Homographies

April 26, 2020

Another special case of 3D reconstruction is when the object we want to reconstruct is a plane. For example, suppose we want to reconstruct the chessboard shown in Figure **??**, left. Given that we only want a particular plane and not the entire 3D world, does it make things easier? Or do we still need two calibrated cameras?

Recall that the 3D point corresponding to a pixel in the image is constrained to lie on a line (passing through the pixel and the camera pinhole). In general, this is not enough, and we would need another camera to locate the 3D point exactly (as discussed in the last lecture). However, in this case, *we know that the 3D point must lie on a particular plane*. Therefore, we can simply intersect the line with this plane to get the location of the 3D point (Figure **??**,right). Of course, this assumes we know how the plane is oriented and located relative to the camera.

Let us concretize this mathematically. We have a camera capturing an image of the plane of interest. Without loss of generality, we can assume a world coordinate system such that the equation of the plane is $Z = 0$. Thus, points on the plane take the form $(X, Y, 0)$.

Suppose that the camera projection matrix is $P = \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix}$. Then, a point $\mathbf{Q} = \begin{bmatrix} X \\ Y \\ 0 \end{bmatrix}$ on the plane gets projected as follows:

$$\vec{\mathbf{q}} \equiv P\vec{\mathbf{Q}} \tag{1}$$

$$\equiv \begin{bmatrix} P_{11} & P_{12} & P_{13} & P_{14} \\ P_{21} & P_{22} & P_{23} & P_{24} \\ P_{31} & P_{32} & P_{33} & P_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 0 \\ 1 \end{bmatrix} \tag{2}$$

$$\equiv \begin{bmatrix} P_{11} & P_{12} & P_{14} \\ P_{21} & P_{22} & P_{24} \\ P_{31} & P_{32} & P_{34} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{3}$$

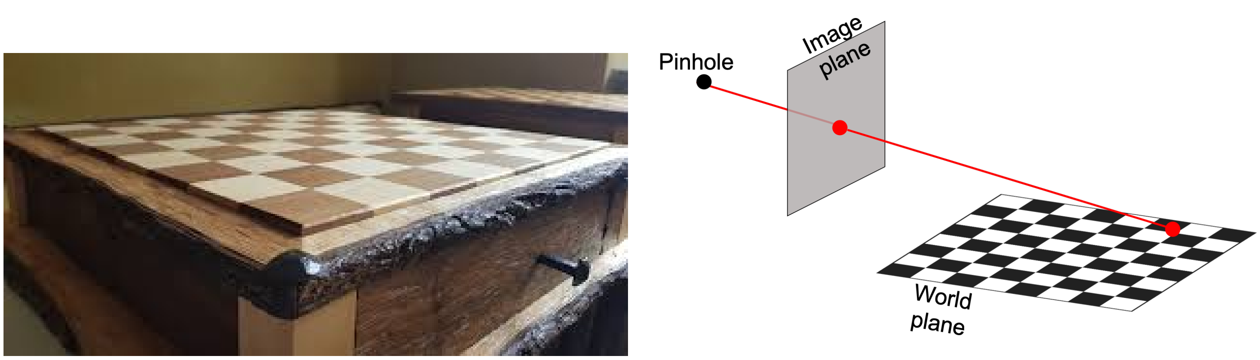

Figure 1: Left: An example of a plane in perspective projection.Right: In this case, a single camera is enough to precisely identify the 3D location of a given pixel.

Where the last line follows from the previous because of the fact that the Z coordinate of points on the plane are identically 0. So the third column of P does not come into play at all.

We can adorn the plane with a *2D* coordinate system with the X and Y axes aligned with the world X and Y axis. In this 2D coordinate system, the point $\mathbf{Q} = \begin{bmatrix} X \\ Y \\ 0 \end{bmatrix}$ is the same as the 2D point $\mathbf{r} = \begin{bmatrix} X \\ Y \end{bmatrix}$. Thus, the mapping between the 2D coordinate system of the plane, and the coordinate system of the image is given by:

$$\vec{\mathbf{q}} \equiv \begin{bmatrix} P_{11} & P_{12} & P_{14} \\ P_{21} & P_{22} & P_{24} \\ P_{31} & P_{32} & P_{34} \end{bmatrix} \vec{\mathbf{r}} \tag{4}$$

$$\equiv H\vec{\mathbf{r}} \tag{5}$$

where the matrix H is a 3×3 matrix and is called a *homography*.

Unlike P, H is a *square* matrix. It is thus possible that H is invertible. In fact, the only circumstance where H is not invertible is if it maps the world plane to a *line* in the image: i.e., when the camera is viewing the plane edge on. In all other cases, H is invertible. This allows us to map any pixel $\vec{\mathbf{q}}$ in the image to its corresponding location in the plane, $\vec{\mathbf{r}}$:

$$\vec{\mathbf{r}} = H^{-1}\vec{\mathbf{q}} \tag{6}$$

1 Estimating H

How do we estimate H? Again, we can use correspondences between points on the plane and points in the image. As with camera calibration, each correspondence between a plane point $\mathbf{r} = \begin{bmatrix} X \\ Y \end{bmatrix}$ and an image point $\mathbf{q} = \begin{bmatrix} x \\ y \end{bmatrix}$ gives us two equations:

$$\vec{\mathbf{r}} \equiv H\vec{\mathbf{q}} \tag{7}$$

$$\Rightarrow \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} \equiv \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{8}$$

$$\Rightarrow \lambda \begin{bmatrix} x \\ y \\ 1 \end{bmatrix} = \begin{bmatrix} H_{11} & H_{12} & H_{13} \\ H_{21} & H_{22} & H_{23} \\ H_{31} & H_{32} & H_{33} \end{bmatrix} \begin{bmatrix} X \\ Y \\ 1 \end{bmatrix} \tag{9}$$

$$\begin{aligned} \lambda x &= H_{11}X + H_{12}Y + H_{13} \\ \Rightarrow \lambda y &= H_{21}X + H_{22}Y + H_{23} \\ \lambda &= H_{31}X + H_{32}Y + H_{33} \end{aligned} \tag{10}$$

$$\Rightarrow \begin{aligned} (H_{31}X + H_{32}Y + H_{33})x &= H_{11}X + H_{12}Y + H_{13} \\ (H_{31}X + H_{32}Y + H_{33})y &= H_{21}X + H_{22}Y + H_{23} \end{aligned} \tag{11}$$

Given a set of correspondences, we can set up a system of linear equations. As with camera calibration, these equations are not enough to identify a unique H, since H and αH will both satisfy the system of equations ($\alpha \neq 0$). So we need to add an additional constraint, $\|H\|_F = 1$. Together, as with camera ccalibration, we can solve for H using SVD.

How many correspondences do we need to estimate H? H is a 3×3 matrix, so it has 9 entries that we must estimate. The constraint $\|H\|_F = 1$ gives us one equation, so we need 8 more. Each correspondence yields 2 equations, so we need a minimum of *4 correspondences*.