Image Analysis and Formation
(Formation et Analyse d'Images)

James L. Crowley

ENSIMAG 3 - MMIS Option MIRV First Semester 2010/2011

Lesson 2 4 Oct 2010

Projective Camera Models

Lesson Outline:

Coordinate Transforms in 2D

Homogeneous coordinates allow us to unify projective transformations a matrix multiplication. This includes transformations for:

- Translation
- Rotation
- Sheer
- Scale Changes
- Projective Transformation.

Translation in Homogeneous Coordinates.

$x_2 = x_1 + t_x,$
$y_2 = y_1 + t_y$

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & t_x \\ 0 & 1 & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}$$

In tensorial notation:

$P^B = \mathbf{T}_A^B P^A$ for A, B = 1, 2, 3.

Thus $\mathbf{T}_A^B$ is a translation from the A reference frame to the B reference frame.

Rotation

Rotation by an angle of θ can be written as:

$x_2 = \mathrm{Cos}(\theta)\, x_1 + \mathrm{Sin}(\theta)\, y_1,$
$y_2 = \mathrm{Sin}(\theta)\, x_1 - \mathrm{Cos}(\theta) y_1$

In matrix form.

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} \mathrm{Cos}(\theta) & \mathrm{Sin}(\theta) & 0 \\ -\mathrm{Sin}(\theta) & \mathrm{Cos}(\theta) & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}$$

In tensor form: $P^B = \mathbf{R}_A^B P^A$

Translation and Rotation

Combining translation and rotation:

$$x_2 = \mathrm{Cos}(\theta)\ x_1 + \mathrm{Sin}(\theta)\ y_1 + t_x$$
$$x_2 = \ \mathrm{Sin}(\theta)\ x_1 - \mathrm{Cos}(\theta)\ y_1 + t_y$$

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} \mathrm{Cos}(\theta) & \mathrm{Sin}(\theta) & t_x \\ -\mathrm{Sin}(\theta) & \mathrm{Cos}(\theta) & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}$$

Note, express the position of the origin of the source coordinates O^1 in the destination coordinates.

Scale Change

Change scale by a factor of 1/s:

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & s \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}$$

Rotation, translation and scale (s_x and s_y) in the same matrix

$$\boxed{\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} s_x\,\mathrm{Cos}(\theta) & s_y\,\mathrm{Sin}(\theta) & t_x \\ -s_x\,\mathrm{Sin}(\theta) & s_y\,\mathrm{Cos}(\theta) & t_y \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}}$$

$$x_2 = s_x\mathrm{Cos}(\theta)\ x_1 + s_y\ \mathrm{Sin}(\theta) y_1 + t_x,$$
$$y_2 = s_y\mathrm{Sin}(\theta)\ x_1 - s_x\ \mathrm{Cos}(\theta) y_1 + t_y$$

Affine transformation (rotation, translation and sheer)

$$\begin{pmatrix} x_2 \\ y_2 \\ 1 \end{pmatrix} = \begin{pmatrix} a & b & c \\ d & e & f \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_1 \\ y_1 \\ 1 \end{pmatrix}$$

Projection between two planes (Homographies)

The projective transformation from one plane to another is called a homography.
A homography is bijective (reversible).

as a matrix :

$$\begin{pmatrix} x_b \\ y_b \\ 1 \end{pmatrix} = \begin{pmatrix} wx_b \\ wy_b \\ w \end{pmatrix} = \begin{pmatrix} h_1^1 & h_2^1 & h_3^1 \\ h_1^2 & h_2^1 & h_3^1 \\ h_1^3 & h_2^3 & 1 \end{pmatrix} \begin{pmatrix} x_a \\ y_a \\ 1 \end{pmatrix}$$

$$x_B = \frac{wx_B}{w} = \frac{h_{11}x_A + h_{12}y_A + h_{13}}{h_{31}x_A + h_{32}y_A + h_{33}}$$

$$y_B = \frac{wy_B}{w} = \frac{h_{21}x_A + h_{22}y_A + h_{23}}{h_{31}x_A + h_{32}y_A + h_{33}}$$

In tensor notation

$$\boxed{Q^B = \mathbf{H}_A^B \; P^A}$$

$$\begin{pmatrix} q^1 \\ q^2 \\ q^3 \end{pmatrix} = \mathbf{H}_A^B \begin{pmatrix} p^1 \\ p^2 \\ p^3 \end{pmatrix} = \begin{pmatrix} h_1^1 & h_2^1 & h_3^1 \\ h_1^2 & h_2^2 & h_3^2 \\ h_1^3 & h_2^3 & h_3^3 \end{pmatrix} \begin{pmatrix} p^1 \\ p^2 \\ p^3 \end{pmatrix}$$

