A Geometric Model for 3D Imaging

Reference : Ch. 2 Schalkoff, R. Digital Image Processing and Computer Vision © 1989 John Wiley & Sons

The Perspective-Projection Transform (World Points to Image Points and Trying to Get Back)

Goal : Model the physical processes involved in the geometric aspects of image formation.

Object or World Points P_o – space points in the real-world, therefore 3-D , $\begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}$.

Image Points P_i – space points in an image plane, therefore 2-D , $\begin{bmatrix} x_i \\ y_i \end{bmatrix}$.

Geometric image formation is the mapping 3-D object points in a 2-D image plane

$$\boldsymbol{P_i = G\, P_o} \tag{1}$$

G is generally non-invertible and is a function of the imaging geometry, lens model and coordinate system chosen.

Finding this model allows us to:
1. Determine the 3-D structure of scenes;
2. Develop stereo or multiple camera systems for range measurement;
3. Identify and correct image distortions (whether optical or due to viewing geometry).

Our modeling approach proceeds from simple to sophisticated progressively building up in sophistication.
1. First, we develop a **pinhole camera model** with object and image points measured with respect to a coordinate system centered in the image plane.
2. Then we explore the **properties and limitations** of this model.
3. Next we revise this model to allow measurements of object points with respect to an arbitrary **world coordinate system**.
4. Finally we explore the **benefits of the revised model**.

Pinhole Camera Model

Consider the pinhole camera model in Figure 1 and initially assume that the coordinate system of the world and image points **coincide.**

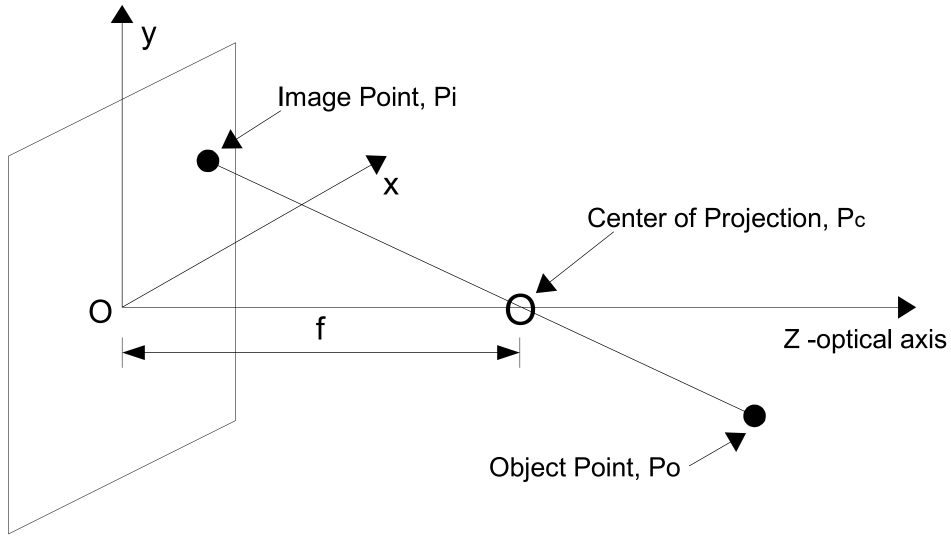

Figure 1: Object-image point transformation by a pinhole camera model.

Based from Figure 1, we identify the following:

Image point, $\boldsymbol{P_i} = \begin{bmatrix} x_i \\ y_i \\ 0 \end{bmatrix}$

Center of projection, $\boldsymbol{P_c} = \begin{bmatrix} 0 \\ 0 \\ f \end{bmatrix}$

Object point, $\boldsymbol{P_o} = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix}$.

These three points are ***collinear*** such that if k is a constant factor

$$k(\boldsymbol{P_i} - \boldsymbol{P_c}) = (\boldsymbol{P_c} - \boldsymbol{P_o}) \tag{2}$$

Expanding this equation yields

$$k\left\{\begin{bmatrix} x_i \\ y_i \\ 0 \end{bmatrix} - \begin{bmatrix} 0 \\ 0 \\ f \end{bmatrix}\right\} = \begin{bmatrix} 0 \\ 0 \\ f \end{bmatrix} - \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \tag{3}$$

Therefore,

$$k = \frac{z_o - f}{f} = \frac{z_o}{f} - 1 \tag{4}$$

$$x_i = \frac{-x_o}{k} = \frac{-x_o f}{z_o - f} \tag{5}$$

$$y_i = \frac{-y_o}{k} = \frac{-y_o f}{z_o - f} \tag{6}$$

Implication and Limitations of the Pinhole Camera Model

1. It is nonlinear in object point parameters $(x_o, y_o, z_o)^T$
2. It is nonlinear in camera parameter *f*.
3. It imposes that the object point be measured from the coordinate system of the image plane.
4. Since x_i and y_i are nonlinear, their relationships are often approximated. For example, in many instances $z_o >> f$ thus

$$x_i \approx \frac{-x_o f}{z_o} = \frac{-x_o}{z_o / f} \tag{7}$$

$$y_i \approx \frac{-y_o f}{z_o} = \frac{-y_o}{z_o / f} \tag{8}$$

   The quantity z_o/f is called the magnification ratio and is generally >>1.

