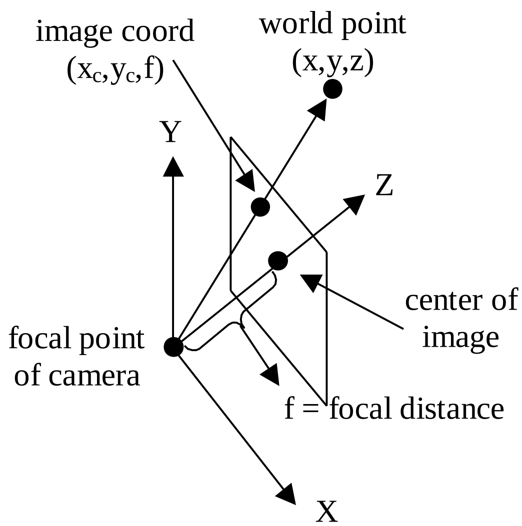
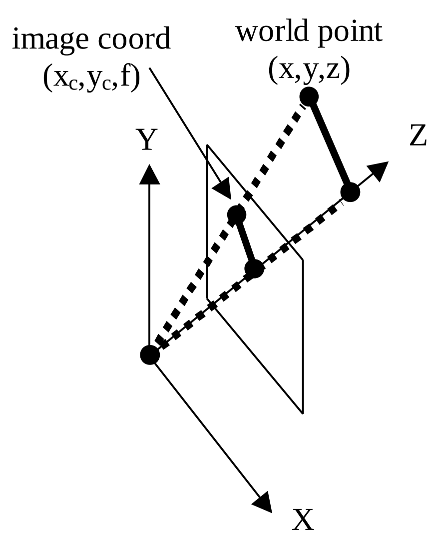
Lecture notes: Stereo

Before discussing stereo, we must review the **perspective projection camera model.** Imagine viewing the world through a square hole cut into a piece of paper. The hole is the image, and your eye is the focal point of a set of rays that see the world:



Place a Cartesian (X,Y,Z) coordinate system centered at the focal point, with the XY-plane parallel to the image. (Yes, I know it is left-handed as drawn – does't matter.) The distance from the focal point to the image plane is called "f" the focal distance, so any point in the image plane is at (x_c, y_c, f).

How are the image coordinates (x_c, y_c, f) related to the world coordinates (x, y, z)? Notice the similar trianges:



This sets up the similar triangle equations:

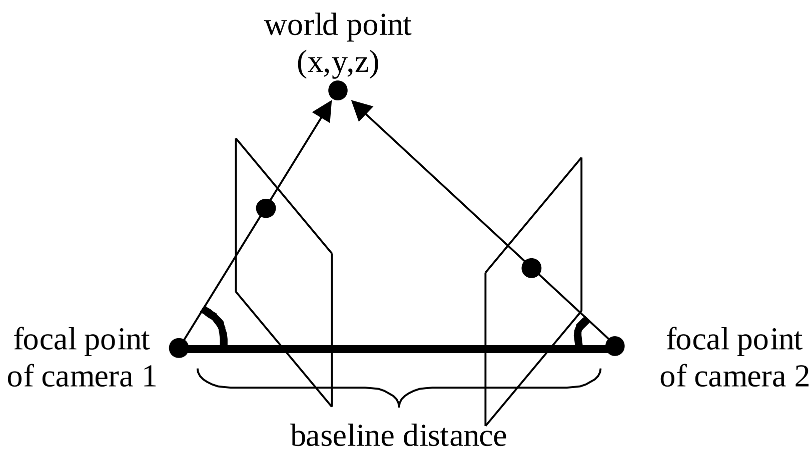
$$\frac{z}{f} = \frac{y}{y_c} = \frac{x}{x_c}$$

or solving for the image coordinates:

$$y_c = f\frac{y}{z} \quad \text{and} \quad x_c = f\frac{x}{z}$$

Suppose we know the image coordinates, but want to determine the original world coordinates (x,y,z) of the point? There are two equations and three unknowns. All we can know is the point is at some **depth** along the ray defined by (x_c, y_c).

