cs676: Lecture 4: Image Formation

Vinay P. Namboodiri

Department of Computer Science and Engineering, IIT Kanpur

Note: This lecture note includes figures from lecture slides by Steve Seitz, Derek Hoiem, James Hays, Jitendra Malik, Kristen Grauman and also from [1]

1 Basic Image Formation

Consider the method to obtain images. Images are obtained by projecting the 3D world onto a 2D image plane. So, how can a 3D scene be projected? A naive attempt could involve exposing the 2D film to the 3D world. All the rays from all over impinge on the film if not controlled and the film so obtained would just be washed out with all the exposure.

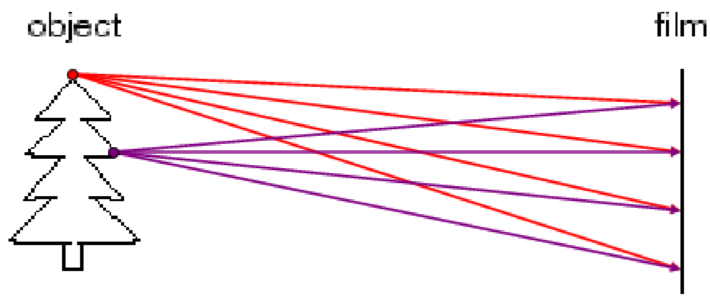

Fig. 1. Can we image an object by just exposing the film to the world?

The first attempt to make a photograph has been by using a pin hole camera. In this method, the rays of light are controlled by exposing a pin-hole aperture to the world as shown in figure 2. These were the first cameras invented by man. Termed *camera obscura*, these were used to create the first images known to man as shown in figure 3. We show below some examples of such cameras. The smaller the aperture, the sharper the image. However, making the aperture narrower reduces the amount of light that impinges and requires longer exposure.

1.1 Depth in the image

The distance is implicitly recorded in the image when it is perspectively projected. As illustrated in the figure 4, the depth in the image is taken into ac-

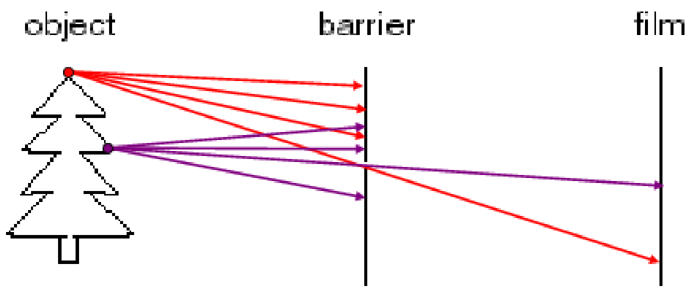

Fig. 2. Imaging using a pinhole camera

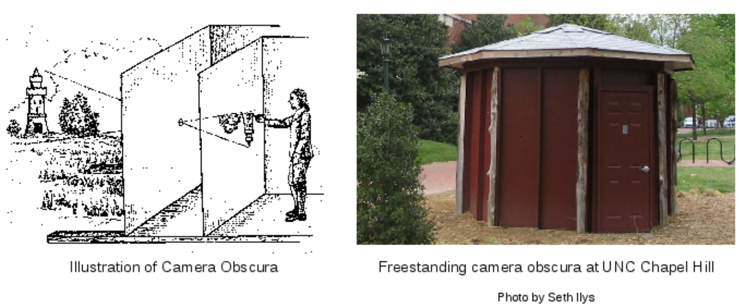

Fig. 3. *Camera obscura*

count during perspective projection by dividing the x, y co-ordinates by the z co-ordinate.

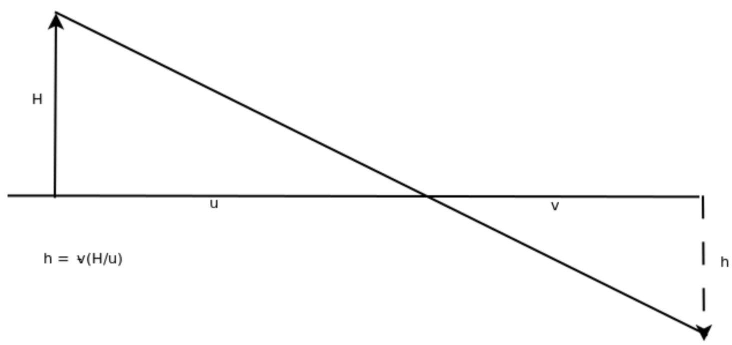

Fig. 4. Relation between the height of an object in the image and the scene depth

2 Perspective projection

Quoting [1], "The fundamental property of perspective is that all image points are collinear with the eyepoint and their corresponding world point." This principle is illustrated in this figure 5 . As can be seen from the figure 6, in perspective projection, very few things are preserved. Basically, length, area, angles between lines, all such quantities are not preserved. What is basically preserved is collinearity, i.e. if points lie on a particular line, then they continue to lie on a line. One crucial aspect is the absence of parallelism. Parallel lines when projected perspectively, meet at a point. Parallel lines on a plane meet at a point that is termed 'a vanishing point'. We see this aspect in real world images as well as can be seen in figure 7

2.1 Vanishing point

The vanishing point represents a point in 3D at infinity that is projected onto a 2D plane. How can we represent this point? This point can be characterised by the property that it is the point where parallel lines in the world meet when projected perspectively.

