

Imaging Optics

Yuhao Zhu

Department of Computer Science
Department of Brain and Cognitive Sciences
University of Rochester

yzhu@rochester.edu
<https://yuhaozhu.com/>
<https://horizon-lab.org/>

Contents

1	Overview	2
2	Pinhole Model	3
2.1	(Why) Do We Need Optics in an Imaging System?	3
2.2	Pinhole Imaging	4
2.3	Diffraction Limit	5
3	Lenses and Aberrations	7
3.1	Image Formation with an Ideal Lens	7
3.2	Magnification vs. Field-of-View	11
3.3	Magnifying Glasses and Projection Lenses	12
3.4	Depth of Field	14
3.5	Aberration	15
3.6	Not All Blurs are Created Equal	20
3.7	Radiometric Analysis of Lens	21
4	Computational Modeling Using Linear System Theory	21
4.1	Basic Idea of a Linear System Theory	22
4.2	Modeling Image Formation in LSI Systems	23
4.3	Fourier Perspectives: OTF and MTF	24

1 Overview

This chapter focuses the first stage in an imaging system: the optics, i.e., optics that are used for image formation. The goal of this chapter is to build a good understanding of the image formation process in optics. Optics manipulates/transforms optical signals, so the signal after optics is still in the optical domain. In later chapters, we will discuss how the optical signals are transformed into electrical signals (first to analog and then to digital signals) and how such electrical signals are further processed.

Imaging optics is important for human vision (because the ocular media of our eyes form an image on the retina), cameras (almost all of which have some form of optics), and graphics (where modeling optics is important for photorealistic rendering). Optics can also be used for ostensibly non-imaging purposes such as communication and computation. But of course the distinction is not black and white. You can argue that imaging is simultaneous communication (transferring signals from one side of the imaging system to the other side) and computation (the output signal is the result of a transfer function, usually not an identity function, applied to the input signal).

We will generally assume that the goal of the optics design is form visually pleasing images as well as possible. The thinking is that if we provide a high-quality image, we are giving the downstream consumer, whether a human observer or a machine vision algorithm, the best chance to extract information from it.

This might not always be necessary. For instance, in machine vision/robotics applications the consumer of an image is a computer vision algorithm such as object detection; so long as the algorithm can detect the object the quality of the image itself is of no significance. In fact, one might argue that it is beneficial to design the imaging system so that output image is obfuscated to protect privacy as long as essential features pertaining to downstream algorithms are preserved — this is an active area of research.

There is also a burgeoning area of research, which this chapter is largely unconcerned with, called **computational imaging**, where a significant amount of computation is involved to form a final image [Bhandari et al., 2022]. In many cases under such a paradigm, the initial image form by the optics is rather unintelligible, and the name of the game is to design computational algorithms that can recover the “clean” image. This is usually formulated as an **inverse problem**: the optics (which could be anything, even a duct tape [Antipa et al., 2017]) transforms information in the physical world into a set of observations, and the algorithm inverts that forward model to obtain the original physical information from the observations. Even in this case, understanding and modeling the forward image formation process of the optics is crucial: only with that knowledge can we invert that process to obtain the physical information. In fact, one usually *co-designs* the image formation process (e.g., optics) with the inversion algorithm to maximize the overall performance.

In this sense, imaging is a form of sensing and the ultimate goal of imaging is to obtain information about the physical world. A visually pleasing image is one way such information can be represented, but there are other forms of information we might be interested: depth, geometry, spectral radiance, polarization, absorption/scattering coefficient of the media, etc.

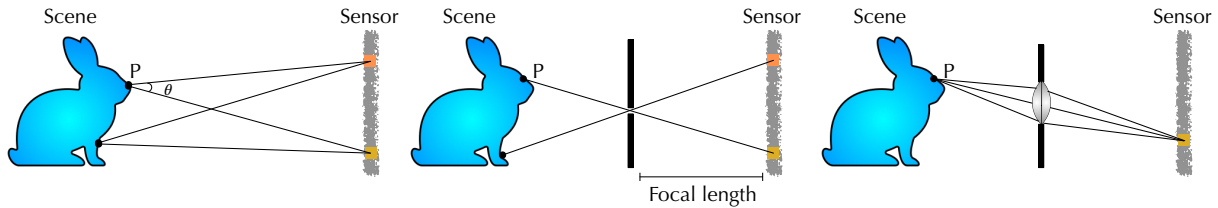


Figure 1: Left: without a pinhole different pixels get roughly the same signal and the differences between pixels are due mostly to noise. Middle: with a pinhole, the signal received by each pixel is restricted to a small area in the physical scene. Right: with a lens, the signal received by each pixel is still restricted to a small physical area but the signal is much stronger.

Many imaging systems are designed to obtain such non-visual information, which is beyond our scope. For instance, an X-ray CT scanner is an essentially computational imaging device; it captures a set of raw images, which by themselves are not directly useful. Subsequent computational algorithms are used to obtain the actual information of interest, the absorption/scattering coefficient of the medium, from the raw images. We have actually covered the gist of the forward process in this imaging device when we discuss volume scattering.

We will assume that there is a sensor plane on the other side of the imaging system to capture observations. An actual sensor has many pixels (along with many other components, some of which are optics!), each of which has a small, but non-zero size, which plays a role in signal processing. In this chapter, however, we will assume that each pixel is infinitesimal. Therefore, the image formed on the sensor plane, for now, is assumed to be a continuous 2D function: for any (x, y) point on the sensor, there is an irradiance value. A retina is a sensor, and the continuous image on the retina usually called the **optical image** in vision science. An actual image captured by the sensor, whether biological (retina) or engineered, is necessarily discretized — by pixels or photoreceptors.

2 Pinhole Model

We will start from discussing the pinhole system, which is very simple and not commonly used but carries interesting properties and implications for more complex imaging systems that we will turn to later.

2.1 (Why) Do We Need Optics in an Imaging System?

What if we just expose the image sensor or our retina to lights? We will get garbage, because each pixel/photoreceptor receive lights from everywhere in the scene. Figure 1 (left) illustrates the geometry of this imaging system.

Each pixel receives lights from everywhere in space. Assuming that each point in space is an ideal Lambertian emitter/scatterer, the two highlighted pixels will receive slightly different energy from the same point, because of the cosine fall-off as a function of the incident direction.

But if the sensor is much smaller relative to the distance to the physical space, the differences in fall-offs between different pixels are small, so we can say that each pixel roughly receives the same energy. In this case, the differences in pixel values are due to noise. So that's why the image looks like a random garbage.

2.2 Pinhole Imaging

What we need is for each pixel to receive information only from a small spatial region in the scene. This is what a pinhole camera does, as illustrated in Figure 1 (middle). If the pinhole is infinitesimally small such that it allows only a single ray direction to go through, each pixel (which, again, for now is assumed to be an infinitesimal point on the sensor plane) captures lights from only a single point in the scene.

As the pinhole size shrinks, the information captured by two adjacent pixels becomes more distinct, which is desirable, but if the pinhole size is too small, there are two issues. First, a pinhole that is too small requires a long exposure time. We will discuss this in the image sensor lecture, but a pixel is very much like a photoreceptor in that it is a photon counting device. Intuitively, the amount of photons a pixel collects (which we care about because it relates to the brightness of the captured image) is, roughly, proportional to both the pinhole area and the exposure time, so if we reduce the pinhole size we need to increase the exposure time to maintain the pixel brightness.

An excessively long exposure time not only poses challenges to actually taking the photo, but also leads to **motion blurs**. Figure 2 (b) shows an image captured by a pinhole camera where, during exposure, objects are moving. As a result, each pixel receives lights from different points in the scene and, visually, the resulting image carries motion blurs.

Second, as the pinhole size gets smaller and smaller eventually we get to the diffraction limit, which means we cannot use geometric optics any more and a single point in the scene does not translate to a single point in the image plane. We will discuss this shortly in Chapter 2.3.

What happens if we increase the pinhole size? We get a blurrier image. Figure 2 (a) shows one such image captured by a pinhole camera using a pinhole size of 0.5 mm. The blur can be easily explained by the geometry of pinhole imaging, as shown in Figure 2 (c), where the information of a point in the scene is spread or "smeared" across multiple pixels if the pinhole size is too large, leading to the blurs. For this reason, the blur here is a form of **defocus blur**; we will later see how a lens-based imaging system can also have a defocus blur with the same mechanism: information at a physical point in the scene is spread across multiple pixels even when the point itself is stationary.

Even in a lens-based imaging system, we do not technically have a pinhole but we usually still have an aperture, which acts like a pinhole in the sense that it limits the amount of light that is allowed into the rest of the system, so the aperture size certainly dictates the imaging quality. Our eye is certainly a lens-based imaging system, and the pupil acts as the aperture. The pupil sizes changes from roughly 2 mm in relatively high ambient light levels to about 8 mm under low light intensities.

Amazingly, pinhole-only imaging is used in some animals. The most famous one is perhaps Nautilus, which has a pinhole eye without lenses [Zhang et al., 2021]. The pinhole size is



Figure 2: (a): blur from large aperture/pinhole (defocus blur); from [Dave Haynie \[2019\]](#). (b): blur from long exposure (motion blur); from [Daniel King \[2017\]](#). (c): defocus blur arises from large pinholes, but when the pinhole becomes very small (on the order of light wavelength) diffraction occurs; adapted from [Dominic Alves \[2006\]](#).

relatively large; the diameter varies between 0.4 to 2.8 mm [[Hurley et al., 1978](#)], so you can imagine the imaging quality is not great.