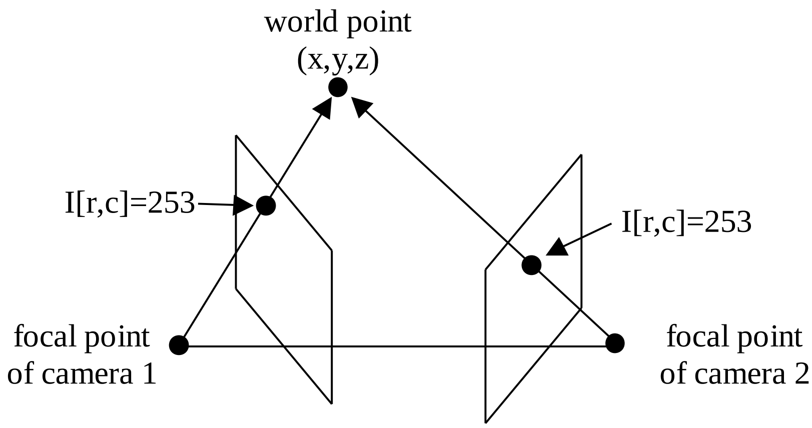
The thing most obviously missing in a color or greyscale image is explicit information about **3D location.** Although we know the color of a scene point, we do not know its **depth.** One way to approach this problem is to use a second camera to observe the same world point. This is called **stereo imaging**:



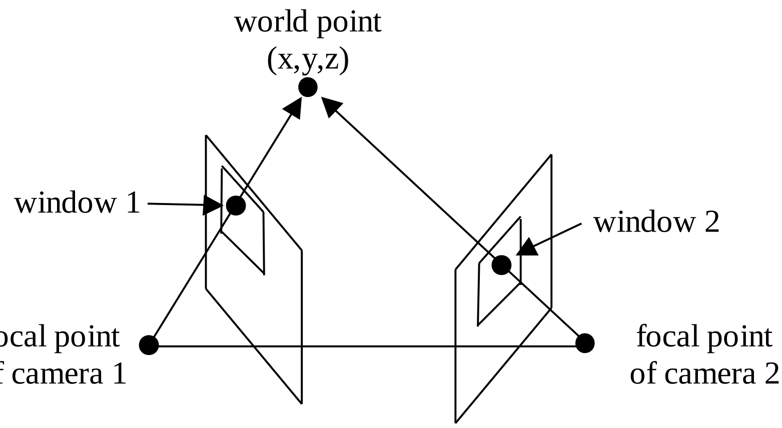
If we have a ray from two cameras observing the same point, we can triangulate. We can measure the **baseline distance** – the distance between the two cameras. We also know the angle each ray (pixel) makes towards the world point. From these three quantities we can find all properties of the triangle, and hence the location of the world point.

Camera calibration is the problem of determining a camera's position, orientation, and focal distance. We will look at this problem in a separate lecture, and here assume it is solved.

How is it that we can determine that two rays from different cameras are observing the same world point? Suppose they see the same color:



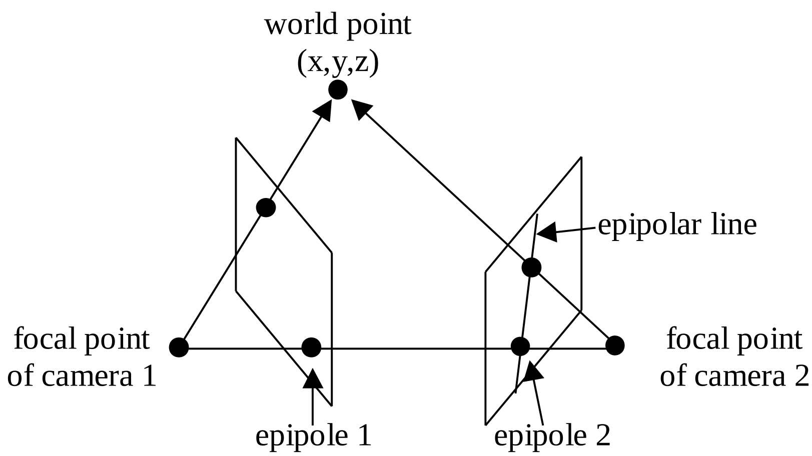
This is a prerequisite, but does it guarantee they are seeing the same world point? No it does not. A better guarantee could be had if a whole window of pixels saw the same thing:



The **correspondence problem** is searching for matches between images for pixels that see the same world point. Often it is approached using window methods, as just drawn. Sometimes it is approached using features, such as corners, lines, or even regions.

However the correspondence problem is approached, it is **doomed to failure** some of the time. For some scenes, there is simply nothing to match (for example a plain wall). The correspondence problem can only be solved when the world contains “interesting things”, meaning features that can be matched. Based on stereo evaluations I have read (e.g. JISCT), in a typical scene, about 30% of the scene is non-matchable.

Epipolar geometry provides some constraints that are useful for solving correspondence problems:



The **epipoles** are the intersections of the baseline with the image planes. An **epipolar line** is the line (in the image plane) for each pixel that passes through the epipole.

How is this useful? Given an image coordinate in camera 1, all possible locations of its world point appear on a single epipolar line in the image of camera 2. We can determine the epipolar line in camera 2 for any pixel in camera 1 by hypothesizing any possible world point for that pixel. This means that we can constrain the search for the correspondence problem to a 1D search along the epipolar line, as opposed to a 2D search in the whole image.