CS 7321 Computer Vision I (Winter 1997)
Notes on Image Formation
Irfan Essa

1. Image Formation:

A. Imaging Geometry:

Monocular Imaging:

Point projection is the fundamental model for the transformation wrought on by our eye, by cameras, or by numerous other imaging devices. A *pinhole camera* is a first order approximation used to describe the formation of an image by projecting a scene (containing objects) through a single point onto an *image plane*.

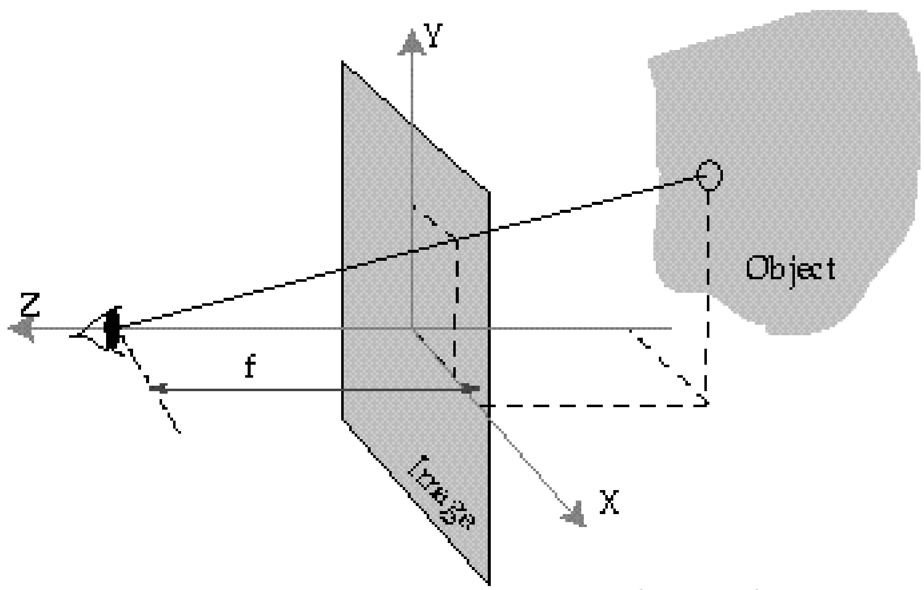

Figure 1: A pinhole camera model, with definition of terms.

Consider an object and a point on the object, $\mathbf{x}_o=(x_o, y_o, z_o)$ to be viewed and projected on the image $\mathbf{x}_i=(x_i, y_i, z_i)$. Using similar triangles, we can relate a point on the image to a point on the object.

$$\frac{x_o}{f - z_o} = \frac{x_i}{f} \quad (1)$$

Using the same for *y* and *z*, we get the following equation. Note that in the above model, the projected image has *z=0* everywhere, However, projecting away the *z* component is best considered a separate transformation; the projective transform is usually thought to distort the *z* component just as it does the *x* and *y* components. The perspective distortion thus maps:

$$(x_i, y_i, z_i) = \left(\frac{fx_o}{f-z_o}, \frac{fy_o}{f-z_o}, \frac{fz_o}{f-z_o}\right) \quad (2)$$

The perspective transformation yields *orthographic projection* as a special case when the viewpoint is the point at infinity in the z direction. Then all objects are projected onto the viewing plane with no distortion of their x and y coordinates.

Note that the difference between the above model and the one described in the Nalwa Book is that the image is at origin in Cartesian coordinates and the viewpoint is at focal length f from the origin.

Binocular Imaging:

Assume a system with two viewpoints, with the eyes not converging; they are aimed in parallel at the point at infinity in the $-z$ direction. The depth information about a point is then encoded only by its different positions (disparity) in the image planes. See the stereo arrangement below:

