IMAGE FORMATION

From 3-D physical world to 2-D image:

• Analog vs digital images
• Camera geometry
• Radiometry and sensing
An image can be understood as a 2D light intensity function $f(x,y)$ where:

- $x$ and $y$ are spatial coordinates
- The value of $f$ at any point $(x, y)$ is proportional to the brightness or gray value of the image at that point.

→ Cannot be stored as such on a digital computer.
A digitized image is one in which:

- Spatial and grayscale values have been made discrete.
- Intensities measured across a regularly spaced grid in x and y directions are sampled to
  - 8 bits (256 values) per point for black and white,
  - 3x8 bits per point for color images.

They are stored as a two dimensional arrays of gray-level values. The array elements are called pixels and identified by their x, y coordinates.
Projection from surfaces to 2-D sensor.

- Where: Geometry
- How bright: Radiometry
- Stored how: Sensing
PINHOLE CAMERA MODEL

Idealized model that defines perspective projection:

- All rays go through a hole and form a star of lines
- The hole acts as a selector of rays that allows the formation of an inverted image.
VIRTUAL IMAGE

Diagram showing a virtual image formation with a pinhole and image plane.
CAMERA GEOMETRY

Pinhole geometry without image reversal
COORDINATE SYSTEMS

World, Camera, Image Coordinate Systems

**World Coordinate System:**

\((X_w, Y_w, Z_w)\)

**Camera Coordinate System:**

\((X_c, Y_c, Z_c)\)

**Image Coordinate System:**

\((X_i, Y_i, Z_i)\)
• The center of the projection coincides with the origin of the world.
• The camera axis (optical axis) is aligned with the world’s z-axis.
• To avoid image inversion, the image plane is in front of the center of projection.
1D IMAGE

\[
\frac{X_f}{f} = \frac{X_c}{Z_c}
\]

\[
X_f = f \frac{X_c}{Z_c}
\]
\[ X_i = f \frac{X_c}{Z_c} \]

\[ Y_i = f \frac{Y_c}{Z_c} \]
Reformulate it as a linear operation.

\[ u = X_i = f \frac{X_c}{Z_c} \]

\[ v = Y_i = f \frac{Y_c}{Z_c} \]
VANISHING POINTS

- The projections of parallel lines all meet at one point, called the vanishing point.
- As focal length and distance to camera increase, the image remains the same size but perspective effects diminish.
HOMOGENEOUS COORDINATES

Homogeneous representation of 2D point:

\[ \mathbf{x} = (x_1, x_2, x_3) \text{ represents } \left( \frac{x_1}{x_3}, \frac{x_2}{x_3} \right) \]

Homogeneous representation of 3D point:

\[ \mathbf{x} = (x_1, x_2, x_3, x_4) \text{ represents } \left( \frac{x_1}{x_4}, \frac{x_2}{x_4}, \frac{x_3}{x_4} \right) \]

\[ \Rightarrow \text{ Projections become linear transformations.} \]
SIMPLE PROJECTION MATRIX

\[
\begin{bmatrix}
x
y
z
\end{bmatrix} =
\begin{bmatrix}
f & 0 & 0 & 0 \\
0 & f & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
X_c \\
Y_c \\
Z_c \\
1
\end{bmatrix}
\]

with \( X_i = \frac{x}{z} = f \frac{X_c}{Z_c} \) and \( Y_i = \frac{y}{z} = f \frac{Y_c}{Z_c} \)

\[
\begin{bmatrix}
f & 0 & 0 & 1 & 0 & 0 & 0 \\
f & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & f & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 \\
0 & 0 & 1 & 0 & 0 & 1 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}
\begin{bmatrix}
X_c \\
Y_c \\
Z_c \\
1
\end{bmatrix}
\]

\[
\begin{bmatrix}
X_c \\
Y_c \\
Z_c \\
1
\end{bmatrix}
\]
INTRINSIC AND EXTRINSIC PARAMETERS

- Camera may not be at the origin, looking down the z-axis
  → Extrinsic parameters
- One unit in camera coordinates may not be the same as one unit in world coordinates
  → Intrinsic parameters

\[
\begin{pmatrix}
x
y
z
\end{pmatrix} = \begin{pmatrix}
1 & 0 & 0 & 0
0 & 1 & 0 & 0
0 & 0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
X
Y
Z
1
\end{pmatrix}
\]
LINEAR CAMERA MODEL

\[
\begin{pmatrix}
  x \\
  y \\
  z
\end{pmatrix} = \begin{pmatrix}
  \text{Transformation} & 1 & 0 & 0 & 0 \\
  \text{representing} & 0 & 1 & 0 & 0 \\
  \text{intrinsic parameters} & 0 & 0 & 1 & 0
\end{pmatrix}
\begin{pmatrix}
  \text{Transformation} & X \\
  \text{representing} & Y \\
  \text{extrinsic parameters} & Z
\end{pmatrix}
\]

\[
= K \begin{pmatrix}
  1 & 0 & 0 & 0 \\
  0 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0
\end{pmatrix} \begin{pmatrix}
  X \\
  Y \\
  Z
\end{pmatrix},
\]

