + s_2^2}} . \tag{8.13}$$

Contour plots of the radiance distribution in the gradient space are shown in Fig. 8.9a for a light source with an incidence angle of $\theta_i = 0°$. In the case of the light source at the zenith, the contour lines of equal radiance mark lines with constant absolute slope $s = (s_1^2 + s_2^2)^{1/2}$. However, the radiance changes with surface slope are low, especially for low surface inclinations. An oblique illumination leads to a much higher contrast in the radiance (Fig. 8.9b). With an oblique illumination, however, the maximum surface slope in the direction opposite to the light source is limited to $\pi/2 - \theta$ when the surface normal is perpendicular to the light direction.

a

b
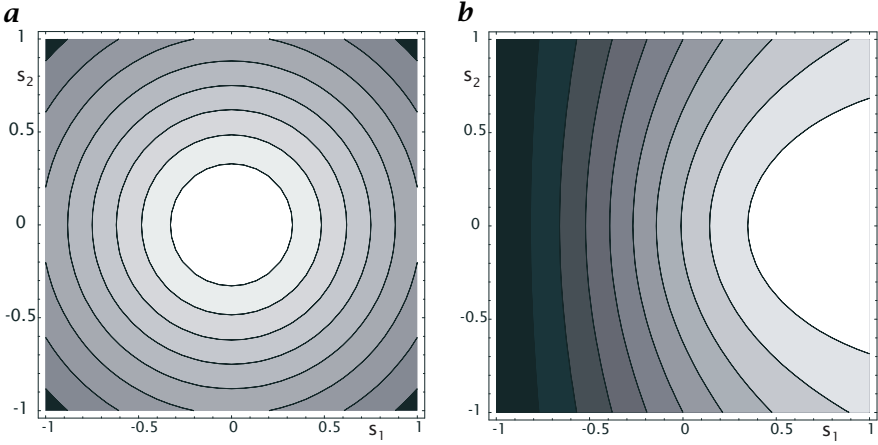
Figure 8.9: *Contour plot of the radiance of a Lambertian surface with homogeneous reflectivity illuminated by parallel light shown in the gradient space for surface slopes between −1 and 1. The radiance is normalized to the radiance for a flat surface.* **a** *Zero incidence angle $\theta_i = 0°$; the spacing of the contour lines is 0.05.* **b** *Oblique illumination with an incidence angle of 45° and an azimuthal angle of 0°; the spacing of the contour lines is 0.1.*

With a single illumination source, the information about the surface normal is incomplete even if the surface reflectivity is known. Only the component of the surface normal in the direction of the illumination change is given. Thus surface reconstruction with a single illumination source constitutes a complex mathematical problem that will not be considered further here. In the next section we consider how many illuminations from different directions are required to solve the shape from shading problem in a unique way. This technique is known as *photometric stereo*.

8.5.2 *Photogrammetric Stereo*

The curved contour lines in Fig. 8.9 indicate that the relation between surface slope and radiance is nonlinear. This means that even if we take two different illuminations of the same surface (Fig. 8.10), the surface slope may not be determined in a unique way. This is the case when the curved contour lines intersect each other at more than one point. Only a third exposure with yet another illumination direction would make the solution unique.

Using three exposures also has the significant advantage that the reflectivity of the surface can be eliminated by the use of *ratio imaging*. As an example, we illuminate a Lambertian surface with the same light source from three different directions

$$
\begin{aligned}
\boldsymbol{l}_1 &= [0,\ 0,\ 1]^T, \\
\boldsymbol{l}_2 &= [\tan\theta_i,\ 0,\ 1]^T, \\
\boldsymbol{l}_3 &= [0,\ \tan\theta_i,\ 1]^T.
\end{aligned}
\tag{8.14}
$$
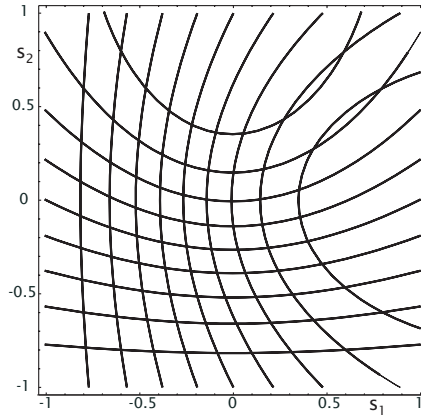
Figure 8.10: Superimposed contour plots of the radiance of a Lambertian surface with homogeneous reflectivity illuminated by a light source with an angle of incidence of 45° and an azimuthal angle of 0° and 90°, respectively.