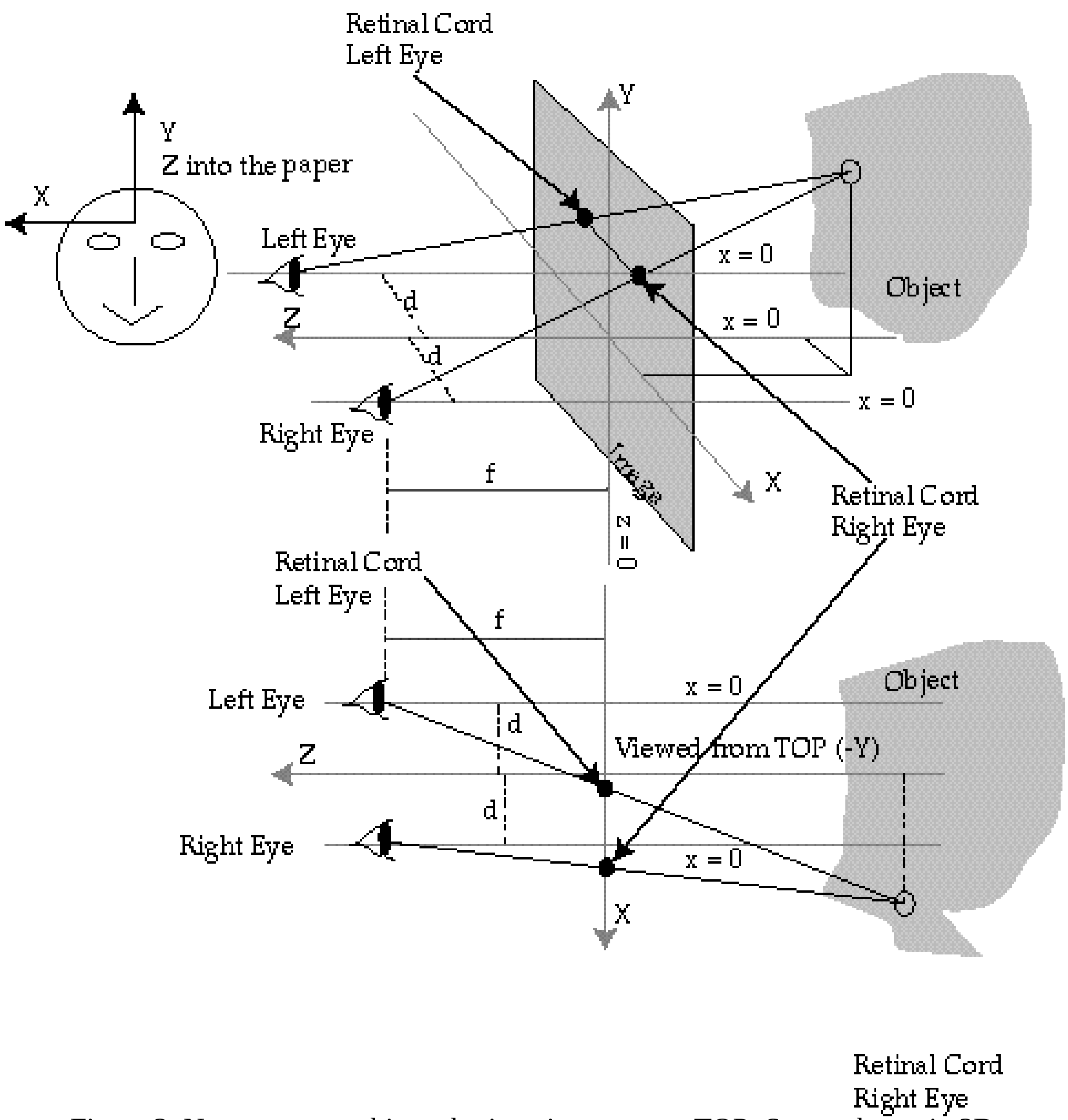

Figure 2: Nonconvergent binocular imaging system. TOP: Set up shown in 3D, BOTTOM: Looking from top down -y axis.

Again using similar triangles for the above arrangement with $(x_{i,L}, y_{i,L})$ and $(x_{i,R}, y_{i,R})$ as retinal coordinates for the world point s imaged by the left and right eyes, we get:

$$x_{i,L} = \frac{(x_o - d)}{f - z_o} f$$
$$x_{i,R} = \frac{(x_o + d)}{f - z_o} f \qquad (3)$$

The baseline of the *binocular system* is *2d.* Solving for *z*, we get:

$$z_o = f - \frac{2df}{x_{i,R} - x_{i,L}} \qquad (4)$$

Thus if points can be matched to determine the disparity, $(x_{i,R} - x_{i,L})$ and the baseline and the focal length are known, the *z* coordinate (depth) can be computed.

If the system can converge its directions of view to a finite distance, convergence angle may also be used to compute depth. The hardest part of extracting depth information from stereo is the matching of points for disparity calculations.

