

Basic 3D Geometry for One and Two Cameras

James Coughlan

Aug. 18, 2009

1 Introduction

This note discusses 3D geometry for a single camera in Sections 2-6, and then covers the basics of stereo in Sections 7-9. It combines most of the contents of two older documents, “Notes on 3D Geometry” and “Enforcing a Ground Plane Bias by Image Shearing.”

2 Basic Projection Equations

The (u, v) image coordinates are:

$$u = c - w/2 \tag{1}$$

$$v = r - h/2 \tag{2}$$

where c is the column and r is the row. (This is a Point Grey convention, and it assumes the camera center is exactly in the center column and row of the image. In fact, proper camera calibration can be done to get a better estimate of the center.)

The (X, Y, Z) coordinate system is shown in Fig. 1, and the (u, v) system in Fig. 2: Z points along line of sight, away from camera; X (and also u) points from left to right as you face the image plane (your eyes are pointing in the Z direction); similarly, Y (and also v) points from top to down as you face the image plane. $(0, 0, 0)$ is the camera center (the pinhole in a pinhole camera).

The projection equations are:

$$u = fX/Z \tag{3}$$

$$v = fY/Z \tag{4}$$

where f is the camera focal length. (If u and v are expressed in pixels, then f should also be measured in pixels.)

3 Vanishing Points

Any 3D direction gives rise to a vanishing point (VP) in the image. Given a 3D direction vector d , we can parameterize a line in that direction as all points

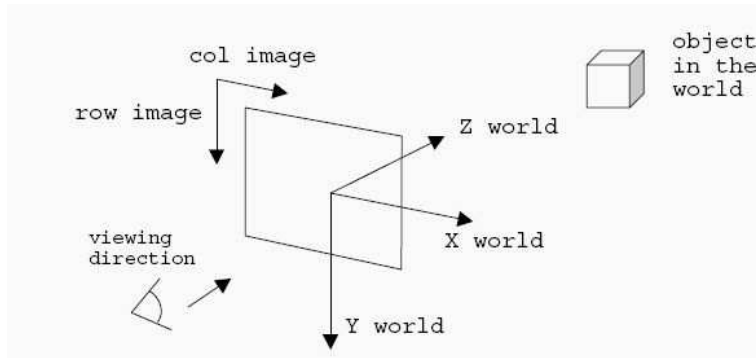


Figure 1: XYZ camera-centered coordinate system, following the Point Grey convention. Note that the positive Z-axis is the camera line of sight, pointing away from the eye.

$p(t) = p_0 + td$, where t is a real number. It is easy to show that the VP, defined as the value of (u, v) in the limit of t approaching $-\infty$ or $+\infty$, is given by $(u, v) = (fd_x/d_z, fd_y/d_z)$. Notice that this doesn't depend on p_0 – in other words, all parallel lines sharing the same direction have the same VP.

4 Calculating Horizon

A plane defines a 2D space of directions, and each direction has its own VP; you can show that the set of VPs forms a line in the image, called the horizon.

Assume we are given a ground plane defined by $r \cdot n = k$, where $r = (X, Y, Z)$, $n = (n_x, n_y, n_z)$ = ground plane unit vector. Then you can show that the horizon is defined by all points (u, v) such that $n \cdot (u, v, f) = 0$. Notice that this equation doesn't depend on the value of k .

The way to prove this is to suppose that we are given a vector a that is perpendicular to n , i.e. $a \cdot n = 0$. The VP of a is $(u, v) = (fa_x/a_z, fa_y/a_z)$. Then $n \cdot (u, v, f) = (f/a_z)(n_x a_x + n_y a_y + n_z a_z) = 0$.

5 Unwarping Ground Plane

Suppose the ground plane is defined by $r \cdot n = k$, where $k < 0$ and $|k|$ is the height of the camera above the ground. (We assume that n is the unit-length *upward*-facing normal coming out of the ground, the opposite of the down direction.) Then any point in the image *below* the horizon (below in the image sense) maps to a 3D location on the ground plane, and we can describe this 3D location in 2D ground coordinates.

First we find the mapping from (u, v) to the corresponding 3D location on ground plane:

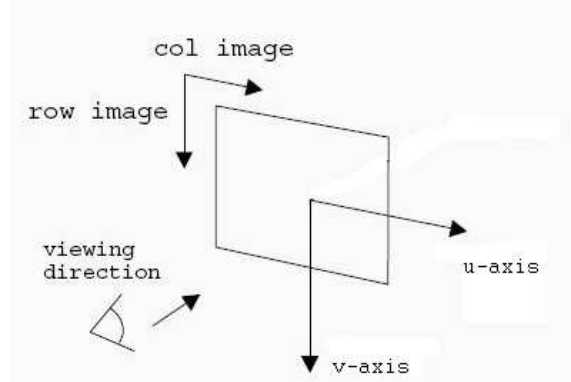


Figure 2: uv coordinate system that corresponds to projecting XYZ coordinates to the image plane, again following the Point Grey convention.