2.3 Diffraction Limit

When the pinhole becomes very small, diffraction becomes visible. The diffraction pattern is called the Airy disk. Figure 3 (left) shows a computer-simulated airy disk, and Figure 3 (right) shows how the intensity of the airy disk falls off from the center.

Diffraction is usually thought of as a wave phenomenon, where the light wave propagated from a small pinhole gets expanded spatially and form the airy disk pattern. But perhaps a more principled way to understand diffraction is through quantum mechanics, which says that the more certain we are of the position of a photon we are less certain of the direction of its travel, and vice versa. When the pinhole is infinitesimal we know for certain where a photon is so we are uncertain where it is going to go: the result is the airy disk pattern. In contrast, when the pinhole is large, we are less certain of the spatial position of a photon so we are more certain of its direction of travel; as a result, diffraction contributes little to the overall imaging.

Imaging through a small pinhole can be thought of as a “single-slit” experiment. When we have a “double-slit” experiment with two small pinholes, the diffraction patterns from the two pinholes interfere, and we get the beautiful interference pattern that you perhaps have seen in middle-school physics class. Interestingly, there is a sequential version of the double-slit experiment, where photons are sent to the two slits sequentially, one by one. Amazingly, if we wait long enough we will still see the interference pattern. This firmly establishes the fact that lights do behave like particles, not waves, just in a probabilistic manner.

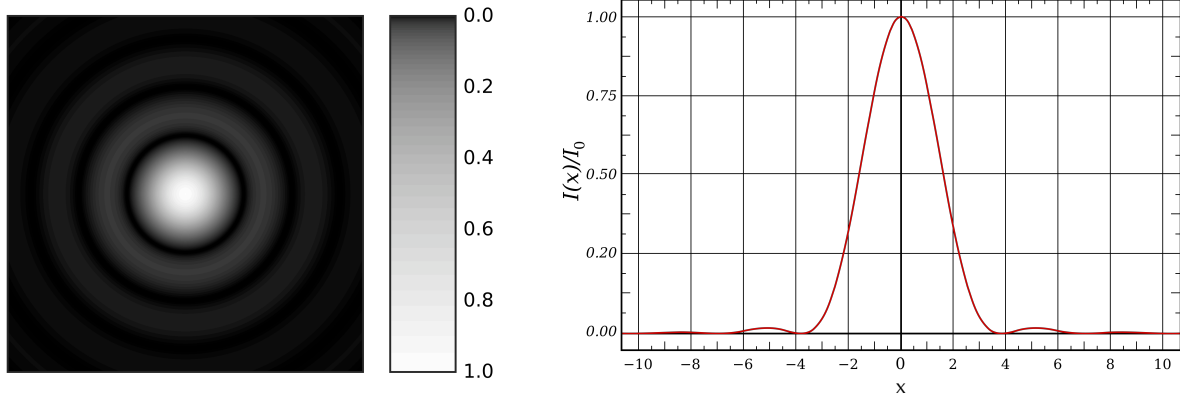


Figure 3: Left: compute-simulated airy disk (contrast is slightly exaggerated); from [Sakurambo \[2007\]](#). Right: the intensity of an airy disk pattern as a function of the spatial position (0 being the center); adapted from [Inductiveload \[2009\]](#).

Theoretical Maximum Resolving Power

Diffraction is a form of blur, because the optical power of a power in the scene is spread spatially on the detector plane. Therefore diffraction limits the maximum resolving power of an imaging system. The way to quantify that is to imagine that we have two different points in space imaged through a pinhole. Each point, of course, will cause a diffraction pattern. The two airy disks add up linearly in the power domain in the captured image, but when the two points are sufficiently apart spatially, the peaks of the two airy disks will be sufficiently apart on the detector plane as well, which means we can tell the two points apart from the image (because the power of the airy disks falls off very quickly with the distance to the center). When the two points are closer the peaks of the airy disks are closer; when the two peaks are sufficiently close the superposition of the two airy disks will result in an image where we cannot easily tell the two peaks apart, and that is when we know we have reached the resolution limit of the imaging system.

A common criterion used to quantify such a limit is called the **Rayleigh criterion**, first defined by Lord Rayleigh (who won the Nobel Prize in Physics in 1904 and the namesake of the Rayleigh scattering), which says the two points are *regarded* as just resolvable when the center of the airy disk of one point coincides with the first minimum of the other [[Rayleigh, 1879](#)]. If you go through the math, this translates to:

$$\theta \approx 1.22 \frac{\lambda}{D}, \quad (1)$$

where λ is the light wavelength, D is the diameter of the pinhole, and θ is the angular resolution of the imaging system, i.e., the angle subtended by the two points and the pinhole. As an example, assuming a 550 nm typical visible light, when the pupil size is about 2 mm, which is typical size under normal daylight, the resolvable angular resolution between two points is about 0.02 degree.

Note that I italicize “regarded” in the text above. There is no reason why one cannot distinguish between two points separated less than the Rayleigh criterion in a given scenario or train a deep neural network to do so. The Rayleigh criterion for the most part serves as an intuitive criterion that empirically well with observations.

How do we improve the resolving power of an imaging system? One way is to use shorter wavelength lights, which, according to Equation 1, would allow us to resolve objects that are closer. Optical microscopes use visible lights, whereas electro microscopes advantage of the wave nature of *electrons* to achieve much higher resolution than optical microscopes. The *de Broglie wavelength* of an electron is inversely proportional to its momentum. An electron microscope accelerates electrons to very high speeds, which reduces their wavelengths to below 1 nm (c.f., hundreds of nm for visible lights) and increases the overall resolving power of the imaging system.

3 Lenses and Aberrations

A convex lens brings many rays from a point together, as shown in Figure 1 (right). If the sensor plane is placed such as the image is in focus (which we will discuss shortly), the captured image is the geometrically the same as the one captured by a pinhole camera, but much brighter given the same exposure time. Both a pinhole imaging system and a convex-lens imaging systems perform a perspective projection, which is basically the camera model used in computer vision when a camera needs to be modeled and in simple graphics rendering pipelines.

3.1 Image Formation with an Ideal Lens

What is the imaging process of a convex lens? How do we model the behavior of a (convex) lens? We can model this using the basic geometrical optics. Figure 4 shows the setup. Assume we have a convex lens, which is made of two spherical surfaces combined together. The curvature of the two surfaces are R_1 (right surface) and R_2 (left surface). The two surfaces are separated by a distance d , which we call the thickness of the lens. The refractive indices of the air and the lens is n_1 and n_2 , respectively. The goal is to calculate, for a ray originated from a distance u on the optical axis in the scene and traveling in a direction that subtends an angle θ with the optical axis, what happens when it reaches the other side.

We will apply the Snell’s law at the two interfaces, essentially tracing the ray through the lens. At the first interface, we have:

$$n_1 \sin \alpha = n_2 \sin \beta, \tag{2a}$$

$$\alpha = \theta_1 + \mu_1, \tag{2b}$$

$$\beta = \mu_1 - \phi, \tag{2c}$$

$$\sin \mu_1 = \frac{h_1}{R_1}, \tag{2d}$$

$$\tan \theta_1 = \frac{h_1}{u}. \tag{2e}$$

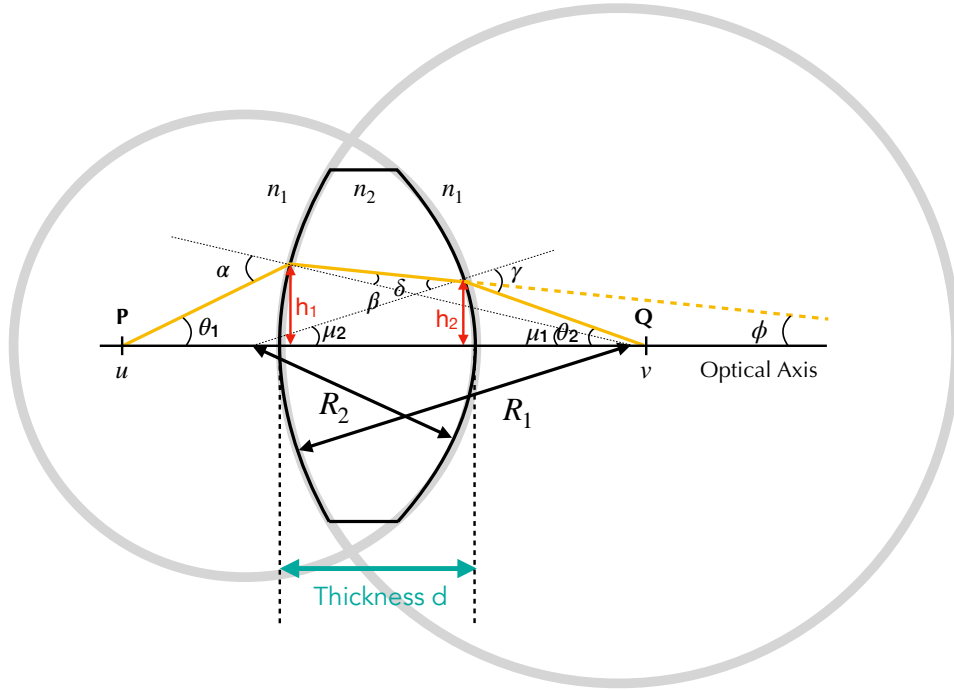


Figure 4: The setup to derive the lens maker's equation, which requires two assumptions: thin lens assumption and paraxial assumption.

