ECE-161C
Cameras

Nuno Vasconcelos

ECE Department, UCSD
Image formation

- all image understanding starts with understanding of image formation:
  - projection of a scene from 3D world into image on 2D plane
Image formation

first of all, why do we care about this?

1) allows us to create (“render”) imaginary scenes
   • special effects, games, architecture/visualization, etc.
   • build a CAD model of the scene and then render from different views
Image formation

▶ e.g. a “fly-through” camera that allows you to see a sports event from new angles

▶ we know
  • where the camera is
  • knowing the projection equations allows us to recreate the image from a 3D model of the scene
  • computer graphics is mostly about this
Image formation

2) even better, we could reconstruct the 3D model from images

rendering:
- 3D world to 2D image

scene reconstruction:
- 2D images to 3D model
- this turns out to be much harder because a 2D projection is consistent with many 3D scenes
- one image is usually not enough, but can be done from a collection of images
Image formation

- when multiple images are available, it is possible to
  - register them
  - deduce the mapping from 2D to 3D
  - extract a 3D model of the scene
  - render from different angles
Image formation

- even when we do not care about the 3D scene per se
  - knowing the geometry is important for many tasks
  - note the appearance changes as the cars move
  - this is due to perspective
  - has to be accounted for even though the goal is tracking not reconstruction
Image formation

- **even for rigid scenes** that do not change that much
  - a change of perspective will create massive pixel changes
  - to compensate for this, one has to understand the projection equations

- **this turns out to be quite complicated**
  - as usual in science, we simplify as much as we can
  - for example, we adopt the pinhole camera model
Pinhole camera

we assume that a camera is

• a black box
• with an infinitesimally small hole on one face
• the hole is so small that only one ray of light passes through it and hits the other side
Pinhole camera

- by placing photo-sensitive material in the back wall you will get an upside-down replica of the scene
  - this is the image plane
  - to avoid the mathematical inconvenience of this inversion
  - we consider a plane outside of the camera
  - this is called the virtual image plane
Pinhole camera

- the virtual image plane

- it is an abstraction
- exactly the same as the image plane, with the exception that there is no inversion
Pinhole camera

One important property:

- Objects that are far away become smaller in the image plane.

We suspect that distance to the camera plays an important role in perspective projection.

In particular, we would expect image size proportional to $1/d$. 
Coordinates

- To relate world point $P$ to image point $P'$
  - We need a coordinate system
  - The $1/d$ dependence suggests using pinhole as origin
  - We also make two coordinate axes $(i,j)$ a basis of the image plane and the third $(k)$ orthogonal to it (measures depth)
Pinhole camera

 definitions:

• line perpendicular to image plane, through pinhole, is the optical axis
• point where optical axis intersects image plane is the image center
• distance f between image plane and pinhole is the focal length
Projection equations

- note that
  - P, O, P' are on the same line
  - this implies that there is a \( \lambda \) such that \( OP' = \lambda OP \), and

\[
\begin{align*}
  x' &= \lambda x \\
  y' &= \lambda y \Rightarrow \lambda = \frac{f}{z} \Rightarrow \left( \begin{array}{c}
  x' \\
  y' \\
  f = \lambda z
\end{array} \right) = f \left( \begin{array}{c}
  x/z \\
  y/z
\end{array} \right)
\end{align*}
\]
Perspective projection

- this is the basic equation of perspective projection

\[
\begin{bmatrix}
    x' \\
    y' \\
    z'
\end{bmatrix} = f \begin{bmatrix}
    x/Z \\
    y/Z \\
    z/Z
\end{bmatrix}
\]

- note that
  - there is indeed an inverse dependence on the depth \( Z \)
  - far objects become small
Perspective projection

- this is a very powerful cue for scene understanding
  - and fun too!
  - note that the visual system infers all sorts of properties from perspective cues
  - e.g. size
Perspective projection

- or shape
- or proximity
Perspective projection

is conceptually very simple

\[
\begin{pmatrix}
x' \\
y'
\end{pmatrix}
= f
\begin{pmatrix}
x/z \\
y/z \\
z
\end{pmatrix}
\]

- but is highly non-linear and usually hard to work with
- e.g. assume you have a big plane on the scene, e.g. a wall

\[z = ax + by\]

- then

\[
\begin{pmatrix}
x' \\
y'
\end{pmatrix}
= f
\begin{pmatrix}
x/(ax + by) \\
y/(ax + by)
\end{pmatrix}
\]

- the image coordinates depend highly non-linearly on the world coordinates
Prospective projection

this is the reason why we see this
Prospective projection

instead of this
Projective projection

- since the size of the wall is constant
- far away (large z) distances appear to shrunk in the image
- in many cases, this non-linearity is too much to handle
- we look for approximations
Affine projection

Consider a plane parallel to the image plane

- This plane has equation $z = C$ and the projection equation becomes

\[
\begin{pmatrix}
    x' \\
    y'
\end{pmatrix} = \frac{f}{C} \begin{pmatrix}
    x \\
    y
\end{pmatrix} = m \begin{pmatrix}
    x \\
    y
\end{pmatrix}, \quad m = \frac{f}{C}
\]