From $u = fX/Z$ and $v = fY/Z$ we have that $(uZ/f, vZ/f, Z) \cdot n = k$. Then $Z = k/[n \cdot (u/f, v/f, 1)]$ and we can also solve for X and Y using the $u = fX/Z$ and $v = fY/Z$ equations.

Next we describe the 3D location in 2D ground coordinates:

We need to find a 2D basis in the ground plane, i.e. two vectors a and b lying in the ground plane that are orthogonal. To do this we will first project the X -axis onto the ground plane by calculating how much of it projects onto the ground normal, and subtracting that off:

$a \propto \hat{X} - (n \cdot \hat{X})n = (1, 0, 0) - n_x n$, where \hat{X} is the unit vector in the X -direction. Next we normalize a to unit length (not shown here). Then we calculate $b = n \times a$, where \times is the 3D vector product. If the camera is roughly horizontal, then a will be approx. parallel to the X -direction and b will be in the forward direction along the ground.

Finally, to project $r = (X, Y, Z)$ into 2D ground coordinates, first define $p_0 = kn$, which is the 3D location of the user's shoes (i.e. the spot on the ground directly below the camera). Then we get 2D ground coordinates (u', v') , where $u' = (r - p_0) \cdot a$ and $v' = (r - p_0) \cdot b$. These coordinates will have units of meters if k is defined in meters.

6 Mapping Image Pixels onto Ground Plane

Given the ground plane defined as above, $r \cdot n = k$, then given a pixel with image coordinates (u, v) we would like to determine the (X, Y, Z) coordinates of this pixel *assuming it corresponds to a point lying on the ground plane*.

The solution is to rewrite the projection equations as $X = uZ/f$, $Y = vZ/f$ and to rewrite the plane equation in terms of Z (but not X or Y): $(uZ/f, vZ/f, Z) \cdot n = k$. We can then solve for Z as $Z = fk/[(u, v, f) \cdot n]$, and can use the rewritten projection equations, $X = uZ/f$, $Y = vZ/f$, to

determine X and Y .

7 Basic Stereo Equations

Now we discuss stereo for a pair of rectified cameras, for which a row in the left camera matches the same row in the right camera.

We use the same (u, v) and (X, Y, Z) coordinate systems as before. The reference (right) camera's center is at $(0, 0, 0)$ as before, but the center of the other camera is at $(-B, 0, 0)$ and is shifted by a baseline distance $B > 0$ from the reference camera. Note that the image coordinates (u, v) refer to the reference image, not the other image.

The projection equations relate (u, v) , (X, Y, Z) and the disparity d :

$$u = fX/Z \tag{5}$$

$$v = fY/Z \tag{6}$$

$$Z = fB/d \tag{7}$$

In the Point Grey convention, the right image $R(u, v)$ is the reference image, and the left image is $L(u, v)$. This convention gives us the following interpretation for the disparity map $d(u, v)$: (u, v) in the right image maps to $(u+d(u, v), v)$ in the left image. In other words,

$$L(u + d(u, v), v) \approx R(u, v) \tag{8}$$

provided the camera gain is similar in the left and right images.

8 Disparity of Ground Plane

As before, the ground plane is defined by $r \cdot n = k$, where $r = (X, Y, Z)$, $n = (n_x, n_y, n_z)$ =ground plane unit vector.

If we assume every point in the scene lies on the ground plane then we can project any point onto this plane:

Since $X = uZ/f$ and $Y = vZ/f$ we get $Z = fk/(n \cdot (u, v, f))$.

Finally, since in general $Z = fB/d$ we have the predicted disparity:

$$d_g(u, v) = fB/Z = Bn \cdot (u, v, f)/k \tag{9}$$

i.e. the predicted disparity (assuming every point in the scene lies on the ground plane) as a function of u and v , called $d_g(u, v)$. Note that the disparity is *linear* in the image row and column coordinates.

9 Converting Between Disparity and Elevation

This is useful for implementing an MRF in terms of elevations instead of disparities.

Given a unit normal direction vector n , define the elevation of any point $r = (X, Y, Z)$ to be $E(r) = n \cdot r$. (Sometimes we measure the elevation relative to a ground plane defined by $n \cdot r = k$, in which case we subtract k from the above expression for $E(r)$. In this section we will continue to use $E(r)$, which is by definition equal to zero at the camera center.)

Converting disparity to elevation: $E(u, v, d) = r \cdot n = (Zu/f, Zv/f, Z) \cdot n$ where $Z = fB/d$.

Converting elevation to disparity: this is a bit more work. $d(u, v, E) = fB/Z$ so let's solve for Z . We know that $E = (X, Y, Z) \cdot n = (Zu/f, Zv/f, Z) \cdot n = Z/f(u, v, f) \cdot n$, so $Z = \frac{fE}{(u, v, f) \cdot n}$. Therefore

$$d(u, v, E) = \frac{B}{E}(u, v, f) \cdot n \quad (10)$$

Note that these equations hold whether or not the baseline is roughly horizontal (i.e. $n_x \approx 0$).