As the light travels inside the lens and reaches the second interface, we have:

$$n_2 \sin \delta = n_1 \sin \gamma, \quad (3a)$$

$$\delta = \mu_2 + \phi, \quad (3b)$$

$$\gamma = \mu_2 + \theta_2, \quad (3c)$$

$$\sin \mu_2 = \frac{h_2}{R_2}, \quad (3d)$$

$$\tan \theta_2 = \frac{h_2}{v}. \quad (3e)$$

Now, we are going to make two assumptions. First, we will assume that the lens is very thin; the thickness d is very small. As a result, $h_1 \approx h_2$. This is called the *thin-lens assumption*. Second, we will assume that the ray stays close to the optical axis as it travels. Such rays are **paraxial rays**, and this assumption is called the *paraxial assumption*. That is, $\theta_1, \theta_2, \alpha, \beta, \gamma, \delta, \mu_1$, and μ_2 are very small angles, for which we can apply the usual small-angle approximation in trigonometry, e.g., $\sin(\alpha) \approx \tan(\alpha) \approx \alpha$ and $\cos(\alpha) = 1$.

Using these two assumptions and through a little algebra, we will get:

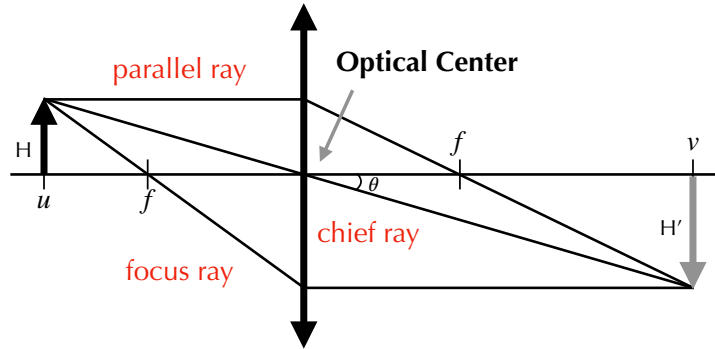


Figure 5: Under an ideal thin lens and paraxial approximation, the ray-tracing diagram can be simplified so that points at the same depth (S) are all in-focus at the same depth (S') on the other side of the lens. The magnification of the lens is $M = H'/H$, and the Field-of-View (FoV) of the system is 2θ (assuming that H' is half of the sensor, which is symmetric).

$$\frac{n_2 - n_1}{n_1} \left(\frac{1}{R_1} + \frac{1}{R_2} \right) = \left(\frac{1}{u} + \frac{1}{v} \right). \quad (4)$$

This is called the **Lens Maker's Equation**. Critically, observe that v depends only on u regardless of the path the ray takes (for a given lens with a particular set of n_1, n_2, R_1 , and R_2). Therefore, all rays originated from an on-axis point will converge at the same point on the other side on the optical axis. This is crucial, because it means we can place a single detector \mathbf{Q} (e.g., a pixel) on the imaging side (right side of the lens in this diagram) to capture all the rays from the point \mathbf{P} . In other words, if we place the detector at \mathbf{Q} , the point \mathbf{P} would be *in focus*. In fact, you can show that this is true for all points in space, not just on-axis points: all the rays originated from a point on one side of the lens will converge to another point on the other side of the lens. In reality of course, only paraxial rays with a thin lens follow this.

Now, if the ray originates from infinity (as if it is parallel to the optical axis), where $u = \infty$, we have

$$\frac{1}{v} = \frac{n_2 - n_1}{n_1} \left(\frac{1}{R_1} + \frac{1}{R_2} \right) := \frac{1}{f}. \quad (5)$$

This allows us to derive the position v where a parallel ray intersects with the optical axis. We define that position as the **focal length** f of the imaging system.

Plugging Equation 5 to Equation 4 gives us the familiar **Gaussian Lens Equation**:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}. \quad (6)$$

Under the ideal thin lens and paraxial approximation, the ray-tracing diagram is simplified to the one depicted in Figure 5, where:

- rays parallel to the optical axis always pass through the focal point (which has a distance f to the optical center) on the other side;
- rays passing through the focal point will be parallel on the other side;
- a ray passing through the optical center does not change its direction if the lens is symmetric; otherwise the incident ray at the first interface is parallel to that leaving the second interface as if the ray has been shifted.

Figure 5 shows that if we place an image sensor in the image space (right side of the lens) at v , all the points at the depth u in the world space (left side of the lens) will be in focus. Another important point Figure 5 makes clear is that, if in focus, the captured image is the geometrically the same as the one captured by a pinhole camera (but of course brighter given the same exposure time because more rays are captured): the optical center of the lens is the pinhole here *geometrically*.

Accommodation

Let's assume that $R_1 = R_2 = R$; Equation 5 suggests that if we reduce R (increase the curvature of the lens surface), f reduces as well. Then look at Equation 6; if f reduces and fix the object at the distance u , for that object to be in-focus we have to reduce v , i.e., move the sensor plane closer to the lens. That is, if we curve the lens surfaces more, rays focus closer to the optical center as if the lens bends lights more, and vice versa.

Another way to think of this is that if we cannot move the relative distance between the sensor and the lens, to focus on an object (in the world space) closer to the lens (u reduces), the lens focal length f has to reduce too. This is exactly what our eye lens does: to focus on closer objects, the lens curves more to gain more light-bending power. For that to take place, the ciliary muscle would have to contract. Conversely, to focus on farther objects, the ciliary muscle relax, which reduces the curvature of the lens, which now bends lights less and, thus, allows us to focus on farther objects. Changing the focal length through changing the curvature is called **accommodation**. As one gets older, the ciliary muscle is not as effective in contracting the lens. That is why one uses the reading glasses, which provide additional light bending power to assist that of the eye lens. Note, from our human visual system lecture, that while the lens is flexible, most of the light refraction was done at the air-cornea interface because of the large difference in the refractive index there.

In cameras, unless you are using liquid lenses, the curvature of each lens surface stays fixed once fabricated, so how do we focus on objects closer or farther than we are currently focused on? The answer is we move the lens, essentially solving for v given a new u using Equation 6. This is essentially how auto-focus works in cameras. Alternatively, we could also move the sensor, but in practice the sensor stays fixed (e.g., attached to the back of the camera housing) and it is the lens that is movable.

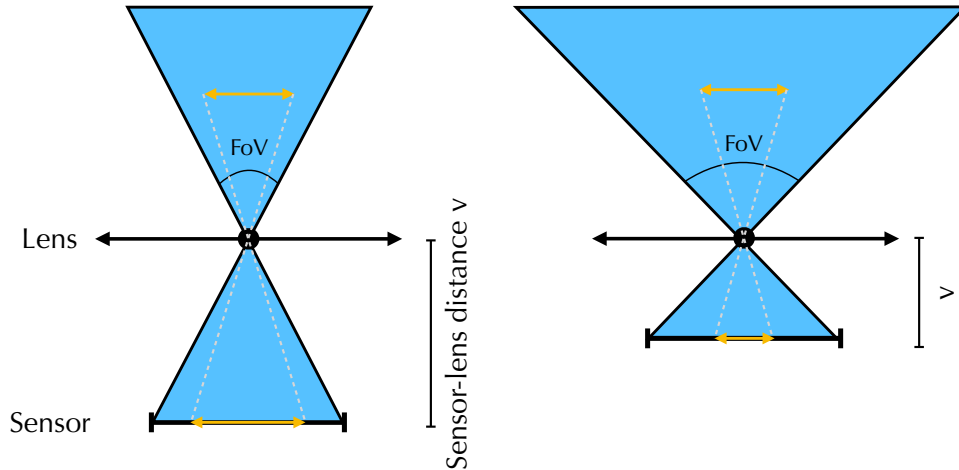


Figure 6: Increasing the focal length increases the magnification but reduces the FoV. For simplicity we use a pinhole geometry here, but both the pinhole camera and the ideal thin lens (with the paraxial approximation) performing the same perspective projection, leading to the same image formation geometrically.

3.2 Magnification vs. Field-of-View

Using simple trigonometry in Figure 5, we can relate the size of an object in the world space (H) and that in the image space (H'):

$$M = \frac{H'}{H} = \frac{f}{u - f} = \frac{1}{\frac{u}{f} - 1}, \quad (7)$$

where M is the magnification of the imaging system. We can see that M increases as f does. That is why telephoto cameras, those that you see in, for instance, sports broadcasting, are very long: they need to be long to accommodate a large focal length so that they can magnify objects that are very small (far away).

What do we sacrifice when we increase magnification by increasing the focal length? The Field-of-View (FoV) reduces. The FoV of an imaging system is the extent of the observable world that can be captured by the sensor. Let's use Figure 5 to derive this, and for simplicity sake, let's just assume that the sensor size is $2H'$ and is symmetric about the optical axis, i.e., the object at u is just fully captured by the sensor. The figure omits the upper half of the sensor. The FoV is defined as $2\theta = 2 \times \arctan(H'/v)$.

