EE 576 - Camera & Image Formation

H.I. Bozma

Electric Electronic Engineering
Bogazici University

February 17, 2020

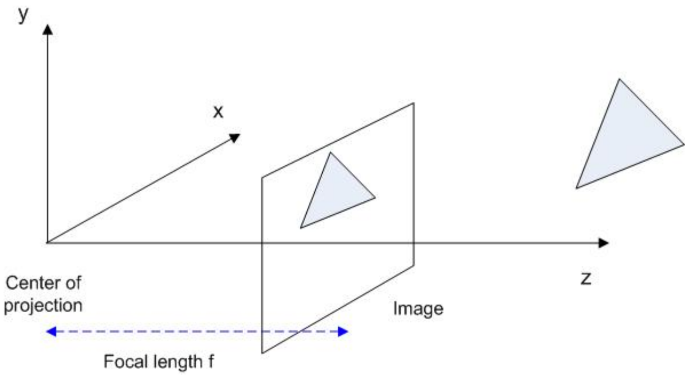
Camera models

- Goal: To understand the image acquisition process.
- Function of the camera   Similar to that of the eye in biological systems.

Pinhole camera model

- A pinhole   Infinitesimally small hole through which light enters before forming an inverted image on the camera surface facing the hole
- A pinhole camera   Modeled by placing the image plane between the focal point of the camera and the object, so that the image is not inverted.
- Perspective projection   This mapping of three dimensions onto two
- Perspective geometry   Fundamental to image formation.

Pinhole camera model: Perspective projection



Perspective geometry

- Perspective geometry $\rightarrow$ Mathematical modeling of the visual computational processes possible
- Perspective projection: projection of a 3D object onto a 2D surface by straight lines that pass through a single point.
- Simple geometry
  - $[X_1; X_2; X_3]^T$ - object coordinates
  - f - the distance of the image plane to the center of projection
  - $x = [x_1\ x_2]^T$ - image coordinates

$$x_1 = \frac{f}{X_3} X_1 \qquad x_2 = \frac{f}{X_3} X_2$$

- Non-linear equations.

Perspective geometry

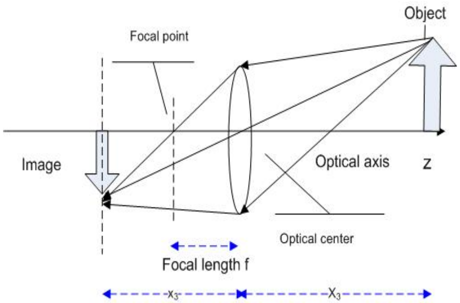
- Linear by introducing homogeneous transformations as:

$$\begin{bmatrix} U \\ V \\ S \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{bmatrix}$$

- $x_1 = \frac{U}{S}$ and $x_2 = \frac{V}{S}$ if $S \neq 0$.
- In general, a point in $R^n \rightarrow [Sx, S]^T \in R^{n+1}$, where $S \neq 0$.

Simple Lens Model

- Use lenses to focus an image onto the camera's focal plane.
- Incoming light parallel to optical axis - Focused on focal point
- Limitation – Only bring into focus those objects that lie on one particular plane that is parallel to the image plane.



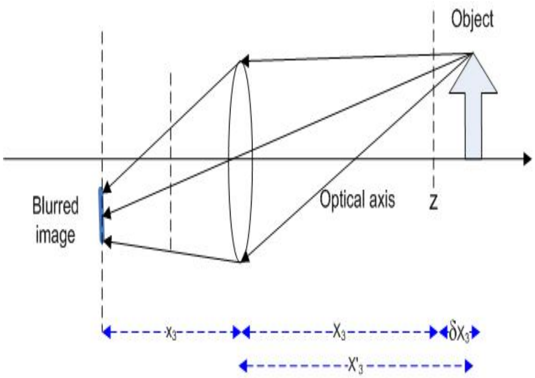
Simple Lens Model(cont)

