Image formation

1 Optics and imaging systems

Optical systems produce two effects on an image: projection and degradation due to the effects of diffraction and lens aberrations. Physical optics provides the tools to describe degradation due to

- the wave nature of light
- the aberrations of imperfectly designed and manufactured optical systems.

The following figure shows a simple optical system:



A point-source at the origin of the focal (object) plane produces a spot image at the origin of the image plane. The image produced by a point-source is called the **point-spread function** (PSF). For a high-quality lens the PSF, though not a impulse, is nonzero only over a small region. It takes on its smallest size when the system is in focus, namely when

$$\frac{1}{d_f} + \frac{1}{d_i} = \frac{1}{f},$$

where f is the focal length of the lens. The focal plane is the plane in the object space that forms an in-focus image in the image plane.

If the point source moves off-axis to a position (x_0, y_0) , then the spot image moves to a new position given by

$$x_i = -Mx_0, \qquad y_i = -My_0,$$

where $M = d_i/d_f$ is the magnification.

1.1 Linearity

Increasing the intensity of the point source causes a proportional increase in the intensity of the spot image. Thus the lens is a 2-D linear system: two point sources produce an image in which the two spots combine by addition.

An opaque object in the scene can be thought of as a 2-D distribution of point sources of light. The image of the object is a summation of spatially distributed PSF spots.

1.2 Shift invariance

For reasonably small off-axis distances in good optical systems, the shape of the spot undergoes essentially no change. Thus the system can be assumed to be shift invariant (or **isoplanatic**).

To a good approximation, an optical imaging system is therefore a 2-D shift-invariant linear system. The point-spread function is then the impulse response, and the image (after projection) can be described as a convolution of the object with the PSF of the system.

The counterpart in the frequency domain to the PSF is the two-dimensional **optical transfer function** (OTF), which is the Fourier transform of the point-spread function. In the context of linear systems theory the OTF therefore plays the role of the transfer function of the system. The **modulation transfer function** (MTF) is the modulus or magnitude of the OTF; high-quality lenses are designed to be phaseless, so the OTF and the MTF are equivalent.

1.3 Diffraction-limited optical systems

Since an optical system is essentially LSI, it can be completely described by either the PSF or the transfer function.

A point source in the focal plane will produce an expanding spherical wave, part of which enters the lens. The refractive action of the lens slows and delays axial rays more near the centre of the lens than at the edges, converting the expanding spherical wave into another spherical wave converging toward the image point.



Pupil plane

Using the Huygens-Fresnel principle the PSF of an optical system can be derived. It is useful to define the **pupil function** as the function that takes on a value of one inside the aperture, and zero outside. There are two cases:

• **Coherent illumination:** If the point sources being imaged vary in synchrony, then stable patterns of constructive and destructive interference exist. The PSF can then be shown to be equivalent to the (scaled) Fourier transform of the pupil function:

$$h(x, y) = \mathcal{F}\{p(\lambda d_i x_a, \lambda d_i y_a)\}.$$

The optical system is then linear in complex amplitude.

• Incoherent illumination: If the illumination is incoherent, and varies randomly from one point to another, then the system is linear in **intensity**, and the PSF is the squared modulus of *h*(*x*, *y*), the coherent PSF. Thus the incoherent PSF is the power spectrum of the pupil function.

For a lens with a *circular* aperture of diameter a in narrow-band incoherent

light of wavelength λ , the PSF is

$$h(r) = \left[2\frac{J_1(\pi[r/r_0])}{\pi[r/r_0]}\right]^2.$$

The constant r_0 is

$$r_0 = \lambda d_i / a$$

and r is the radial distance measured from the optical axis in the image plane

$$r = \sqrt{x_i^2 + y_i^2}.$$

The relevant imaging quantities are shown below, both for the case of a circular and a rectangular aperture:



Note that as the aperture size *a* increases, the PSF becomes narrower. This allows objects to be imaged with higher resolution, and is (part of) the reason for telescopes having such a large diameter. (They also have a large aperture to capture a larger portion of the incoming light.) Synthetic aperture techniques also make use of this property by synthesising large apertures by using many spatially separated imaging systems to increase the effective aperture.

