Announcement

• Please send me *three* topics you want me to review next class for the midterm
• The more requests I get for each topic, the more time I will devote to it
• If no one asks me to review a topic, I won’t review it
Exam Format

• Multiple-choice and true/false
• Short answers
• Demonstrating an algorithm on a particular example
Last class vs this class

- Last class: Same camera center, but camera rotates
- This class: Camera center is not the same (we have multiple cameras)

- Epipolar geometry
  - Relates cameras from two positions/cameras

- Stereo depth estimation
  - Recover depth from disparities between two images

Adapted from Derek Hoiem
Why multiple views?

- Structure and depth are inherently ambiguous from single views.

- Multiple views help us to perceive 3d shape and depth.
Stereo photography and stereo viewers

Take two pictures of the same subject from two slightly different viewpoints and display so that each eye sees only one of the images.

Invented by Sir Charles Wheatstone, 1838

Image from fisher-price.com
Stereo photography and stereo viewers
Depth from stereo

Two cameras, simultaneous views

Single moving camera and static scene

Kristen Grauman
Depth from stereo

• Goal: recover depth by finding image coordinate $x'$ that corresponds to $x$
Depth from stereo

• Goal: recover depth by finding image coordinate $x'$ that corresponds to $x$

• Sub-Problems
  1. Calibration: How do we recover the relation of the cameras (if not already known)?
  2. Correspondence: How do we search for the matching point $x'$?
Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). **What is expression for Z?**

Similar triangles \((p_l, P, p_r)\) and \((O_l, P, O_r)\):

\[
\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}
\]

Depth is inversely proportional to disparity.

Adapted from Kristen Grauman
Depth from disparity

We have two images taken from cameras with different intrinsic and extrinsic parameters.
• How do we match a point in the first image to a point in the second?

So if we could find the corresponding points in two images, we could estimate relative depth…
Summary (Introduction of terms)

• **Epipolar geometry**
  
  – Epipoles are intersection of baseline with image planes
  – Matching point in second image is on a line passing through its epipole
  – Epipolar constraint limits where points from one view will be imaged in the other, which makes search for correspondences quicker
  
  – Essential (E) and fundamental (F) matrices map from a point in one image to a line (its epipolar line) in the other
  – Can solve for E, F given corresponding points (e.g., interest points)

• **Stereo depth estimation**
  
  – Find corresponding points along epipolar scanline
  – Estimate disparity (depth is inverse to disparity)

Adapted from Kristen Grauman and Derek Hoiem
Stereo correspondence constraints

- Given $p$ in left image, where can corresponding point $p'$ be?
Stereo correspondence constraints
Epipolar constraint

Geometry of two views constrains where the corresponding pixel for some image point in the first view must occur in the second view.

- It must be on the line carved out by a plane connecting the world point and optical centers.
- Potential matches for $p'$ have to lie on the corresponding line $l$.
- Potential matches for $p$ have to lie on the corresponding line $l'$. 
Epipolar geometry: notation

- **Baseline** – line connecting the two camera centers
- **Epipoles**
  - intersections of baseline with image planes
  - projections of the other camera center
- **Epipolar Plane** – plane containing baseline
- **Epipolar Lines** - intersections of epipolar plane with image planes (always come in corresponding pairs)
  - Note: All epipolar lines intersect at the epipole.
Epipolar constraint

The epipolar constraint is useful because it reduces the correspondence problem to a 1D search along an epipolar line.
Summary (Introduction of terms)

• **Epipolar geometry**
  - Epipoles are intersection of baseline with image planes
  - Matching point in second image is on a line passing through its epipole
  - Epipolar constraint limits where points from one view will be imaged in the other, which makes search for correspondences quicker
  - Essential (E) and fundamental (F) matrices map from a point in one image to a line (its epipolar line) in the other
  - Can solve for E, F given corresponding points (e.g., interest points)

• **Stereo depth estimation**
  - Find corresponding points along epipolar scanline
  - Estimate disparity (depth is inverse to disparity)

Adapted from Kristen Grauman and Derek Hoiem
If the stereo rig is calibrated, we know:

- how to rotate and translate camera reference frame 1 to get to camera reference frame 2.