where \( K \) is a 3x3 matrix and \( R_t \) a 4x4 matrix.
PRINCIPAL POINT

\[ u = X_i + p_u = \frac{fX}{Z} + p_u \]

\[ v = Y_i + p_v = \frac{fY}{Z} + p_v \]

\[
\begin{bmatrix}
    f & p_u \\
    f & p_v \\
    1 & 1 
\end{bmatrix}
\]
INHOMOGENEOUS SCALING

\[ u = \alpha_u X_i + p_u = \alpha_u X / Z + p_u \]
\[ v = \alpha_v Y_i + p_v = \alpha_v Y / Z + p_v \]

\[ K = \begin{bmatrix} \alpha_u & p_u \\ \alpha_v & p_v \\ 1 & 1 \end{bmatrix} \]

The pixels are not necessarily square.
**AXIS SKEW**

\[
K = \begin{bmatrix}
\alpha_u & s & p_u \\
\alpha_v & p_v & 1
\end{bmatrix}
\]

\(s\) encodes the non-orthogonality of the \(u\) and \(v\) directions. Very close to zero in modern cameras.
Rotations and translations also expressed in terms of matrix multiplications in projective space.

\[
\begin{bmatrix}
X_c \\
Y_c \\
Z_c
\end{bmatrix}
= \begin{bmatrix}
X_w \\
Y_w \\
Z_w
\end{bmatrix} - \tilde{C}
\]

with \( R^tR = I \)

\[
\begin{bmatrix}
X_c \\
Y_c \\
Z_c
\end{bmatrix}
= \begin{bmatrix}
R & T \\
0 & 1
\end{bmatrix}
\begin{bmatrix}
X_w \\
Y_w \\
Z_w
\end{bmatrix}
\]

with \( T = -R\tilde{C} \)
FULL PROJECTION MATRIX

\[
\begin{bmatrix}
  x \\
  y \\
  z
\end{bmatrix} =
\begin{bmatrix}
  \alpha_u & s & p_u & 1 & 0 & 0 & 0 \\
  0 & \alpha_v & p_v & 0 & 1 & 0 & 0 \\
  0 & 0 & 1 & 0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
  R & T \\
  0 & 1 \\
  0 & 1
\end{bmatrix}
\]

with \( T = -R\tilde{C} \) and \( R'R = I \)

\[
= \begin{bmatrix}
  p_{11} & p_{12} & p_{13} & p_{14} \\
  p_{21} & p_{22} & p_{23} & p_{24} \\
  p_{31} & p_{32} & p_{33} & 1
\end{bmatrix}
\begin{bmatrix}
  X \\
  Y \\
  Z \\
  1
\end{bmatrix}
\]

\[
x = K[R \mid -R\tilde{C}]X = KR[I \mid -\tilde{C}]X = [M \mid m]X = PX
\]

Hartley, Chap 6.
CAMERA CALIBRATION

Internal Parameters:
- Horizontal and vertical scaling (2)
- Principal points (2)
- Skew of the axis (1)

External Parameters:
- Rotations (3)
- Translations (3)

→ 11 free parameters.
LIMITATIONS

Idealization because the hole cannot be infinitely small

- Image would be infinitely dim
- Diffraction effects

→ Use of Lenses.
Thin lens Equation:

\[ \frac{1}{u} + \frac{1}{v} = \frac{1}{f} \]

→ Lens with focal distance \( f \) equivalent to pinhole camera with similar focal distance but larger aperture.
DISTORTIONS

The lens is not exactly a “thin lens:”

- Different wave lengths refracted differently
- Barrel Distortion
Once the image is undistorted, the camera projection can be formulated as a projective transform.

→ The pinhole camera model applies.
**RADIOMETRY**

**Scene Radiance:** Amount of light radiation from a surface point (Watt / m² / Steradian)

**Image Irradiance:** Amount of light incident at the image of the surface point. (Watt / m²)

Fundamental Radiometric Equation:

\[
Irr = \frac{\pi}{4} \left(\frac{d}{f}\right)^2 \cos^4(\alpha) \text{Rad}
\]
DE VIGNETTING

Y. Zheng, S. Lin, and S.B. Kang, CVPR’08
SENSOR ARRAY

- Photons free up electrons that are then captured by a potential well.
- Charges are transferred row by row wise to a register.
- Pixel values are read from the register.
SENSING

Conversion of the “optical image” into an “electrical image”:

\[ E(x, y) = \iiint \text{Irrad}(x, y, t, \lambda) s(\lambda) \tau(t - t_0) d\lambda dt \]

\[ I(m, n) = \text{Quantize}(\iiint E(x, y) \omega(x - m, y - n) dxdy) \]

\[ \rightarrow \text{Quantization in} \]
- Time
- Space
IN SHORT

- Camera geometry can be modeled in terms of the pinhole camera model, which is linear in projective space.

- Image radiance is roughly proportional to surface radiance and the two can be used interchangeably for our purposes.