Homogeneous Transformation and Image-Object Mapping

Part 2 – Approach to the Solution

Written By Edward Gu for the Senior Design Class of Winter 2005

1. Unique Solution to the Image-Object Mapping

As we discussed in Part 1, the inverse mapping from a pair of 2D pixel coordinates (x_p, y_p) on the image screen to the real-world object coordinates (x, y, z) does not have a unique solution. The relationship of the mapping is given by

$$x_p = \frac{x_i D}{z_i} \text{ and } y_p = \frac{y_i D}{z_i} , \quad (1)$$

where x_i, y_i and z_i are the three coordinates of a point of object referred to the image frame, and *D* is a common zoom ratio in both x- and y-directions. It can be further shown that if the real-world object point to be determined is constrained on the **floor surface** that is also parallel to the x-axis of the image frame, then the inverse mapping can have a **unique solution**.

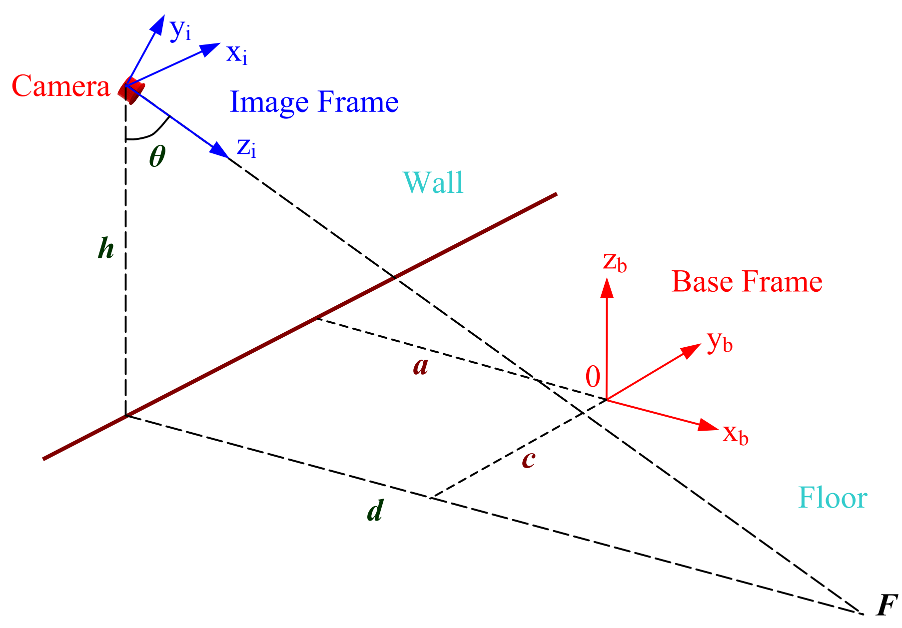

Figure 1 – Geometry of the Image-Object Mapping

With a careful observation on the 3D geometry, as depicted in Figure 1, we can clearly see that the z-coordinate of each point on the floor has a **linear relation** with its y-coordinate, both of which are referred to the image frame, i.e.,

$$z_i = ky_i + b$$

with a constant slope *k* and a constant intercept *b*. In fact, based on the triangle geometry displayed in Figure 1,

$$k = \tan\theta = \frac{d}{h} \quad \text{and} \quad b = \frac{h}{\cos\theta}. \tag{2}$$

Namely, the intercept *b* is actually the length from the camera to the intersection point *F* between the z_i-axis of the image frame and the floor.