B. Reflectance:

Terminology:

1.	Φ	Light Energy Flux Brightness is measured with respect to area and solid angle	watts
2.	$I = \frac{d\Phi}{d\omega}$	Radiant Intensity of a source (exitant flux per unit solid angle)	watts/steradian
3.	$d\omega$	Incremental solid angle	steradian
4.	$d\omega = \frac{dA}{r^2}$	Solid angle of a small areas dA measured perpendicular to a radius r.	
5.	$E = \frac{d\Phi}{dA}$	Irradiance (flux incident on surface element dA).	watts/sq. meter
6.	$L = \frac{d^2\Phi}{dA \cos\theta \, d\omega}$	Radiance (flux emitted per unit foreshortened surface area per unit solid angle)	watts/(sqm steradian)

Effects of Geometry on an Imaging System

Assume that an imaging device (Figure 3) is properly focused; rays originating in the infinitesimal area dA_o on the object's surface are projected into some small area dA_p in the image plane and no rays from other projections of the object's surface reach this area of the image (ideal system, observing laws of simple geometrical optics).

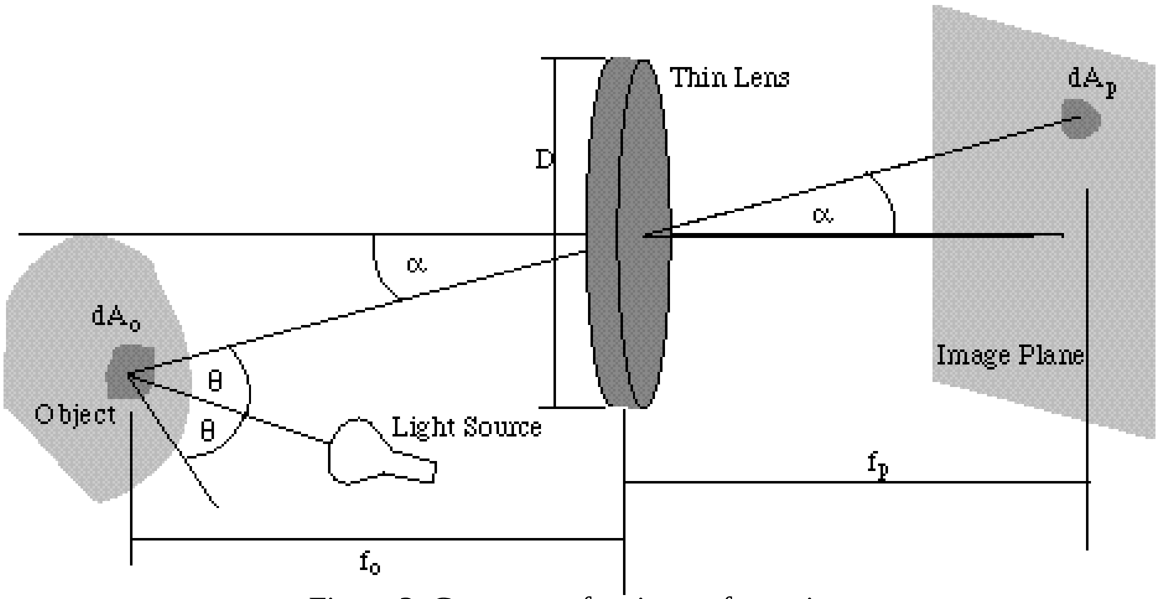

Figure 3: Geometry of an image formation system.

First consider the flux arriving at the lens from a small surface area dA_o based on point 6 above:

$$d\Phi = dA_o \int L \cos\theta \, d\omega \tag{5}$$

This flux is assumed to arrive at an area dA_p on the imaging plane. The irradiance on the imaging plane, E_p is given by:

$$E_p = \frac{d\Phi}{dA_p} \tag{6}$$

Now by relating the respective solid angles we can relate dA_o and dA_p as follows:

$$dA_o \frac{\cos\theta}{f_o^2} = dA_p \frac{\cos\alpha}{f_p^2} \tag{7}$$

we get:

$$E_p = \cos\alpha \left[\frac{f_o}{f_p}\right]^2 \int L \, d\omega \tag{8}$$

Since L can be considered constant and approximating $d\omega$ by the area of the lens foreshortened by $\cos\alpha$, that is $(\pi/4)D^2 \cos\alpha$ divided by $f_o/\cos\alpha$ squared.

$$d\omega = \frac{\pi}{4} D^2 \frac{\cos^3\alpha}{f_o^2} \tag{9}$$

which yields:

$$E = \frac{1}{4}\left[\frac{D}{f_p}\right]^2 \cos^4\alpha \; \pi L \qquad (10)$$

References:

Ballard and Brown 1982, *Computer Vision,* Prentice Hill.

Nalwa 1993, A Guided Tour of Computer Vision, Addison Wesley

Horn 1986, Robot Vision, MIT Press