CGR2025 — Lecture 3: Cameras

Kartic Subr

September 23, 2025

1 Pinhole Camera

A pinhole camera is a simple imaging model in which a single small aperture admits light without using lenses. Rays from scene points pass through the aperture and project an inverted image onto a plane opposite the pinhole. Smaller pinholes yield sharper images but admit less light. This forms the foundation for modeling lens-based cameras.

The concept of the pinhole camera dates back thousands of years. The earliest known description comes from the Chinese philosopher Mozi (5th century BCE), who observed that light travels in straight lines and projects an inverted image through a small hole. Similar ideas were later explored by Aristotle and Alhazen (Ibn al-Haytham, 10th–11th century), whose detailed experiments with the *camera obscura* laid the foundations of modern optics. During the Renaissance, artists such as Leonardo da Vinci used the camera obscura as a drawing aid, projecting scenes onto surfaces to achieve accurate perspective.

The pinhole camera reemerged as a scientific and artistic tool in the 19th century with the development of light-sensitive photographic materials. Unlike lens-based cameras, pinhole cameras eliminate lens aberrations and produce images with infinite depth of field, though at the cost of longer exposure times and softer focus.

Today, pinhole photography remains popular among experimental photographers, educators, and hobbyists. Worldwide events such as *Worldwide Pinhole Photography Day* celebrate its simplicity and creativity. Interestingly, NASA has even used pinhole imaging principles in solar observations, while DIY enthusiasts construct pinhole cameras from everyday objects like shoeboxes or soda cans, blending ancient optical knowledge with modern curiosity.

2 Virtual Pinhole Camera Model and Projection Matrix

2.1 Why Virtual Pinhole Cameras?

A physical pinhole camera is a simple device: a dark box with a tiny aperture that projects an inverted image onto a screen or sensor. While useful for intuition, it is not practical for most applications in computer vision and graphics. Real cameras have lenses, finite sensors, and arbitrary positions in 3D space.

A *virtual pinhole camera* is a mathematical abstraction based on the pinhole principle. It models the mapping from 3D world coordinates to 2D pixel coordinates, incorporating both the camera's pose (extrinsic parameters) and sensor geometry (intrinsic parameters). This abstraction is essential for rendering, reconstruction, and calibration.

2.2 Step 1: World to Camera Coordinates

A point in world coordinates $\mathbf{X}_w = (X_w, Y_w, Z_w)^\top$ is expressed in the camera coordinate system by applying a rigid transformation:

$$\mathbf{X}_c = \mathbf{R}\,\mathbf{X}_w + \mathbf{t},$$

where $\mathbf{R} \in SO(3)$ is the rotation matrix and $\mathbf{t} \in \mathbb{R}^3$ is the translation vector. In homogeneous coordinates:

$$\begin{bmatrix} \mathbf{X}_c \\ 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ \mathbf{0}^\top & 1 \end{bmatrix} \begin{bmatrix} \mathbf{X}_w \\ 1 \end{bmatrix}.$$

2.3 Step 2: Perspective Projection

In the camera frame, a 3D point $\mathbf{X}_c = (X_c, Y_c, Z_c)^\top$ projects to the image plane at distance f:

$$x = f\frac{X_c}{Z_c}, \quad y = f\frac{Y_c}{Z_c}.$$

In homogeneous form this can be written as:

$$\begin{bmatrix} \tilde{x} \\ \tilde{y} \\ \tilde{z} \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_c \\ Y_c \\ Z_c \\ 1 \end{bmatrix},$$

with the final coordinates obtained by normalizing: $(x, y) = (\tilde{x}/\tilde{z}, \tilde{y}/\tilde{z})$.

