CSE 252A: Geometric Image Formation

Lecturer: David Kriegman

Scribed by Owen Jow on October 04, 2018

We can look at image formation from either a geometric perspective (*what is the pixel corresponding to a 3D point?*) or a photometric perspective (*how is a pixel's brightness determined?*). Today we'll take a geometric view of image formation, and study a few of the main projection models.

1 Perspective Projection

In perspective projection, (barring degenerate cases) points map to points, lines map to lines, and planes map to planes. However, angles, distances, and ratios of distances are not preserved.

In camera coordinates (with the z-axis pointing out into the world), the projection is $(-fx/z, -fy/z, -f)$. To avoid the sign change, we can look at a "virtual image plane" out in front of the camera. Note that x' and y' are nonlinear functions of the 3D point, since z is in the denominator. As points get further away ($z \to \infty$), projected values get smaller and thus closer to the center.

2 Projective Geometry

Projective geometry describes the geometry of projective transformations by adding points at infinity ("vanishing points") and allowing for transformations between these new points and good old Euclidean points. Unlike Euclidean geometry, projective geometry allows two parallel lines to meet at a vanishing point.

To represent points and lines in projective spaces, we use homogeneous coordinates.

2.1 Projective Plane Axioms

1. Every two points define a line.
2. **Every two distinct lines intersect at a point.**
3. If we define a line AB, there exists another point C which isn't on that line.

Thus the projective plane is just a little bit larger than the infinite Euclidean plane. It's a Euclidean plane *plus* the vanishing points at which parallel lines intersect. The collection of such points is called the *line at infinity*.

We have to go just a little bit farther than infinity to get to these vanishing points.

2.2 Homogeneous Coordinates

Since the projective plane is bigger than the Euclidean plane, we need extra numbers to represent our newfound points at infinity. If previously we had two coordinates, we now have three (homogeneous coordinates)! In homogeneous coordinates, a point at infinity is one with a final coordinate of 0.

There is an equivalence relation defined over homogeneous coordinates, i.e. all nonzero scalings of a point in homogeneous coordinates are identical. To obtain the canonical Euclidean form, we divide out by the final coordinate. (*Accordingly, there is no corresponding Euclidean point for a vanishing point which has final coordinate 0.*)

By the equivalence relationship, every point $\lambda \cdot (x, y, 1) = (\lambda x, \lambda y, \lambda)$ is identical. Such points lie on a line through the origin and $(x, y, 1)$. In other words, a single point is a 1D linear subspace in homogeneous coordinates.

Now, the intersection of parallel lines is a point at infinity $(x, y, 0)$.

Homogeneous coordinates also let us write perspective projection as a matrix transformation:

$$\begin{bmatrix} x' \\ y' \\ z' \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & -1/f & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

even though we'd previously outed the perspective projection equations as being nonlinear!

2.3 Projective Transformations

How to map a projective plane (e.g. 3D plane) to another projective plane (e.g. image plane)? The simplest plane-plane mapping is a linear matrix relationship:

$$\begin{bmatrix} u_1 \\ u_2 \\ u_3 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{bmatrix} \begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix}$$

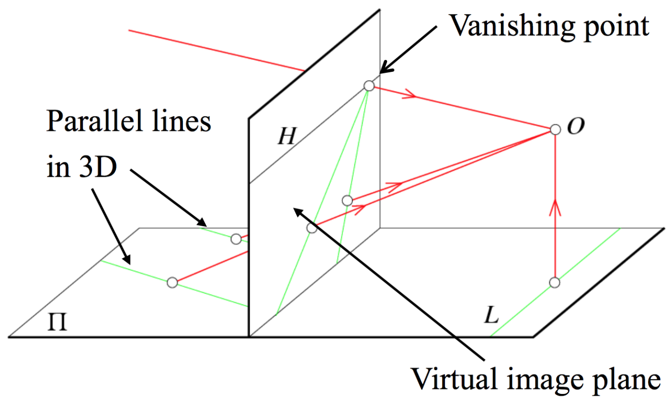
Under this formulation, points map to points (substituting $\mathbf{x}$ with $\lambda\mathbf{x}$ results in $\lambda\mathbf{u}$), lines map to lines, and we can map between non-infinite and infinite points. We might end up with a point for which the final coordinate is 0 where we didn't have one before.

2.4 Vanishing Points

A vanishing point's location in the image is determined by the line parallel in 3D to the given line(s) which also passes through the center of projection. By this construction, a single line also has a vanishing point (unless it is parallel to the image plane).

And of course *every* set of parallel lines defines a vanishing point, meaning there can be many in an image!

Note: even if lines in 3D intersect at a point at infinity, when projected onto the image their intersection can be a regular point again! In other words, we can go from a 3D point with final coordinate 0 to a 2D point with final coordinate $\neq 0$ using perspective projection.



