Image formation

- A Bayer filter for constructing colour data has twice as many green, as red and blue to mimic how the human eye responds to light. Green light is detected more due to both L and M cones being used during daylight vision & are most sensitive to green light.

Camera Model I

- Pinhole cameras- exactly one ray passes through each point in the image plane, the pinhole and the scene
- Pinhole perspective is also known as central perspective
- Size of objects in image plane is related to their distance from the focal point.
- The point C' which passes through the pinhole along a vector $\mathbf{k}$ and is perpendicular to the image plane Π' is called the **Image centre** and plays an important role in camera calibration

Camera Model II

- When using Cartesian coordinates to represent projection

  If we have a ray travelling along a vector from $P = \begin{bmatrix} x \\ y \\ z \end{bmatrix}$ which passes through the pinhole, we can calculate its projected position $P' = \begin{bmatrix} x' \\ y' \\ z' \end{bmatrix}$ as $\begin{bmatrix} f'\frac{x}{z} \\ f'\frac{y}{z} \\ f' \end{bmatrix}$ Where f' is the distance between the image plane and the pinhole (perpendicular to the image plane)

  While this method is *easy* it is only an approximation and this only holds for scenes in which magnification across all objects can be taken as constant. I.e. scenes which have little to no depth.

  This is known as **Scaled Orthography**

Camera Model III

We can generalise this further by assuming the constant magnification, $m = -1$, allowing us to simply directly map $x = x'$ and $y = y'$

Size of the pinhole

- Pinhole too big $\rightarrow$ blurry image
- Pinhole the correct size $\rightarrow$ a dim but sharp image
- Pinhole too small $\rightarrow$ blurry image

Generally pinhole cameras produce dark images as very little light is able to get through to the screen.
This is solved through the use of lenses which allow cameras to gather more light (pinhole too big) but focus it s.t. the image is not blurry.
See Snell's law for lens equations:

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2$$

Thin lens

For a thin *lens* the following equation holds:

$$\frac{1}{z'} - \frac{1}{z} = \frac{1}{f}$$

Where:
f is the focal length of the lens:

$$f = \frac{R}{2(n-1)}$$

Geometric properties of projection

- Points $\mapsto$ points
- Lines $\mapsto$ lines
- Planes $\mapsto$ whole image
- Polygons $\mapsto$ polygones
- **Degenerate cases:**
  - Lines through focal point $\mapsto$ point
  - Plane through focal point $\mapsto$ line
- 3D objects (polyhedra) project to polygons as only their outermost edge is captured and lines $\mapsto$ lines.

Camera Parameters I

Camera parameters include: Intrinsic parameters:

- Focal Length
- Principal point
- Aspect ratio
- angle between axis

Extrinsic parameters:

- Position + orientation in space

These parameters differ from camera to camera.

Camera Parameters II

There are two kinds of projection:

1. Extrinsic projections which project from the 3D world to a 3D camera space
2. Intrinsic projections which project from a 3D camera to a 2D image.

Aside: Homogeneous coordinates

Projective geometry uses Homogeneous coordinates.

$$\overbrace{\begin{bmatrix} x \\ y \end{bmatrix}}^{CartesianForm} \Rightarrow \overbrace{\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}}^{Homogeneousform}$$

To convert from Homogeneous to Euclidean: divide by last coordinate & remove it.

Camera Parameters III

Camera coordinate system

- Principal axis: A line from the camera centre perpendicular to the image plane
- Principal point, p: A point where the principal axis punctures the image plane
- Normalised camera coordinate system: A coordinate system in which the origin is at the principal point.

Camera Parameters IV

We can now rewrite our projection equations for a pinhole camera from before:
To project a 3D point in the world coordinate system to a 2D point in the image plane:

$$\begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \mapsto \begin{bmatrix} f\,\frac{x}{z} \\ f\,\frac{y}{z} \\ 1 \end{bmatrix}$$

This can also be written as a vector-matrix multiplication:

$$\overbrace{\begin{pmatrix} fX \\ fY \\ Z \end{pmatrix}}^{\mathbf{x}} = \underbrace{\begin{bmatrix} f & & & 0 \\ & f & & 0 \\ & & 1 & 0 \end{bmatrix}}_{\mathbf{P_0}} \overbrace{\begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}}^{\mathbf{X}}$$