Now for the same object at u , if f increases, v has to increase as well for the object to be captured in-focus. As a result, θ reduces, so does the FoV. This intuitively makes sense: if an object is magnified more on the sensor, which has a fixed size, the amount of *other* objects that can be captured naturally reduces, hence the reduction of the FoV. As an example, the two imaging systems in Figure 6 differ only in the focal length: the one on the left has a shorter

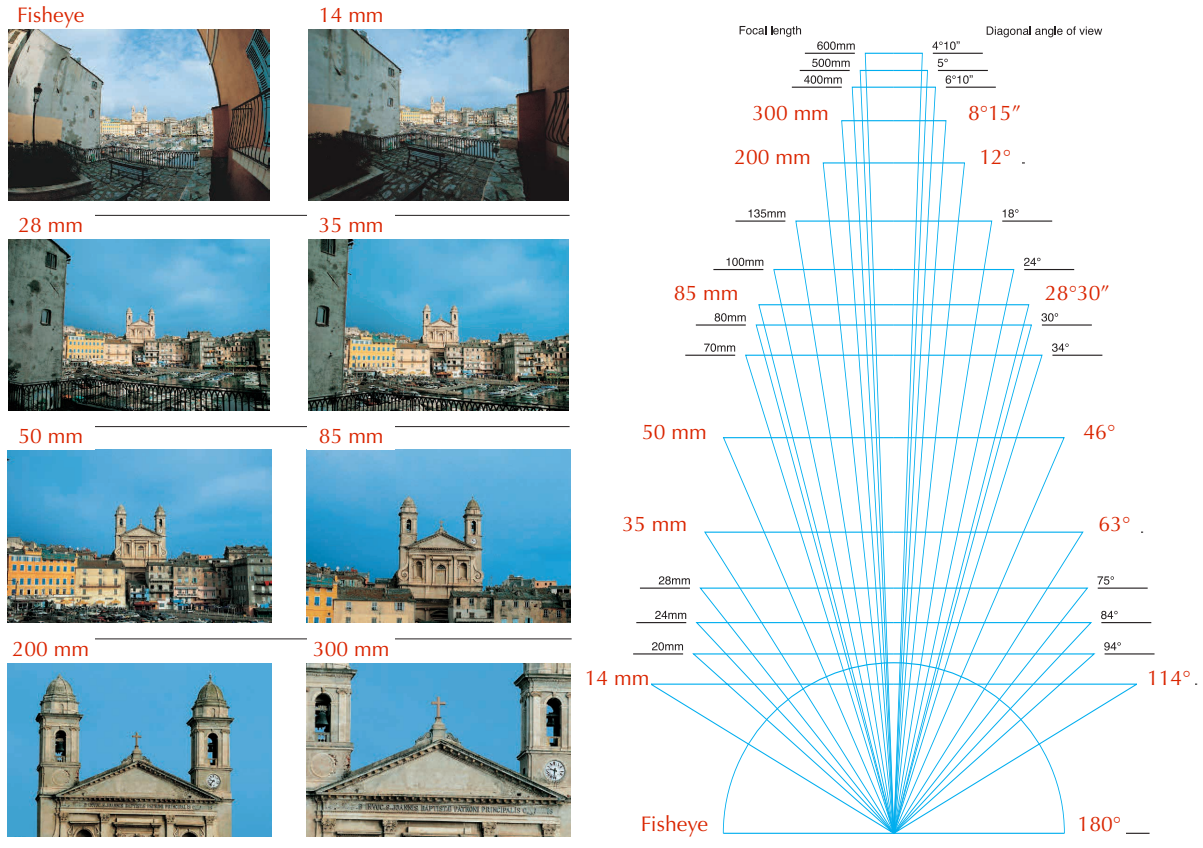


Figure 7: The effect of focal length on both magnification and field-of-view; adapted from Canon [2006, p. 123-125]. The fisheye lens does not perform a perspective projection, so its FoV is not directly comparable with others.

focal length f and hence a shorter sensor-lens distance v (for the same object to be captured in-focus), which translates to a larger magnification and narrower FoV.

Figure 7 shows a few concrete examples of how the focal length affects magnification and FoV. The fisheye lens does not perform a perspective projection (straight lines in the world space are not straight in the image space) so its image formation is not directly comparable, but we can see that it has the widest FoV. Other seven photons are taken with the same sensor but different lenses that differ in their focal lengths.

3.3 Magnifying Glasses and Projection Lenses

The Gaussian lens equation also helps us understand the geometry behind magnifying glasses and the projection lenses in AR/VR devices and cinematography.

When $u < f$ in Equation 6, v is negative. Figure 8 (top) shows the geometric of this case. The object does not form a physical image in the image space, because rays from a point on

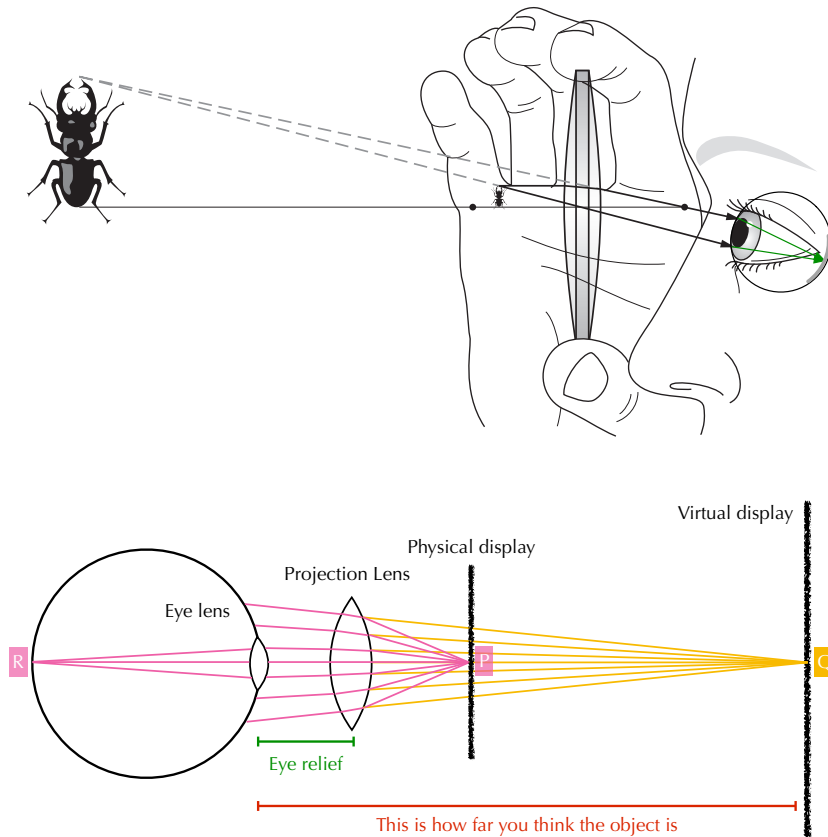


Figure 8: Top: the geometry of a magnifying glass; from Hecht [2016, Fig. 5.102 (c)]. Bottom: VR devices all have a pair of magnifying glasses to create a virtual display that our eye lenses can focus on.

the object do not converge to a point in the image space. Instead, those rays *diverge*, and the extension of those rays actually converge at a point farther away from the lens in the world space. Now, if our eye is at the right place, i.e., the diverging rays converge on the retina after traveling through the eye lens, as is the case in Figure 8 (top), we will see a magnified object. In this case, the lens acts as a magnifying glass.

The magnifying glass functionally 1) projects a small physical object to an apparently larger virtual object that is 2) farther away from the eye. These two functionalities are exactly what a project lens in AR/VR need. Figure 8 (bottom) illustrates a projection lens in VR; the optics in AR are much more complicated, but the basic idea of a projection lens applies there too. In AR/VR devices, the actual display is very close to the eye to the point that no eye lens can actually be accommodated to focus on the display (the lens would have to be curved so unrealistically much). Of course the display itself is very small, so seeing details is hard, too. The solution is to place a convex lens between the display and the eye, and the three components are so position that the display is closer to the lens than a focal length. As a result, the actual,

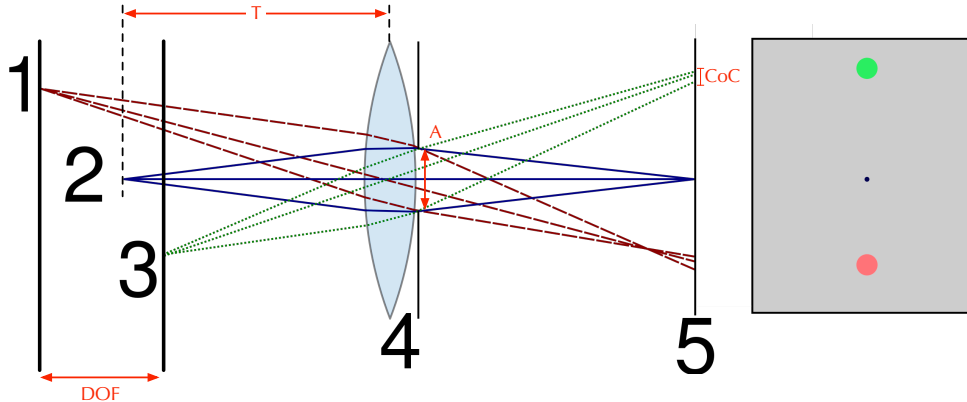


Figure 9: The setup to derive the depth-of-field (DoF) equation; adapted from [BenFrantzDale \[2010\]](#). CoC: circle of confusion.

physical display is projected to a much larger virtual display that is also farther away, to which our eye lens could actually accommodate. When you watch a movie in a cinema on a large screen or use a home projector, there is a projection lens sitting at the back doing the same thing.

3.4 Depth of Field

What if the sensor is not correctly positioned according to The Gaussians lens equation (Equation 6)? The object/point being imaged will be out of focus, and the result is a blur on the image. Figure 9 shows three cases, where the sensor (5) and the lens are fixed in position (4), under which objects at plane 2 (with a distance T to the lens) would be in-focus but both object 1 and object 3 would be out-of-focus, because rays originated from them will be spread across a small area on the sensor plane, looking like blurs. The shape of the blur is called the **bokeh**, which is mostly determined by the aperture shape (and also aberrations introduced by the imaging system, which we will see later). If the aperture is a circle, the bokeh would be one too, and we call such a blur the **circle of confusion** (CoC).

