Lecture 5: Cameras and Projection
Reading

• Szeliski 2.1.3-2.1.6
Announcements

• Project 1 assigned, see projects page:

• Quiz 1 on Wednesday
Let’s design a camera
- Idea 1: put a piece of film in front of an object
- Do we get a reasonable image?
Pinhole camera

- Add a barrier to block off most of the rays
  - This reduces blurring
  - The opening known as the **aperture**
  - How does this transform the image?
Camera Obscura

- Basic principle known to Mozi (470-390 BC), Aristotle (384-322 BC)
- Drawing aid for artists: described by Leonardo da Vinci (1452-1519)

Gemma Frisius, 1558

Source: A. Efros
Camera Obscura
Home-made pinhole camera

Why so blurry?

http://www.debevec.org/Pinhole/
Pinhole photography


6-month exposure
Shrinking the aperture

- Why not make the aperture as small as possible?
  - Less light gets through
  - *Diffraction* effects...
Shrinking the aperture
• A lens focuses light onto the film
  – There is a specific distance at which objects are “in focus”
    • other points project to a “circle of confusion” in the image
  – Changing the shape of the lens changes this distance
Lenses

- A lens focuses parallel rays onto a single focal point
  - focal point at a distance \( f \) beyond the plane of the lens (the *focal length*)
    - \( f \) is a function of the shape and index of refraction of the lens
  - Aperture restricts the range of rays
    - aperture may be on either side of the lens
    - Lenses are typically spherical (easier to produce)
Thin lenses

- Thin lens equation: \[
\frac{1}{d_o} + \frac{1}{d_i} = \frac{1}{f}
\]
  - Any object point satisfying this equation is in focus
  - What is the shape of the focus region?
  - How can we change the focus region?
  - Thin lens applet: [http://www.phy.ntnu.edu.tw/java/Lens/lens_e.html](http://www.phy.ntnu.edu.tw/java/Lens/lens_e.html) (by Fu-Kwun Hwang)
Depth of Field

- Changing the aperture size affects depth of field
  - A smaller aperture increases the range in which the object is approximately in focus

Depth of Field

Large aperture opening

Small aperture opening
The human eye is a camera

- **Iris** - colored annulus with radial muscles
- **Pupil** - the hole (aperture) whose size is controlled by the iris
- What’s the “film”?
  - photoreceptor cells (rods and cones) in the **retina**
Before Film

Lens Based Camera Obscura, 1568
Film camera

*Still Life*, Louis Jaques Mande Daguerre, 1837
Silicon Image Detector

Silicon Image Detector, 1970
Digital camera

- A digital camera replaces film with a sensor array
  - Each cell in the array is a Charge Coupled Device
    - light-sensitive diode that converts photons to electrons
    - other variants exist: CMOS is becoming more popular
Color

• So far, we’ve talked about grayscale images
• What about color?
• Most digital images are comprised of three color channels – red, green, and, blue – which combine to create most of the colors we can see

• Why are there three?
Color perception

- Three types of cones
  - Each is sensitive in a different region of the spectrum
    - but regions overlap
    - Short (S) corresponds to blue
    - Medium (M) corresponds to green
    - Long (L) corresponds to red
  - Different sensitivities: we are more sensitive to green than red
    - varies from person to person (and with age)
  - Colorblindness—deficiency in at least one type of cone
Field sequential
Field sequential
Field sequential
Prokudin-Gorskii (early 1900’s)

http://www.loc.gov/exhibits/empire/
Prokudin-Gorskii (early 1990’s)
Color sensing in camera: Prism

- Requires three chips and precise alignment
- More expensive
Color filter array

Bayer grid

Estimate missing components from neighboring values (demosaicing)

Why more green?

Human Luminance Sensitivity Function

Source: Steve Seitz
Bayer’s pattern
Color filter array

Mosaic Capture

In conventional systems, color filters are applied to a single layer of photodetectors in a tiled mosaic pattern.

The filters let only one wavelength of light—red, green or blue—pass through to any given pixel, allowing it to record only one color.