- Assuming
  - Lens relatively thin
  - optical axis perpendicular to the image plane,
- Lens law: Image formation

$$\frac{1}{x_3} + \frac{1}{X_3} = \frac{1}{f}$$

  - X_3 - Distance of an object point from the plane of the lens,
  - x_3 Distance of the focused image from this plane,
  - f - Focal length of the lens

Simple Lens Model - Blurring and depth of field



Simple Lens Model(cont)

- Ideal lens Brings focus light from points at distance X_3.
- Points at other distances imaged as little circles
- If objects at distance X_3 are correctly focused, an object at distance X_3'.

$$x_3' - x_3 = \frac{f^2(X_3' - X_3)}{(X_3' - f)(X_3 - f)}$$

- The blurring will be with circles whose radius is

$$\propto \frac{d}{x_3}\left|x_3' - x_3\right|$$
$$\propto \frac{d}{x_3}\left|X_3' - X_3\right|$$

- The depth of field = The range of distances $|X_3' - X_3|$ which the objects are focused sufficiently well

Lens Selection - **Focal Length**

Amount of subject matter that can be viewed relative to the subject distance.

- Length: Short (Wide angle), normal, long (telephoto)
- Zoom - Fixed, variable

28 mm lens

50 mm lens

70 mm lens

210 mm lens

Selection of Lens

- **Aperture:** This refers to the size of the opening in the lens for light (iris).
  - Depth-of-field (the amount of scene in sharp focus): As aperture $\downarrow$, depth-of-field $\uparrow$
  - f-stop: The focal length divided by the "effective" aperture diameter. As f-stop $\downarrow$, light $\uparrow$. The amount of light transmitted to the film (or sensor) decreases with the f-number squared. Doubling the f-number increases the necessary exposure time by a factor of four.
  - Speed: Fast (f1.2, f1.4) $\Longrightarrow$ Slow (f3.5)

Selection of Camera

- **Format:** This refers to the size of the camera pickup device ($\frac{1}{2}, \frac{2}{3}, 1$ inch). As the format decreases, the pickup becomes darker.
- **Mount type:** C-mount standard
- Installation (size & weight)
- Vibration (Locking screws)

Coordinate Systems

- Imagine we have a three dimensional coordinate system
  - Origin at the center of projection (also called the optical center)
  - X_3 axis is along the optical axis.
- This coordinate system is called the standard coordinate system of the camera.
- These coordinates are with respect to a coordinate system where
  - The origin is at the intersection of the optical axis and the image plane,
  - x_1 and x_2 axes are parallel to the X_1 and X_2 axes.

World Image Coordinate Transformations

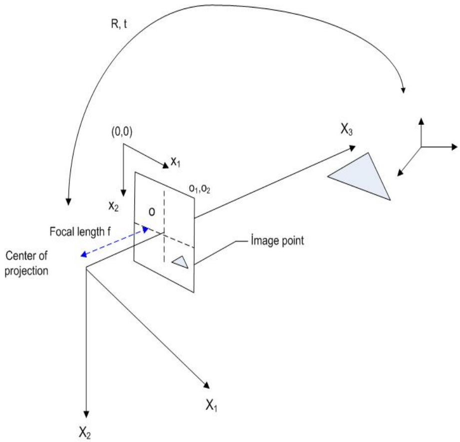


Figure: Coordinate transformations between world and image coordinate

Image Formation(cont.)

- This can be written in homogeneous coordinates as:

$$\begin{vmatrix} sx_1 \\ sx_2 \\ s \end{vmatrix} = \begin{vmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix} \begin{vmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{vmatrix}$$

- $s \neq 0$ and f is the effective focal length1 of the camera and is often measured in pixel units.
- Not the same as the focal length (e.g., a 9mm lens) marked on the lens of the camera.

Image Formation(cont.)

- Now, the actual pixel coordinates $u = [u_1, u_2]^T$ are defined with respect to an origin in the top left hand corner of the image plane, and will satisfy

$$\begin{aligned} u_1 &= o_1 + x_1 \\ u_2 &= o_2 + x_2 \end{aligned}$$

- Express the transformation from three dimensional world coordinates to image pixel coordinates using a 3×4 matrix.

$$\begin{aligned} X_3 u_1 &= X_3 o_1 + X_3 x_1 = X_3 o_1 + X_1 f \\ X_3 u_2 &= X_3 o_2 + X_3 x_2 = X_3 o_2 + X_2 f \end{aligned}$$

Image Formation(cont.)