5. Object points are converted into image points through division by the magnification ratio and by a sign inversion.
6. Measuring of object points with respect to an image plane coordinate system is tedious and not practical.
7. Instead of having a single coordinate system for both image and object points it is more convenient to have a separate world coordinate system and image coordinate system.

Since these equations are nonlinear, we introduce a representation in computer graphics modeling that helps further modeling.

Homogeneous Coordinates

The homogenous coordinates for the physical object point $(x_o, y_o, z_o)^T$ is now represented by a 4 x1 vector $(wx_o, wy_o, wz_o, w)^T$ where w is a nonzero arbitrary constant.

There is no unique representation of a physical point in homogeneous coordinates.

To convert a point represented as an n×1 vector from a homogenous coordinate representation to physical coordinates of dimension (n-1), we divide all components by the nth element and delete the nth component thus forming a new (n-1)×1 vector.

An image point is represented in homogeneous coordinates by augmenting the 2×1 image point vector by one dimension (the scale factor) and multiplying the image point coordinates by this nonzero factor.

With the ^ denoting homogeneous coordinates our object and image points become

$$\boldsymbol{P}_i = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \rightarrow \hat{\boldsymbol{P}}_i = \begin{bmatrix} wx_i \\ wy_i \\ w \end{bmatrix} \tag{9}$$

$$\boldsymbol{P}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \end{bmatrix} \rightarrow \hat{\boldsymbol{P}}_o = \begin{bmatrix} wx_o \\ wy_o \\ wz_o \\ w \end{bmatrix} \tag{10}$$

Where w ≠ 0.

This augmentation of the vector space dimension by one allows the nonlinear relationship in the pinhole camera model to be written as

$$\hat{\boldsymbol{P}}_i = \boldsymbol{F}_1 \hat{\boldsymbol{P}}_o \tag{11}$$

where the matrix $\boldsymbol{F}_1$ is written as

$$\boldsymbol{F}_1 = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & f & 0 \\ 0 & 0 & -1 & f \end{bmatrix} \tag{12}$$

As a consequence of the homogeneous representation, F1 is not unique. F1 may be multiplied by any nonzero scalar and the relationship still holds. For example, we may instead use $\boldsymbol{F}_2$

$$\boldsymbol{F}_2 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & -1/f & 1 \end{bmatrix} \tag{13}$$

To illustrate the use of F1 choose the object point in homogeneous coordinates as

$$\hat{\boldsymbol{P}}_o = \begin{bmatrix} x_o \\ y_o \\ z_o \\ 1 \end{bmatrix} \tag{14}$$

Applying Equation 12 and 11 we therefore get

$$\hat{\boldsymbol{P}}_i = \begin{bmatrix} fx_o \\ fy_o \\ f z_o \\ f - z_o \end{bmatrix} \tag{15}$$

Therefore in physical image coordinates

$$\boldsymbol{P}_i = \begin{bmatrix} x_i \\ y_i \\ z_i \end{bmatrix} = \begin{bmatrix} f x_o / (f - z_o) \\ f y_o / (f - z_o) \\ f x_o / (f - z_o) \end{bmatrix} \tag{16}$$

Now the desired image point actually has $z_i=0$ but Equation 16 does not achieve this. We can chop off the third row of $\boldsymbol{F_1}$ and assume no knowledge of z_i that is, we now have a new matrix

$$\boldsymbol{F} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & -1 & f \end{bmatrix} \tag{17}$$

So now we have

$$\hat{\boldsymbol{P}}_i = \boldsymbol{F} \hat{\boldsymbol{P}}_{\boldsymbol{o}} \tag{18}$$

which is the desired relationship that can yield $z_i=0$.
Rewriting Equations 5 and 6 we get

$$\frac{x_o}{x_i} = \frac{y_o}{y_i} = \frac{z_o - f}{-f} \tag{19}$$

which is the equation of a line in 3-D ratio form. The line passes through the point $(0,0,f)^T$ as expected.

Measurement of object points with respect to an arbitrary coordinate system

The choice of the coordinate system with the origin at the image plane is not practical. A user-selected coordinate system is more useful. The goal of this section is to relate by a transformation the image plane-centric coordinate system to the user-defined coordinate system.

In general, a user-defined global coordinate system can transform through any number of translation and rotations to match the image-plane centric coordinate system as in Figure 2.

Figure 2: Global coordinate system and image-plane coordinate system.