As the CoC increases, eventually it becomes objectionable to the human visual system. Exactly what that CoC threshold depends on a number of factors that we will omit here (e.g., how the image will be scaled when being viewed, the contrast sensitivity of the human visual system, etc.), but let's just use C to denote that threshold for now. You can see that if an object is placed slightly before or after the depth T (where the object is perfectly in-focus), as long as the resulting CoC is smaller than C , our visual system would still regard it as in-focus. The distance between the nearest and farthest objects whose CoCs are still within C is called the Depth-of-Field (DoF) of the system.

Using geometrical optics and with a few assumptions, we can show that the DoF is given by:

$$DoF \approx \frac{2CT^2N}{f^2} = \frac{2CT^2}{fA}, \quad (8)$$

where T is the distance of the object that is perfectly in-focus, f is the focal length, and $N = f/A$ is called the **F-number** of the camera, which is defined as the ratio between the focal length and the aperture size (A).

Given Equation 8, there are three ways to increase the DoF. First, we can increase T , i.e., focus on objects that are farther way (e.g., landscape photography). Second, we can decrease the focal length, but just keep in mind that changing the focal length will also affect the magnification and FoV as discussed in Chapter 3.1. Finally, we can also reduce the aperture size, which would increase the F-number. Changing the aperture size, however, will have implications on other aspects of the imaging quality. Specifically, a small aperture increases the exposure time and, thus, motion blur.

A larger DoF would mean that objects within a larger depth range could be simultaneously in-focus. A shallow DoF, however, is at many times desirable. The “portrait mode” in many modern smartphone cameras essentially captures photos with a shallow DoF. Intuitively, one can invert all three methods above to obtain a shallow DoF, but what if the hardware does not permit us to do that? For instance, what if we cannot increase the aperture size and focal length but want to capture a close object with a shallow DoF?

Computation comes to the rescue. There is a notion of **Synthetic DoF**, which uses post-processing algorithms to emulate a shallow DoF. For instance, one might first capture an all-in-focus photo, estimate the depth for each pixel in the photo (including the pixels that correspond to the object that we do want to have in-focus), then selectively blur pixels that are farther or closer than the objects of interest. This is what the portrait mode in Google’s Pixel phone does [Wadhwa et al., 2018].

Synthetic DoF is a classic example of computational photography, where the imaging system is largely assisted by computational algorithms (to reduce the design complexities of the imaging hardware). How one estimates the per-pixel depth is actually of critical importance and is best left to another lecture. It turns out that auto-focus in cameras is all about depth estimation, which, again, usually involves some form of collaboration between software and hardware.

3.5 Aberration

When building an imaging system, we ideally want a point in the physical space to be captured as a single point in the image space. In our derivation of the Gaussian lens equation in Chapter 3.1, this is indeed the case, so if the sensor is correctly positioned we will capture a sharpe image of the point. This derivation, however, assumes an ideal thin lens and considers only paraxial rays. It turns out that the equation still holds even if the lens is thick (i.e., the distance between the two surfaces is not negligible), even though the definition of the focal length would have to be slightly more complicated than Equation 5 [Hecht, 2016, Chpt. 6.1].

The real complication is that in practice we cannot ignore non-paraxial rays (i.e., rays that do not stay close to the optical axis), in which case rays from a single point (or from infinitely

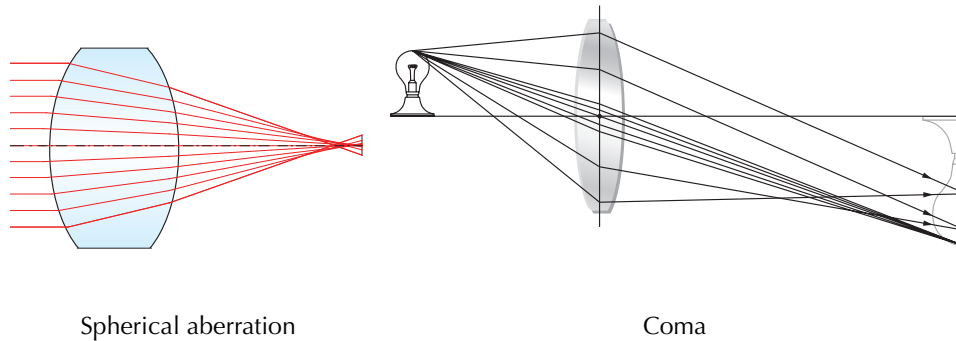


Figure 10: Left: spherical aberration; from [Hollows and James \[2024\]](#). Right: coma; from [Hecht \[2016, Fig. 6.22 \(a\)\]](#).

far away) will not all converge at a single point in the image space, resulting in a blur. Mathematically, this means we cannot invoke the small angle approximations. For instance, using Taylor expansion we have:

$$\sin \theta = \theta - \frac{\theta^3}{3!} + \frac{\theta^5}{5!} - \dots \quad (9)$$

When consider paraxial rays we can afford to consider only the first term, but when θ is large we have to include other terms. When considering the second term of Taylor series expansion (compared to considering only the first term), five forms of **aberrations** show up: spherical aberration, coma, astigmatism, field curvature, distortion. Geometrical optics that consider the second term is called the **third-order theory**, as opposed to the **first-order theory** or **Gaussian optics** that considers only the first term.

Spherical Aberration

It turns out that for a spherical lens, non-paraxial rays originated from a point on the optical axis will not converge at the same point. This can be shown by going through the derivation of the Gaussian lens equation (Chapter 3.1) but this time without the small angle approximations. We would then see that v depends not only on u and f but also on the direction of the ray leaving u . By extension, not all rays parallel to the optical axis (especially those that are far away from the optical axis) will focus at the same point. This is called the spherical aberration, which is illustrated in Figure 10 (left).

Mirrors have spherical aberrations too. Perfectly spherically curved mirrors cannot focus parallel lights; parabolic mirrors are free from spherical aberrations. One might venture to guess that's why Archimedes could not have used mirrors to burn Roman ships, because they could not have the skills to make parabolic mirrors. From modern perspectives, the burning mirror story is more likely a story than fact. There are just too many technical reasons why that would have been very hard. For instance, it would have taken a very large mirror given the

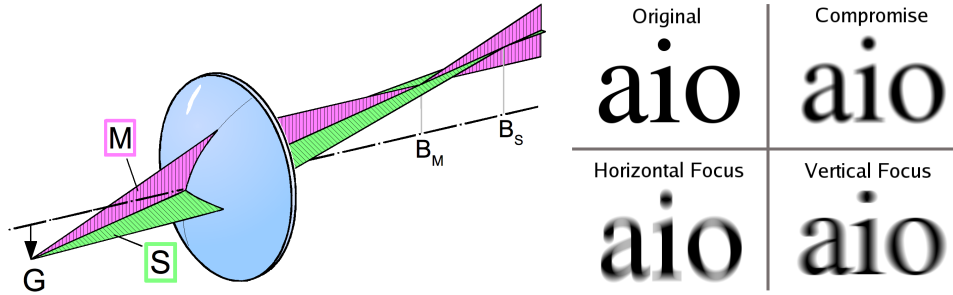


Figure 11: Left: an illustration of astigmatism; adapted from Michael Schmid [2008]. Right: the images captured when the sensor plane is placed at different positions; from Tallfred [2005].

intensity of sunlight and the distance of the ships, and the ship will have to perfectly positioned at the focal point [Chris Rorres; Mills and Clift, 1992].

Coma

While spherical aberration is concerned with rays from on-axis points or parallel rays that are also parallel to the optical axis, another aberration called coma or comatic aberration, as illustrated in Figure 10 (right), is concerned with rays from off-axis points or parallel rays that have an oblique incident angle w.r.t. the optical axis. We can show that rays from an off-axis point focus on different points and, by extension, parallel rays that are not parallel to the optical axis do not focus on the same point. This aberration is called coma because the resulting blur looks like a coma.

Astigmatism

Yet another form of aberration is called astigmatism. It is also concerned with points off the optical axis. In particular, we are concerned with rays propagated in two planes. The first plane is one defined by the object point and the optical axis, and is called the **tangential plane** or the **meridional plane**. The other plane is one that is orthogonal to the meridional plane, and is called the **sagittal plane**. It turns out that rays from the two planes focus on different points on the optical axis. This is illustrated in Figure 11 (left), where all the rays in the meridional (M) plane focus at B_M and all the rays in the sagittal (S) plane focus at B_S .

The blur we get depends on where we place the sensor plane, and some examples are shown in Figure 11 (right). If we place the sensor at B_M , a single point source get imaged as a horizontal/lateral “line”, because of the spread of rays in the S plane. We say a line, but it is not actually a line because rays in other planes (other than the M and S planes) will not focus at B_M and still contribute to the image formation, so the resulting image is really a very much elongated ellipse. If the object is not a point but, say, spans a plane (top-left), the resulting image has a somewhat horizontal/lateral blur as if the in-focus image is smeared laterally (bottom-right).

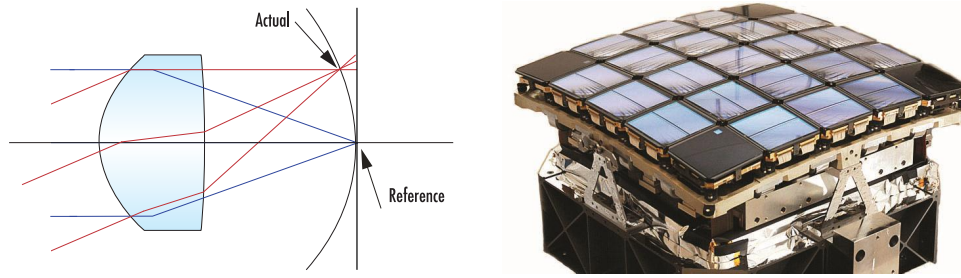


Figure 12: Left: a manifestation of field curvature; from [Hollows and James \[2024\]](#). Right: focal plane of Keplers space telescope is curved to mitigate field curvature; from [HandWiki \[2024\]](#).