- Substituting

$$\begin{aligned} X_3 u_1 &= X_3 o_1 + X_1 f \\ X_3 u_2 &= X_3 o_2 + X_2 f \end{aligned}$$

$$\begin{vmatrix} su_1 \\ su_2 \\ s \end{vmatrix} = \begin{vmatrix} f & 0 & o_1 & 0 \\ 0 & f & o_2 & 0 \\ 0 & 0 & 1 & 0 \end{vmatrix} \begin{vmatrix} X_1 \\ X_2 \\ X_3 \\ 1 \end{vmatrix}$$

where the scaling factor s has value X_3.

Image Formation(cont.)

- In short hand notation, we write this as

$$\widetilde{u} = K\widetilde{X}$$

where

  - $\widetilde{u}$ is the homogeneous vector of image pixel coordinates
  - K is the perspective projection matrix
  - $\widetilde{X}$ is the homogeneous vector of world coordinates.

Intrinsic Camera Parameters

There are four intrinsic camera parameters:

- Two are for the position of the origin of the image coordinate frame, and
- Two are for the scale factors of the axes of this frame.
- A camera $\rightarrow$ A system that performs a linear projective transformation from the projective space that is subset of $R^3 \times S^3$ into the projective plane X.
- There are three camera parameters – :
  1. The focal length f,
  2. o_1 and o_2
- Intrinsic parameters: Do not depend on position and orientation of the camera in space

Aspect Ratio

- Some old-fashioned CCD cameras   non-square pixels.
  - Aspect ratio $\neq 1 \rightarrow$ Different scalings in the u and v-axes (e.g., a sphere would appear as an ellipse in the image).
  - Two terms f_1 and f_2 : To describe the effective focal length.
  - The term f_1 - The effective focal length in the u pixel units
  - f_2 - The effective focal length in the v pixel units

As all modern cameras have unit aspect ratio, assume $f_1 = f_2 = f$.

Imaging Distortions

- Radial distortion: Distortion as a function of r (Fish-eye efffect)

$$\begin{aligned} x_c &= x(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \\ y_c &= y(1 + k_1 r^2 + k_2 r^4 + k_3 r^6) \end{aligned}$$

- Tangential distortion: Lenses not perfectly parallel to the imaging plane.

$$\begin{aligned} x_c &= x + (2p_1 xy + p_2(r^2 + 2x^2)) \\ y_c &= y + (p_1(r^2 + 2y^2) + 2p_2 xy) \end{aligned}$$

- Distortion coefficients: $[k_1 k_2 p_1 p_2 k_3]$

Extrinsic Camera Parameters

There are six extrinsic camera parameters:

- Three are for the position of the center of projection,
- Three are for the orientation of the image plane coordinate frame.

Coordinate Transformations

- In general, 3D world coordinates of a point – Not specified in frame with origin at the center of projection and X_3 axis lies along the optical axis.
- A more convenient frame – more likely be specified,

Coordinate Transformations - 2

- Consider the transformation of coordinate system from this other frame to the standard coordinate system

$$\widetilde{u} = KT\widetilde{X}$$

where T is a 4×4 homogeneous transformation matrix.

$$T = \begin{vmatrix} R & t \\ 0_3^T & 1 \end{vmatrix}$$

1. The top 3×3 part - Rotation matrix R and encodes camera orientation
2. The final column - Homogeneous vector containing t corresponding camera translation from the world frame origin

Extrinsic Parameters

- 6 degrees of freedom - **Extrinsic camera parameters.**
- the camera calibration matrix C: Combine 3×4 matrix K and 4×4 matrix $T \rightarrow 3 \times 4$ as:

$$C = \begin{vmatrix} fr_1 + o_1 r_3 & ft_1 + o_1 t_3 \\ fr_2 + o_2 r_3 & ft_2 + o_2 t_3 \\ r_3 & t_3 \end{vmatrix}$$

$$R = \begin{vmatrix} r_1 \\ r_2 \\ r_3 \end{vmatrix} t = \begin{vmatrix} t_1 \\ t_2 \\ t_3 \end{vmatrix}$$

- The vectors r_i - Row vectors of the rotation matrix
- The matrix C, like the matrix K, – Rank three.