  Rotation: 3x3 matrix $R$; translation: 3x1 vector $T$. 
If the stereo rig is calibrated, we know:
how to **rotate** and **translate** camera reference frame 1 to get to camera reference frame 2.

\[ X'_c = RX_c + T \]
An aside: cross product

\[ \vec{a} \times \vec{b} = \vec{c} \]
\[ \vec{a} \cdot \vec{c} = 0 \]
\[ \vec{b} \cdot \vec{c} = 0 \]

Vector cross product takes two vectors and returns a third vector that’s perpendicular to both inputs.

So here, \( \vec{c} \) is perpendicular to both \( \vec{a} \) and \( \vec{b} \), which means the dot product equals 0.
From geometry to algebra

\[ X' = RX + T \]

\[ T \times X' = T \times RX \]

normal to the plane

\[ X' \cdot (T \times X') = X' \cdot (T \times RX) = 0 \]

Kristen Grauman
Aside: Matrix form of cross product

\[
\vec{a} \times \vec{b} = \begin{bmatrix}
0 & -a_3 & a_2 \\
a_3 & 0 & -a_1 \\
-a_2 & a_1 & 0
\end{bmatrix}
\begin{bmatrix}
b_1 \\
b_2 \\
b_3
\end{bmatrix}
= \vec{c}
\]

\[
\vec{a} \cdot \vec{c} = 0 \\
\vec{b} \cdot \vec{c} = 0
\]

Can be expressed as a matrix multiplication.

\[
[a_x] = \begin{bmatrix}
0 & -a_3 & a_2 \\
a_3 & 0 & -a_1 \\
-a_2 & a_1 & 0
\end{bmatrix}
\]

\[
\vec{a} \times \vec{b} = [a_x] \vec{b}
\]
Essential matrix

\[ X' \cdot (T \times RX) = 0 \]
\[ X' \cdot ([T_x]RX) = 0 \]

Let \( E = [T_x]R \)

\[ X' \cdot EX = X''^T EX = 0 \]

\( E \) is called the **essential matrix**, and it relates corresponding image points between both cameras, given the rotation and translation.

Before we said: If we observe a point in one image, its position in other image is constrained to lie on line defined by above.

• Turns out \( Ex' \) is the epipolar line through \( x \) in the first image, corresp. to \( x' \).

Note: these points are in **camera coordinate systems**.
Essential matrix example: parallel cameras

For the parallel cameras, image of any point must lie on same horizontal line in each image plane.
We can also find correspondences and depth if cameras’ optical axes are not parallel (not discussed here).

Adapted from Kristen Grauman
What if we don’t know camera parameters $R, T$?

- Want to estimate world geometry without requiring calibrated cameras
  - Archival videos
  - Photos from multiple unrelated users

- **Weak calibration:**
  - Estimate epipolar geometry from a (redundant) set of point correspondences between two uncalibrated cameras
Computing $\mathbf{F}$ from correspondences

Each point correspondence generates one constraint on $\mathbf{F}$

$$
\mathbf{p}_{im, right}^T \mathbf{F} \mathbf{p}_{im, left} = 0
$$

$$
\begin{bmatrix}
    u' & v' & 1
\end{bmatrix}
\begin{bmatrix}
    f_{11} & f_{12} & f_{13} \\
    f_{21} & f_{22} & f_{23} \\
    f_{31} & f_{32} & f_{33}
\end{bmatrix}
\begin{bmatrix}
    u \\
    v \\
    1
\end{bmatrix} = 0
$$

