CSci 6974 and ECSE 6966 Math. Tech. for Vision, Graphics and Robotics
Lecture 14, March 9, 2006
The Perspective Camera

Overview

Most of this material is covered in Chapter 5 and Sections 7.1 through 7.4 of Hartley and Zisserman, but our discussion will not be as detailed. Here's what we will emphasize:

- Basic perspective projection as a matrix multiplication.
- Internal (intrinsic) and external (extrinsic) camera parameters.
- Camera properties.
- Projections and backprojections of geometric objects.

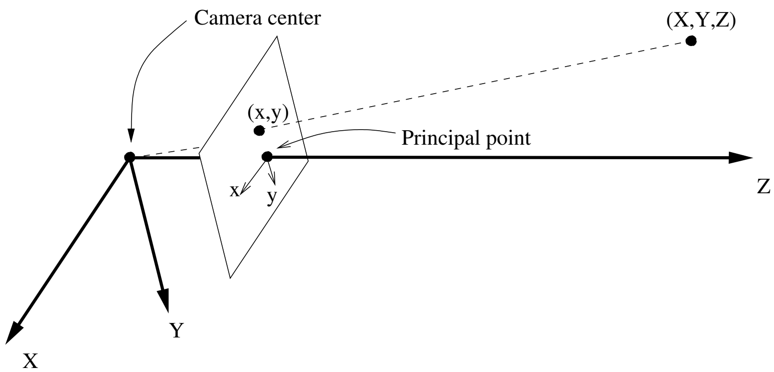
The notes also include a discussion of simpler, approximate camera models. We probably will not have time to cover these during lecture.

Notation

- For the notes here, coordinates of points in 3-space (the world) are written using capital letters and coordinates of points in 2-space (the image) are written using small letters.
- The symbol = here takes on its usual meaning when affine (non-homogeneous) coordinates are involved and means "equality up to a scale factor" when homogeneous coordinates are involved. The difference should be clear from the context.

Basic perspective projection

- Here is a drawing of the basic perspective projection model using a right-handed coordinate system with the image plane (schematically) in front of the lens.



In establishing our coordinate system this way we will not need to invert the y axis when converting to image (pixel array) coordinates. Image coordinates usually have the origin in the upper left corner of the image plane, with the y axis pointing down, although in the picture, the origin is in the center.

- Affine point $(X, Y, Z)^\top$ in the world projects to

$$(x, y)^\top = \left(\frac{fX}{Z}, \frac{fY}{Z} \right)^\top \tag{1}$$

in the image.

- Using homogeneous coordinates, we can use matrix multiplication to describe the projection operation:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{2}$$

- Affine image coordinates are recovered by division following projection:

$$x = \frac{u}{w} \qquad \text{and} \qquad y = \frac{v}{w}. \tag{3}$$

Terminology

Camera Center: The point through which all projection lines pass. In other words, the line formed by each image point with its corresponding world coordinate point passes through the camera center. In the simple camera model, the camera center is at the origin.

Optical (principal) axis: The direction from the camera center perpendicular to the image plane.

Principal plane: The plane parallel to the image plane and containing the camera center point.

Principal point: The location, in *image coordinates*, where the optical axis pierces the image plane.

Camera coordinate frame: The world coordinate frame whose origin is the camera center and whose Z axis is the optical axis.

Internal Camera Parameters

These are also called "calibration parameters".

- We need a conversion from world coordinate units (e.g. cm) to pixel units. If s_x and s_y are the pixel sizes then we have

$$(x, y)^\top = \left(\frac{fX}{s_x Z}, \frac{fY}{s_y Z}\right)^\top = \left(\frac{\alpha_x X}{Z}, \frac{\alpha_y Y}{Z}\right)^\top, \tag{4}$$

  where $\alpha_x = f/s_x$ and $\alpha_y = f/s_y$.

