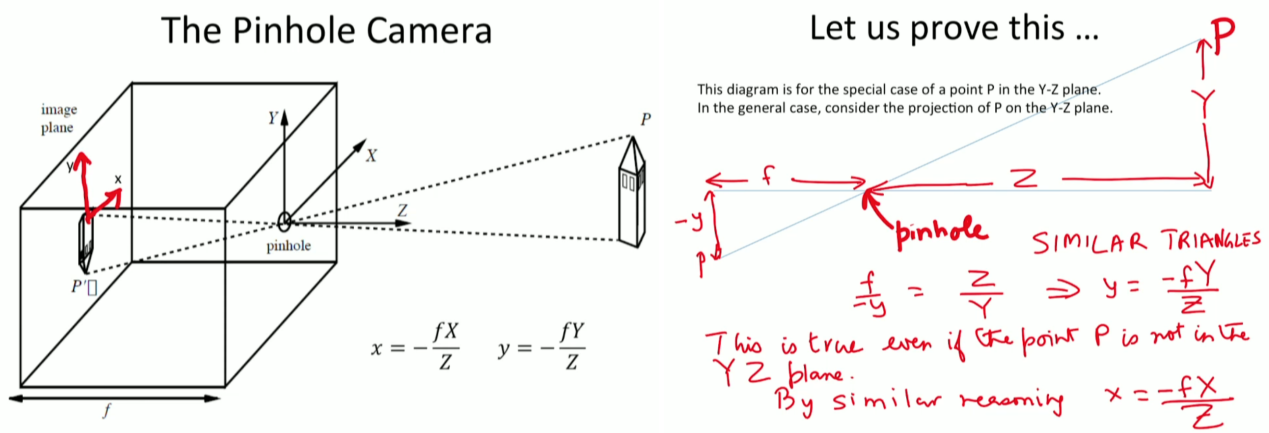
CS 280 Computer Vision
Spring 2018 Efros, Malik, Yu Lecture 2

1 Lecture

The *ABC*s of vision: understanding how an image is formed. An image $I(x, y)$ is created by the camera, which measures how much light is captured at each pixel (x, y). We would like to know where points in world space get imaged (i.e. what is the geometry of the conversion from (X, Y, Z) world coordinates to (x, y) image coordinates?) and how bright each resulting point is.

Geometry

We will first simplify the camera to be a pinhole camera (no lenses involved). We will put one coordinate system at the pinhole of the camera (the center of projection). It has an X, Y, and Z axis. Typically, the Z-axis is along the optical axis of the camera (a line which is perpendicular to the image plane and which goes through the pinhole), while the X- and Y-axes are parallel to those of the image plane. (We will also have a coordinate system (x, y) for the image plane, related to the uppercase coordinates by $x = -fX/Z$ and $y = -fY/Z$. In this way, (X, Y, Z) will correspond to the coordinates of a point P in the world, while (x, y) will correspond to the projection p of that point onto the image plane. Note that f is the z-distance from the image plane to the pinhole, while Z is the z-distance from the pinhole to the point in the world.)



source: slides from Jitendra Malik's CS 280 lecture

To avoid inversion, we can also put the image plane (imagining it as a transparent sheet) at a distance f in *front* of the pinhole, i.e. on the world side. In this case, $x = fX/Z$ and $y = fY/Z$.

We refer to the mapping from 3D world to 2D image as **perspective projection**. In perspective projection, parallel lines (parallel in the world) converge to a vanishing point in the image. Each family of parallel lines will have its own vanishing point. All vanishing points lie on a line which we call the horizon.

Why do parallel lines converge to a vanishing point? A general point on a line can be written as

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

where A is a point on the line and D is the direction of the line.

We then have

$$x = \frac{fX}{Z} = \frac{f \cdot (A_x + \lambda D_x)}{A_z + \lambda D_z}$$

What is the [x-coordinate of the] projection of a point which is infinitely far away? Let $\lambda \to \infty$. Then

$$x = \frac{f\lambda D_x}{\lambda D_z} = \frac{fD_x}{D_z}$$

because the λD_x and λD_z terms dominate (L'Hopital's rule). We can see that a line parallel to this original line (same direction, but goes through another point) would have the same projection, since the expression does not depend on A! Thus we see that parallel lines will converge to the same point.

Note: if D_z is 0, i.e. if the line is parallel to the image plane, the calculation is incorrect. Thus parallel *vertical* lines will remain parallel in the image.

Vanishing Point in Vector Notation

$$\mathbf{p} = f\frac{\mathbf{X}}{Z}$$

$\mathbf{p}$ is the projection of the point in the image plane. (By convention, lowercase signifies image quantities and uppercase signifies scene quantities.)

A line of points in 3D can be represented as $\mathbf{X} = \mathbf{A} + \lambda\mathbf{D}$, where $\mathbf{A}$ is a fixed point, $\mathbf{D}$ a unit vector parallel to the line, and λ a measure of distance along the line. As points grow increasingly further away,

$$\lim_{\lambda\to\infty} p = f\frac{\mathbf{A} + \lambda\mathbf{D}}{A_z + \lambda D_z} = f\frac{\mathbf{D}}{D_z}$$

and we see that the image of the line terminates in a vanishing point with coordinates $(fD_x/D_z, fD_y/D_z)$.