Then

$$L_2/L_1 = \frac{-s_1 \tan \theta_i + 1}{\sqrt{1 + \tan^2 \theta_i}}, \quad L_3/L_1 = \frac{-s_2 \tan \theta_i + 1}{\sqrt{1 + \tan^2 \theta_i}}. \qquad (8.15)$$

Now the equations are linear in s_1 and s_2 and — even better — they are de-coupled: s_1 and s_2 depend only on L_2/L_1 and L_3/L_1, respectively (Fig. 8.11). In addition, the normalized radiance in Eq. (8.15) does not depend on the reflectivity of the surface. The reflectivity of the surface is contained in Eq. (8.10) as a factor and thus cancels out when the ratio of two radiance distributions of the same surface is computed.

8.5.3   *Shape from Refraction for Specular Surfaces*

For specular surfaces, the shape from shading techniques discussed in Section 8.5.1 do not work at all as light is only reflected towards the camera when the angle of incidence from the light source is equal to the angle of reflectance. Thus, extended light sources are required. Then, it turns out that for transparent specular surfaces, *shape from refraction* techniques are more advantageous than shape from reflection techniques because the radiance is higher, steeper surface slopes can be measured, and the nonlinearities of the slope/radiance relationship are lower.

A shape from refraction technique requires a special illumination technique, as no significant radiance variations occur, except for the small fraction of light reflected at the surface. The base of the shape from refraction technique is the *telecentric illumination system* which converts a spatial radiance distribution into an angular radiance distribution. Then, all we have to do is to compute the relation between the surface slope and the angle of the refracted beam and to use a light source with an appropriate spatial radiance distribution.

Figure 8.12 illustrates the optical geometry for the simple case when the camera is placed far above and a light source below a transparent surface of a medium
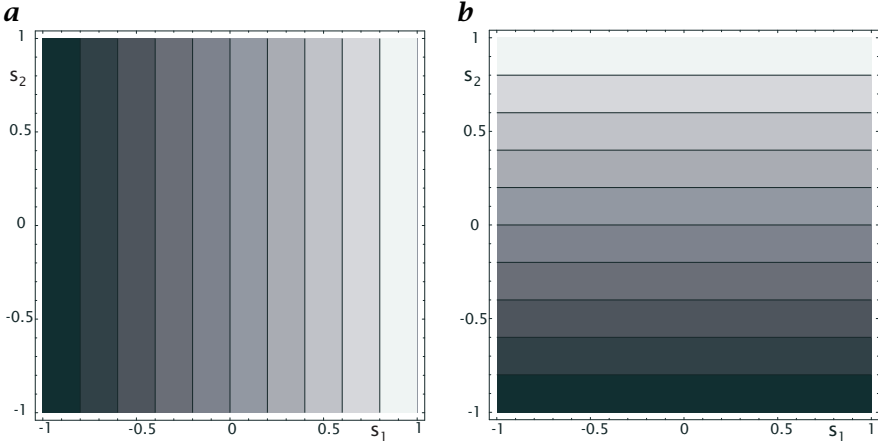
Figure 8.11: *Contour plots of the radiance of a Lambertian surface illuminated by parallel light with an incidence angle of 45° and an azimuthal angle of 0° (**a**) and 90° (**b**), respectively, and normalized by the radiance of the illumination at 0° incidence according to Eq. (8.15). The step size of the contour lines is 0.1. Note the perfect linear relation between the normalized radiance and the x and y surface slope components.*

with a higher index of refraction. The relation between the surface slope s and the angle γ is given by Jähne et al. [97] as

$$s = \tan\alpha = \frac{n\tan\gamma}{n - \sqrt{1 + \tan^2\gamma}} \approx 4\tan\gamma\left(1 + \frac{3}{2}\tan^2\gamma\right) \qquad (8.16)$$