As a result, mosaic sensors capture only 25% of the red and blue light, and just 50% of the green.

red  green  blue  output

YungYu Chuang's slide
Color images

• We’ll treat color images as a vector-valued function:

\[
f(x, y) = \begin{bmatrix} r(x, y) \\ g(x, y) \\ b(x, y) \end{bmatrix}
\]

• We’ll often convert to grayscale (e.g., \(0.3 \times r + 0.59 \times g + 0.11 \times b\))
Dynamic range

• What is the range of light intensity that a camera can capture?
  – Called *dynamic range*
  – Digital cameras have difficulty capturing both high intensities and low intensities in the same image
Light response is nonlinear

• Our visual system has a large *dynamic range*
  – We can resolve both light and dark things at the same time
  – One mechanism for achieving this is that we sense light intensity on a *logarithmic scale*
    • an exponential intensity ramp will be seen as a linear ramp
  – Another mechanism is *adaptation*
    • rods and cones adapt to be more sensitive in low light, less sensitive in bright light.
### Visual dynamic range

<table>
<thead>
<tr>
<th>Background</th>
<th>Luminance (candela per square meter)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Horizon sky</strong></td>
<td></td>
</tr>
<tr>
<td>Moonless overcast night</td>
<td>0.00003</td>
</tr>
<tr>
<td>Moonless clear night</td>
<td>0.0003</td>
</tr>
<tr>
<td>Moonlit overcast night</td>
<td>0.003</td>
</tr>
<tr>
<td>Moonlit clear night</td>
<td>0.03</td>
</tr>
<tr>
<td>Deep twilight</td>
<td>0.3</td>
</tr>
<tr>
<td>Twilight</td>
<td>3</td>
</tr>
<tr>
<td>Very dark day</td>
<td>30</td>
</tr>
<tr>
<td>Overcast day</td>
<td>300</td>
</tr>
<tr>
<td>Clear day</td>
<td>3,000</td>
</tr>
<tr>
<td>Day with sunlit clouds</td>
<td>30,000</td>
</tr>
<tr>
<td><strong>Daylight fog</strong></td>
<td></td>
</tr>
<tr>
<td>Dull</td>
<td>300–1,000</td>
</tr>
<tr>
<td>Typical</td>
<td>1,000–3,000</td>
</tr>
<tr>
<td>Bright</td>
<td>3,000–16,000</td>
</tr>
<tr>
<td><strong>Ground</strong></td>
<td></td>
</tr>
<tr>
<td>Overcast day</td>
<td>30–100</td>
</tr>
<tr>
<td>Sunny day</td>
<td>300</td>
</tr>
<tr>
<td>Snow in full sunlight</td>
<td>16,000</td>
</tr>
</tbody>
</table>

**Figure 1.13**

Dynamic range

• Our total dynamic range is high (~$10^9$)
• Our dynamic range at a given time is still pretty high (~$10^4$)
• A camera’s dynamic range for a given exposure is relatively low ($2^8 = 256$ tonal values, range of about ~$10^3$)
High dynamic range imaging
Historical context

- **Pinhole model**: Mozi (470-390 BC), Aristotle (384-322 BC)
- **Principles of optics (including lenses)**: Alhacen (965-1039)
- **Camera obscura**: Leonardo da Vinci (1452-1519), Johann Zahn (1631-1707)
- **First photo**: Joseph Nicephore Niepce (1822)
- **Daguerréotypes** (1839)
- **Photographic film** (Eastman, 1889)
- **Cinema** (Lumière Brothers, 1895)
- **Color Photography** (Lumière Brothers, 1908)
- **Television** (Baird, Farnsworth, Zworykin, 1920s)
- **First consumer camera with CCD**: Sony Mavica (1981)
- **First fully digital camera**: Kodak DCS100 (1990)
Questions?

• 3-minute break
Projection

MAKE POVERTY HISTORY

CoolOpticalIllusions.com
Projection
Müller-Lyer Illusion

http://www.michaelbach.de/ot/sze_muelue/index.html
Modeling projection

• The coordinate system
  – We will use the pinhole model as an approximation
  – Put the optical center (Center Of Projection) at the origin
  – Put the image plane (Projection Plane) in front of the COP
    • Why?
  – The camera looks down the negative z axis
    • we need this if we want right-handed-coordinates
Modeling projection

- Projection equations
  - Compute intersection with PP of ray from \((x, y, z)\) to COP
  - Derived using similar triangles (on board)
    \[
    (x, y, z) \rightarrow \left(-\frac{d x}{z}, -\frac{d y}{z}, -d\right)
    \]
  - We get the projection by throwing out the last coordinate:
    \[
    (x, y, z) \rightarrow \left(-\frac{d x}{z}, -\frac{d y}{z}\right)
    \]
Modeling projection

- Is this a linear transformation?
  - no—division by $z$ is nonlinear

Homogeneous coordinates to the rescue!