As we move the sensor beyond B_M , the elongated ellipse gradually expands vertically, and then become circular, and then shrinks laterally; eventually when the sensor is placed as B_S , we get a vertical “line” (an elongated ellipse along the vertical axis) because, mainly, of the rays in the M plane. The resulting image would appear to have a somewhat vertical blur as if the in-focus is smeared vertically (bottom-left). The somewhat circular blur when the sensor plane is in-between B_M and B_S means that the resulting image (top-right) appears as if the in-focus image is smeared in all directions.

Field Curvature

If an imaging system is free of all the previous aberrations, a single point in the world space corresponds to a single point in the image space. However, a plane of points in the world space would not correspond to a plane in the image space. In fact, it would correspond to a curved surface. If we use a planar sensor for imaging, we would get a blurred image. This form of aberration is called field curvature, as illustrated in Figure 12 (left).

While it might be difficult to build a single curved sensor, it is relatively easy to assemble a set of sensors on a curved surface. The image-sensor array of the Kepler space observatory is curved to compensate for the field curvature, as shown in Figure 12 (right) Interestingly, you might recall that the human retina is not planar either; it is curved. This to some extent helps mitigate the effect of field curvature.

Distortion

Even when all the previous aberrations are somehow corrected, the image would look sharp but distorted. Distortion does not introduce blurs. Rather, it is a result of the variation of magnification as a function of the distance to the optical axis (object height).

Equation 7 suggests that magnification depends only on the object distance to the lens u , but in reality the magnification depends also on the object height. We can imagine for a point that is distant from the optical axis, rays originated that point will not be paraxial rays. If the magnification increases with the height, we have a positive or **pincushion distortion**; otherwise we have a negative or **barrel distortion**. The two forms of distortion are illustrated

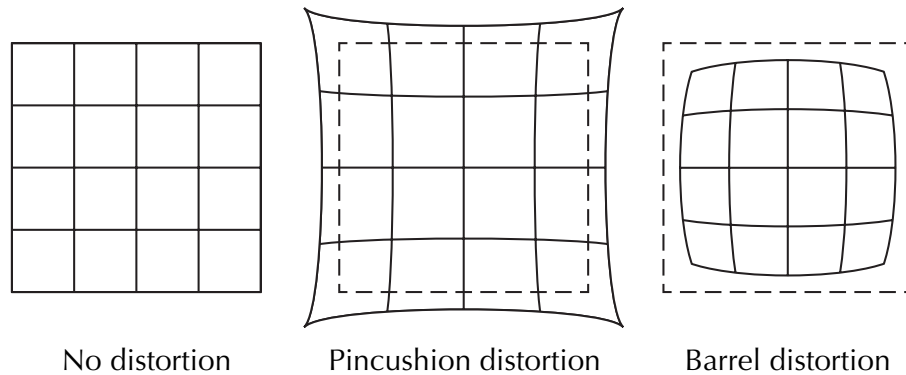


Figure 13: A comparison of no distortion, pincushion (positive) distortion, and barrel (negative) distortion; from Hecht [2016, Fig. 6.34].

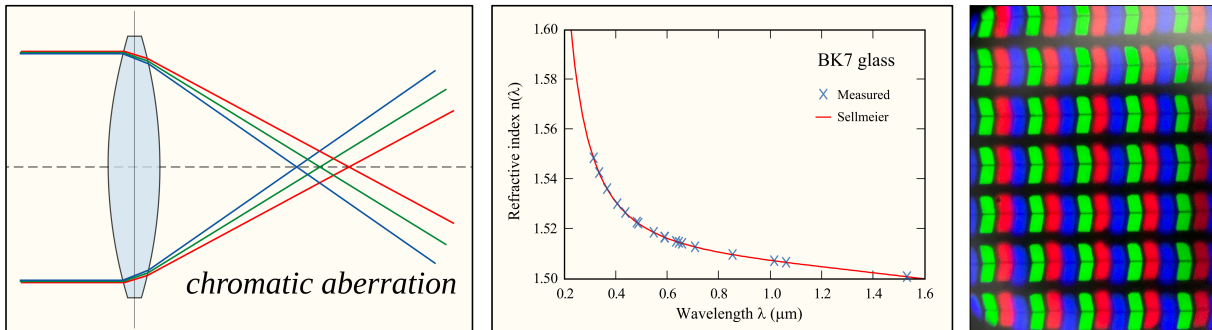


Figure 14: Left: illustration of chromatic adaptation; adapted from Bob Mellish [2006]. Middle: refractive index vs. wavelength for BK7 glass; from DrBob [2007]. Right: sRGB white as displayed by a 4th-generation iPad Pro; the image is captured when focusing on the green subpixel so other subpixels are out of focus.

in Figure 13.

Chromatic Aberration

All the aberrations we have discussed before are present even if we consider only a single wavelength; they are called monochromatic aberrations. When we consider lights that comprise a mixture of different wavelengths, **chromatic aberration** shows up. Chromatic aberration arises fundamentally because the refractive index is a function of wavelength; after all, that is how Newton was able to disperse white light and show the spectrum. Figure 14 (left) illustrates the issue of chromatic adaptation, which introduces “colorful” blurs. Figure 14 (right) shows how the refractive index of BK7 glass (which is commonly used in lenses) changes with wavelength.

As another example, I took a picture of my 4th-gen iPad Pro when display sRGB white.

I intentionally focused on the green subpixels. As a result, the other two subpixels are out of focus — due to chromatic aberration.

Correction for Aberrations

One of the main tasks of optical design, especially for imaging lenses, is to correct for aberrations. There are two main approaches: non-spherical (aspherical) lenses and compound lenses.

Optical designers often use multiple (compound) lenses in combination to correct various aberrations. For instance, chromatic doublets or apochromatic triplets are specifically designed to counteract chromatic aberration. One obvious downside of compound lenses is form factor, which becomes an issue for systems like Augmented Reality that needs to be very compact.

One promising technology that people are currently investigating is called **freeform optics**. Traditional aspherical lenses, while deviating away from a spherical design and can in many cases avoid compound lenses, are still rotationally symmetric so are still limited in what they can do. Freeform optics take this concept further by allowing surfaces that lack rotational symmetry, providing additional degrees of freedom in optical design. This enables better correction of higher-order aberrations, such as coma and astigmatism, which aspheric lenses alone may not fully eliminate.

3.6 Not All Blurs are Created Equal

Ideally a point source in the scene should really be captured as a single point in the image plane, but we have seen a few ways that a blur can occur. But not all blurs are created equal; it is perhaps useful to review the different causes of a blur.

Blurs can be resulted from aberrations, diffraction, defocus, and motion. We have just seen blurs from aberrations, but just note that not all aberrations result in blurs, an example of which would be distortion. Assuming an aberration-free imaging system, if the sensor is not placed as the focal plane, we could get a de-focus blur, as we have seen in the DoF section (Chapter 3.4). Note that a pinhole camera would never have defocus blur, because its DoF is infinite (using $A = 0$ in Equation 8).

Even if the sensor is placed as the focal plane, but if the object is motion, we would most likely get motion blur, because the exposure time is finite — unless of course the exposure time is so short that the object motion, when projected on the sensor plane, is within the pixel width. The longer the exposure time, the more pronounced the motion blur becomes.

Finally, we have blurs from diffraction, which, as we have discussed in Chapter 2.3, is fundamentally a result of the quantum nature of light. If an imaging system is free from all previous forms of blur¹, we say it is “diffraction limited”, because its imaging capability (the ability to avoid blurs) is limited only by diffraction.

¹mostly aberration, because defocus can be easily fixed and motion blur is out of the hands of an imaging system

3.7 Radiometric Analysis of Lens

What does a convex lens do to the radiance of incident light? We know that the radiance of a ray does not change as the ray propagates through space along a particular direction, but what does a lens do to the radiance? This is an important problem in practice: the lens essentially transforms the light field in the physical scene to the light field inside the camera, which means if we know the latter *and* the radiance transformation done by the lens, we can infer the light field in the scene.

The way to reason about is to think of a lens as performing a sequence of two refractions at its two surfaces, so we will have to first reason about, at each surface, what happens to the radiance and then consider the composite effect of the two surfaces.

With a little radiometry (which we will omit but refer to [Bohren and Clothiaux \[2006, Chpt. 4.1.6\]](#) for the derivation), we can show that the radiance after refraction L_r relates to the incident radiance L_i by:

$$L_r = n^2 L_i, \quad (10)$$

where n is the relative refractive index of the lens/medium to the air. Usually $n > 1$, which means after refraction the radiance increases. This makes sense because after refraction from the air to, say, glass, the set of incident rays map to a smaller solid angle.

What happens in the second surface? The same thing except the relative refractive is now $1/n$, since we are now going from the medium to the air:

$$L_o = \frac{1}{n^2} L_r, \quad (11)$$

where L_o is the radiance leaving the lens. Combining the two equations above, we can see $L_o = L_i$, meaning the lens does not change the radiance. This is a nice result, because it essentially means we can simply trace rays through a lens and be reasonably sure that the ray radiance does not change.

