Image Formation & Geometric Camera Models

Adapted from: Computer Vision Lecture 2.2 (video)
and Geometric Image Formation notes

Prepared by Your Name

December 8, 2025

Outline

1. Learning Objectives
2. Why geometric image formation?
3. Pinhole Camera Model
4. Homogeneous Coordinates
5. Projection Matrix
6. Common Camera Models
7. Lens Distortion
8. Camera Calibration
9. Worked Example (Projection)
10. Takeaways
11. References

Learning Objectives

- Understand the pinhole camera model and perspective projection.
- Represent points and transformations using homogeneous coordinates.
- Build the camera projection matrix and separate intrinsics/extrinsics.
- Recognize common lens distortions and basics of calibration.

Why geometric image formation?

- Relates 3D scene geometry to 2D images.
- Foundation for stereo, structure-from-motion, pose estimation, calibration.
- Mathematical models let us invert or reason about scene geometry.

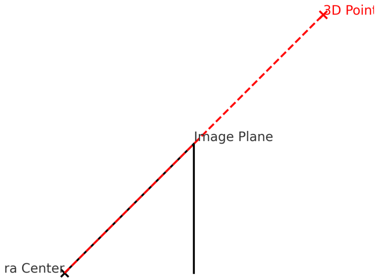
Pinhole camera model

Intuition: A pinhole camera projects 3D points onto an image plane through a single center of projection.

$$\text{3D point } \mathbf{X} = (X, Y, Z)^\top$$

$$\text{Image point } (x, y) = \left(\frac{fX}{Z}, \frac{fY}{Z}\right)$$

where f is the focal length.



Pinhole geometry (replace with own figure).

Homogeneous coordinates

- Represent Euclidean 2D/3D points with extra coordinate for linear transformations.
- 3D point: $\tilde{\mathbf{X}} = (X, Y, Z, 1)^\top$.
- 2D image (homogeneous): $\tilde{\mathbf{x}} = (x, y, 1)^\top$.
- Perspective projection becomes linear in homogeneous coordinates:

$$s\tilde{\mathbf{x}} = \mathbf{P}\,\tilde{\mathbf{X}}$$

where s is scale and $\mathbf{P}$ is the 3×4 projection matrix.

Camera projection matrix

- Full camera matrix: $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$
- $\mathbf{K}$ (intrinsics) and $[\mathbf{R} \mid \mathbf{t}]$ (extrinsics):

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

where f_x, f_y are focal lengths (pixels), s is skew, and c_x, c_y is principal point.

Interpretation of extrinsics

- World-to-camera: $\mathbf{X}_c = \mathbf{R}\mathbf{X}_w + \mathbf{t}$.
- $\mathbf{R}$ is a 3×3 rotation, $\mathbf{t}$ is translation — together they place the camera in the world.
- Combined into projection: $\tilde{\mathbf{x}} \sim \mathbf{K}[\mathbf{R} \mid \mathbf{t}]\tilde{\mathbf{X}}_w$.

Other camera approximations

- **Orthographic / scaled orthographic:** approximates perspective when depth variation is small.
- **Affine camera:** linear mapping, drops division by Z.
- **Omnidirectional / generalized cameras:** spherical, catadioptric or multi-view sensors.

Lens distortion (radial & tangential)

- Real lenses introduce radial distortion r (barrel/pincushion) and tangential terms due to misalignment.
- Simple radial correction (2–3 coefficients):

$$x_{\text{dist}} = x(1 + k_1 r^2 + k_2 r^4 + \dots)$$

where $r^2 = x^2 + y^2$ in normalized image coordinates.

Camera calibration

- Estimate **K**, **R**, **t** (and distortion) from known 3D–2D correspondences.
- Common toolboxes: Bouguet's Camera Calibration Toolbox, OpenCV calibrateCamera.
- Calibration yields parameters used to undistort images and recover metric geometry.

Example: projecting a 3D point

Given $\mathbf{X} = (X, Y, Z)^\top$, compute image point:

$$s \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} = \begin{bmatrix} f_x & s & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mathbf{R} \mid \mathbf{t} \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix}$$

Solve for $u = \frac{s_u}{s_w},\ v = \frac{s_v}{s_w}$.

Takeaways

- The pinhole model and homogeneous coordinates give a compact, linear framework for projection.
- Intrinsics vs extrinsics: intrinsics describe the camera sensor/lens, extrinsics describe camera pose.
- Calibration and distortion modeling are essential for accurate geometric tasks.

References

- Lecture video: "Computer Vision - Lecture 2.2 (Image Formation: Geometric ...)" (YouTube).
- Lecture notes: Geometric Image Formation (example course notes).
- R. Szeliski, *Computer Vision: Algorithms and Applications* and Hartley & Zisserman, *Multiple View Geometry*.

Appendix: Useful formulas

- $\mathbf{P} = \mathbf{K}[\mathbf{R} \mid \mathbf{t}]$
- Projection (homogeneous): $s\tilde{\mathbf{x}} = \mathbf{P}\tilde{\mathbf{X}}$
- Intrinsics matrix $\mathbf{K}$ shown earlier.