Once we have equation (2), substituting it into (1) yields

$$x_p = \frac{x_i D}{ky_i + b} \quad \text{and} \quad y_p = \frac{y_i D}{ky_i + b}. \tag{3}$$

Therefore, the inverse mapping has the following **unique solution for every point on the floor:**

$$y_i = \frac{by_p}{D - ky_p}, \quad x_i = -\frac{x_p(ky_i + b)}{D} \quad \text{and} \quad z_i = ky_i + b, \tag{4}$$

where *k* and *b* are given by (2) and a minus sign is imposed into the second equation for x_i, because the image shown on the computer screen is often flipped over with respect to the x-axis from the reception signal on the CCD panel of a digital camera. Since equation (4) gives the coordinates of an object point with respect to the image frame that is fixed on the camera, we therefore need a homogeneous transformation H_b^i to convert them into the coordinates referred to the base, i.e.,

$$\begin{pmatrix} x \\ y \\ z \\ 1 \end{pmatrix} = H_b^i \begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix}. \tag{5}$$

2. How to Determine the Camera Shooting Axis?

This question is virtually equivalent to finding both the position and orientation of a camera, which is also a **Calibration Process**. The easiest way is to first determine the center pixel of the image window on your computer screen, i.e., the pixel of both $x_p = 0$ and $y_p = 0$. Then, you may ask your teammate to locate and match the corresponding real point on the floor while you are monitoring the center pixel on the image window, as shown in Figure 2. As a matter of fact, the real point that you just located on the floor is exactly the point F in Figure 1, which is the intersection point between the camera shooting axis and the floor. Now, you may measure the vertical distance from the wall to Point F and it gives you a value of d. Of course, the height h of the camera hanging on the wall is given. Using equation (2), you can readily determine both the slope k and the intercept b.

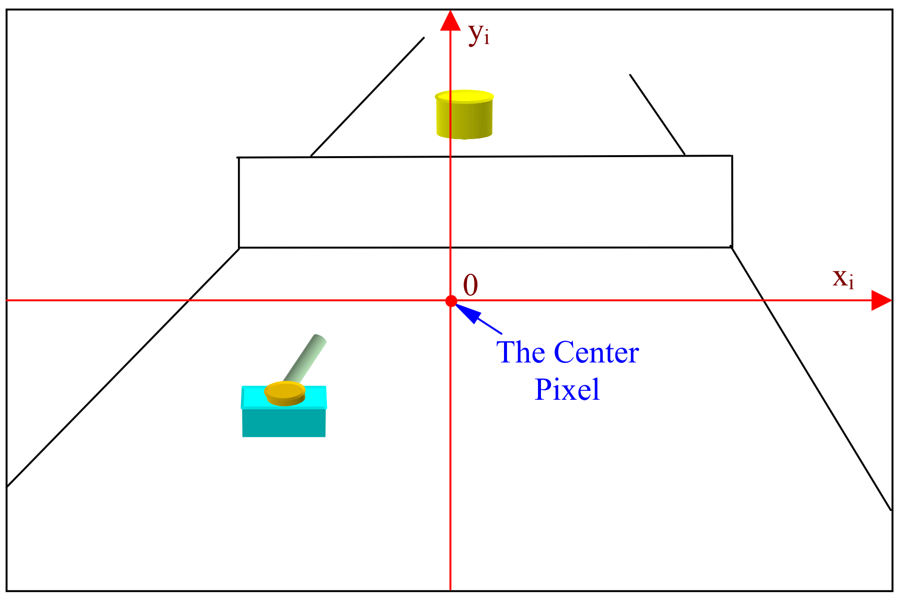

Figure 2 – The Center Pixel in an Image Window on Computer Screen

3. How to Determine the Homogeneous Transformation Matrix?

The 4 by 4 homogeneous transformation matrix H_b^i is required to transform a point referred to the image frame to the same point referred to the base according to equation (5). To find it, you have to decide where your base frame is and how it orients. Although the choice of your base frame is arbitrary, a better definition of the base coordinate system will often provide you with much more convenient way to your application. In our project, I bet you would like to place the base frame at your ball-launching device center. If so, the origin of the base is at the rotating center of both the azimuth and elevation, while the three base axes could be defined to make your future projectile planning easier.