- The principal point is at location $(x_0, y_0)^\top$ in the image pixel coordinate system. This implies that

$$(x, y)^\top = \left(\frac{\alpha_x X}{Z} + x_0, \frac{\alpha_y Y}{Z} + y_0\right)^\top. \tag{5}$$

- Writing this in matrix form yields:

$$\begin{pmatrix} u \\ v \\ w \end{pmatrix} = \begin{pmatrix} \alpha_x & 0 & x_0 & 0 \\ 0 & \alpha_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \begin{pmatrix} X \\ Y \\ Z \\ 1 \end{pmatrix}. \tag{6}$$

- The 3×4 camera matrix may be factored as

$$\mathbf{P} = \begin{pmatrix} \alpha_x & 0 & x_0 & 0 \\ 0 & \alpha_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} = \mathbf{K}\big(\mathbf{I} \mid \mathbf{0}\big) \tag{7}$$

  where

$$\mathbf{K} = \begin{pmatrix} \alpha_x & 0 & x_0 \\ 0 & \alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \tag{8}$$

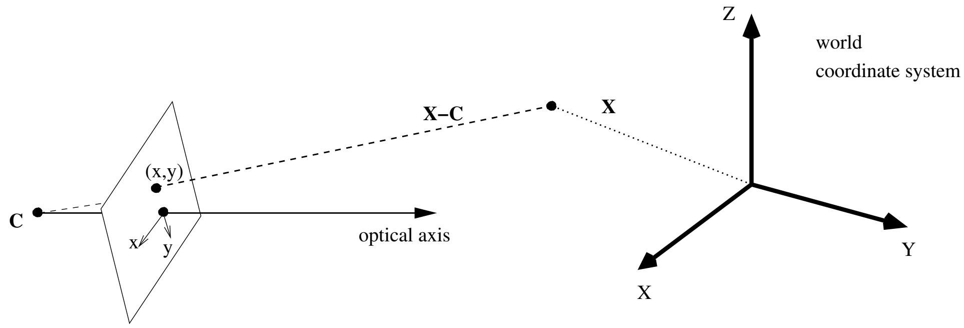
  is a 3×3 upper triangular matrix of internal parameters and

$$\begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{pmatrix} \tag{9}$$

  is often thought of as the *projection matrix*.

- Aside: sometimes a skew factor, s, is placed in the first row, second column of $\mathbf{K}$.

External Parameters



- Consider a different coordinate system for points in the world, with the camera center at $\mathbf{C}$ and a rotation of $\mathbf{R}$ from the world to the camera coordinate system.
- Converting a point $\mathbf{X}$ from the world coordinate system into the camera coordinate system requires a centering and a rotation:

$$\mathbf{X}_c = \mathbf{R}(\mathbf{X} - \mathbf{C}). \tag{10}$$

- Writing this in matrix form and treating $\mathbf{X}$ as homogeneous yields:

$$\mathbf{X}_c = \begin{pmatrix} \mathbf{R} & -\mathbf{RC} \\ \mathbf{0}^\top & 1 \end{pmatrix} \mathbf{X}. \tag{11}$$

- Multiplying the camera matrix $\mathbf{P}$ on the right by the Euclidean transformation matrix yields a new $\mathbf{P}$:

$$\mathbf{P} = \mathbf{K}\big(\mathbf{R} \mid -\mathbf{RC}\big) \tag{12}$$

- Rotation matrix $\mathbf{R}$, which is 3×3 but has only 3 degrees of freedom, and $\mathbf{C}$ are the external parameters.

Recovering Internal and External Parameters from P

- Through a calibration process we will be given $\mathbf{P}$, and we will want to know $\mathbf{K}$, $\mathbf{R}$ and $\mathbf{C}$.
- Write

$$\mathbf{P} = \big(\mathbf{M} \mid \mathbf{p}_4\big). \tag{13}$$

- Then $\mathbf{M} = \mathbf{KR}$ and $\mathbf{p}_4 = -\mathbf{KRC}$.
- $\mathbf{K}$ and $\mathbf{R}$ may be recovered from an RQ decomposition.
- $\mathbf{C} = -\mathbf{M}^{-1}\mathbf{p}_4$.

Camera Properties

Several of these properties may also be found from the previous decomposition.

- The camera center is the right nullspace of $\mathbf{P}$, which gives $\mathbf{C} = -\mathbf{M}^{-1}\mathbf{p}_4$ as above.
- The image vanishing points of the X, Y, and Z axes of the world coordinate system are the first three columns of $\mathbf{P}$. The fourth column is the image of the world coordinate system origin.
- The principal plane π_p is the 3^{rd} row of $\mathbf{P}$.
- The normal to this plane is the optical axis. This is the third row of $\mathbf{M}$, which we will denote as $\mathbf{m}_3^\top$.
- The principal point is $\mathbf{M}\mathbf{m}_3$.

