Announcements

• HW1 assigned, due Thursday at midnight
Stereo Vision Outline

- Offline:
  B  Calibrate cameras & determine epipolar geometry

- Online
  1. Acquire stereo images
  C  2. Rectify images to convenient epipolar geometry
  D  3. Establish correspondence
  A  4. Estimate depth

Epipolar geometry
Epipolar matching

- Potential matches for $p$ have to lie on the corresponding epipolar line $L'$
- Epipolar line $L'$ passes through epipole $e'$, the intersection of the baseline with the image plane
- Potential matches for $p'$ have to lie on the corresponding epipolar line $L$

Epipolar Geometry Terminology

- **Baseline**: line connecting two centers of projection $O$ and $O'$
- **Epipoles $(e,e')$**: Two intersection points of baseline with image planes
- **Epipolar Plane**: Any plane that contains the baseline
- **Epipolar Lines $(l, l')$**: Pair of lines from intersection of an epipolar plane with the two image planes
Epipolar Constraint: Calibrated Case

The vectors $\overrightarrow{Op} \cdot \overrightarrow{OO'}$ and $\overrightarrow{O'p'}$ are coplanar

$$\overrightarrow{Op} \cdot [\overrightarrow{OO} \times \overrightarrow{O'p}] = 0$$

$1p \cdot [1t_2 \times (2R \cdot p')] = 0$

Essential Matrix
(Longuet-Higgins, 1981)

$1p' E' = 0$ with $E = [(1t_2)_{skw} \cdot 1R$

How do we use the Essential Matrix $E$?
Given a pixel coordinates $q$ in image 1, what is epipolar line equation in image 2?

$1p' E' p' = 0$ with $E = [(1t_2)_{skw} \cdot 1R$

- Given a point in Image 1 with homogenous pixel coordinates $q$, convert it to a direction as $1p = (K_1^{-1}) q$
- Let $a = 1p^T E$ where $a$ is 1 X 3 vector.
- Then apply the epipolar constraint, we have $a \cdot 2p' = 0$
- And converting direction to image coordinates with $2p' = K_2^{-1} q'$, we have that

$$aK_2^{-1}) q' = 0$$

is the equation of the epipolar line in Image 2.
- Likewise, given a point $q'$ in Image 2, we can obtain a line equation in Image 1.
Computing the epipoles given $E$

$$p^T E p' = 0 \text{ with } E = [t_x] R$$

- The epipole $e'$ in the right image is the Eigenvector of $E$ corresponding to the zero eigenvalue.
- The epipole $e$ in the left image is the Eigenvector of $E^T$ corresponding to the zero eigenvalue.

Why?

1. $\det(E) = \det([t_x])\det(R) = 0$ because $\det([t_x]) = 0$ since $[t_x]$ is skew symmetric. Consequently $E$ has a zero Eigenvalue.

---

**Stereo Vision Outline**

- Offline: Calibrate cameras & determine
  - “epipolar geometry”
- Online
  1. Acquire stereo images
  2. **Rectify images to convenient epipolar geometry**
  3. Establish correspondence
  4. Estimate depth
• Epipolar geometry reduces matching complexity from $O(n^4)$ to $O(n^3)$.
• But, matching requires comparing points across pairs of epipolar lines which may have arbitrary orientation. That can be costly to index.
• Is there a more convenient epipolar geometry.

Cameras with a convenient epipolar geometry

• When two cameras have parallel optical axes and these axis are orthogonal to the baseline, the epipolar line are parallel.

• When rows of the two images are parallel to the baseline, the epipolar lines are horizontal rows of the two images.
Cameras with a convenient epipolar geometry

- When two cameras have parallel optical axes and these axes are orthogonal to the baseline, the epipolar line are parallel.

- When rows of the two images are parallel to the baseline, the epipolar lines are horizontal rows of the two images.

What if stereo geometry isn’t convenient?
Rectification: Given a pair of images, transform both images so that epipolar lines are image rows.
Rectification

Under perspective projection, the mapping from a plane to a plane is given by a linear transformation of homogeneous coordinates (called a projective transformation or homography).

\[
\begin{bmatrix}
  x_L \\
  y_L \\
  w_L
\end{bmatrix}
= H_L
\begin{bmatrix}
  u_L \\
  v_L \\
  1
\end{bmatrix}
\]

Two images
Two homographies
\( H_L, H_R \)

Question: Where are the epipoles in the rectified images?
At a point at infinity.
Rectification

Under perspective projection, the mapping from a plane to a plane is given by a linear transformation of homogeneous coordinates called a projective transformation or homography.

\[
\begin{bmatrix}
  x_L \\
  y_L \\
  w_L
\end{bmatrix} = \begin{bmatrix}
  u_L \\
  v_L \\
  1
\end{bmatrix}
\]

Two images
Two homographies $H_L, H_R$

\[
\begin{bmatrix}
  x_R \\
  y_R \\
  w_R
\end{bmatrix} = \begin{bmatrix}
  u_R \\
  v_R \\
  1
\end{bmatrix}
\]

See text, blog posts, OpenCV for methods to compute $H_L$ and $H_R$
Rectification

Given a pair of images, transform both images so that epipolar lines are scan lines.