Illustration: Translation Matrix

Figure 2 shows the vector **d** which is the translation of the origin of the global coordinate system to the image plane origin. The translation of the object point global coordinates to an arbitrary 3-D coordinate system may be expressed as

$$\boldsymbol{P} = \boldsymbol{P}_o^g - \begin{bmatrix} d_x \\ d_y \\ d_z \end{bmatrix} \tag{20}$$

Converting P and P^g_o into homogeneous representation yields

$$\hat{\boldsymbol{P}} = \begin{bmatrix} 1 & 0 & 0 & -d_x \\ 0 & 1 & 0 & -d_y \\ 0 & 0 & 1 & -d_z \\ 0 & 0 & 0 & 1 \end{bmatrix} \hat{\boldsymbol{P}}_o^g \tag{21}$$

where the 4×4 matrix is the **TRANSLATION MATRIX T.** Thus,

$$\hat{\boldsymbol{P}} = \boldsymbol{T}\,\hat{\boldsymbol{P}}_o^g \tag{22}$$

This transformation may be applied to any object point.

General Transformation

In the same manner, the coordinate systems may be related by a rotation described by **ROTATION MATRIX R**

$$\hat{\boldsymbol{P}} = \boldsymbol{R}\,\hat{\boldsymbol{P}}_o^g \tag{23}$$

In general, the global coordinate system may undergo any number of translation and rotation matrices until it coincides with the image-centric coordinate system. We may represent the product of all these transformations as the matrix **V**, thus ,in general, we have

$$\hat{\boldsymbol{P}} = \boldsymbol{V}\,\hat{\boldsymbol{P}}_o^g \tag{24}$$

Recall that we introduced a perspective transformation matrix **F** in Equation 17 to transform a 3-D point $\hat{\boldsymbol{P}}$ to the image plane coordinates $\hat{\boldsymbol{P}}_i$. Thus Equation (24) becomes

$$\hat{\boldsymbol{P}}_i = \boldsymbol{F}\boldsymbol{V}\,\hat{\boldsymbol{P}}_o^g \tag{25}$$

or

$$\hat{\boldsymbol{P}}_i = \boldsymbol{A}\,\hat{\boldsymbol{P}}_o^g \tag{26}$$

where **A** has the form

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \tag{27}$$

The components of matrix **A** is not unique due to the use of homogeneous coordinates. Its 12 components are nonlinear functions of 7 parameters (3 translations, 3 rotation angles, f) that determines the camera geometry.

Camera Calibration

Expressing Equation (26) in component form yields

$$\begin{bmatrix} wx_i \\ wy_i \\ w \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} & a_{14} \\ a_{21} & a_{22} & a_{23} & a_{24} \\ a_{31} & a_{32} & a_{33} & a_{34} \end{bmatrix} \begin{bmatrix} x_o^g \\ y_o^g \\ z_o^g \\ 1 \end{bmatrix} \tag{28}$$

The process of getting the parameters a_{ij} is called **CAMERA CALIBRATION**. It is typical to set $a_{34} = 1.0$ to normalize Equation (28) and to remove the ambiguity caused by the use of homogeneous coordinates. Rewriting Equation (28) in physical coordinates we get

$$x_i = \frac{a_{11}x_o + a_{12}y_o + a_{13}z_o + a_{14}}{a_{31}x_o + a_{32}y_o + a_{33}z_o + a_{34}} \tag{29}$$

and

$$y_i = \frac{a_{21} x_o + a_{22} y_o + a_{23} z_o + a_{24}}{a_{31} x_o + a_{32} y_o + a_{33} z_o + a_{34}} \tag{30}$$

Setting a_{34} = 1.0 and after some algebraic manipulation yields

$$\begin{bmatrix} x_o & y_o & z_o & 1 & 0 & 0 & 0 & 0 & -(x_i x_o) & -(x_i y_o) & -(x_i z_o) \\ 0 & 0 & 0 & 0 & x_o & y_o & z_o & 1 & -(y_i x_o) & -(y_i y_o) & -(y_i z_o) \end{bmatrix} \begin{bmatrix} a_{11} \\ a_{12} \\ a_{13} \\ a_{14} \\ a_{21} \\ a_{22} \\ a_{23} \\ a_{24} \\ a_{31} \\ a_{32} \\ a_{33} \end{bmatrix} = \begin{bmatrix} x_i \\ y_i \end{bmatrix} \tag{31}$$

A single pair of points yields 2 equations and 11 unknowns. Given n such points we can formulate Equation (31) as

$$\boldsymbol{Q}\boldsymbol{a} = \boldsymbol{p} \tag{32}$$

where **a** is the matrix of camera paramters, **p** are the image coordinates and **Q** is the form given in Equation (31).

With enough sample points n >> 11 we can solve for the camera parameters **a** by

$$\boldsymbol{a} = \left(\boldsymbol{Q}^T \boldsymbol{Q}\right)^{-1} \boldsymbol{Q}^T \boldsymbol{p} \tag{33}$$