Projections and Backprojections of Geometric Objects

In order to more fully understand perspective projection, we consider the projections and backprojections of a variety of geometric objects:

- Points.
- Lines.
- Planes.
- Conics.
- Quadrics.

Backprojection of Points

The backprojection of each image point $\mathbf{x} = (x, y, 1)^\top$, where x and y are the image coordinates, is a line in the world (3-space).

- The camera center $\mathbf{C} = -\mathbf{M}^{-1}\mathbf{p}_4$ is on the line.
- A second point on the line is the point at infinity that projects onto $\mathbf{x}$. This point can be written as $(\mathbf{D}^\top, 0)^\top$ for some three-component vector $\mathbf{D}$.
- We need to find $\mathbf{D}$. This is done using the camera matrix:

$$\mathbf{x} = \begin{pmatrix} \mathbf{M} & \mathbf{p}_4 \end{pmatrix} \begin{pmatrix} \mathbf{D} \\ 0 \end{pmatrix}, \tag{14}$$

  so

$$\mathbf{D} = \mathbf{M}^{-1}\mathbf{x}. \tag{15}$$

- The line is now, parametrically,

$$-\mathbf{M}^{-1}\mathbf{p}_4 + \lambda \mathbf{M}^{-1}\mathbf{x} \tag{16}$$

for parameter λ. These are affine coordinates (no homogeneous term).

Lines Map Onto Image Lines

This may be proved both geometrically and algebraically.

- Geometrically, the line in the world and the camera center form a plane. The intersection of this plane and the image plane gives a line.
- Algebraically, one way we can prove this is using the parametric form for the line.
- We can also work with Plucker coordinates. If $\mathbf{L}$ is the Plucker matrix corresponding to the line in the world then the parameters of the image line are found from

$$\mathbf{P}\mathbf{L}\mathbf{P}^\top, \tag{17}$$

which is a skew-symmetric matrix encoding the image line parameters:

$$\mathbf{P}\mathbf{L}\mathbf{P}^\top = \begin{pmatrix} 0 & l_3 & -l_2 \\ -l_3 & 0 & l_1 \\ l_2 & -l_1 & 0 \end{pmatrix}. \tag{18}$$

Image Lines Map Onto 3-Space Planes

These planes must include the camera center and the line on the image plane. The equation of this plane is derived as follows.

- Let $\mathbf{l}$ be a line in the image, and let $\mathbf{x}$ be any point on the line.
- Let $\mathbf{X}$ be a 3-space point mapping onto $\mathbf{x}$, so that $\mathbf{x} = \mathbf{P}\mathbf{X}$.
- Then

$$0 = \mathbf{l}^\top \mathbf{x} = \mathbf{l}^\top \mathbf{P}\mathbf{X} = (\mathbf{l}^\top \mathbf{P})\mathbf{X} = (\mathbf{P}^\top \mathbf{l})^\top \mathbf{X}. \tag{19}$$

- This implies the equation of the plane is

$$\pi = \mathbf{P}^\top \mathbf{l}. \tag{20}$$

Points from a Plane Map Projectively Onto the Image Plane

- The simplest means of proving this is to fix the world coordinate system so that the $X - Y$ plane coincides with the plane of interest π. Points on π then map onto the image plane using a 3×3 projective transformation.
- In general, the mapping of an arbitrary plane requires establishing a two-dimensional coordinate system for points on the plane.
- Planes mapping projectively onto planes implies that a conic, which is a planar object, will map onto a conic in the image.

Conics Backproject to (Generalized) Cones

- Remember, a cone is one of the rank 3 quadrics.
- If $\mathbf{C}$ is the conic matrix, then the cone matrix is easily shown to be

$$\mathbf{P}^\top \mathbf{C} \mathbf{P}. \tag{21}$$

- Clearly this has rank at most 3.
- Geometric intuitions are straightforward.

The Occluding Contour of a Quadric Maps Onto a Conic

The points on the surface of a quadric map to an image plane region. We show here that the outline of this region is a conic.