Imaging systems may also exhibit **aberrations** which cause the exit wave to depart from its ideal spherical shape. Common aberrations are defocus,

astigmatism, coma, and image distortion. These are due to imperfect lenses.



Note also that the equations given depend on the wavelength λ of the light. For polychromatic imaging chromatic aberration can also occur. Good lenses make use of many additional elements (using lenses with different compositions) to try to minimise this effect.

2 Charge-coupled devices

The predominant method of sampling a 2-D distribution of light intensity is by means of a charge-coupled device (CCD). CCDs are manufactured on a light-sensitive crystalline silicon chip. A rectangular array of photodetector sites (potential wells) is built into the silicon substrate.



As long as the CCD is exposed to light, charge accumulates in the potential wells. Apart from any effects due to saturation, the charge is proportional to the number of light photons incident on the specific location on the CCD.

In the simplest case (the full-frame CCD), readout is performed by shuttering the CCD. The entire array of charges is then clocked downward by one position, and the bottom line of charges enters a serial register. This register is in turn read out by clocking the pixel charges out one at a time into an accumulator, where a digital readout of the magnitude of the charge is obtained. The entire array of charges is then clocked down one position further, and the next line read out. Once all the lines have been processed, the device is ready to integrate another image.

CCDs can be scanned at television rates (25 frames per second) or much more slowly. Since they can integrate for periods of seconds to hours to create low-light images, they are used in astronomy and florescence microscopy, for example. The long integration times require that the sensor be cooled to reduce dark current effects. CCDs exhibit *readout noise*, generated by the on-chip electronics, and *photon noise* resulting from the quantum nature of light.

3 Perspective projection

A camera produces a 2-D representation of a 3-D scene.



The mapping from 3-D to 2-D is well-defined, but the operation cannot in general be reversed. In many applications it is important to be able to relate points in 3-D to their corresponding points in 2-D, so that inferences regarding the 3-D scene can be made from image data.

Using the standard coordinate system



the relationship between image coordinates (u, v) and 3-D space coordinates (x, y, z) can be written as

$$u = f\frac{x}{z}, \qquad v = f\frac{y}{z}$$

This can be rewritten linearly as

$$\begin{pmatrix} U \\ V \\ S \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

where u = U/S, v = V/S if $s \neq 0$. Further defining the projective coordinates x = X/T, y = Y/T, and z = Z/T the entire system can be written as a linear equation in projective coordinates

$$\begin{pmatrix} U \\ V \\ S \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ T \end{pmatrix}$$

A camera may therefore be considered as a system that performs a linear projective transformation from the projective space \mathcal{P}^3 into the projective plane \mathcal{P}^2 . This constitutes an **affine** projection in rectangular coordinates.



A strongly calibrated 3-D environment can allow predictions to be made regarding the appearance of 3-D objects in a 2-D image. For example, for person tracking research in the DIP laboratory at UCT it became necessary to quantify the appearance of (elliptical) people in a 2-D camera view. This was done by finding the parameters of a suitable affine projection from 3-D to 2-D:



Using one of the properties of affine projections, namely that the projection of an ellipsoid is an ellipse, this allowed the appearance of people to be modelled:



Note that distortion also had to be accommodated to make this relationship accurate, due to poor lenses in the cameras used.

4 Computer vision

The field of computational **computer vision** often makes use of detailed information regarding the image formation process (both geometric and optic) to make inferences regarding objects in the real world. In contrast to traditional image processing techniques, computer vision usually involves **camera calibration** to make the relationship between objects in the world and their images precise.

This is important for many applications:

• **Stereo imaging** techniques attempt to build up 3-D models from multiple views of the scene. One needs to know how the views of many cameras

correspond to one another in order to achieve this.

- **Depth from defocus** methods construct a 3-D model of a scene by making use of the fact that the degree of defocus of an object in the scene is directly proportional to the distance of the object from the focal plane.
- **Structure from motion** takes multiple images of a moving object from the *same* camera, and constructs a 3-D model of the object.
- Shape from shading constructs a 3-D model of an object from the variation in shading across the surface of the object.

These are all active research areas, and new methods and applications are being developed all the time. These methods are highly relevant in the field of robotics and machine vision.