2.4 Step 3: From Metric to Pixel Coordinates

To map image-plane metric coordinates to pixel units, we introduce the *intrinsic matrix*:

$$\mathbf{K} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix},$$

where f_x, f_y are focal lengths in pixel units, (c_x, c_y) is the principal point, and s is a skew parameter (often zero). Thus pixel coordinates are

$$\tilde{\mathbf{u}} = \mathbf{K} \begin{bmatrix} x \\ y \\ 1 \end{bmatrix}.$$

2.5 Step 4: Full Projection Matrix

Combining all steps, the mapping from a homogeneous world point to image homogeneous coordinates is:

$$\tilde{\mathbf{u}} \sim \mathbf{K}\,[\mathbf{R} \mid \mathbf{t}] \begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix}.$$

This 3×4 matrix

$$\mathbf{P} = \mathbf{K}\,[\mathbf{R} \mid \mathbf{t}]$$

is the **camera projection matrix**. After homogeneous normalization, the pixel coordinates are $(u, v) = (\tilde{u}_1/\tilde{u}_3,\ \tilde{u}_2/\tilde{u}_3)$.

3 Thin Lens

A *thin lens* is an idealized optical element whose thickness is negligible compared to its focal length. Under the thin-lens approximation one assumes that refraction occurs at a single plane (or that the two principal planes coincide), which simplifies geometric analysis while capturing essential imaging behavior of real lenses and simple lens systems.

3.1 Geometric derivation and the thin-lens equation

Consider a thin lens with focal length f. Let an object point lie on the optical axis at distance u from the lens (object distance) and let the corresponding image form at distance v on the opposite side (image distance). Using paraxial (small angle) geometry and similar triangles (see figure), one obtains the celebrated thin-lens equation:

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}. \tag{1}$$

This relation follows because rays emanating from the object that are parallel to the optical axis are focused at the focal point, and rays through the lens centre are undeviated in the thin-lens model; comparing the intercepts of these rays yields equation (1).

3.2 Sign conventions and special cases

Adopt the common sign convention: distances measured in the direction of incident light are positive. For a real object in front of the lens, $u > 0$. If $v > 0$ the image is real and forms on the side opposite the object; if $v < 0$ the image is virtual (appears on the same side as the object). When the object is at infinity ($u \to \infty$), equation (1) gives $v = f$: parallel rays focus at the focal plane. When $u = f$, $v \to \infty$ and no finite image plane exists (collimated output).

3.3 Lateral magnification

The lateral (transverse) magnification m of the lens is the ratio of image height y_i to object height y_o:

$$m = \frac{y_i}{y_o} = -\frac{v}{u}. \tag{2}$$

The negative sign indicates image inversion for a single thin lens forming a real image.

3.4 Lens maker's formula (thin lens approximation)

For a lens formed by two spherical surfaces with radii R_1 and R_2 and refractive index n (in air $n \approx n_{\text{lens}}/n_{\text{air}}$), the focal length under the thin-lens approximation is given by the lens maker's formula:

$$\frac{1}{f} = (n-1)\left(\frac{1}{R_1} - \frac{1}{R_2}\right). \tag{3}$$

This formula assumes the lens thickness is negligible; for thick lenses the principal planes shift and the effective focal length must be computed including thickness.

3.5 Practical considerations

The thin-lens model is widely used in graphics and vision because it yields simple analytic relations for focus, depth of field, and magnification. Real lenses exhibit aberrations (spherical, chromatic, astigmatism) and finite aperture effects (diffraction, circle of confusion). Depth of field and defocus blur depend on aperture diameter D and acceptable circle of confusion; these refine the idealized behavior predicted by the thin-lens approximation but the core imaging geometry is governed by equations (1) and (2).

4 Cameras with Thin Lenses

Real cameras employ lenses with finite aperture diameters, allowing control over focus and depth of field. Though practical lenses are thick and complex, they approximate thin lens behavior. Aperture size, focal length, and sensor distance govern image sharpness, while aberrations such as distortion or chromatic separation are corrected using compound lens elements.

4.1 From Pinhole to Thin Lens Models

While the pinhole and virtual pinhole camera models provide mathematically elegant projections, real cameras employ lenses to collect more light and to control focus. The thin-lens model refines the ideal pinhole by introducing a finite aperture and a focusing mechanism, enabling brighter images, adjustable depth of field, and reduced diffraction limits.

4.2 Aperture, f-number, and Exposure

The aperture of a lens is the opening that controls how much light enters the camera. Its size is often expressed by the *f-number* (or f-stop):

$$N = \frac{f}{D},$$

where f is the focal length and D is the aperture diameter. Smaller f-numbers (e.g., $f/1.8$) correspond to larger apertures, which admit more light, enabling shorter exposure times but reducing depth of field. Larger f-numbers (e.g., $f/16$) increase depth of field at the expense of brightness.