Different vanishing points exist for different set of parallel lines. The line at horizon is determined by connecting several such vanishing points on a plane.

3 Homogeneous coordinates

We now address the question of representing the points. When we represent points, one convenient way to represent them is using homogeneous co-ordinates.

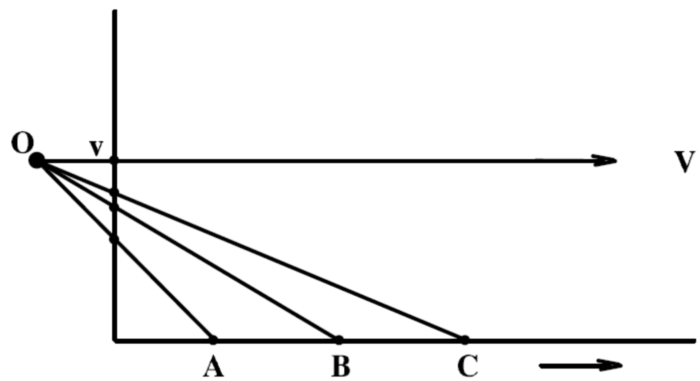

Fig. 5. A one dimensional construction of perspective viewing which illustrates the formation of a vanishing point, image from [1]

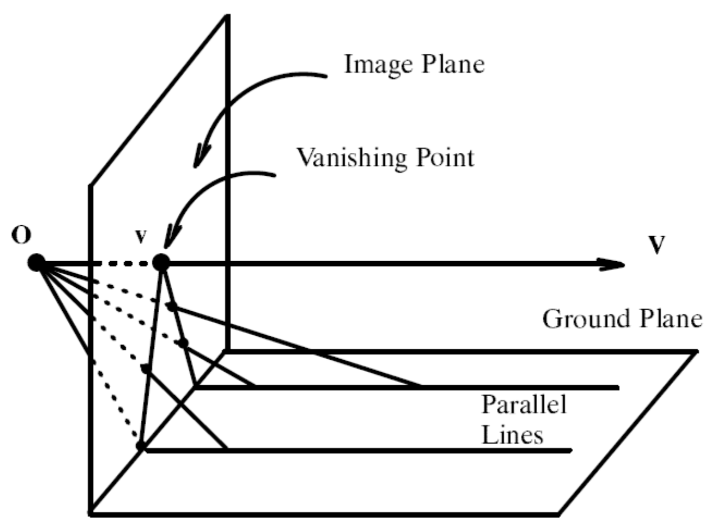

Fig. 6. A two dimensional construction of perspective viewing which illustrates the formation of a vanishing point, image from [1]

Fig. 7. Example of a perspective view where the railway tracks appear to converge to a point

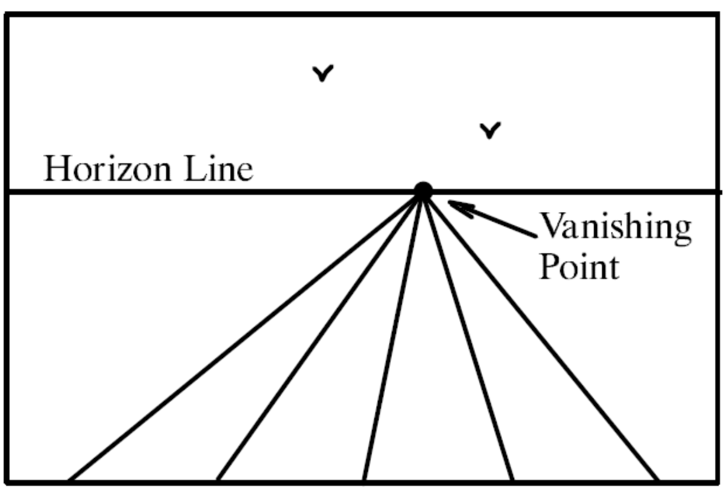

Fig. 8. A perspective view of a set of parallel lines in the plane. All lines converge to a single vanishing point, image from [1]