with $n = n_2/n_1$. The inverse relation is

$$\tan\gamma = s\frac{\sqrt{n^2 + (n^2 - 1)s^2} - 1}{\sqrt{n^2 + (n^2 - 1)s^2} + s^2} \approx \frac{1}{4}s\left(1 - \frac{3}{32}s^2\right). \qquad (8.17)$$
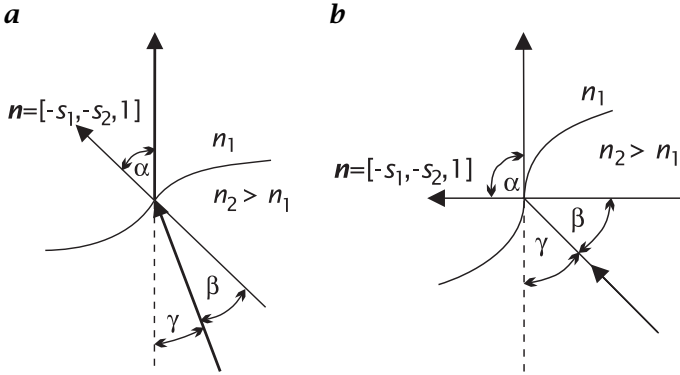
In principle, the shape from refraction technique works for slopes up to infinity (vertical surfaces). In this limiting case, the ray to the camera grazes the surface (Fig. 8.12b) and

$$\tan\gamma = \sqrt{n^2 - 1}. \qquad (8.18)$$

The refraction law thus causes light rays to be inclined in a certain direction relative to the slope of the water surface. If we make the radiance of the light source dependent on the direction of the light beams, the water surface slope becomes visible. The details of the construction of such a system are described by Jähne et al. [97]. Here we just assume that the radiance of the light rays is proportional to $\tan\gamma$ in the x_1 direction. Then we obtain the relation

$$L \propto s_1\frac{\sqrt{n^2 + (n^2 - 1)s^2} - 1}{\sqrt{n^2 + (n^2 - 1)s^2} + s^2}. \qquad (8.19)$$

Of course, again we have the problem that from a scalar quantity such as the radiance no vector component such as the slope can be inferred. The shape

Figure 8.12: Refraction at an inclined surface as the basis for the shape from refraction technique. The camera is far above the surface. **a** Rays emitted by the light source at an angle γ are refracted in the direction of the camera. **b** Even for a slope of infinity (vertical surface, $\alpha = 90°$), rays from the light source meet the camera.
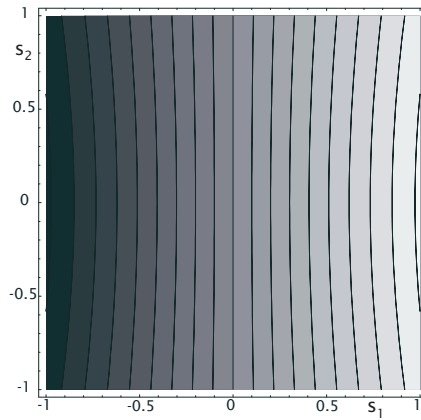
Figure 8.13: Radiance map for the shape from refraction technique where the radiance in a telecentric illumination source varies linearly in the x_1 direction.

from refraction technique, however, comes very close to an ideal setup. If the radiance varies only linearly in the x_1 direction, as assumed, the radiance map in the gradient space is also almost linear (Fig. 8.13). A slight influence of the cross slope (resulting from the nonlinear terms in Eq. (8.19) in s^2) becomes apparent only at quite high slopes.

Ratio imaging can also be used with the shape from refraction technique. Color images have three independent primary colors: red, green, and blue. With a total of three channels, we can identify the position in a telecentric illumination system — and thus the inclination of the water surface — uniquely and still have one degree of freedom left for corrections. With color imaging we also have the

advantage that all three illuminations are taken simultaneously. Thus moving objects can also be observed.

A unique position coding with color can be achieved, for example, with the following color wedges:

$$
\begin{aligned}
G(\boldsymbol{s}) &= (1/2 + cs_1)E_0(\boldsymbol{s}) \\
R(\boldsymbol{s}) &= [1/2 - c/2(s_1 + s_2)]E_0(\boldsymbol{s}) \\
B(\boldsymbol{s}) &= [1/2 - c/2(s_1 - s_2)]E_0(\boldsymbol{s}).
\end{aligned}
\tag{8.20}
$$

We have again assumed a linear relation between one component of the slope and the radiance, with nonlinear isotropic corrections of the form $s_1E_0(\boldsymbol{s})$; c is a calibration factor relating the measured radiance to the surface slope.