- Consider a quadric described by matrix $\mathbf{Q}$ and, as before, let the camera center be $\mathbf{C}$.
- The points on $\mathbf{Q}$ that project onto the outline (image contour) are called the *contour generator* and the image contour is called the *apparent contour*.
- For $\mathbf{Q}$ and any point $\mathbf{X}$ on the contour generator, the line between $\mathbf{X}$ and $\mathbf{C}$ is tangent to the quadric.
- Hence $\mathbf{C}$ is in the tangent plane at each point on the contour generator.
- In fact, the contour generator is the intersection between the quadric $\mathbf{Q}$ and the polar of $\mathbf{C}$ with respect to $\mathbf{Q}$.
- Now, consider the dual (plane) quadric $\mathbf{Q}^*$. Planes tangent to this satisfy $\pi^\top \mathbf{Q}^* \pi = 0$.
- Since the planes tangent to the contour generator contain $\mathbf{C}$, they must project to image lines.
- They therefore satisfy $\pi = \mathbf{P}^\top \mathbf{l}$ for some image line, $\mathbf{l}$, as above.
- Putting these results together implies

$$\mathbf{l}^\top \mathbf{P} \mathbf{Q}^* \mathbf{P}^\top \mathbf{l} = 0. \tag{22}$$

  so the resulting image conic is $\mathbf{P} \mathbf{Q}^* \mathbf{P}^\top$.

- This is a line conic. This illustrates the power of using dual conics.

Approximations to the Perspective Camera

As we have discussed in class, perspective effects aren't strong in all images. Therefore, we should consider simplified approximations to perspective cameras. These are especially useful because they do not include the non-linearities involved in converting from homogeneous to image coordinates.

Affine Camera — Definition

We will discuss a variety of special cases of an *affine camera* which is defined in terms of a 3×4 matrix

$$\mathbf{P}_{\text{aff}} = \begin{pmatrix} m_{11} & m_{12} & m_{13} & m_{14} \\ m_{21} & m_{22} & m_{23} & m_{24} \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{23}$$

Note that this camera has 8 degrees of freedom. The perspective camera matrix has 11 (only 10 if the skew parameter is fixed at 0).

Derivation — Scaled Orthographic Projection

Hartley and Zisserman have a derivation of the general affine camera where they move the camera center to infinity while simultaneously zooming the lens.

Here is a more traditional derivation.

- Let's return to the original equations for perspective projection:

$$x = \frac{fX}{Z} \qquad \text{and} \qquad y = \frac{fY}{Z}. \tag{24}$$

- Now, suppose the depths of all observed image points are within a small range around Z_0. Write the depth of a point as $Z = Z + \Delta Z$.
- Then, we write

$$x = \frac{fX}{Z_0 + \Delta Z} \qquad \text{and} \qquad y = \frac{fY}{Z_0 + \Delta Z}. \tag{25}$$

- When $\Delta Z / Z_0 \rightarrow 0$, this simplifies to

$$x = \frac{f}{Z_0} X = mX \qquad \text{and} \qquad y = \frac{f}{Z_0} Y = mY, \tag{26}$$

  where $m = f/Z_0$ is a *magnification factor*.
- In matrix form this is

$$\begin{pmatrix} m & 0 & 0 & 0 \\ 0 & m & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{27}$$

- This matrix describes *scaled orthographic projection*.
- If $m = 1$ the matrix describes *orthographic projection*.

Combining with Internal Camera Parameters

- Multiplying on the left with internal parameter matrix $\mathbf{K}$ as above this becomes

$$\begin{pmatrix} m\alpha_x & 0 & 0 & x_0 \\ 0 & m\alpha_y & 0 & y_0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \tag{28}$$

  which can be rearranged as

$$\begin{aligned} \begin{pmatrix} m\alpha_x & 0 & x_0 \\ 0 & m\alpha_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} &= \begin{pmatrix} \alpha'_x & 0 & x_0 \\ 0 & \alpha'_y & y_0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \\ &= \mathbf{K}' \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix} \end{aligned}$$

  where $\alpha'_x = m\alpha_x$ and $\alpha'_y = m\alpha_y$.

- Algebraically, the fundamental difference here over perspective projection is the shift of the 1 in the 3^{rd} row of the projection matrix!
- More differences will emerge.
- For the rest of the discussion we will use $\mathbf{K}$ instead of $\mathbf{K}'$, since folding in the scale parameter does not change the structure of the internal parameter matrix.