Converting from homogeneous coordinates

$\begin{bmatrix} x \\ y \\ w \end{bmatrix} \Rightarrow (x/w, y/w)$

Converting from homogeneous coordinates

$\begin{bmatrix} x \\ y \\ z \\ w \end{bmatrix} \Rightarrow (x/w, y/w, z/w)$
Perspective Projection

Projection is a matrix multiply using homogeneous coordinates:

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1/d & 0 \\
0 & 0 & -z/d & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
= 
\begin{bmatrix}
x \\
y \\
-z/d \\
z
\end{bmatrix}
\Rightarrow (-d \frac{x}{z}, -d \frac{y}{z})
\]

divide by third coordinate

This is known as **perspective projection**

- The matrix is the **projection matrix**

- (Can also represent as a 4x4 matrix – OpenGL does something like this)
Perspective Projection

• How does scaling the projection matrix change the transformation?

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & -1/d & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
= \begin{bmatrix}
x \\
y \\
-z/d \\
1
\end{bmatrix}
\Rightarrow (-d\frac{x}{z}, -d\frac{y}{z})
\]

\[
\begin{bmatrix}
-d & 0 & 0 & 0 \\
0 & -d & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
= \begin{bmatrix}
-dx \\
-dy \\
-z \\
1
\end{bmatrix}
\Rightarrow (-d\frac{x}{z}, -d\frac{y}{z})
\]
Orthographic projection

• Special case of perspective projection
  – Distance from the COP to the PP is infinite

\[
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
x \\
y \\
z \\
1
\end{bmatrix}
= \begin{bmatrix}
x \\
y \\
1
\end{bmatrix} \Rightarrow (x, y)\]
Orthographic projection
Perspective projection
Perspective distortion

• What does a sphere project to?
Perspective distortion

- The exterior columns appear bigger
- The distortion is not due to lens flaws
- Problem pointed out by Da Vinci
Perspective distortion: People
Distortion

- Radial distortion of the image
  - Caused by imperfect lenses
  - Deviations are most noticeable for rays that pass through the edge of the lens
Correcting radial distortion

from Helmut Dersch
Modeling distortion

To model lens distortion

- Use above projection operation instead of standard projection matrix multiplication

\[ x_n' = \frac{\hat{x}}{\hat{z}} \]
\[ y_n' = \frac{\hat{y}}{\hat{z}} \]

Apply radial distortion

\[ r^2 = x_n'^2 + y_n'^2 \]
\[ x_d' = x_n'(1 + \kappa_1 r^2 + \kappa_2 r^4) \]
\[ y_d' = y_n'(1 + \kappa_1 r^2 + \kappa_2 r^4) \]

Apply focal length

\[ x' = f x_d' + x_c \]
\[ y' = f y_d' + y_c \]
Other types of projection

- Lots of intriguing variants...
- (I’ll just mention a few fun ones)
360 degree field of view...

• Basic approach
  – Take a photo of a parabolic mirror with an orthographic lens (Nayar)
  – Or buy one a lens from a variety of omnicam manufacturers...
    • See http://www.cis.upenn.edu/~kostas/omni.html
Tilt-shift

http://www.northlight-images.co.uk/article_pages/tilt_and_shift_ts-e.html

Tilt-shift images from Olivo Barbieri and Photoshop imitations
Dimensionality Reduction Machine
(3D to 2D)

What have we lost?

- Angles
- Distances (lengths)

Point of observation

Slide by A. Efros
Figures © Stephen E. Palmer, 2002
Projection properties

• Many-to-one: any points along same ray map to same point in image
• Points → points
• Lines → lines (collinearity is preserved)
  – But line through focal point projects to a point
• Planes → planes (or half-planes)
  – But plane through focal point projects to line
Projection properties

- Parallel lines converge at a vanishing point
  - Each direction in space has its own vanishing point
  - But parallels parallel to the image plane remain parallel
Camera parameters

• How can we model the geometry of a camera?