We now have three illuminations to determine two slope components. Thus, we can take one to compensate for unwanted spatial variation of E_0. This can be done by normalizing the three color channels by the sum of all channels $G + R + B$:

$$
\begin{aligned}
\frac{G}{G + R + B} &= \frac{2}{3}\left(\frac{1}{2} + cs_1\right), \\
\frac{B - R}{G + R + B} &= \frac{2}{3}cs_2.
\end{aligned}
\tag{8.21}
$$

Then the position on the wedge from which the light originates is given as

$$
s_1 = \frac{1}{2c}\frac{2G - R - B}{G + R + B}, \quad s_2 = \frac{3}{2c}\frac{B - R}{G + R + B}.
\tag{8.22}
$$

From these position values, the x and y components of the slope can be computed according to Eq. (8.19).

8.6 *Depth from Multiple Projections: Tomography*

8.6.1 *Principle*

Tomographic methods do not generate a 3-D image of an object directly, but allow reconstruction of the 3-D shape of objects using suitable methods. Tomographic methods can be considered as an extension of stereoscopy. With stereoscopy only the depth of surfaces can be inferred, but not the 3-D shape of transparent objects. Intuitively, we may assume that it is necessary to view such an object from as many directions as possible.

Tomographic methods use radiation that penetrates an object from different directions. If we use a point source (Fig. 8.14b), we observe a perspective or *fan-beam projection* on the screen behind the object just as in optical imaging (Section 7.3). Such an image is taken from different projection directions by rotating the point source and the projection screen around the object. In a similar way, we can use parallel projection (Fig. 8.14a) which is easier to analyze but harder to realize. If the object absorbs the radiation, the intensity loss measured in the projection on the screen is proportional to the path length of the ray in the object. The 3-D shape of the object cannot be reconstructed from
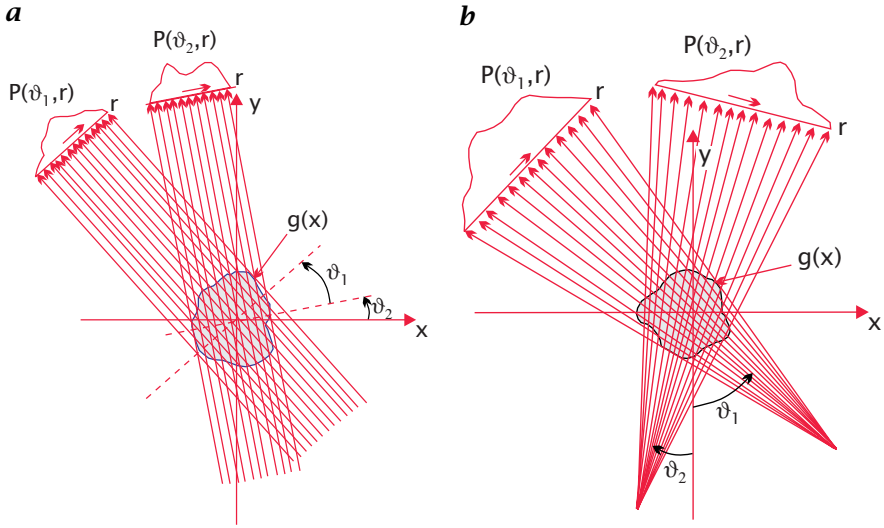
Figure 8.14: a *Parallel projection and* **b** *fan-beam projection in tomography.*

one projection. It is necessary to measure projections from all directions by turning the radiation source and projection screen around the object.

As in other imaging methods, tomography can make use of different interactions between matter and radiation. The most widespread application is *transmission tomography*. The imaging mechanism is by the absorption of radiation, e. g., x-rays. Other methods include emission tomography, reflection tomography, and time-of-flight tomography (especially with ultrasound), and complex imaging methods using *magnetic resonance* (*MR*).

### 8.6.2   *Radon Transform and Fourier Slice Theorem*

With respect to reconstruction, it is important to note that the projections under all the angles ϑ can be regarded as another 2-D representation of the image. One coordinate is the position in the projection profile, r, the other the angle ϑ (Fig. 8.15). Consequently, we can regard the parallel projection as a transformation of the image into another 2-D representation. Reconstruction then just means applying the inverse transformation. The critical issue, therefore, is to describe the tomographic transform mathematically and to investigate whether the inverse transform exists.