- Image coordinates are simply a re-scaling of the 3D coordinates
Affine projection

scaling:

• if $m < 1$ image points are closer than 3D points,
• else they are further away
• this can be seen by noting that, for $P = (x_p, y_p), Q = (x_q, y_q)$

$$d(P', Q') = \sqrt{(x'_p - x'_q)^2 + (y'_p - y'_q)^2}$$

$$= \sqrt{m^2 (x_p - x_q)^2 + m^2 (y_p - y_q)^2}$$

$$= |m| \ d(P, Q)$$

• this is also captured by the relation through a scaling matrix

$$\begin{pmatrix} x' \\ y' \end{pmatrix} = \begin{bmatrix} m & 0 \\ 0 & m \end{bmatrix} \begin{pmatrix} x \\ y \end{pmatrix}$$
Affine projection

**when** can we use this approximation?

- we are assuming $z$ constant
- this is acceptable if the variation of depth in the scene is much smaller than the average depth
- e.g. an airplane taking aerial photos
Orthographic projection

if the camera is always at (approximately) the same distance from the scene:

• $m$ only contributes a change of scale that we do not care much about (e.g. measure in centimeters vs meters)
• it is common to normalize to $m = 1$
• this is orthographic projection

\[
\begin{pmatrix}
  x' \\
  y'
\end{pmatrix} =
\begin{pmatrix}
  x \\
  y
\end{pmatrix}
\]
Orthographic projection

this is, of course, very convenient:

• “image coordinates = world coordinates”
• there are not that many scenarios in which it is a good approximation
• nevertheless can be a good model for a preliminary solution
• which is then refined with a more complicated model

\[
\begin{pmatrix}
\begin{bmatrix}
    x' \\
    y' 
\end{bmatrix}
\end{pmatrix} = \begin{pmatrix}
\begin{bmatrix}
    x \\
    y 
\end{bmatrix}
\end{pmatrix}
\]
Lenses

- so far we have assumed the pinhole camera
- in practice we cannot really build such a camera and obtain decent quality
- problems:
  - when pinhole is too big
  - many directions are averaged, blurring the image
Lenses

- Pinhole problems:
  - If the pinhole is too small
  - We have diffraction effects which also blur the image

- There is a correct pinhole size from an image distortion point of view, but that introduces other problems
Lenses

- pinhole problems:
  - for the "correct" pinhole size
  - we cannot get enough light in the camera to sufficiently excite the recording material
  - generally, pinhole cameras are dark,
  - a very small set of rays from a particular point hits the screen

- this is the reason why we need camera lenses
Lenses

the basic idea is:

• let's make the aperture bigger so that we can have many rays of light into the camera

• to avoid blurring we need to concentrate all the rays that start in the same 3D point

• so that they end up on the same image plane point
Lenses

- the geometry is as follows

- **image plane**
- light ray
- lens

\[ R: \text{radius of curvature of the lens} \]
\[ d_1: \text{distance from 3D point to lens} \]
\[ d_2: \text{distance to image plane} \]
Lenses

we assume all angles are small (d₂, h are in microns):

- \(\alpha = \sin \alpha = \tan \alpha\)

\[\alpha_1 \approx \frac{h}{R} + \frac{h}{d_1}\]
Lenses

we assume all angles are small (\(d_2, h\) are in microns):

- \(\alpha = \sin \alpha = \tan \alpha\)

\[
\alpha_2 \approx \frac{h}{R} - \frac{h}{d_2}
\]
Lenses

Snell’s law

- for light propagating between two media of indexes of refraction $n_1$ and $n_2$

$$n_1 \sin \alpha_1 = n_2 \sin \alpha_2$$

\[
\alpha_1 \approx \frac{h}{R} + \frac{h}{d_1} \\
\alpha_2 \approx \frac{h}{R} - \frac{h}{d_2} \\
\Rightarrow n_1 h \left(\frac{1}{R} + \frac{1}{d_1}\right) \approx n_2 h \left(\frac{1}{R} - \frac{1}{d_2}\right)
\]
Lenses

which means that

\[
\frac{1}{d_1} \approx \frac{1}{n_1} \left( \frac{n_2 - n_1}{R} + \frac{n_2}{d_2} \right)
\]

- given distance \(d_2\), we can compute distance \(d_1\) of the 3D point
Lenses

which means that

\[
\frac{1}{d_1} \approx \frac{1}{n_1} \left( \frac{n_2 - n_1}{R} - \frac{n_2}{d_2} \right)
\]

- note that it does not depend on the vertical position of P
- we can show that it holds for all rays that start in the plane of P
Lenses

- note that, in general,
  - we can only have in focus objects that are in a certain depth range
  - this is why the background is sometimes out of focus on photographs
  - by controlling the focus you are effectively changing the plane of the rays that converge on the image plane without blur
Any questions?