Camera Parameters

- Camera calibration
  - In order to deduce three-dimensional geometric information from an image $\rightarrow$ The parameters that relate the position of a point in a scene to its position in the image.
- Estimating the intrinsic and extrinsic parameters of a camera.

Calibration Procedure

- Calibration - Estimating the intrinsic and extrinsic parameters of the camera.
- **Calibration Problem**: Asume that we are given N points $X^j \in R^3$ as well as their corresponding x^j, $j = 1, \cdots, N$.
- A two stage process:
  1. Estimating the matrix C,
  2. Estimating the intrinsic and extrinsic parameters from C.
- In many cases, the second stage is not necessary.
- Solution: Linear and non-linear methods

Solving for Calibration Matrix

The approaches vary in the numerical methods used to solve the setup problem.

- Least Squares Approach ⇒ Matrix inverse finding & matrix algebra
- Lagrangian Formulation ⇒ Multi-dimensional optimization

General Formulation

Transforming the transformation equations into a linear form
In homogeneous coordinates, the relationship between the image points with coordinates x^j and the 3D reference point coordinates X^j

$$\begin{vmatrix} s^j u_1^j \\ s^j u_2 \\ s^j \end{vmatrix} = \begin{vmatrix} q_{11} & q_{12} & q_{13} & q_{14} \\ q_{21} & q_{22} & q_{23} & q_{24} \\ q_{31} & q_{32} & q_{33} & q_{34} \end{vmatrix} \begin{vmatrix} X_1^j \\ X_2^j \\ X_3^j \\ 1 \end{vmatrix} \rightarrow s^j \begin{vmatrix} u_1^j \\ u_2^j \\ 1 \end{vmatrix} = \begin{vmatrix} q_1^T & q_{14} \\ q_2^T & q_{24} \\ q_3^T & q_{34} \end{vmatrix} \widetilde{X^j}$$

where $q_k^T = [q_{k1} q_{k2} q_{k3}]$.

General Formulation

Set $q_{34} = 1$.
Substituting for $s = X_1^j q_{31} + X_2^j q_{32} + X_3^j q_{33} + 1$,

$$\begin{aligned}\left(X_1^j q_{31} + X_2^j q_{32} + X_3^j q_{33} + 1\right) u_1^j &= X_1^j q_{11} + X_2^j q_{12} + X_3^j q_{13} + q_{14} \\ \left(X_1^j q_{31} + X_2^j q_{32} + X_3^j q_{33} + 1\right) u_2^j &= X_1^j q_{21} + X_2^j q_{22} + X_3^j q_{23} + q_{24}\end{aligned}$$

Define $x \in R^{2N}$ as $x = \sum e^j \otimes x^j$.
Let $q \in R^{11}$ be the concatenated vector
$q^T = [q_{11} q_{12} q_{13} q_{14} q_{21} q_{22} q_{23} q_{24} q_{31} q_{32} q_{33}]$
Letting D to be the large matrix consisting of known image and world coordinates with dimensions $2N \times 11$
$d \in R^{2N}$ as $d^T = \left[u_1^1 u_2^1 \cdots u_1^N u_2^N\right]$

General Formulation

$$Dq = d$$

$$D = \begin{vmatrix} X_1^1 & X_2^1 & X_3^1 & 1 & 0 & 0 & 0 & 0 & -u_1^1 X_1^1 & -u_1^1 X_2^1 & -u_1^1 X_3^1 \\ 0 & 0 & 0 & 0 & X_1^1 & X_2^1 & X_3^1 & 1 & -u_2^1 X_1^1 & -u_2^1 X_2^1 & -u_2^1 X_3^1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ X_1^N & X_2^N & X_3^N & 1 & 0 & 0 & 0 & 0 & -u_1^N X_1^N & -u_1^N X_2^N & -u_1^N X_3^N \\ 0 & 0 & 0 & 0 & X_1^N & X_2^N & X_3^N & 1 & -u_2^N X_1^N & -u_2^N X_2^N & -u_2^N X_3^N \end{vmatrix}$$