A projection beam is characterized by the angle ϑ and the offset r (Fig. 8.15). The angle ϑ is the angle between the projection plane and the x axis. Furthermore, we assume that we slice the 3-D object parallel to the xy plane. Then, the scalar product between a vector $\boldsymbol{x}$ on the projection beam and a unit vector

$$\bar{\boldsymbol{n}} = [\cos\vartheta, \sin\vartheta]^T \tag{8.23}$$

normal to the projection beam is constant and equal to the offset r of the beam

$$\boldsymbol{x}\bar{\boldsymbol{n}} - r = x\cos\vartheta + y\sin\vartheta - r = 0. \tag{8.24}$$
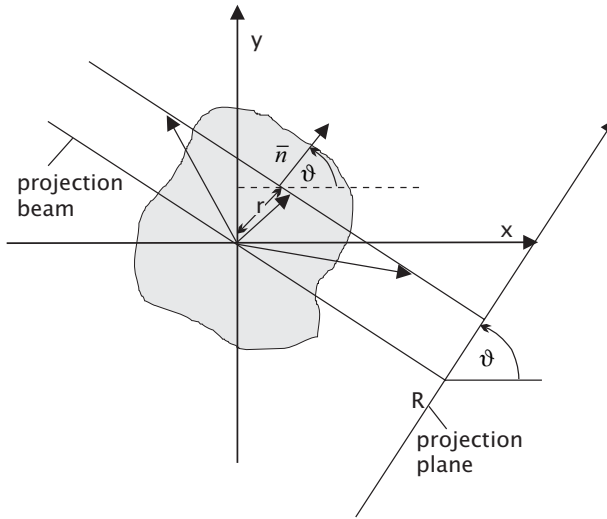
Figure 8.15: *Geometry of a projection beam.*

The projected intensity $P(r, \vartheta)$ is given by integration along the projection beam:

$$P(r, \vartheta) = \int\limits_{\text{path}} g(\boldsymbol{x}) \mathrm{d}\boldsymbol{s} = \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g(\boldsymbol{x}) \delta(x_1 \cos \vartheta + x_2 \sin \vartheta - r) \mathrm{d}^2 x. \qquad (8.25)$$

The δ distribution in this equation reduces the double integral to a projection beam in the direction ϑ that has a distance r from the center of the coordinate system. The projective transformation of a 2-D function $g(\boldsymbol{x})$ onto $P(r, \vartheta)$ is named after the mathematician Radon as the *Radon transform*.

To better understand the properties of the Radon transform, we analyze it in the Fourier space. The Radon transform can be understood as a special case of a linear shift-invariant filter operation, the *projection operator*. All gray values along the projection beam are added up. Therefore the point spread function of the projection operator is a δ line in the direction of the projection beam. In the Fourier domain this convolution operation corresponds to a multiplication with the transfer function, which is a δ line (2-D) or δ plane (3-D) normal to the δ line in the spatial domain (see ≻R5). In this way, the projection operator slices a line or plane out of the spectrum that is perpendicular to the projection beam.

This elementary relation can be computed most easily, without loss of generality, in a rotated coordinate system in which the projection direction coincides with the y' axis. Then the r coordinate in $P(r, \vartheta)$ coincides with the x' coordinate and ϑ becomes zero. In this special case, the Radon transform reduces to an integration along the y' direction:

$$P(x', 0) = \int\limits_{-\infty}^{\infty} g(x', y') \mathrm{d}y'. \qquad (8.26)$$

The Fourier transform of the projection function can be written as

$$\hat{P}(k_{x'}, 0) = \int\limits_{-\infty}^{\infty} P(x', 0) \exp(-2\pi i k_{x'} x') dx'. \tag{8.27}$$

Replacing $P(x', 0)$ by the definition of the Radon transform, Eq. (8.26) yields

$$\hat{P}(k_{x'}, 0) = \int\limits_{-\infty}^{\infty} \left[\int\limits_{-\infty}^{\infty} g(x', y') dy' \right] \exp(-2\pi i k_{x'} x') dx'. \tag{8.28}$$