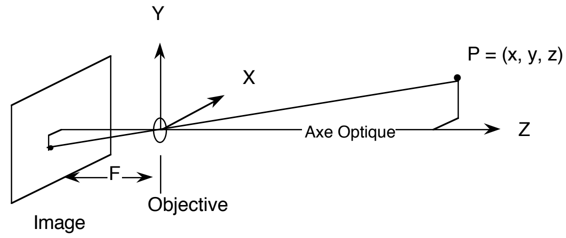
$$x_B = \frac{q^1}{q^3} \qquad y_B = \frac{q^2}{q^3}$$

The Camera Model

A "camera" is a closed box with an aperature (a "camera obscura"). Photons are reflected from the world, and pass through the aperture to form an image on the retina. Thus the camera coordinate system is defined with the aperture at the origin.

The Z (or depth) axis runs perpendicular from the retina through the aperture.
The X and Y axes define coordinates on the plane of the aperture.

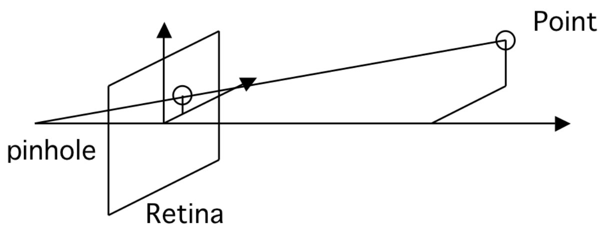
The Pinhole Camera



Points in the scene are projected to an "up-side down" image on the retina.

This is the "Pin-hole model" for the camera.

The scientific community of computer vision often uses the "Central Projection Model". In the Central Projection Model, the retina is placed in Front of the projective point.



We will model the camera as a projective transformation from scene coordinates, S, to image coordinates, i.

$$\vec{Q}^i = \mathbf{M}_s^i \vec{P}^s$$

This transformation is expressed as a 3x4 matrix:

$$M_s^i = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ m_{31} & m_{32} & m_{33} & m_{34} \end{pmatrix}$$

composed from 3 transformations between 4 reference frames.

Extrinsic and Intrinsic camera parameters

The camera model can be expressed as a function of 11 parameters.
These are often separated into 6 "extrinsic" parameters and 5 "intrinsic" parameters:

Thus the "Extrinsic" parameters of the camera describe the camera position and orientation in the scene. These are the six parameters:

Extrinsic Parameters = $(x, y, z, \theta, \varphi, \gamma)$

The intrinsic camera parameters express the projection to the retina, and the mapping to the image. These are :

F : The "focal" length
C_x, C_y : the image center (expressed in pixels).
D_x, D_y : The size of pixels expressed in pixels/mm.

Coordinate Systems

This transformation can be decomposed into 3 basic transformations between 4 reference frames. The reference frames are:

Coordinate Systems:

Scene Coordinates:

Point Scène: $P^s = (x_S, y_S, z_S, 1)^T$

Camera Coordinates:

external world: $P^c = (x_C, y_C, z_C, 1)^T$

Retina: $Q^r = (x_r, y_r, 1)^T$

Image Coordaintes

Image: $Q^i = (i, j, 1)^T$

The transformations are represented by Homogeneous projective transformations.

$$\vec{Q}^i = C_r^i \vec{Q}^r \qquad \vec{Q}^r = P_c^r \vec{P}^c \qquad \vec{P}^c = T_s^c \vec{P}^s$$

These express

1) A translation/rotation from scene to camera coordinates: T_s^c
2) A projection from scene points in camera coordinates to the retina: P_c^r

3) Sampling scan and A/D conversion of the retina to give an image: C_r^i

When expressed in homogeneous coordinates, these transformations are composed as matrix multiplications.

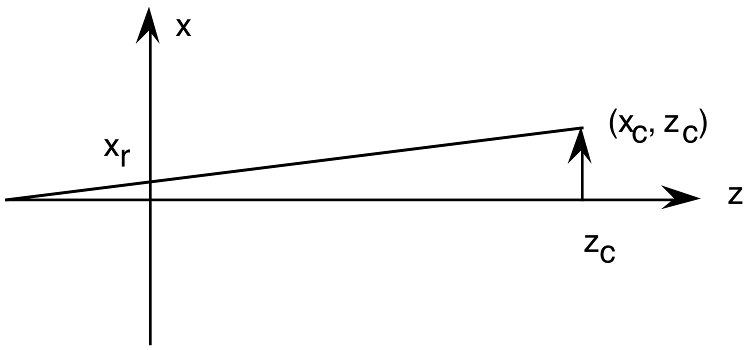
$$\vec{Q} = M_s^i \vec{P} = C_r^i P_c^r T_s^c \vec{P}$$

We will use "tensor notation" to keep track of our reference frames:

Projective Transforms: from the scene to the retina

Projection through an aperture is a projective transformation

Consider the central projection model for a 1D camera:



In camera coordinates:

$P^c = (x_c, y_c, z_c, 1)$ is a scene point in camera (aperture centered) coordinates
$Q^r = (x_r, y_r, 1)$ is a point on the retina.