Combining with External Camera Parameters

- Multiplying the camera matrix on the right by the 4×4 rigid transformation matrix yields a camera matrix of

$$\mathbf{K} \begin{pmatrix} \mathbf{r}_1^\top & t_x \\ \mathbf{r}_2^\top & t_y \\ \mathbf{0}^\top & 1 \end{pmatrix} \tag{29}$$

  where $\mathbf{r}_1^\top$ and $\mathbf{r}_2^\top$ are the first two rows of a 3×3 rotation matrix.

- Note that the external parameter matrix (on the right) has 5 degrees of freedom and the internal parameter matrix, $\mathbf{K}$, on the left has 4, which makes the overall matrix appear to have 9. We are about to discover that this is not true.

The Principal Point is Undefined for an Affine Camera

- Multiply the internal and external matrix factors to obtain the full matrix:

$$\begin{pmatrix} \alpha_x \mathbf{r}_1^\top & \alpha_x t_x + x_0 \\ \alpha_y \mathbf{r}_2^\top & \alpha_y t_y + y_0 \\ \mathbf{0}^\top & 1 \end{pmatrix}. \tag{30}$$

- Now, consider $t'_x = \alpha_x t_x + x_0$ and $t'_y = \alpha_y t_y + y_0$. Remember, given a 3×4 projection matrix, t'_x and t'_y will be in the (1,4) and (2,4) positions.
- We would like to be able to recover t_x, x_0, t_y, y_0 from t'_x and t'_y, but we can't! (We can recover α_x and α_y though. Do you know how?)
- This is a manifestation of the problem that the principal point location $(x_0, y_0)^\top$ holds no meaning for the affine camera.
- Hence, for an affine camera

$$\mathbf{K} = \begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & 1 \end{pmatrix}, \tag{31}$$

and the full projection matrix can be described as

$$\begin{pmatrix} \alpha_x & 0 & 0 \\ 0 & \alpha_y & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mathbf{r}_1^\top & t_x \\ \mathbf{r}_2^\top & t_y \\ \mathbf{0}^\top & 1 \end{pmatrix} \tag{32}$$

- This has only 7 degrees of freedom — 8 if the skew factor s in $\mathbf{K}$ is allowed.

Hierarchy of Cameras

In summary, here's a hierarchy of camera models:

Orthographic: $\alpha_x = \alpha_y = 1$, so $\mathbf{K}$ is the identity. There are 5 degrees of freedom, all due to the external parameters.

Scaled orthographic: $\alpha_x = \alpha_y$. 6 degrees of freedom.

Weak perspective: $\alpha_x \neq \alpha_y$ (in general). 7 degrees of freedom.

Affine: As defined above. 8 degrees of freedom.

Perspective: General 3×4 matrix, 11 degrees of freedom.

More complex, non-linear cameras, generally include extra terms for radial lens distortion and, perhaps, tangential distortions.

Practice Problems

1. Give conditions under which the rank of $\mathbf{M}$, the leftmost 3×3 submatrix of $\mathbf{P}$, is full-rank.

2. Prove that parallel lines do remain parallel under the general affine camera (23).

3. We have proved that a conic C_w on a plane in the world projects onto a conic C_i in the image. Now, prove that the center of C_w does not necessarily map onto the center of C_i. To do this, construct as simple an example as possible.

4. Suppose $\boldsymbol{\pi}$ is a plane in the world, and $\mathbf{M}$ is the matrix of a given camera. We know that there is a homography mapping points on $\boldsymbol{\pi}$ onto the image plane. Sometimes, however, an affine transformation is a good approximation to this plane-to-plane mapping, especially when we are only interested in a small region of the plane. Suppose $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are non-collinear image points and these points are projections of their corresponding points on $\boldsymbol{\pi}$. Find the approximate 2D affine transformation mapping image points onto points on $\boldsymbol{\pi}$ that is exactly correct for $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$. Hint: you will need to establish a coordinate system on $\boldsymbol{\pi}$.

Homework 7 Problems; Due Monday March 20, 2006

Solutions to these problems, together with the two problems at the end of Lecture 12, are to be turned in on the Monday after Spring Break.

1. (**15 points**) Give an algebraic proof that parallel lines do not remain parallel under perspective projection. What lines do remain parallel?

2. (**10 points**) Give general conditions under which a line in the world maps onto a single point in the image.