If we insert the factor $\exp(-2\pi i 0 y') = 1$ in this double integral, we recognize that the integral is a 2-D Fourier transform of $g(x', y')$ for $k_{y'} = 0$:

$$\begin{aligned} \hat{P}(k_{x'}, 0) &= \int\limits_{-\infty}^{\infty} \int\limits_{-\infty}^{\infty} g(x', y') \exp(-2\pi i k_{x'} x') \exp(-2\pi i 0 y') dx' dy' \\ &= \hat{g}(k_{x'}, 0). \end{aligned} \tag{8.29}$$

Back transformation into the original coordinate system finally yields

$$\hat{P}(q, \vartheta) = \hat{g}(\boldsymbol{k}) \delta(\boldsymbol{k} - (\boldsymbol{k}\bar{\boldsymbol{n}})\bar{\boldsymbol{n}}), \tag{8.30}$$

where q is the coordinate in the k space in the direction of ϑ and $\bar{\boldsymbol{n}}$ the normal vector introduced in Eq. (8.23). The spectrum of the projection is identical to the spectrum of the original object on a beam normal to the direction of the projection beam. This important result is called the *Fourier slice theorem* or *projection theorem*.

8.6.3 *Filtered Back-Projection*

If the projections from all directions are available, the slices of the spectrum obtained cover the complete spectrum of the object. Inverse Fourier transform then yields the original object. Filtered back-projection uses this approach with a slight modification. If we just added the spectra of the individual projection beams to obtain the complete spectrum of the object, the spectral density for small wave numbers would be too high as the beams are closer to each other for small radii. Thus, we must correct the spectrum with a suitable weighting factor. In the continuous case, the geometry is very easy. The density of the projection beams goes with $|\boldsymbol{k}|^{-1}$. Consequently, the spectra of the projection beams must be multiplied by $|\boldsymbol{k}|$. Thus, filtered back-projection is a two-step process. First, the individual projections must be filtered before the reconstruction can be performed by summing up the back-projections.

In the first step, we thus multiply the spectrum of each projection direction by a suitable weighting function $\hat{w}(|\boldsymbol{k}|)$. Of course, this operation can also be performed as a convolution with the inverse Fourier transform of $\hat{w}(|\boldsymbol{k}|)$, $w(r)$. Because of this step, the procedure is called the *filtered back-projection*.

In the second step, the back-projection is performed and each projection gives a slice of the spectrum. Adding up all the filtered spectra yields the complete spectrum. As the Fourier transform is a linear operation, we can add up the

filtered projections in the space domain. In the space domain, each filtered projection contains the part of the object that is constant in the direction of the projection beam. Thus, we can back-project the corresponding gray value of the filtered projection along the direction of the projection beam and add it up to the contributions from the other projection beams.

After this illustrative description of the principle of the filtered back-projection algorithm we derive the method for the continuous case. We start with the Fourier transform of the object and write the inverse Fourier transformation in polar coordinates (q, ϑ) in order to make use of the Fourier slice theorem

$$g(\boldsymbol{x}) = \int\limits_{0}^{2\pi} \int\limits_{0}^{\infty} q\hat{g}(q, \vartheta) \exp[2\pi i q(x_1 \cos\vartheta + x_2 \sin\vartheta)]\mathrm{d}q\mathrm{d}\theta. \qquad (8.31)$$

In this formula, the spectrum is already multiplied by the wave number, q. The integration boundaries, however, are not yet correct to be applied to the Fourier slice theorem (Eq. (8.30)). The coordinate, q, should run from $-\infty$ to ∞ and ϑ only from 0 to π. In Eq. (8.31), we integrate only over half a beam from the origin to infinity. We can compose a full beam from two half beams at the angles ϑ and $\vartheta + \pi$. Thus, we split the integral in Eq. (8.31) into two over the angle ranges $[0, \pi[$ and $[\pi, 2\pi[$ and obtain

$$\begin{aligned} g(\boldsymbol{x}) \quad &= \quad \int\limits_{0}^{\pi} \int\limits_{0}^{\infty} q\hat{g}(q, \vartheta) \exp[2\pi i q(x_1 \cos\vartheta + x_2 \sin\vartheta)]\mathrm{d}q\mathrm{d}\vartheta \\[2mm] &+ \quad \int\limits_{0}^{\pi} \int\limits_{0}^{\infty} q\hat{g}(-q, \vartheta') \exp[-2\pi i q(x_1 \cos\vartheta' + x_2 \sin\vartheta')]\mathrm{d}q\mathrm{d}\vartheta' \end{aligned}$$

using the following identities:

$$\vartheta' = \vartheta + \pi, \; \hat{g}(-q, \vartheta) = \hat{g}(q, \vartheta'), \; \cos(\vartheta') = -\cos(\vartheta), \; \sin(\vartheta') = -\sin(\vartheta).$$

