The Mathematics of Working with Digital 3D Imagery

Figure 1: A Photograph Taken at the Army-Navy Game

The goal of this series of papers is to analyze images, like the one shown in Figure 1, taken with digital cameras at sporting events and in laboratory settings. We work with two coordinate systems.

- A three-dimensional real world coordinate system. See Figure 2 on page 2. We use the center of the field as the origin. The x-axis runs along the length of the field midway between the sidelines and the y-axis runs along the 50 yard line. The z-axis represents height above the field. Since football fields are crowned, we use the height around the goal lines and sidelines as the zero point on the z-axis. Thus, the origin is actually a bit below the playing surface at the center of the field. We use yards as units. We use (x, y, z) to denote real world coordinates.

- A two-dimensional image coordinate system. See Figure 3 on page 2. We use pixels as units. The origin is in the center of the image. The x-axis runs horizontally and the y-axis runs vertically. We use (s, t) to denote image coordinates.

Figure 2: Real World Axes on a Photograph Taken at the Army-Navy Game

Figure 3: Image Axes on a Photograph Taken at the Army-Navy Game

The key to analyzing the three-dimensional real world from two-dimensional digital images is the ability to convert back-and-forth between three-dimensional real world coordinates and two-dimensional image coordinates. Because we are interested in the three-dimensional real world coordinates of features we will often work with pairs of images, like the ones shown in Figure 4 on page 3, that show the same scene but are taken from different viewpoints. The techniques we develop enable us to estimate the three-dimensional real world coordinates of features that appear in both images and are at the same real world

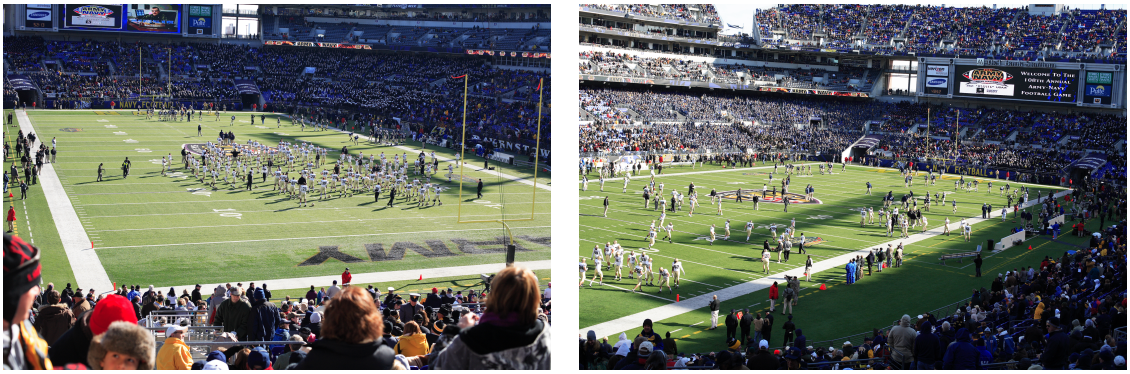

Figure 4: Two Photographs Taken at the Army-Navy Game

location in both images. For example, the people on the field in these two photographs moved between the time the two photographs were taken but the goalposts at the far end of the field did not move. It is ideal to work with pairs of images that were taken simultaneously.

Using the techniques we develop in this series of papers, we estimated from these two photographs that the real world coordinates of a particular point on these goalposts were $(-60.6, -3.2, 3.4)$. According to the regulations the coordinates of this point should be $(-60, -3.06, 3.33)$. Our estimate is remarkably good since the camera was over 150 yards from the point for each of the two photographs,

The mathematics we discuss in this paper is a nice subject of study in a multivariable calculus class. It involves three-dimensional geometry – the dot and cross products and projections – and minimizing functions of several variables. We minimize one function of seven variables using the *Mathematica* **FindMinimum** procedure or the Excel **Solver** tool and we minimize one function of two variables by solving for the point at which the two partial derivatives are zero. This latter minimization is a great multivariable calculus exercise.

Converting back-and-forth between real world and image coordinates, I

We work with a simplified model in which the light rays from the camera's sensor to the real world are straight lines that pass through the optical center of the lens. The implies that the image of a straight line will be a straight line. This is not always true. Modern lenses often have more complex optical paths and can produce significant distortion – for example, pinhole distortion or barrel distortion. Such distortion is often particularly evident in wide angle lenses and zoom lenses. If you have a digital SLR then the best lens is a prime (single focal length) lens of "normal" focal length. Normal focal length for a full frame digital SLR is 50mm. Normal focal length for an APS digital SLR is 30-35mm.