Other Observations

- **Nearer objects are lower in the image.**
  Proof. The equation of the ground plane is $Y = -h$, where h is the height of the camera, and a point on the ground plane will have y-coordinate $y = -fh/Z$. As Z increases, y grows closer to 0.
- **Nearer objects look bigger.**
- **The natural measure of image size is visual angle.** This is a measure of the size of objects in terms of their projection on the image plane, *as the angle an object subtends at the eye.*

Perspective projection is a mapping from 3D points to rays through the center of projection. It doesn't matter whether the imaging surface is a sphere or a plane. Image formation is merely about rays going through the eye of the observer and intersecting a sheet.

That being the case, projection of a line in planar perspective (onto a plane) is a line, while projection of a line in spherical perspective (onto a sphere) is a great circle!

Two Main Effects of Perspective Projection

- **Distance:** farther objects project to smaller sizes on the image plane. The scaling factor is $1/Z$.
- **Foreshortening:** objects that are slanted with respect to the line of sight project to smaller sizes on the image plane. The scaling factor is $\cos\sigma$, where σ is the angle between the line of sight and the surface normal of the object. Think of disks being viewed from very near the ground at a significant distance away, versus disks being viewed straight-on. Obviously, the former case will involve disks of smaller area (even if the shapes for the two are scaled to the same width or height). Foreshortening is also what accounts for the seasons: in the summer, the rays of the sun are more perpendicular to the Earth's surface than they are in the winter.

Orthographic Projection

Orthographic projection is an approximation to perspective projection, which applies when the object is relatively far away compared to its internal depth variation (i.e. change in Z is negligible). If the depth Z of points on the object varies within a very small range, then the perspective scaling factor f/Z can be approximated by a constant $s = f/Z_0$. The equations for projection from the scene coordinates (X, Y, Z) to the image plane will then become $x = sX$ and $y = sY$.

This is just a linear transformation, so it makes life a lot easier. It is used a lot in computer vision. Note that it does not give vanishing points.

On Transformations

- **Pose:** The position and orientation of the object with respect to the camera, specified by six numbers (three for translation, three for rotation). Pose exists in the camera's frame.
- **Shape:** The coordinates of the points of an object relative to a coordinate frame on the object. These remain invariant when the object is rotated or translated. Shape exists in the object's frame.

When I look at coordinates (X, Y, Z) of some point on an object, those coordinates are affected by both the pose (where the object is, relative to the camera) and the shape.

Rigid objects: if we have a point A and a point B on the object, the distance between the points A and B does not change even if we move the object around. In other words, if ψ is a transformation on the object, then $\|A - B\| = \|\psi(a) - \psi(b)\|$. As a whole, a human is not a rigid object. A chair is. Many objects in the world can be modeled as rigid.

Euclidean transformations (also known as **isometries**) are transformations that preserve distances between pairs of points. In physics, this is known as rigid body motion.

$$\|\psi(\mathbf{a}) - \psi(\mathbf{b})\| = \|\mathbf{a} - \mathbf{b}\|$$

For example, translations $\psi(\mathbf{a}) = \mathbf{a} + \mathbf{t}$ are isometries:

$$\|\psi(\mathbf{a}) - \psi(\mathbf{b})\| = \|\mathbf{a} + \mathbf{t} - (\mathbf{b} + \mathbf{t})\| = \|\mathbf{a} - \mathbf{b}\|$$

Rotation and reflection are also isometries (more specifically orthogonal transformations). **Orthogonal transformations**, a major class of isometries, are linear transformations which preserve inner products:

$$\mathbf{a} \cdot \mathbf{b} = \psi(\mathbf{a}) \cdot \psi(\mathbf{b})$$

Note that **linear transformations** are defined as $\psi(\mathbf{a}) = \mathbf{A}\mathbf{a}$ for some matrix $\mathbf{A}$.

Every isometry is of the following form (i.e. an orthogonal transformation followed by a translation):

$$\psi(\mathbf{a}) = \mathbf{A}\mathbf{a} + \mathbf{t}$$

Since $\mathbf{A}\mathbf{a}$ represents an orthogonal transformation, $\mathbf{A}$ must be an orthogonal matrix.

Orthogonal Transformations

- Orthogonal transformations preserve norms.

  $$\mathbf{a} \cdot \mathbf{a} = \psi(\mathbf{a}) \cdot \psi(\mathbf{a}) \implies \|\mathbf{a}\| = \|\psi(\mathbf{a})\|$$

  ($\mathbf{a} \cdot \mathbf{a}$ gives us the norm squared, or the distance from the origin squared.)
- Orthogonal transformations are isometries.

Orthogonal Matrices

Let ψ be an orthogonal transformation whose action we can represent by matrix multiplication, i.e. $\psi(\mathbf{a}) = \mathbf{Aa}$. Because it preserves inner products,

$$\psi(\mathbf{a}) \cdot \psi(\mathbf{b}) = \mathbf{a}^T\mathbf{b}$$

By substitution,

$$\begin{aligned}\psi(\mathbf{a}) \cdot \psi(\mathbf{b}) &= (\mathbf{Aa})^T(\mathbf{Ab}) \\ &= \mathbf{a}^T\mathbf{A}^T\mathbf{Ab}\end{aligned}$$

Therefore,

$$\mathbf{a}^T\mathbf{b} = \mathbf{a}^T\mathbf{A}^T\mathbf{Ab} \implies \mathbf{A}^T\mathbf{A} = \mathbf{I} \implies \mathbf{A}^T = \mathbf{A}^{-1}$$

Meanwhile, $\det(\mathbf{A})^2 = 1$, implying that $\det(\mathbf{A}) = +1$ or -1, and at the same time each column of $\mathbf{A}$ has norm 1 and is orthogonal to the other columns.