Now we can recompose the two integrals again, if we substitute q by $-q$ in the second integral and replace $\hat{g}(q, \vartheta)$ by $\hat{P}(q, \vartheta)$ because of the Fourier slice theorem Eq. (8.30):

$$g(\boldsymbol{x}) = \int\limits_{0}^{\pi}\int\limits_{-\infty}^{\infty} |q|\hat{P}(q, \vartheta) \exp[2\pi i q(x_1 \cos\vartheta + x_2 \sin\vartheta)]\mathrm{d}q\mathrm{d}\vartheta. \qquad (8.32)$$

Equation (8.32) gives the inverse Radon transform and is the basis for the *filtered back-projection* algorithm. The inner integral performs the back-projection of a single projection:

$$P' = \mathcal{F}^{-1}(|q|\mathcal{F}P). \qquad (8.33)$$

$\mathcal{F}$ denotes the 1-D Fourier transform operator. P' is the projection function P multiplied in the Fourier space by $|\boldsymbol{q}|$. If we perform this operation as a convolution in the space domain, we can formally write

$$P' = [\mathcal{F}^{-1}(|q|)] * P. \qquad (8.34)$$

The outer integral in Eq. (8.32) over the angle ϑ,

$$g(\boldsymbol{x}) = \int_0^\pi P'(r, \vartheta)\, \mathrm{d}\vartheta, \tag{8.35}$$

sums up the back-projected and filtered projections over all directions and thus forms the reconstructed image. Note that the filtered projection profile $P'(r, \vartheta)$ in Eq. (8.35) must be regarded as a 2-D function to built up a 2-D object $g(\boldsymbol{x})$. This means that the projection profile is projected back into the projection direction.

8.6.4 *Discrete Filtered Back-Projection*

There are several details we have not yet discussed that cause serious problems for the reconstruction in the infinite continuous case. First, we observe that it is impossible to reconstruct the mean of an object. Because of the multiplication by $|\boldsymbol{k}|$ in the Fourier domain (Eq. (8.32)), $\hat{g}(\boldsymbol{0})$ is eliminated. Second, it is altogether impossible to reconstruct an object of infinite size, as any projection beam will result in infinite values.

Fortunately, all these difficulties disappear where we turn from the infinite continuous case to the finite discrete case where the objects are of limited size. In practice, the size limit is given by the distance between the radiation source and the detector. The resolution of the projection profile is limited by the combined effects of the extent of the radiation source and the resolution of the detector array in the projection plane. Finally, we can only take a limited number of projections. This corresponds to a sampling of the angle ϑ in the Radon representation of the image.

We illustrate the discussion in this section with an example. We can learn much about projection and reconstruction by considering the reconstruction of the simplest object, a point, because the Radon transform (Eq. (8.25)) and its inverse are linear transforms. Then, the projections from all directions are equal (Fig. 8.16a) and show a sharp maximum in the projection functions $P(r, \vartheta_i)$. In the first step of the filtered back-projection algorithm, P is convolved with the $|k|$ filter. The result is a modified projection function P' which is identical to the point spread function of the $|k|$ filter (Fig. 8.16b).

In a second step, the back-projections are added up in the image. From Fig. 8.16c, we can see that at the position of the point in the image the peaks from all projections add up. At all other positions in the images, the filtered back-projections are superimposed on each other in a destructive manner, because they show negative and positive values. If the projection directions are sufficiently close to each other, they cancel each other out except for the point at the center of the image. Figure 8.16c also demonstrates that an insufficient number of projections leads to star-shaped distortion patterns.