Intuitively, this conclusion is obviously *wrong*: some energy of incident lights is absorbed/reflected away by the lens, so the energy leaving the lens is definitely smaller than that entering the lens. So the derivation above is a bit of a simplification, because we have assumed that no reflection takes place at each surface and no absorption by the lens. That said, this invariance largely still holds if we confine ourselves to near-normal angle of incidence and assume typical materials for lenses (which are mostly transparent with little absorption).

4 Computational Modeling Using Linear System Theory

How do we model the image formation process of an image system? We could trace rays out of a point in the world space, but it has many limitations. First, we could afford to sample only a few rays for a point. Second, we could afford to sample only a few points on an object. Third, things like diffraction that go beyond geometrical optics need special treatment if not straight up impossible. How do we effectively model the image formed on the sensor plane?

4.1 Basic Idea of a Linear System Theory

A common modeling strategy is to first characterize the response of the imaging system against a *single point source*. If we assume that the system is **linear and shift invariant** (LSI), we can derive how the system response to an arbitrarily complex object, which is treated nothing more than a collection of (infinitely many) points, using the linear system theory. Let's unpack this step by step.

Point Spread Function

The response of a single point source is called the **Point Spread Function** (PSF) of the imaging system. How does the PSF look like? Ideally, a single point in the world space would be imaged as a single point in the image plane, so the corresponding PSF would be a Dirac delta function, but as we have discussed in Chapter 3.6, in reality the image of a single point would be blurred, whether it is because of diffraction, defocus, or aberration (assuming the point source is stationary).

Figure 15 (left) shows a few examples of the PSFs. The bottom-right corner shows a diffraction limited PSF, which is essentially the Airy disk (but visualized as a 2D grayscale map). As we move vertically up, we add more spherical aberration to the system, and we move to the right, we add more defocus to the system. Figure 15 (middle) shows a PSF of a system with astigmatism; this time the PSF is visualized in 3D rather than a 2D grayscale map. We can see that the PSF is not rotationally symmetric; rather, it is elongated along one dimension, indicating that the sensor is placed somewhere closer to one of the two focal points.

Linear System

Informally, if there are two inputs x and y to the system, say two points in the world space, and the responses to these two inputs, i.e., their respective PSFs, are $H(x)$ and $H(y)$, the response of a linear system to a new input $\alpha x + \beta y$ would be $\alpha H(x) + \beta H(y)$. αx means to scale the input x 's value (e.g., irradiance of a point) by a factor of α .

A linear system essentially means when we image two points simultaneously the resulting image is equal to the sum of the individual image of each point. You can imagine how this would simplify our modeling later. In practice, an imaging system is linear when it interrogates non-coherent light, e.g., sun light or OLEDs rather than lasers.

Shift-Invariant System

An imaging system is shift invariant if its PSF of a point is invariant to the shifts of the point in the world space. This property allows us to use a single PSF to characterize the system.

Of course in reality a system is hardly shift-invariant. For instance, if we move a point away from or closer to the lens, we get different kinds of defocus blurs, so the point response depends on depth. Even if we shift a point within a single depth plane, the rays incident on the lens, leaving the lens, and, by extension, hitting the sensor plane would be different. Even ignoring aberrations, different incident angles result in different irradiance captured by the sensor plane

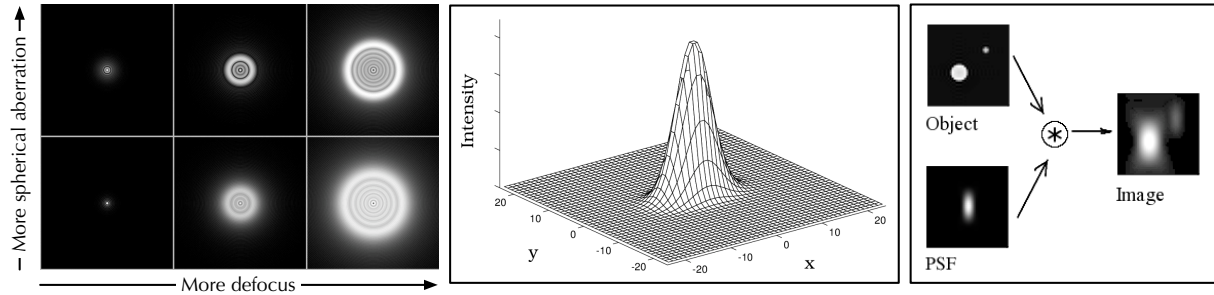


Figure 15: Left: PSF of a diffraction-limited system (bottom-left), with spherical aberration (vertically), and with defocus (horizontally); adapted from Mdf [2005]. Middle: PSF (visualized in 3D) with astigmatism; from Wandell [1995, Fig. 2.22]. Right: in a linear and shift-invariant imaging system, image formation is equivalent to a convolution using the PSF; from Mdf [2006].

(because of the Lambertian cosine law and is a form of vignetting, which we will discuss later in the sensor lecture).

In general, however, shift invariance approximately holds if we assume that the object to be imaged is very far away from the lens (so the depth variation within an object is negligible with respect to the overall distance to the lens) and the imaging system has a relatively small FoV.

4.2 Modeling Image Formation in LSI Systems

Under the linear and shift-invariance assumptions, we can derive a simple but incredibly useful computational model for the image formation process, which is decoupled into two conceptual steps.

In the first step, we calculate an ideal image I_{in} formed by the imaging system assuming the imaging system PSF is a delta function. This in turns assumes that the imaging system has no diffraction/aberration and the sensor plane is properly positioned so every scene point is sharply in-focus (no defocus blur)². I_{in} is a 2D continuous function.

Geometrically, I_{in} is a perspective projection of the 3D scene to the sensor plane. That is, each (x, y) point in this ideal image I_{in} corresponds to a point $P(x', y', z')$ in the scene as if that scene point is captured through a pinhole (recall that geometrically an ideal thin lens performs the same projection as a pinhole system). This is shown in Figure 16 (left). Radiometrically, the value of $I_{in}(x, y)$ is an irradiance quantity, representing the irradiance emitted from $P(x', y', z')$ that is captured at $I_{in}(x, y)$. We will discuss in the sensor lecture exactly how to calculate this irradiance, but you can imagine that it depends on the distance of P to the lens, the FoV of the lens, the sensor-lens distance, etc.

In the second step, we then using the PSF to convolve I_{in} , and the result:

²which is of course impossible because different scene depths would require different sensor positions to be all in-focus.

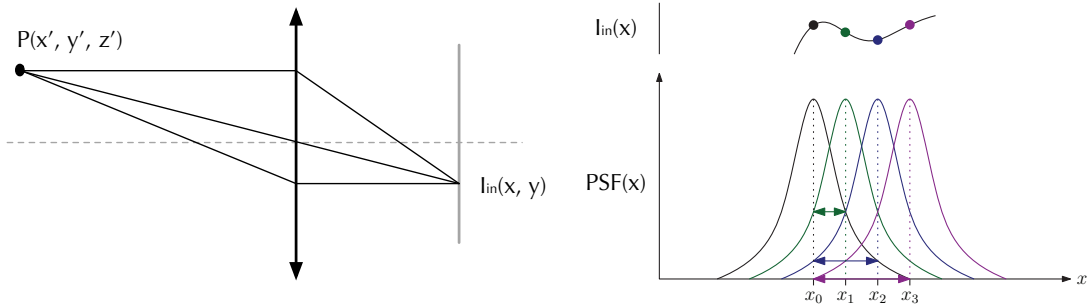


Figure 16: Left: geometrically, the ideal image I_{in} is a perspective projection of the scene, and radiometrically $I_{in}(x, y)$ corresponds to the irradiance from $P(x', y', z')$ that are captured by the imaging system. Right: the intuition behind why imaging in a LSI system is a convolution with the PSF; adapted from Rowlands [2020, Fig. 3.3].

$$I_{out} = I_{in} \star PSF, \quad (12)$$

is the actual image formed by the imaging system. This is illustrated in Figure 15 (right).

The convolution is a natural conclusion once we assume linearity (irradiances add) and shift invariance (constant PSF) of the imaging system. Figure 16 (right) illustrates the intuition using an 1D example. $I_{in}(x, y)$ is the irradiance at (x, y) with a delta PSF. With a non-delta PSF, the irradiance of $I_{in}(x, y)$ is distributed over the sensor plane as defined in the PSF. Each point on the sensor, thus, receives contributions from all the point spreads. Since we assume linearity, the result is a convolution between I_{in} and the PSF.

In the literature, it is common to see people taking images from a dataset, e.g., ImageNet, and simply convolve a PSF against them. The underlying assumption is that those images are captures of distant objects with an ideal system (with no blurs) and, thus, can be treated as essentially irradiance maps I_{in} .

What if the system is not shift invariant? For instance, if we cannot assume that objects are all very far away from the lens, scene points at different depths will have different PSFs. So long as we can still assume linearity, however, we can still relatively easily simulate the image formation process using the exact same principle shown before in Figure 16: by adding up the PSF of each of the point in the scene scaled by the corresponding irradiance except now the PSF is no longer a constant.