Two important coordinate systems:
1. *World* coordinate system
2. *Camera* coordinate system
Camera parameters

- To project a point \((x,y,z)\) in *world* coordinates into a camera
- First transform \((x,y,z)\) into *camera* coordinates
- Need to know
  - Camera position (in *world* coordinates)
  - Camera orientation (in *world* coordinates)
- The project into the image plane
  - Need to know camera *intrinsics*
Camera parameters

A camera is described by several parameters

- Translation $T$ of the optical center from the origin of world coords
- Rotation $R$ of the image plane
- focal length $f$, principle point $(x'_c, y'_c)$, pixel size $(s_x, s_y)$
- blue parameters are called “extrinsics,” red are “intrinsics”

Projection equation

$$
X = \begin{bmatrix}
sx \\
sy \\
s
\end{bmatrix}
= \begin{bmatrix}
* & * & * \\
* & * & * \\
* & * & * \\
\end{bmatrix}
\begin{bmatrix}
X \\
Y \\
Z \\
1
\end{bmatrix}
= \Pi X
$$

- The projection matrix models the cumulative effect of all parameters
- Useful to decompose into a series of operations

$$
\Pi = \begin{bmatrix}
-fs_x & 0 & x'_c \\
0 & -fs_y & y'_c \\
0 & 0 & 1
\end{bmatrix}
\begin{bmatrix}
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0
\end{bmatrix}
\begin{bmatrix}
R_{3\times3} & 0_{3\times1} \\
0_{1\times3} & 1
\end{bmatrix}
\begin{bmatrix}
I_{3\times3} \\
T_{3\times1}
\end{bmatrix}
$$

- The definitions of these parameters are not completely standardized
  - especially intrinsics—varies from one book to another
Extrinsics

• How do we get the camera to “canonical form”?
  – (Center of projection at the origin, x-axis points right, y-axis points up, z-axis points backwards)

Step 1: Translate by -$\mathbf{c}$
Extrinsics

• How do we get the camera to “canonical form”?
  – (Center of projection at the origin, x-axis points right, y-axis points up, z-axis points backwards)

Step 1: Translate by $-\mathbf{c}$

How do we represent translation as a matrix multiplication?

$$
T = \begin{bmatrix}
I_{3 \times 3} & -\mathbf{c} \\
0 & 0 & 0 & 1
\end{bmatrix}
$$
Extrinsics

- How do we get the camera to “canonical form”?
  - (Center of projection at the origin, x-axis points right, y-axis points up, z-axis points backwards)

Step 1: Translate by $-c$
Step 2: Rotate by $R$

\[
R = \begin{bmatrix}
u^T \\ v^T \\ w^T
\end{bmatrix}
\]

3x3 rotation matrix
Extrinsics

• How do we get the camera to “canonical form”? – (Center of projection at the origin, x-axis points right, y-axis points up, z-axis points backwards)

Step 1: Translate by $-c$
Step 2: Rotate by $R$

\[
R = \begin{bmatrix}
  u^T \\
  v^T \\
  w^T
\end{bmatrix}
\]
Perspective projection

\[
\begin{bmatrix} -f & 0 & 0 \\ 0 & -f & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}
\]

**K** (intrinsics) (converts from 3D rays in camera coordinate system to pixel coordinates)

in general, \( \mathbf{K} = \begin{bmatrix} -f & s & c_x \\ 0 & -\alpha f & c_y \\ 0 & 0 & 1 \end{bmatrix} \) (upper triangular matrix)

\( \alpha : \text{aspect ratio} \) (1 unless pixels are not square)

\( s : \text{skew} \) (0 unless pixels are shaped like rhombi/parallelograms)

\((c_x, c_y) : \text{principal point} \) ((0,0) unless optical axis doesn’t intersect projection plane at origin)
Focal length

• Can think of as “zoom”

• Also related to field of view
Projection matrix

\[ \Pi = K \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix} \begin{bmatrix} R & 0 \\ 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} I_{3 \times 3} & -c \\ 0 & 0 & 0 & 1 \end{bmatrix} \]

(t in book’s notation)
Projection matrix

\[ \mathbf{q} = (x, y, z, 1) \]

(in homogeneous image coordinates)
Questions?