The simple example of the reconstruction of a point from its projections is also useful to show the importance of filtering the projections. Let us imagine what happens when we omit this step. Then, we would add up δ lines as back-projections which rotate around the position of the point. Consequently, we would not obtain a point but a rotation-symmetric function that falls off with $|\boldsymbol{x}|^{-1}$. As a result, the reconstructed objects would be considerably blurred.
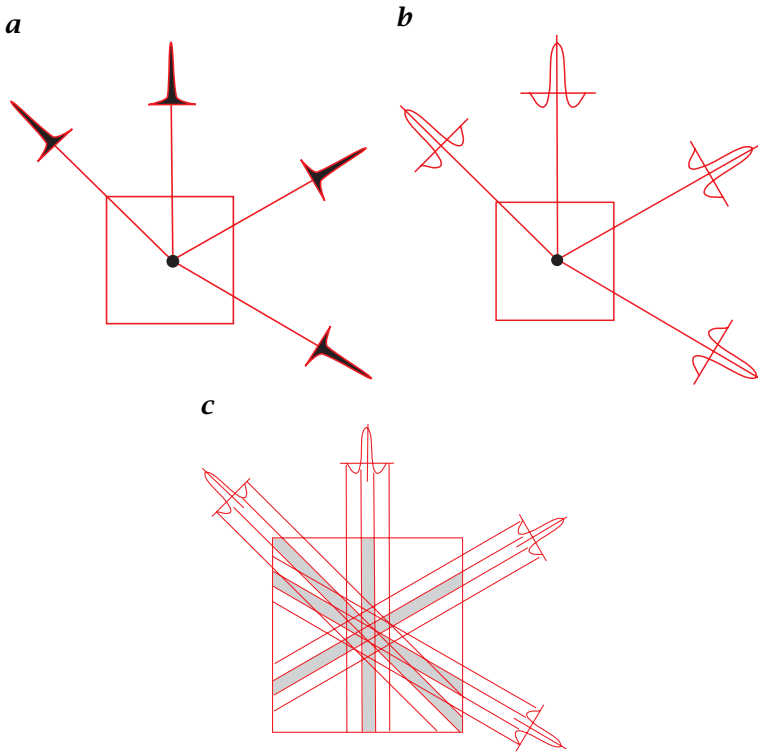
*Figure 8.16: Illustration of the filtered back-projection algorithm with a point object: **a** projections from different directions; **b** filtering of the projection functions; **c** back-projection: adding up the filtered projections.*

8.7   *Exercises*

8.1: **Stereoscopy**

Interactive demonstration of the reconstruction of depth maps from stereo images (dip6ex08.01).

8.2: *Human stereo vision

Estimate how well the human vision system can estimate depth. Assume the focal length of the eye to be 17 mm and the stereo basis to be 65 mm. Answer the following questions:

1. At which distance is the parallax equal to the spatial resolution of the eye? Assume that the eye is a diffraction-limited optical system (Section 7.6.3) with an aperture of 3 mm.
2. How large is the standard deviation of the depth estimate at 0.5 m and 5 m distance if we assume that the standard deviation of the measurement of the parallax is a quarter of the spatial resolution of the eye?

8.3: Depth form focus

Interactive demonstration of the reconstruction of images with large depth of field und of depth maps from focus series (dip6ex08.02).

8.4: Tomography

Interactive demonstration of the radon transform and the tomographic reconstruction using the filtered backprojection; demonstration of reconstruction artifacts (dip6ex08.03).

8.5: **Artifacts with tomography

In practical applications it is often required to carry out a tomography with as few as possible projections. Imagine that the angle intervals become larger and larger. Discuss what happens by using a point object with Gaussian shape and the standard deviation σ:

1. When do artifact commence and how do they look like?
2. Where do these artifacts occur first?
3. What do you conclude from these observations: is the resolution of a tomographic system position independent?

8.6: * Tomography with few projections**

With special classes of objects, it is possible to apply tomographic techniques with only a few projections. Examine the following examples and determine how many projections are required for a complete reconstruction:

1. An arbitrary rotationally symmetric object.
2. An arbitrarily formed object without holes (only one surface) consisting of a homogeneous material.
3. Few small objects that do not superimpose each other in any projection. You only want to determine the center of gravity of thesse objects and their volume.

8.8   *Further Readings*

A whole part with seven chapters of the "Handbook of Computer Vision and Applications" is devoted to 3-D imaging [94, Vol. I, Part IV]. Klette et al. [109] discusses 3-D computer vision focusing on stereo, shape from shading, and photometric stereo.