Subsections

- Orthographic projection

Camera calibration

We will now return to image formation and camera geometry in a bit more detail to determine how one *calibrates* a camera to determine the relationship between what appears on the image (or retinal) plane and where it is located in the 3D world.

Imagine we have a three dimensional coordinate system whose origin is at the centre of projection and whose *Z* axis is along the optical axis, as shown in figure 1. This coordinate system is called the *standard coordinate system* of the camera. A point *M* on an object with coordinates (*X*,*Y*,*Z*) will be imaged at some point *m* = (*x*, *y*) in the image plane. These coordinates are with respect to a coordinate system whose origin is at the intersection of the optical axis and the image plane, and whose *x* and *y* axes are parallel to the *X* and *Y* axes. The relationship between the two coordinate systems (*c*,*x*,*y*) and (*C*,*X*,*Y*,*Z*) is given by

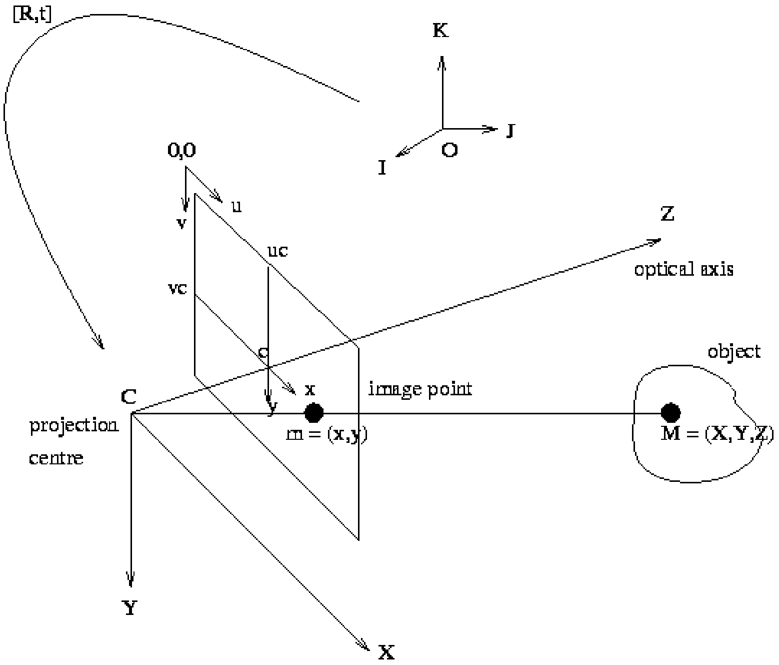
$$x = \frac{Xf}{Z} \qquad \text{and} \qquad y = \frac{Yf}{Z}. \tag{1}$$

This can be written linearly in homogeneous coordinates as

$$\begin{bmatrix} sx \\ sy \\ s \end{bmatrix} = \begin{bmatrix} f & 0 & 0 & 0 \\ 0 & f & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

where $s \neq 0$ is a scale factor.

Figure 1: The coordinate systems involved in camera calibration.



Now, the actual pixel coordinates (*u*,*v*) are defined with respect to an origin in the top left hand corner of the image plane, and will satisfy

$$u = u_c + \frac{x}{\text{pixel width}} \qquad \text{and} \qquad v = v_c + \frac{y}{\text{pixel height}}. \tag{2}$$

We can express the transformation from three dimensional world coordinates to image pixel coordinates using a 3×4 matrix. This is done by substituting equation (1) into equation (2) and multiplying through by *Z* to obtain

$$Zu = Zu_c + \frac{Xf}{\text{pixel width}}$$

$$Zv = Zv_c + \frac{Yf}{\text{pixel height}}.$$

In other words,

$$\begin{bmatrix} su \\ sv \\ s \end{bmatrix} = \begin{bmatrix} \frac{f}{\text{pixel width}} & 0 & u_c & 0 \\ 0 & \frac{f}{\text{pixel height}} & v_c & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix},$$

where the scaling factor *s* has value *Z*. In short hand notation, we write this as

$$\tilde{\mathbf{u}} = \mathbf{P} \cdot \tilde{\mathbf{M}},$$

where $\tilde{\mathbf{u}}$ represents the homogeneous vector of image pixel coordinates, **P** is the perspective projection matrix, and $\tilde{\mathbf{M}}$ is the homogeneous vector of world coordinates. Thus, a camera can be considered as a system that performs a linear projective transformation from the projective space $\mathcal{P}^3$ into the projective plane $\mathcal{P}^2$.