Collect $n$ of these constraints

$$
\begin{bmatrix}
    u'_1u_1 & u'_1v_1 & u'_1 & v'_1u_1 & v'_1v_1 & v'_1 & u_1 & v_1 & 1
\end{bmatrix}
\begin{bmatrix}
    f_{11} \\
    f_{12} \\
    f_{13} \\
    f_{21} \\
    f_{22} \\
    f_{23} \\
    f_{31} \\
    f_{32} \\
    f_{33}
\end{bmatrix} = 0
$$

Solve for $f$, vector of parameters.
Fundamental matrix

- Relates *pixel coordinates* in the two views
- More general form than *essential matrix*: we remove need to know intrinsic parameters

- [Fundamental matrix song](#)
Summary (Introduction of terms)

• **Epipolar geometry**
  - Epipoles are intersection of baseline with image planes
  - Matching point in second image is on a line passing through its epipole
  - Epipolar constraint limits where points from one view will be imaged in the other, which makes search for correspondences quicker
  - Essential (E) and fundamental (F) matrices map from a point in one image to a line (its epipolar line) in the other
  - Can solve for E, F given corresponding points (e.g., interest points)

• **Stereo depth estimation**
  - Find corresponding points along epipolar scanline
  - Estimate disparity (depth is inverse to disparity)

Adapted from Kristen Grauman and Derek Hoiem
Using epipolar geometry for stereo

Fuse a calibrated binocular stereo pair to produce a depth image

Dense depth map
Basic stereo matching algorithm

- For each pixel in the first image
  - Find corresponding epipolar scanline in the right image
  - Search along epipolar line and pick the best match $x'$
  - Compute disparity $x - x'$ and set $\text{depth}(x) = f \times T / (x - x')$
Correspondence search

- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: e.g. Euclidean distance
Geometry for a simple stereo system

- Assume parallel optical axes, known camera parameters (i.e., calibrated cameras). What is expression for $Z$?

Similar triangles $(p_l, P, p_r)$ and $(O_l, P, O_r)$:

$$\frac{T + x_l - x_r}{Z - f} = \frac{T}{Z}$$

$Z = f \left(\frac{T}{x_r - x_l}\right)$

depth

disparity
Results with window search

Data

Left image

Right image

Window-based matching

Ground truth
How can we improve?

- **Uniqueness**
  - For any point in one image, there should be at most one matching point in the other image

- **Ordering**
  - Corresponding points should be in the same order in both views

- **Smoothness**
  - We expect disparity values to change slowly (for the most part)
Many of these constraints can be encoded in an energy function and solved using graph cuts.


For the latest and greatest: [http://vision.middlebury.edu/stereo/](http://vision.middlebury.edu/stereo/)
Projective structure from motion

- Given: $m$ images of $n$ fixed 3D points

  $$x_{ij} = P_i X_j, \quad i = 1, \ldots, m, \quad j = 1, \ldots, n$$

- Problem: estimate $m$ projection matrices $P_i$ and $n$ 3D points $X_j$ from the $mn$ corresponding 2D points $x_{ij}$
Photo synth


http://photosynth.net/
3D from multiple images

Building Rome in a Day: Agarwal et al. 2009
Recap: Epipoles

• Point \( x \) in left image corresponds to **epipolar line** \( l' \) in right image

• Epipolar line passes through the epipole (the intersection of the cameras’ baseline with the image plane)
Recap: Fundamental Matrix

• Fundamental matrix maps from a point in one image to a line in the other
  \[ l' = Fx \quad l = F^\top x' \]

• If \( x \) and \( x' \) correspond to the same 3d point \( X \):
  \[ x'^\top Fx = 0 \]

• Essential matrix is like fundamental matrix but more constrained

Adapted from Derek Hoiem
Recap: stereo with calibrated cameras

• Given image pair, $\mathbf{R}, \mathbf{T}$
• Detect some features
• Compute essential matrix $\mathbf{E}$
• Match features using the epipolar and other constraints
• Triangulate for 3d structure and get depth