As an illustrative example, I first place the origin of my base at a point with a vertical distance $a > 0$ from the wall and a vertical distance $c > 0$ to the left of the projection line of the camera shooting axis on the floor, as shown in Figure 1. Then, my x_b-axis is defined to be leaving away from and perpendicular to the wall, and my z_b-axis is perpendicular to the floor and going upward, see also Figure 1. With such a definition of the base, according to the description of Part 1, the homogeneous transformation matrix has the following form:

$$H_b^i = \begin{pmatrix} 0 & \cos\theta & \sin\theta & -a \\ 1 & 0 & 0 & -c \\ 0 & \sin\theta & -\cos\theta & h \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{6}$$

where θ can be found by the first equation in (2). If the orientation of your base is to be defined differently, then your homogeneous transformation should be modified accordingly. Since H_b^i is a constant matrix, its determination is needed only once.

4. How to Determine the Zoom Ratio *D*?

Even if every group shares a common camera, different computer imaging system has different zoom ratio, depending on the scales in your computer display. To find your zoom ratio, it suffices to test a special point, and only one point on the floor. One of the simplest ways is to pick the origin of your base frame, whose coordinates are obviously (0, 0, 0) seen with respect to the base frame. Then, convert them to the coordinates referred to the image frame via

$$\begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = H_i^b \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = (H_b^i)^{-1} \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix}. \tag{7}$$

On the other hand, the origin of the base can also be seen and measured on the computer screen as an image pixel, say (x_p, y_p). Using equation (1), we have

$$D = \frac{y_p \cdot z_i}{y_i}. \tag{8}$$

This D will be used forever, because it is a constant for each group. Please don't pick any point having $y_i = 0$ that is referred to the image frame in the above testing process to avoid unnecessary singularity.

Note that the above testing process for D is under the assumption that both x- and y-directions of your image have a common zoom ratio D, which is often a usual case. If not, you have to put D_1 in the first equation of (1) and D_2 in the second equation of (1) with $D_1 \neq D_2$, and then determine them separately. The inverse mapping solution (4) has also to be distinguished between x_i and y_i with D_1 and D_2, respectively.

5. Redefine a New Base at the Corner of the Barrier for Better Calibration

We now consider more realistically that a new base coordinate frame is defined at the corner between the floor and one end of the barrier, as shown in Figure 3. Suppose that we have no knowledge of where the intersection point ***F*** is, and we can only find the distance $c > 0$ between the end corner point of the barrier and the point ***Q*** that is the bottom line of the barrier intersecting with the $y_i - z_i$ plane of the image frame (please think about how to determine this c if the total length of the barrier is given).