By similar triangles:

$$\frac{x_r}{F} = \frac{x_c}{z_c} \quad \Leftrightarrow \quad x_r = x_c \frac{F}{z_c} \quad \Leftrightarrow \quad x_r \frac{z_c}{F} = x_c$$

$$\frac{y_r}{F} = \frac{y_c}{z_c} \quad \Leftrightarrow \quad y_r = y_c \frac{F}{z_c} \quad \Leftrightarrow \quad y_r \frac{z_c}{F} = y_c$$

Assume: $w = \dfrac{z_c}{F}$

then: $w\, x_r = x_c$
$w\, y_r = y_c$

$$w = \frac{z_c}{F}$$

As a matrix:

$$\begin{pmatrix} wx_r \\ wy_r \\ w \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{F} & 0 \end{pmatrix} \begin{pmatrix} x_c \\ y_c \\ z_c \\ 1 \end{pmatrix}$$

The transformation from Scene points in camera coordinates to retina points is:

$$Q^r = \begin{pmatrix} q_1 \\ q_2 \\ q_3 \end{pmatrix} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{F} & 1 \end{pmatrix} \begin{pmatrix} p^1 \\ p^2 \\ p^3 \\ 1 \end{pmatrix} = P_c^r \vec{P}^c$$

and $$\begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix} = \begin{pmatrix} q_1 / q_3 \\ q_2 / q_3 \\ 1 \end{pmatrix}$$

thus:

$$P_c^r = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{F} & 1 \end{pmatrix}$$

Note that P_c^r is not invertable.

Remark: If we place the origin in the retina:

$$\frac{x_r}{F} = \frac{x_c}{(F + z_c)} \quad \Rightarrow \quad x_r = \frac{x_c F}{(F + z_c)}$$

Which gives:

$$x_r = \frac{x_c F}{(F+z_c)} \qquad y_r = \frac{y_c F}{(F+z_c)}$$

$$z_r = 0$$

and thus : $\mathbf{P}_c^{\,r} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & \frac{1}{F} & 1 \end{pmatrix}$

From Scene to Camera

The following matrix represents a translation Δx, Δy, Δz and a rotation R.

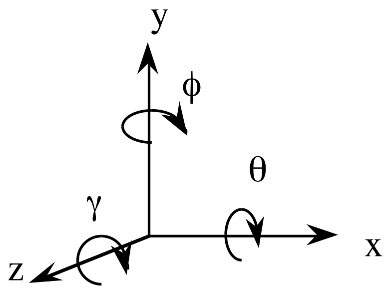
$$T_s^c = \begin{pmatrix} & & & \Delta x \\ & R & & \Delta y \\ & & & \Delta z \\ 0 & 0 & 0 & 1 \end{pmatrix}$$

The transformation is composed by expressing the position of the source reference frame in the destination reference frame.

The rotation part is a 3x3 matrix that can be decomposed into 3 smaller rotations.

$$\mathbf{R} = \mathbf{R}_z(\gamma)\mathbf{R}_y(\varphi)\mathbf{R}_x(\theta)$$

En 3D



$\mathbf{R}_x(\theta)$ is a rotation around the x axis.

$$R_x(\theta) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\theta) & \sin(\theta) \\ 0 & -\sin(\theta) & \cos(\theta) \end{pmatrix}$$

$$R_y(\varphi) = \begin{pmatrix} \cos(\varphi) & 0 & -\sin(\varphi) \\ 0 & 1 & 0 \\ \sin(\varphi) & 0 & \cos(\varphi) \end{pmatrix}$$

$$R_z(\gamma)=\begin{pmatrix}\cos(\gamma) & \sin(\gamma) & 0\\ -\sin(\gamma) & \cos(\gamma) & 0\\ 0 & 0 & 1\end{pmatrix}$$

Scale Change:

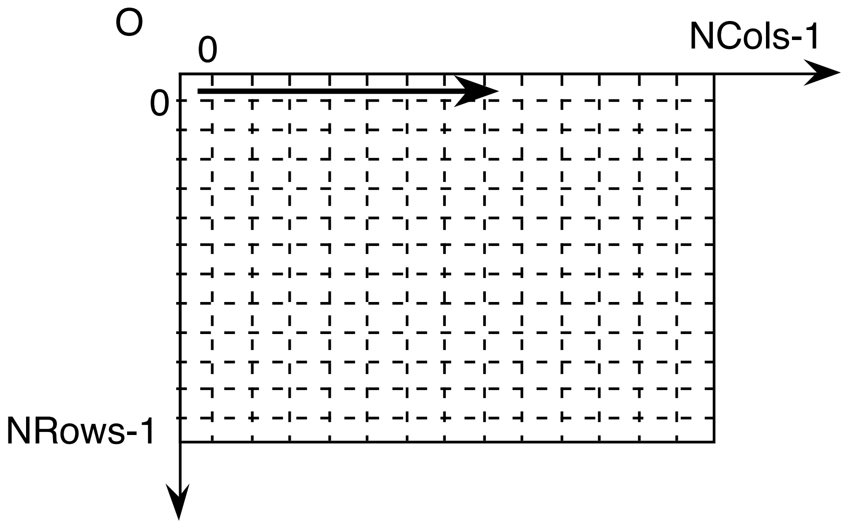
We can change the scale of each axis with a scale transformation

$$S_i^j=\begin{pmatrix}s_x & 0 & 0 & 0\\ 0 & s_y & 0 & 0\\ 0 & 0 & s_z & 0\\ 0 & 0 & 0 & 1\end{pmatrix}$$

From the Retina to Digitized Image

The "intrinsic parameters of the camera are F and C_x, C_y, D_x, D_y

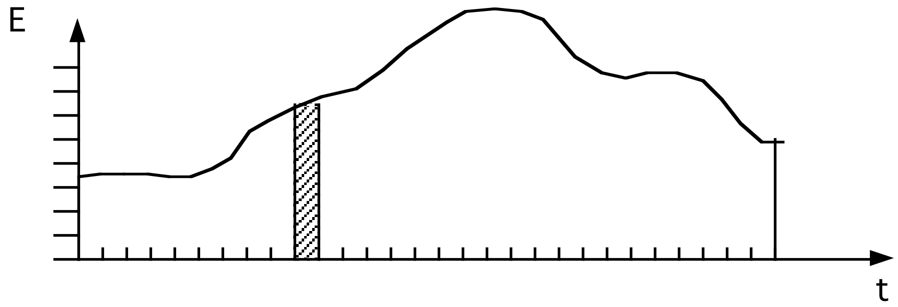
The image frame is composed of pixels (picture elements)



Note that pixels are not necessarily square.

Typical image sizes VGA : 640 x 480

Sampling and A/D Conversion.



The mapping from retina to image can be expressed with 4 parameters:

C_x, C_y : the image center (expressed in pixels).
D_x, D_y : The size of pixels expressed in pixels/mm.

$i = x_r D_i$ (mm · pixel/mm) $+ C_i$ (pixel)

$j = y_r D_j$ (mm · pixel/mm) $+ C_j$ (pixel)

Transformation from retina to image :

$$Q^i = \mathbf{C}_r^i \; Q^r$$

$$\begin{pmatrix} i \\ j \\ 1 \end{pmatrix} = \begin{pmatrix} D_i & 0 & C_i \\ 0 & D_j & C_j \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_r \\ y_r \\ 1 \end{pmatrix}$$

That is:

$$\begin{pmatrix} wi \\ wj \\ w \end{pmatrix} = \begin{pmatrix} D_i & 0 & C_i \\ 0 & D_j & C_j \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} wx_r \\ wy_r \\ w \end{pmatrix}$$

The Complete Camera Model

$$\boxed{P^i = \mathbf{C}_r^i \ \mathbf{P}_c^r \ \mathbf{T}_s^c \, P^s = \mathbf{M}_s^i \ P^s}$$

$$\begin{bmatrix} w\,i \\ w\,j \\ w \end{bmatrix} = \mathbf{M}_s^i \begin{bmatrix} x_s \\ y_s \\ z_s \\ 1 \end{bmatrix}$$

and thus

$$i = \frac{w\,i}{w} = \frac{M_s^1 \cdot P^s}{M_s^3 \cdot P^s} \qquad j = \frac{w\,j}{w} = \frac{M_s^2 \cdot P^s}{M_s^3 \cdot P^s}$$

or

$$i = \frac{w\,i}{w} = \frac{M_{11} X_s + M_{12} Y_s + M_{13} Z_s + M_{14}}{M_{31} X_s + M_{32} Y_s + M_{33} Z_s + M_{34}}$$

$$j = \frac{w\,j}{w} = \frac{M_{21} X_s + M_{22} Y_s + M_{23} Z_s + M_{24}}{M_{31} X_s + M_{32} Y_s + M_{33} Z_s + M_{34}}$$