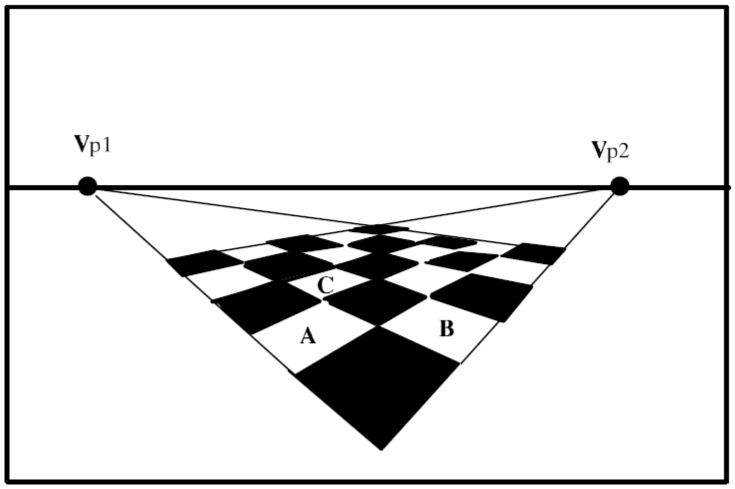

Fig. 9. Different directions determine different vanishing points. A tiled floor in perspective is shown here. Image from [1]

Let $[X, Y, Z]$ be the world co-ordinates for a point and the corresponding image co-ordinates be denoted by $[x, y]$. We can define homogeneous co-ordinates to represent these points. In homogeneous co-ordinates, the points in 2 dimensions are represented by 3 points and points in 3 dimensions are represented by 4 points. So the corresponding homogeneous co-ordinates for the points are $[X, Y, Z, 1]$ and $[x, y, 1]$. Note that the homogeneous co-ordinates are invariant to scaling.

Let $[x, y, w]$ be the homogeneous co-ordinate for a point. Then the corresponding cartesian co-ordinate is obtained by $[x/w, y/w]$. Similarly, if for a point in 3 dimensions the homogeneous co-ordinates are $[X, Y, Z, W]$, then the corresponding cartesian co-ordinates are $[X/W, Y/W, Z/W]$.

They are also a convenient means for representing the vanishing points as these can be represented by the third co-ordinate being zero.

Let there be a point P in 3D given by $[X, Y, Z]$. This can be represented in terms of a point A and a direction vector D by, the following equation

$$\begin{bmatrix} X \\ Y \\ Z \end{bmatrix} = \begin{bmatrix} A_x \\ A_y \\ A_z \end{bmatrix} + \lambda \begin{bmatrix} D_x \\ D_y \\ D_z \end{bmatrix} \tag{1}$$

We know that for the image co-ordinates, the value is given by
$x = v\frac{X}{Z}$ and similarly $y = v\frac{Y}{Z}$.

Therefore

$$x = v\frac{A_x + \lambda D_x}{A_z + \lambda D_z} \tag{2}$$

and similarly for y.

When $\lambda \to \infty$, the equation is given by

$$x = v\frac{D_x}{D_z} \tag{3}$$

and similarly for y.

Therefore the co-ordinates for the vanishing point in homogeneous co-ordinate system is given by

$$[v\frac{D_x}{D_z}, v\frac{D_y}{D_z}, 0]$$

.

3.1 Thin Lens

The pin-hole camera requires smaller and smaller apertures. One way that was discovered was to use a lens. The lens serves to focus the rays over a wider aperture onto the image plane as shown in figure 10. This enables a sharp image to be produced with the exposure time being limited. The basic camera is based on this principle.

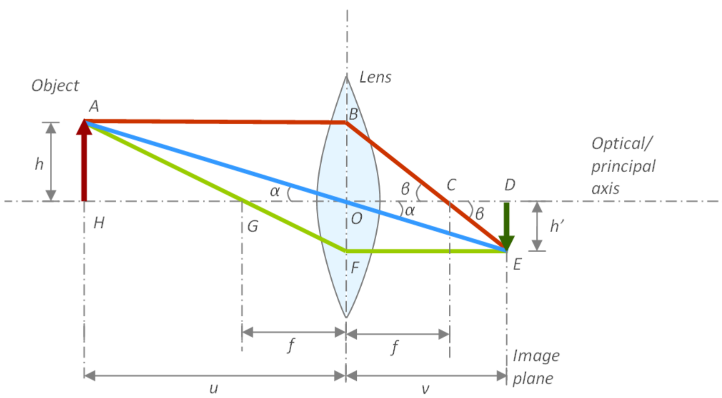

Fig. 10. Thin lens imaging and lens law

References

1. Mundy, J.L., Zisserman, A.: Appendix - projective geometry for machine vision (1992)