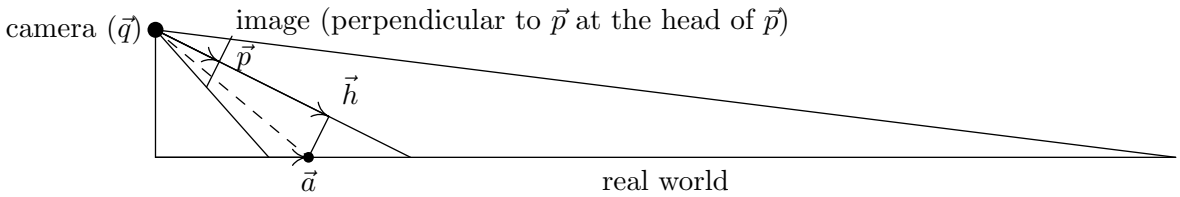

Figure 5: The Optical Center of the Camera, the Image, and the Real World

Figure 5 is the key to understanding how we convert back-and-forth between two-dimensional image coordinates and three-dimensional real world coordinates for each feature. The large dot represents the optical center of the lens but you can also think of it as your eye if you place your eye at the same point as the optical center of the lens when the picture was taken. Think of the image as being transparent and hold the image in front of you so that as you look through the image what you see matches the real world. The vector $\vec{p}$ will go from your eye to the center of the image and will be perpendicular to the image.

Now consider a light ray going from your eye (or the optical center of the lens) through the image and hitting a point $\vec{a}$ in the real world. This light ray is indicated by a dashed line in Figure 5. It is $\vec{a} - \vec{q}$ and the length of its projection on the vector $\vec{p}$ is

$$||\vec{a} - \vec{q}|| \cos\theta = \frac{(\vec{a} - \vec{q}) \cdot \vec{p}}{||\vec{p}||}$$

where θ is the angle between $(\vec{a} - \vec{q})$ and $\vec{p}$. Thus, the projection of $(\vec{a} - \vec{q})$ on $\vec{p}$ is

$$\vec{h} = \left(\frac{(\vec{a} - \vec{q}) \cdot \vec{p}}{||\vec{p}||} \right) \left(\frac{\vec{p}}{||\vec{p}||} \right)$$

Now, the vector $(\vec{a} - \vec{q} - \vec{h})$ is perpendicular to the vector $\vec{p}$ and, obviously,

$$\vec{a} - \vec{q} = \vec{h} + (\vec{a} - \vec{q} - \vec{h}).$$

Finally, the vector

$$\vec{w} = \left(\frac{||\vec{p}||}{||\vec{h}||} \right) (\vec{a} - \vec{q} - \vec{h})$$

is also perpendicular to the vector $\vec{p}$ and $\vec{w} + \vec{p}$ is the vector from $\vec{q}$ to the point on the image through which the light ray from your eye (or the optical center of the lens) to the point $\vec{a}$ passes. See Figure 6 on page 5. We need to find the two-dimensional image coordinates of $\vec{w}$. This is accomplished by using unit vectors $\vec{u}$ and $\vec{v}$ that point in the positive s-direction (horizontally) and the positive t-direction (vertically) on the image.

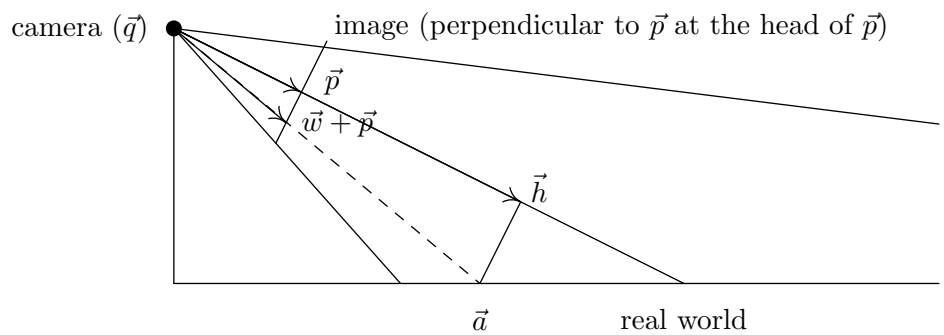

Figure 6: The Optical Center of the Camera, the Image, and the Real World

The vector $\vec{u}$ is determined by how the camera was rotated about the vector $\vec{p}$ when the image was made. It is perpendicular to $\vec{p}$ and since it is a unit vector this vector involves only one new parameter. If you hold your camera in the usual way using your left hand to hold it up to your eye and extend your right arm out to the side then the vector $\vec{u}$ is pointing in the direction of your right hand. The vector $\vec{v}$ is a unit vector perpendicular to $\vec{p}$ and $\vec{u}$ and is completely determined by these vectors and the right hand rule. Operationally, it is computed using the cross product.

Using the four vectors, $\vec{q}$, $\vec{p}$, $\vec{u}$, and $\vec{v}$ we convert from three-dimensional real world coordinates to two-dimensional image coordinates by:

$$s = \vec{w} \cdot \vec{u} \quad \text{and} \quad t = \vec{w} \cdot \vec{v}.$$

The four vectors are determined by seven real parameters. These four vectors are called the **camera vectors**.