4.3 Fourier Perspectives: OTF and MTF

Since we are using convolution to model imaging in LSI systems, it is only natural to take a Fourier perspective. Recall the convolution theorem:

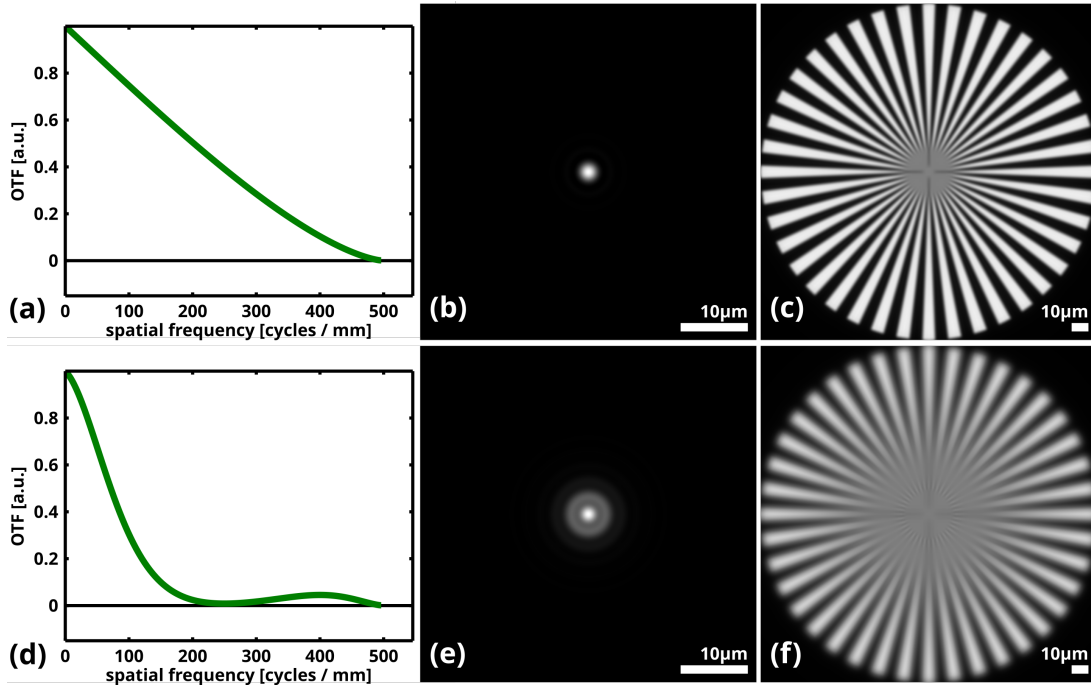


Figure 17: Top: OTF, PSF, and the resulting image of a diffraction-limited system. Bottom: OTF, PSF, and the resulting image of a diffraction-limited but defocused system. From [Tom.vettenburg \[2017\]](#).

$$\mathcal{F}(f \star g) = \mathcal{F}(f)\mathcal{F}(g), \quad (13a)$$

$$f \star g = \mathcal{F}^{-1}(\mathcal{F}(f)\mathcal{F}(g)), \quad (13b)$$

where \mathcal{F} and \mathcal{F}^{-1} denote Fourier transform and inverse Fourier transform. This allows us to reason about the effect of an imaging system in the frequency domain.

The Fourier transform of a PSF is called the Optical Transfer Function (OTF), which is necessarily complex-valued, which has a magnitude and a phase component. The magnitude component of the OTF is called the Modulation Transfer Function (MTF) and the phase component of the OTF is called the Phase Transfer Function (PTF):

$$OTF(\omega) = MTF(\omega)e^{iPTF(\omega)}. \quad (14)$$

What is the OTF of an ideal PSF, i.e., a delta function? It is a constant 1 across all frequencies. This makes sense: an ideal PSF introduces no blur so it does nothing to each spatial frequency.

Figure 17 shows two more examples; the top half shows the OTF, PSF, and the resulting imaging of a diffraction-limited system (i.e., PSF being an Airy disk), and the bottom half shows

the same system with a defocus blur. In both cases, the OTF is the same as the MTF because the Fourier transform of both PSFs have zero phase (PTF is zero at any ω). You can convince yourself of this by taking a Fourier transform of the Airy function and assuming that defocus adds a Gaussian blur to the Airy disk; we will omit the math here. General OTFs do have a phase term because the PSFs of many aberrations, e.g., coma and astigmatism, are not radially symmetric.

We can see that in the diffraction-limited case, the OTF drops to 0 at a frequency of 500, meaning information at any frequency higher than the cut-off is lost. The (first) cut-off for the defocused system is at an even lower frequency (about 200), naturally leading to more blurs in the resulting image.

References

- Nick Antipa, Grace Kuo, Reinhard Heckel, Ben Mildenhall, Emrah Bostan, Ren Ng, and Laura Waller. Diffusercam: lensless single-exposure 3d imaging. *Optica*, 5(1):1–9, 2017.
- BenFrantzDale. Effect of aperture on blur and DOF; CC BY-SA 3.0 license. https://commons.wikimedia.org/wiki/File:Depth_of_field_illustration.svg, 2010.
- Ayush Bhandari, Achuta Kadambi, and Ramesh Raskar. *Computational Imaging*. MIT Press, 2022.
- Bob Mellish. Chromatic aberration diagram; CC BY-SA 3.0 license. https://commons.wikimedia.org/wiki/File:Chromatic_aberration_lens_diagram.svg, 2006.
- Craig F Bohren and Eugene E Clothiaux. *Fundamentals of atmospheric radiation: an introduction with 400 problems*. John Wiley & Sons, 2006.
- Canon. *EF Lens Work III: The Eye of EOS*. Canon Inc. Lens Products Group, 8 edition, 2006.
- Chris Rorres. Burning Mirrors: Refuting the Legend. <https://math.nyu.edu/Archimedes/Mirrors/legend/legend.html>.
- Daniel King. An image created by one of Daniel King’s pinhole cameras during WOUB’s Emily Votaw’s interview with the artist. <https://woub.org/2017/10/20/create-your-own-pinhole-camera-at-daco-with-artist-daniel-king/>, 2017.
- Dave Haynie. What changes occur if the size of pinhole in a pinhole camera is doubled? <https://www.quora.com/What-changes-occur-if-the-size-of-pinhole-in-a-pinhole-camera-is-doubled>, 2019.
- Dominic Alves. Pinhole Size Chart; CC BY 2.0 license. <https://www.flickr.com/photos/dominicpics/4589206921>, 2006.

- DrBob. Refractive index vs. wavelength for BK7 glass; CC BY-SA 3.0 license. <https://en.wikipedia.org/wiki/File:Sellmeier-equation.svg>, 2007.
- HandWiki. Kepler space telescope focal plane; CC BY-SA 3.0 license. <https://handwiki.org/wiki/index.php?curid=2015813>, 2024.
- Eugene Hecht. *Optics*. Pearson, 5 edition, 2016.
- Hollows and James. Imaging Application Notes: Aberration (Edmund Optics). <https://www.edmundoptics.com/knowledge-center/application-notes/imaging/how-aberrations-affect-machine-vision-lenses/>, 2024.
- Ann C Hurley, G David Lange, and Peter H Hartline. The adjustable “pinhole camera” eye of nautilus. *Journal of Experimental Zoology*, 205(1):37–43, 1978.
- Inductiveload. Mathematical function of an airy disk pattern; released into the public domain by the copyright holder. https://commons.wikimedia.org/wiki/File:Airy_Pattern.svg, 2009.
- Mdf. A simulation of spherical aberration in an optical system with a circular, unobstructed aperture admitting a monochromatic point source; released into the public domain by the copyright holder. <https://commons.wikimedia.org/wiki/File:Spherical-aberration-disk.jpg>, 2005.
- Mdf. Imaging as a convolution against the PSF; released into the public domain by the copyright holder. https://commons.wikimedia.org/wiki/File:Convolution_Illustrated_eng.png, 2006.
- Michael Schmid. Graphic illustratic the astigmatism phenomenon; CC BY-SA 3.0 license. https://commons.wikimedia.org/wiki/File:Meridional%2BSaggittalEbene_1.svg, 2008.
- Allan A Mills and R Clift. Reflections of the ‘burning mirrors of archimedes’. with a consideration of the geometry and intensity of sunlight reflected from plane mirrors. *European journal of physics*, 13(6):268, 1992.
- Lord Rayleigh. Xxxi. investigations in optics, with special reference to the spectroscope. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 8(49):261–274, 1879.
- D Andrew Rowlands. *Physics of digital photography*. IOP Publishing, 2 edition, 2020.
- Sakurambo. A computer-generated image of an Airy disk (the grayscale intensities have been adjusted to enhance the brightness of the outer rings of the Airy pattern); released into the public domain by the copyright holder. <https://commons.wikimedia.org/wiki/File:Airy-pattern.svg>, 2007.

- Tallfred. Text blurred by different focal positions of an astigmatic lens; 3-clause BSD License. https://commons.wikimedia.org/wiki/File:Astigmatism_text_blur.png, 2005.
- Tom.vettenburg. Illustration of the optical transfer function and its relation to image quality; CC BY-SA 4.0 license. https://en.wikipedia.org/wiki/File:Illustration_of_the_optical_transfer_function_and_its_relation_to_image_quality.svg, 2017.
- Neal Wadhwa, Rahul Garg, David E Jacobs, Bryan E Feldman, Nori Kanazawa, Robert Carroll, Yair Movshovitz-Attias, Jonathan T Barron, Yael Pritch, and Marc Levoy. Synthetic depth-of-field with a single-camera mobile phone. *ACM Transactions on Graphics (ToG)*, 37(4): 1–13, 2018.
- Brian A. Wandell. *Foundations of vision*. Sinauer Associates, 1995.
- Yang Zhang, Fan Mao, Huawei Mu, Minwei Huang, Yongbo Bao, Lili Wang, Nai-Kei Wong, Shu Xiao, He Dai, Zhiming Xiang, et al. The genome of nautilus pompilius illuminates eye evolution and biomineralization. *Nature ecology & evolution*, 5(7):927–938, 2021.