Observe the parallel lines in 3D. The vanishing point is determined as the red line at the top which is parallel to these lines *and* passes through O.

Image credit: CSE 252A slides.

3 Other Camera Models

A camera model is a mathematical expression that describes where 3D points in the world show up as 2D points in an image. Pinhole perspective projection is one such model, but there are many others, e.g. simpler ones. We will now discuss two additional models which are affine.

3.1 Affine Camera Model

The affine camera model is a simplification of the perspective projection model. It is a *first-order Taylor series approximation* of the perspective projection equation about a particular point (x_0, y_0, z_0). Accordingly, it's only valid in the 3D neighborhood around this point (not over the entire 3D space).

Let the perspective projection model be

$$\begin{bmatrix} u \\ v \end{bmatrix} = \frac{f}{z} \begin{bmatrix} x \\ y \end{bmatrix}$$

Then the affine camera model is

$$\begin{aligned} \begin{bmatrix} u \\ v \end{bmatrix} &= \begin{bmatrix} \frac{f}{z_0} x_0 + \frac{f}{z_0}(x - x_0) - \frac{f x_0}{z_0^2}(z - z_0) \\ \frac{f}{z_0} y_0 + \frac{f}{z_0}(y - y_0) - \frac{f y_0}{z_0^2}(z - z_0) \end{bmatrix} \\ &= \frac{f}{z_0} \begin{bmatrix} x_0 \\ y_0 \end{bmatrix} + \begin{bmatrix} f/z_0 & 0 & -f x_0/z_0^2 \\ 0 & f/z_0 & -f y_0/z_0^2 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \end{aligned}$$

which is in the form $\mathbf{Ap} + \mathbf{b}$ as one would expect from an affine transformation.

Written in terms of homogeneous coordinates,

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f/z_0 & 0 & -f x_0/z_0^2 & f x_0/z_0 \\ 0 & f/z_0 & -f y_0/z_0^2 & f y_0/z_0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix}$$

The main advantage of the affine camera model is simplicity: it's linear in Euclidean coordinates, not just homogeneous coordinates like perspective projection. Thus we can use it with techniques such as linear least squares.

3.2 Scaled Orthographic Projection

The scaled orthographic projection model is a simplification of the affine camera model. The point of the Taylor series expansion is now restricted to lie along the optical axis, i.e. its form is $(0, 0, z_0)$ instead of (x_0, y_0, z_0) as before.

Substituting $(x_0, y_0) = (0, 0)$ into the affine camera model, we get the scaled orthographic model:

$$\begin{aligned} \begin{bmatrix} u \\ v \end{bmatrix} &= \begin{bmatrix} f/z_0 & 0 & 0 \\ 0 & f/z_0 & 0 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \end{bmatrix} \\ &= \frac{f}{z_0} \begin{bmatrix} x \\ y \end{bmatrix} \end{aligned}$$

notably now linear. f/z_0 is the "scale" of scaled orthographic projection.

In homogeneous coordinates,

$$\begin{bmatrix} u \\ v \\ w \end{bmatrix} = \begin{bmatrix} f/z_0 & 0 & 0 & 0 \\ 0 & f/z_0 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x \\ y \\ z \\ t \end{bmatrix}$$

Basically, we get a scaled version of the x, y coordinates only; the depth z doesn't matter any more. Every 3D point with the same (x, y) [everything on the line (x, y, λ)] projects to the same 2D point. Lines of projection don't pass through the origin anymore. If $f/z_0 = 1$, lines of projection are in fact parallel to the optical axis.

Under affine projection (includes scaled orthographic), parallel lines and ratios of distances are preserved. (*In perspective projection, parallel lines map to lines that meet at a vanishing point. In an affine projection, parallel lines map to parallel lines.*)

3.3 When do these simplifications apply?

Recall that these camera models were derived via a Taylor series approximation, a first-order approximation no less, of the perspective projection model about a single point – that single point being either (x_0, y_0, z_0) or $(0, 0, z_0)$. *So they are only valid in the neighborhood around that point.*

All points of interest should be in the neighborhood of the point of expansion. The point of expansion can be anywhere in the scene.

A zoom lens, often looking very far away with a small field of view (FoV), is well approximated by an orthographic model because with a narrow FoV, everything is close to the optical axis. A satellite too, perhaps, looking down at the ground from afar.

To recap: these affine models are of course less accurate than perspective projection – they are approximations missing the higher-order terms. The benefit, when they apply, is a lack of nonlinearities.

3.4 *Other* Camera Models

These models, even perspective projection, are not universally applicable. A camera might be omnidirectional, it might have mirrors... basically, it might not obey these models at all.