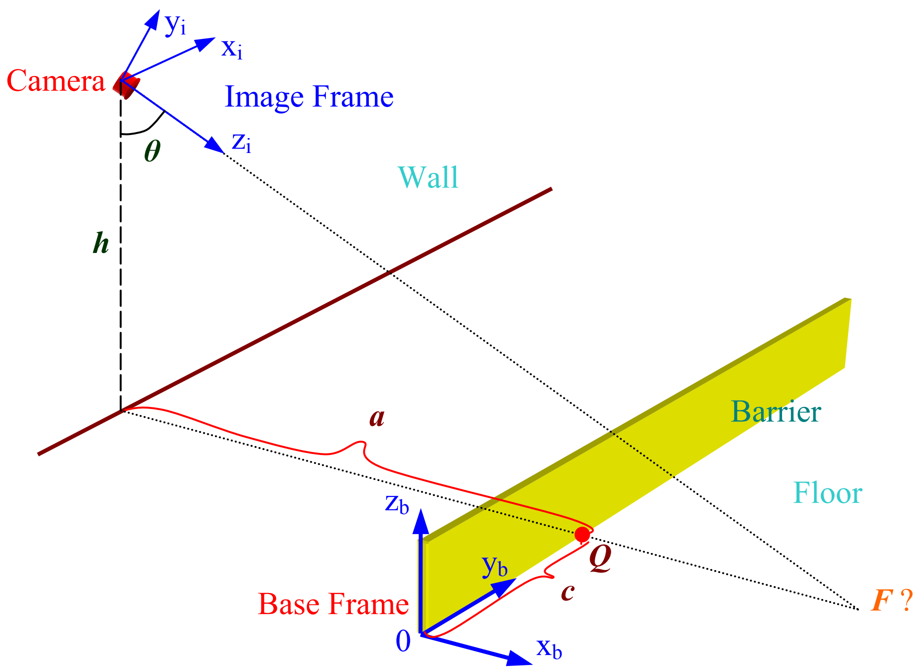

Figure 3 – A New Base Frame is defined at the Corner of the Barrier

It is sure that the distance $a > 0$ from the point ***Q*** to the wall will be given. Then, the homogeneous transformation between the image frame and the new base becomes

$$H_b^i = \begin{pmatrix} 0 & \cos\theta & \sin\theta & -a \\ 1 & 0 & 0 & c \\ 0 & \sin\theta & -\cos\theta & h \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{9}$$

The symbolical form of its inverse can be directly derived without major difficulty by utilizing the orthogonal property of the upper-left 3 by 3 corner of (9), and it turns out to be

$$H_i^b = (H_b^i)^{-1} = \begin{pmatrix} 0 & 1 & 0 & -c \\ \cos\theta & 0 & \sin\theta & a\cos\theta - h\sin\theta \\ \sin\theta & 0 & -\cos\theta & a\sin\theta + h\cos\theta \\ 0 & 0 & 0 & 1 \end{pmatrix}. \tag{10}$$

We now start calibrating the camera orientation by observing the origin of the base, i.e., the end point of the barrier bottom line on the computer screen and measuring its pixel coordinates (x_p, y_p). Based on equation (7), we can determine the coordinates of the base origin referred to the image frame. Substituting (10) into (7), we immediately obtain

$$\begin{pmatrix} x_i \\ y_i \\ z_i \\ 1 \end{pmatrix} = H_i^b \begin{pmatrix} 0 \\ 0 \\ 0 \\ 1 \end{pmatrix} = \begin{pmatrix} -c \\ a\cos\theta - h\sin\theta \\ a\sin\theta + h\cos\theta \\ 1 \end{pmatrix}. \tag{11}$$

Then, substituting each element of (11) into the second equation of the unique solution (4) for the inverse image-object mapping and noticing equations (2) and (8), we have

$$-x_i = c = \frac{x_p\left[\tan\theta(a\cos\theta - h\sin\theta) + \dfrac{h}{\cos\theta}\right]}{y_p \cdot \dfrac{a\sin\theta + h\cos\theta}{a\cos\theta - h\sin\theta}} = \frac{x_p}{y_p}(a\cos\theta - h\sin\theta) \tag{12}$$

This result is clearly equivalent to the following equation:

$$c = \frac{x_p}{y_p}\sqrt{a^2 + h^2} \cdot \cos\left(\theta + \tan^{-1}\frac{h}{a}\right). \tag{13}$$

The solution of (13) for θ becomes

$$\theta = \cos^{-1}\left(\frac{c \cdot y_p}{x_p\sqrt{a^2 + h^2}}\right) - \tan^{-1}\frac{h}{a}. \tag{14}$$

After the above calibration process via the observation of the base origin, equation (14) is now critical to determine not only the constant slope k and intercept b in (2) even if we don't know where the point $\boldsymbol{F}$ on the floor is, but also the constant zoom ratio D given in (8) by the following equation in terms of θ:

$$D = \frac{y_p \cdot z_i}{y_i} = y_p \cdot \frac{a\sin\theta + h\cos\theta}{a\cos\theta - h\sin\theta} = y_p \cdot \frac{a\tan\theta + h}{a - h\tan\theta}. \tag{15}$$