4.3 Depth of Field and Bokeh

Unlike the pinhole model, a thin lens does not bring all depths into sharp focus simultaneously. Points at distances not equal to the focus distance appear blurred into *circles of confusion*. The shape and quality of this blur is known as *bokeh*, influenced by the aperture geometry, lens aberrations, and sensor characteristics. Wide apertures yield pronounced bokeh, often desirable in portrait photography, while narrow apertures produce more uniformly sharp images.

4.4 Mathematical expression for depth of field (DOF)

Depth of field quantifies the range of object distances that appear acceptably sharp for a given focus distance s. Key quantities are the focal length f, the aperture diameter D (or f-number $N = f/D$), and the acceptable circle of confusion diameter c on the image sensor. The *hyperfocal distance* H is commonly defined as

$$H \;=\; \frac{f^2}{N\,c} + f \quad (\text{often approximated as } H \approx \frac{f^2}{N\,c}).$$

When the lens is focused at distance s, the near and far limits of acceptable focus (object distances) are typically expressed as

$$D_N \;=\; \frac{H\,s}{H + (s - f)}, \qquad D_F \;=\; \frac{H\,s}{H - (s - f)},$$

with the convention that if the denominator in D_F is non-positive then $D_F = \infty$ (i.e. the far limit extends to infinity). The depth of field is the interval

$$\text{DOF} \;=\; D_F - D_N.$$

These formulas follow from paraxial thin-lens geometry together with the requirement that the blur circle produced by points off the focus plane must not exceed the chosen circle of confusion c. For practical use, photographers often employ the approximation $H \approx f^2/(Nc)$ (neglecting the additive f when $H \gg f$), and compute DOF numerically using the formulae above. Note that smaller f-numbers (larger apertures) reduce DOF, while larger N and smaller acceptable c increase DOF.

4.5 Lens Types and Trade-offs

Different lens designs serve different imaging needs:

- **Wide-angle lenses** (short focal length) capture large fields of view, useful for landscapes but prone to distortion.
- **Telephoto lenses** (long focal length) magnify distant subjects and compress perspective, but require larger optics and have shallow depth of field.
- **Macro lenses** are optimized for close focusing, enabling detailed imaging of small objects, though they often trade field of view for magnification.
- **Fast lenses** (low f-number) allow low-light photography and shallow depth of field, but are heavier, more expensive, and more prone to aberrations.

4.6 Macro Photography

Macro photography refers to imaging subjects at close range, typically where the image size on the sensor approaches or exceeds the object's actual size (magnification $m \geq 1$). The thin-lens equation

$$\frac{1}{u} + \frac{1}{v} = \frac{1}{f}$$

governs the relationship between object distance u, image distance v, and focal length f. At close focusing distances, u is only slightly larger than f, so v must increase significantly. This pushes the image plane farther from the lens, leading to substantial extension of the camera's optics or bellows.

The magnification is given by

$$m = -\frac{v}{u},$$

so in macro photography both u and v are comparable in scale, producing large values of $|m|$. A practical technique for achieving high magnification without specialized macro lenses is to *reverse* a standard lens: when reversed, lenses designed to focus distant objects can project nearby subjects at large scale onto the sensor. Extension tubes or bellows can similarly increase the effective image distance and magnification.

Depth of field becomes extremely shallow at high magnifications, because the circle of confusion increases with magnification and the effective f-number scales as $N_{\text{eff}} = N(1 + m)$. This leads to millimeter or even sub-millimeter DOF at macro scales, requiring precise focusing, small apertures, or computational techniques such as focus stacking to produce images with acceptable sharpness across the subject.

4.7 Computational and Artistic Considerations

In computer graphics, simulating thin-lens effects adds realism through depth of field and photorealistic bokeh. In photography, lens choice and aperture setting balance competing requirements: light sensitivity, image sharpness, background separation, and portability. The thin-lens model provides a tractable basis for analyzing these trade-offs while capturing much of the richness of real imaging systems.