Least Squares Problem

$$\|Dq - d\|^2 = (Dq - d)^T (Dq - d)$$

Our objective:

$$min_q (Dq - d)^T (Dq - d)$$

Differentiating wrt q and setting to 0 gives

$$\begin{aligned} D^T (Dq - d) &= 0 && (1) \\ \implies D^T Dq &= D^T d && (2) \\ \implies q &= \left(D^T D\right)^{-1} D^T d && (3) \end{aligned}$$

$(D^T D)^{-1} D^T$ is known as the *pseudo-inverse* of D.

Least Squares Problem (cont.)

- q can only be estimated if $D^T D$ is invertible.
- With D being a $2N \times 11$ matrix $\rightarrow$ $D^T D$ must be of full rank.
- At least 11 equations which means that $N = 6$.
- Furthermore, the rank of D ≥ 11.
- The reference points $\{X_i | 1 = i = N, N = 6\}$ must not lie in a certain configuration, which can be defined mathematically.
- If six or more points are chosen at random, and do not lie on a plane, then $\rightarrow$ This situation will not occur

Lagrangian Formulation

$$D'q' = 0$$

D – $2N \times 12$ matrix $D = [D - d]$
$q' = [q \; q_{34}]$ – concatenated vector with unknown q_{34}.
Problem Formulation: Minimizing $(D'q')^T(D'q')$ subject to the constraint that

$$q'^T q' - 1 = 0$$

Let γ be a Lagrange multiplier. The Lagrangian to be minimized:

$$(D'q')^T(D'q') + \gamma(q'^T q' - 1)$$

Lagrangian Formulation (cont.)

Taking gradient wrt q' ,

$$\begin{aligned} D'^T D' q' - \gamma q' &= 0 \\ q'^T q' &= 1 \end{aligned}$$

With some mathematical manipulation

$$\begin{aligned} q'^T D'^T D' q' &= \gamma q'^T q' \\ &= \gamma \end{aligned}$$

Minimize $(D'q')^T(D'q') \rightarrow$ Minimize γ.
γ is an eigenvalue with corresponding q'.
Conclude that q' should be the eigenvector that has the smallest eigenvalue of $D'^T D'$.

Lagrangian Formulation (cont.)

From matrix algebra, $rank(D'^T D') + null(D'^T D') = m \; m = 12.$
In our case, $N > m$ – hence the following three cases:

1. $rank(D'^T D') = 12 \implies$ One solution to the system
2. $rank(D'^T D') = 11$ There is a unique solution up to a scale factor.
3. $rank(D'^T D') < 11$ Infinite solutions

In real applications, $rank(D') = 12$ as noise affects the rank of the matrix.
The smallest eigenvalue of $D'^T D' \neq 0$, but a small positive number.
The amount of noise $\approx$ Ratio of smallest and largest eigenvalues $D'^T D'$.

Solving for Camera Parameters

C – Defined up to a unknown scale factor s, so that $\|q_3\| \neq 1$
First recall that

$$C = \begin{vmatrix} fr_1 + o_1 r_3 & ft_1 + o_1 t_3 \\ fr_2 + o_2 r_3 & ft_2 + o_2 t_3 \\ r_3 & t_3 \end{vmatrix} = s \begin{vmatrix} q_1^T & q_{14} \\ q_2^T & q_{24} \\ q_3^T & q_{34} \end{vmatrix}$$

Solving for Camera Parameters (cont.)