Calibrating images

Although in theory one could take photographs after precisely measuring the seven parameters that determine these four vectors, in practice this is often impractical. If the image has some recognizable features whose three-dimensional real world coordinates are known then we can estimate the seven parameters. This process is called **calibrating** the image. If we have n such features we have n data points

$$(x_1, y_1, z_1, s_1, t_1), (x_2, y_2, z_2, s_2, t_2), \ldots (x_n, y_n, z_n, s_n, t_n).$$

with the known three real world coordinates and the two image coordinates measured on the photograph for each point. The program **3D-Image-Marker** makes it easy to collect this data in a form that can be copied and then pasted either into *Mathematica* or a spreadsheet. Given values for the seven parameters, we use the work above to compute

the predicted values, $\hat{s}_i$ and $\hat{t}_i$, of the image coordinates, s and t. Then we estimate the values of the seven parameters by minimizing

$$\mathcal{E} = \sum_{i=1}^{n} (\hat{s}_i - s_i)^2.$$

The *Mathematica* notebook **AN76.nb** or the Excel spreadsheet **ICTCM_KErickson.xls** can be used together with the corresponding documentation to do this.

Converting back-and-forth between real world and image coordinates, II

Now we are ready to analyze images. The key ingredient now is converting from two-dimensional image coordinates to three-dimensional real world coordinates. The basic idea is simple. A feature whose image coordinates are (s, t) lies on the line

$$L(r) = \vec{q} + r(\vec{p} + \vec{w}) = \vec{q} + r(\vec{p} + s\vec{u} + t\vec{v}).$$

Notice with just one image we cannot determine the three dimensional coordinates. We discuss three situations.

- We have two images made from two different vantage points that show the feature of interest.
- The feature of interest is on the ground – for example, we are tracking a hockey puck on the ice.
- The feature of interest lies in a known plane perpendicular to the ground – for example, if we were tracking the flight of a baseball on a relatively calm day.

Suppose we have two images with camera vectors $\{\vec{q}_1, \vec{p}_1, \vec{u}_1, \vec{v}_1\}$ and $\{\vec{q}_2, \vec{p}_2, \vec{u}_2, \vec{v}_2\}$. Suppose that the feature of interest has coordinates (s_1, t_1) on the first image and (s_2, t_2) on the second image. Then, in theory, the feature is located at the intersection of the two lines

$$L_1(r_1) = \vec{q}_1 + r_1(\vec{p}_1 + s_1\vec{u}_1 + t_1\vec{v}_1) \quad \text{and} \quad L_2(r_2) = \vec{q}_2 + r_2(\vec{p}_2 + s_2\vec{u}_2 + t_2\vec{v}_2).$$

In practice, however, there is usually some measurement or roundoff error and these rays will not intersect. Thus, we look at the function

$$\mathcal{E}(r_1, r_2) = \left(L_1(r_1) - L_2(r_2)\right)^2$$

and minimize this function by solving the pair of equations

$$\frac{\partial \mathcal{E}}{\partial r_1} = 0 \quad \text{and} \quad \frac{\partial \mathcal{E}}{\partial r_2} = 0.$$

This is a straightforward multivariable calculus problem. It has a single solution because the function $\mathcal{E}$ is quadratic in r_1 and r_2.

Our estimate for the three-dimensional real world coordinates of the feature of interest is then

$$\frac{1}{2}\left(L_1(r_1) + L_2(r_2)\right)$$

where (r_1, r_2) is the unique solution of the pair of equations above.

Next we consider the case when the feature of interest is on the ground. In this case it is located on the ray

$$L(r) = \vec{q} + r(\vec{p} + s\vec{u} + t\vec{v}).$$

But, now we know that its z-coordinate is zero and, thus, using q_3, p_3, u_3, and v_3 for the z-coordinates of the vectors $\vec{q}$, $\vec{p}$, $\vec{u}$, and $\vec{v}$, respectively, we have

$$q_3 + r(p_3 + su_3 + tv_3) = 0$$

so

$$r = -\frac{q_3}{p_3 + su_3 + tv_3}.$$

Using this value of r we see that the real world three-dimensional coordinates of the feature of interest are

$$L(r) = \vec{q} + r(\vec{p} + s\vec{u} + t\vec{v}).$$

Next we look at a situation like tracking a baseball from the time it is hit by the batter until the time it is caught by a fielder. If there isn't much wind and the ball is hit squarely the ground track of the ball's trajectory should be close to a straight line that can be determined from the point at which the ball was hit and the point at which the ball was caught. Thus, points (x, y, z) on the ball's trajectory will satisfy an equation of the form $ax+by = c$ and the three-dimensional real world coordinates of each point on the trajectory can be determined by finding the intersection of the line $L(r)$ and the plane $ax + by = c$.

Notice that all the math we've done here is standard multivariable calculus and the kind of three-dimensional geometry that is usually taught in a multivariable calculus course. Thus, analyzing digital imagery is of interest in undergraduate education for two reasons – first, because it can provide interesting real world data for applications – and, second, because the kinds of work we've done here are in themselves interesting applications of the mathematics covered in multivariable calculus courses.