Camera Parameters V

which can also be written

$$\mathbf{x} = \mathbf{P_0}\mathbf{X}$$

We can re-write the projection matrix $\mathbf{P_0}$ to separate the focal lengths:

$$\mathbf{P_0} = \mathrm{diag}([f, f, 1])[\mathbf{I}|0] = \begin{bmatrix} f & & \\ & f & \\ & & 1 \end{bmatrix} \begin{bmatrix} 1 & & & 0 \\ & 1 & & 0 \\ & & 1 & 0 \end{bmatrix}$$

Camera Parameters VI

Converting to image pixels from image plane

The origin of the image plane is the principal point, p. We want to translate these coordinates to have the origin at the image (bottom left) corner.

We write this transformation as:

$$\begin{bmatrix} x \\ y \\ z \end{bmatrix} \mapsto \begin{bmatrix} f\frac{x}{z} + p_x \\ f\frac{y}{z} + y_y \end{bmatrix}$$

Or in vector-matrix multiplication:

$$\begin{pmatrix} fx + zp_x \\ fy + zp_y \\ z \end{pmatrix} = \begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix}$$

Camera Parameters VII

We now want to project onto our sensor of size Ws $\times$ Hs (in metres). We represent pixels in a rectangular $M_x \times M_y$ matrix.
Let $m_x = \frac{M_x}{W_s}$ and $m_y = \frac{M_y}{H_s}$
We now construct the following projection in vector-matrix multiplication form:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \underbrace{\begin{bmatrix} m_x & 0 & 0 \\ 0 & m_y & 0 \\ 0 & 0 & 1 \end{bmatrix}}_{\text{pixel /m}} \underbrace{\begin{bmatrix} f & 0 & p_x & 0 \\ 0 & f & p_y & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}}_{\text{m}} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Which can also be written:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} \alpha_x & 0 & x_0 & 0 \\ 0 & \alpha_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

Camera Parameters VIII

It is often difficult to guarantee a perfectly rectangular sensor, so we also have a case for a skewed sensor, here we simply add a single value to the projection matrix $\mathbf{P_0}$ to form:

$$\begin{pmatrix} x \\ y \\ z \end{pmatrix} = \begin{bmatrix} \alpha_x & s & x_0 & 0 \\ 0 & \alpha_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}$$

We can decompose $\mathbf{P_0}$ into two separate matrices to allow for easier computation and reasoning, we can construct $\mathbf{P_0}$ from the product of a square matrix K and a concatination of the 3×3 identity matrix and a 3D 0-vector:

$$\mathbf{P_0} = \mathbf{K}[\mathbf{I}|0] = \begin{bmatrix} \alpha_x & s & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}$$

Camera Parameters IX

We refer to $\mathbf{K}$ as our **projection matrix** which prescribes the projection of any 3D point in the camera coordinate system onto our pixels.
Where:

- $\alpha_x = m_x \cdot f$
- $\alpha_y = m_y \cdot f$
- $x_0 = p_x \cdot m_x$
- $y_0 = p_y \cdot m_y$
- s is our skewness factor

Binary Images I

Advantages:

- Easy to acquire
- Low storage
- Simple to process

Disadvantages:

- Limited use cases
- Does not generalise to 3D
- Specialised lighting required to capture silhouettes

Binary Images II

Binary images are created from conventional images via **thresholding**. These are *cleaned* using **morphological operators**
There are different methods for thresholding:

- Above a threshold
- Within a threshold
- In a set of valid values

Morphological Operators (MOs):

- Change the shape of foreground regions via intersection or union operations between a strel and a binary image.
- Basic MOs include: Dilation & erosion

Binary Images III

Dilation

- Expands **connected** components
- Grows features
- Fills holes

If currently considered pixel is 1, set all pixels under the strel to 1.

Erosion

- Erodes connected components
- shrinks features
- Removes bridges, branches & noise

If every pixel under the strel is 1, set the currently considered output pixel to 1.

Opening is the process of eroding then dilating an image. It is good for removing small objects whilst keeping the original shape.

Closing is the process of dilating then eroding an image. It is good for filling holes but keeping the original shape.

Binary Images IV

Connected components are determined by their pixel's **connectedness**. There are different measures of connectedness including 4 and 8-connectedness.

Photometric Image Formation I

Images are made up of discrete colour or intensity values. These are related to the lighting of the surrounding environment.