There are five camera parameters, namely the focal length *f*, the pixel width, the pixel height, the parameter u_c which is the *u* pixel coordinate at the optical centre, and the parameter v_c which is the *v* pixel coordinate at the optical centre. However, only four separable parameters can be solved for as there is an arbitrary scale factor involved in *f* and in the pixel size. Thus we can only solve for the ratios $\alpha_u = f/$ pixel width and $\alpha_v = f/$ pixel height. The parameters α_u, α_v, u_c and v_c do not depend on the position and orientation of the camera in space, and are thus called the *intrinsic* parameters.

In general, the three dimensional world coordinates of a point will not be specified in a frame whose origin is at the centre of projection and whose *Z* axis lies along the optical axis. Some other, more convenient frame, will more likely be specified, and then we have to include a change of coordinates from this other frame to the standard coordinate system. Thus we have

$$\tilde{\mathbf{u}} = \mathbf{P} \cdot \mathbf{K} \cdot \tilde{\mathbf{M}},$$

where **K** is a 4×4 homogeneous transformation matrix:

$$K = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0_3^{\mathbf{T}} & 1 \end{bmatrix}.$$

The top 3×3 corner is a rotation matrix **R** and encodes the camera orientation with respect to a given world frame; the final column is a homogeneous vector **t** capturing the camera displacement from the world frame origin. The matrix **K** has six degrees of freedom, three for the orientation, and three for the translation of the camera. These parameters are known as the *extrinsic* camera parameters.

The 3×4 camera matrix **P** and the 4×4 homogeneous transform **K** combine to form a single 3×4 matrix **C**, called the *camera calibration matrix*. We can write the general form of **C** as a function of the intrinsic and extrinsic parameters:

$$\mathbf{C} = \begin{bmatrix} \alpha_u \mathbf{r}_1 + u_c \mathbf{r}_3 & \alpha_u t_x + u_c t_z \\ \alpha_v \mathbf{r}_2 + v_c \mathbf{r}_3 & \alpha_v t_y + v_c t_z \\ \mathbf{r}_3 & t_z \end{bmatrix}, \tag{3}$$

where the vectors $\mathbf{r}_1, \mathbf{r}_2$, and $\mathbf{r}_3$ are the row vectors of the matrix **R**, and **t** = (t_x, t_y, t_z). The matrix **C**, like the matrix **P**, has rank three.

Orthographic projection

Consider a translation of -*f* along the *Z* axis of the standard coordinate frame, so that the focal plane and the image plane are now coincident. Since there is no rotation involved in this transformation, it is easy to see that the camera calibration matrix is just

$$\mathbf{C} = \begin{bmatrix} -f & 0 & 0 & 0 \\ 0 & -f & 0 & 0 \\ 0 & 0 & -1 & -f \end{bmatrix},$$

where we are assuming that the pixel width and height are both 1. Now since **C** is defined up to a scale factor, this is the same as

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -\frac{1}{f} & 1 \end{bmatrix}.$$

Now, if we let *f* go to infinity, the matrix becomes

$$\mathbf{C} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}.$$

This defines the transformation *u* = *X* and *v* = *Y* and is known as an *orthographic projection* parallel to the *Z* axis. It appears as the limit of the general perspective projection as the focal length *f* becomes large with respect to the distance *Z* of the camera from the object.

Next Up Previous

Next: Solving for the calibration **Up:** Computer Vision IT412 **Previous:** Lecture 9

Robyn Owens
10/29/1997