- r_3 – A row of the rotation matrix $R \rightarrow \|sq_3\| = \|r_3\| = 1 \rightarrow$

$$s = \pm \frac{1}{\|q_3\|} \qquad \text{2 solutions}$$

- $r_3 = sq_3$ (2 solutions)
- $$t_3 = sq_{34} = \pm \frac{q_{34}}{\|q_3\|} \qquad \text{2 solutions}$$

Solving for Camera Parameters (cont.)

$$s^2 q_1^T q_3 = (s q_1^T)(s q_3) = (f r_1 + o_1 r_3)^T r_3 = o_1 r_3^T r_3 = o_1 \Longrightarrow o_1 = s^2 q_1^T q_3$$

Similarly

$$s^2 q_2^T q_3 = (s q_2^T)(s q_3) = (f r_2 + o_2 r_3)^T r_3 = o_2 r_3^T r_3 = o_2 \Longrightarrow o_2 = s^2 q_2^T q_3$$

Solving for Camera Parameters (cont.)

Consider $s^2 q_1 \times q_3$ and first note that $\|r_1 \times r_3\| = 1$

$$s^2 q_1 \times q_3 = (sq_1) \times (sq_3) = (fr_1 + o_1 r_3) \times r_3 = fr_1 \times r_3 \implies f = \pm s^2 \|q_1 \times q_3$$

Also noting that $\|r_2 \times r_3\| = 1$

$$s^2 q_2 \times q_3 = (sq_2) \times (sq_3) = (fr_2 + o_1 r_3) \times r_3 = fr_2 \times r_3 \implies f = \pm s^2 \|q_2 \times q_3$$

These two values must be identical ideally. If not, reconsider the accuracy of the data and the calibration computation procedure

Solving for Camera Parameters (cont.)

$$
\begin{aligned}
r_1 &= \frac{s}{f}(q_1 - o_1 q_3) \\
r_2 &= \frac{s}{f}(q_2 - o_2 q_3) \\
t_1 &= \frac{1}{f}(s q_{14} - o_1 t_3) \\
t_2 &= \frac{1}{f}(s q_{24} - o_2 t_3)
\end{aligned}
$$

Also to be noted:

$$
\begin{aligned}
(q_1 \times q_3)^T (q_2 \times q_3) &= ((f r_1 + o_1 r_3) \times r_3^T)((f r_2 + o_2 r_3) \times r_3 \\
&= (f r_1 \times r_3)^T (f r_2 \times r_3) \\
&= 0
\end{aligned}
$$

Solving for Camera Parameters (cont.)

In regards to the found parameters,

- Due to the $\pm$ in the computations of s and f, four sets of solutions exist.
- Each correspond to whether the origin of the coordinates is in front of or behind the camera and the choice of the optical axis.
- Two solutions can be eliminated if a right-hand coordinate system is adapted.

Orthographic Projection

Consider translation of f along the X_3-axis of standard coordinate frame,

- the origin and the center of the image plane - coincident
- The focal point - Positioned at $X_3 = -f$.
- No rotation involved in this transformation, hence $\rightarrow$

$$C = \begin{vmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & f \end{vmatrix}$$

Orthographic Projection

Consider translation of f along the X_3-axis of standard coordinate frame,

- Assuming that the pixel width and height are both 1.

$$C = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{f} & 1 \end{vmatrix} \text{ Letting } f \to \infty \; C = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{vmatrix}$$

- Orthographic projection parallel to X_3 axis $\to$ $x_1 = X_1$ and $x_2 = X_2$

- D. G. Lowe, “Object recognition from local scale-invariant features,” in *ICCV*, 1999, pp. 1150–1157.
- J. Sivic and A. Zisserman, “Efficient visual search of videos cast as text retrieval,” *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 31, no. 4, pp. 591–606, April 2009.
- A. Torralba, K. P. Murphy, W. T. Freeman, and M. A. Rubin, “Context-based vision system for place and object recognition,” *Computer Vision, IEEE Int. Conf. on*, vol. 1, p. 273, 2003.
- O. Erkent and H. I. Bozma, “Bubble Space and Place Representation in Topological Maps,” *The Int. J. of Rob. Res.*, vol. 32, no. 6, pp. 671 – 688, 2013.