5 Camera Sensors

Modern cameras capture images by converting light into electrical signals using semiconductor-based sensors. The two dominant technologies are the *Charge-Coupled Device* (CCD) and the *Complementary Metal–Oxide–Semiconductor* (CMOS) sensor. Both rely on the photoelectric effect: incident photons generate electron–hole pairs in a silicon substrate, and the accumulated charge is read out and digitized.

5.1 Pixel Structure and Color Filter Arrays

A sensor consists of a two-dimensional grid of pixels, each integrating incident light over an exposure period. Since bare silicon is largely color-blind, most cameras employ a *color filter array* (CFA), such as the Bayer mosaic, which places red, green, and blue filters over different pixels. Demosaicing algorithms then interpolate to reconstruct full-color images. Some high-end or scientific sensors use alternative designs, such as Foveon layers (stacked photodiodes) or monochrome sensors with separate optical filters.

5.2 Resolution, Dynamic Range, and Noise

Sensor Sensor size plays a critical role in image quality by influencing both resolution and noise performance. For a fixed number of pixels, a larger sensor has bigger individual pixels (larger pixel pitch), which allows each pixel to collect more photons during exposure. This improves the signal-to-noise ratio (SNR), reduces photon shot noise, and enhances low-light performance. Conversely, a smaller sensor with the same pixel count has smaller pixels that capture fewer photons, increasing noise and reducing sensitivity. Regarding resolution, sensor size determines the total imaging area available for capturing spatial detail. Larger sensors can achieve higher effective resolution with the same optical system because they reduce diffraction effects and allow wider apertures without excessively shallow depth of field. Smaller sensors require lenses with shorter focal lengths to achieve the same field of view, which can introduce optical aberrations and limit sharpness. Additionally, sensor size affects depth of field and bokeh: larger sensors produce shallower depth of field for the same field of view and f-number, making background blur more pronounced. In summary, increasing sensor size generally improves noise performance and optical quality, while small sensors trade off sensitivity and low-light capability for compactness and cost savings.

Resolution. Sensor resolution refers to the total number of pixels and their individual sizes. Higher resolution allows finer spatial detail to be captured, but smaller pixels collect fewer photons per exposure, which can increase noise. Conversely, larger pixels capture more light, improving signal quality but reducing spatial sampling. For a sensor of fixed physical size, increasing the number of pixels reduces pixel pitch, which may increase diffraction effects and limit optical sharpness.

Dynamic Range. Dynamic range quantifies the ratio between the maximum and minimum detectable light intensities. It is typically expressed in stops or decades of intensity. High dynamic range allows the sensor to capture both very bright and very dark regions in a single exposure without saturation or underexposure. The dynamic range is constrained by the full well capacity of each pixel (maximum number of electrons it can hold) and the sensor noise floor (minimum detectable signal above noise).

Noise. Noise in camera sensors arises from several sources:

- **Photon shot noise:** inherent randomness of photon arrival, scales with the square root of the number of photons.

- **Readout noise:** introduced during charge conversion and amplification.
- **Dark current noise:** thermal generation of electrons in the absence of light.

The total noise determines the signal-to-noise ratio (SNR) and limits the effective sensitivity of the sensor. Noise characteristics are critical in low-light conditions, long exposures, and high-speed imaging.

Trade-offs. Designing sensors involves balancing resolution, dynamic range, and noise. For example, increasing resolution may reduce pixel size, increasing noise and limiting dynamic range. Similarly, maximizing dynamic range can require larger pixels or lower amplification, affecting spatial resolution. Understanding these trade-offs is essential for applications in photography, computer vision, and scientific imaging.

5.3 Rolling vs. Global Shutter

In CMOS sensors, pixels may be read out sequentially (*rolling shutter*) or simultaneously (*global shutter*). Rolling shutters can produce distortions in fast motion (e.g., skewed shapes in video), while global shutters preserve geometry but are more costly to implement. The choice of shutter mode depends on the application, with global shutters preferred in scientific and industrial vision systems.

5.4 Implications for Graphics and Vision

In computer graphics, accurate models of sensors enable photorealistic simulation of noise, exposure, and color responses. In computer vision, understanding sensor behavior is critical for calibration, high-dynamic-range imaging, and denoising. The sensor thus forms the final stage of the imaging pipeline, bridging optical projection with digital representation.