Input Images

Rectified images

Notes:

• The rectified images may not be rectangular
• A homography maps a square to an arbitrary quadrilateral

\[
\begin{bmatrix}
x_L \\ y_L \\ w_L \\
\end{bmatrix} = H_L \begin{bmatrix}
u_L \\ v_L \\ 1 \\
\end{bmatrix}
\]
Rectification
Given a pair of images, transform both images so that epipolar lines are scan lines.

Rectified Images

Stereo Vision Outline

- Offline: Calibrate cameras & determine “epipolar geometry”
- Online
  1. Acquire stereo images
  2. Rectify images to convenient epipolar geometry
  3. Establish correspondence
  4. Estimate depth
Features on same epipolar line

Two approaches to finding correspondence

1. Feature-Based (Sparse)
   - Process each image “monocularly” to detect image features (e.g. corners or SIFT features)
   - Establish correspondence between features, using appearance info and descriptors

2. Area-Based (Dense)
   - Directly compare image regions between the two images.
Human Stereopsis: Binocular Fusion

How are the correspondences established?

Julesz (1971): Is the mechanism for binocular fusion a monocular process or a binocular one??

- Random dot stereograms provide an objective answer

Random Dot Stereograms
Random Dot Stereograms

Rembrandt

The Nightwatch (1642)

Self-Portrait with Beret and Turned-Up Collar (1659)
Was Rembrandt Stereo Blind?
• Detail of a 1639 etching.

• In Rembrandt's painted self-portraits (left panel) in which the eyes are clearly visible, his left eye frequently looks straight out and the right off to the side. It is the opposite in his etchings (right panel).
Using epipolar & constant Brightness constraints for stereo matching

For each epipolar line
  For each pixel in the left image
    • compare with every pixel on same epipolar line in right image
    • pick pixel with most similar brightness.
This will never work, so: Match windows
  (Seitz)

Finding Correspondences

\[ W(p_l) \quad W(p_r) \]
Greedy Correspondence Search Algorithm

Input: $i_1, i_2$
Output: disparities

for $i = 1:nrows$
  for $j=1:ncols$
    best($i,j$) = -1
    for $k =$ mindisparity:maxdisparity
      $c =$ Match_Metric($i_1(i,j), i_2(i,j+k), \text{winsize}$)
      if ($c > \text{best}(i,j)$)
        best($i,j$) = $c$
        disparities($i,j$) = $k$
      end
    end
  end
end

$O(nrows \times ncols \times \text{disparities} \times \text{winx} \times \text{winy})$

Simple match metrics

- SSD (Sum of Squared Differences)
  $$\sum_{x,y} |W_1(x,y) - W_2(x,y)|^2$$

- NCC (Normalized Cross Correlation)
  $$\frac{\sum_{x,y}(W_1(x,y) - \overline{W_1})(W_2(x,y) - \overline{W_2})}{\sigma_{W_1} \sigma_{W_2}}$$
  where $\overline{W_i} = \frac{1}{n} \sum_{x,y} W_i$, $\sigma_{W_i} = \sqrt{\frac{1}{n} \sum_{x,y} (W_i - \overline{W_i})^2}$

- What advantages might NCC have over SSD?
### More Match Metrics

<table>
<thead>
<tr>
<th>MATCH METRIC</th>
<th>DEFINITION</th>
</tr>
</thead>
</table>
| Normalized Cross-Correlation (NCC)        | \[
\frac{\sum \{I_u (u, v) \cdot I_v (u + d, v) \}}{\sqrt{\sum \{I_u (u, v) \cdot I_u (u + d, v) \}} \cdot \sqrt{\sum \{I_v (u, v) \cdot I_v (u + d, v) \}}} \]
| Sum of Squared Differences (SSD)          | \[
\sum \{I_u (u, v) - I_u (u + d, v) \}^2
\]                                                                |
| Normalized SSD                            | \[
\frac{\sum \{I_u (u, v) \cdot I_v (u + d, v) \}}{\sqrt{\sum \{I_u (u, v) \cdot I_u (u + d, v) \}} \cdot \sqrt{\sum \{I_v (u, v) \cdot I_v (u + d, v) \}}} \]
| Sum of Absolute Differences (SAD)         | \[
\sum |I_u (u, v) - I_u (u + d, v)|
\]                                                                |
| Zero Mean SAD                             | \[
\sum |I_u (u, v) - I_u (u + d, v)|
\]                                                                |
| Rank                                      | \[
\sum \frac{I_u (u, v) \cdot I_v (u + d, v)}{I_u (u, v) - I_u (u + d, v)}
\]                                                                |
| Census                                    | \[
\sum \text{HAMMING}(I_u (u, v), I_v (u + d, v))
\]                                                                |

These two are actually the same.

---

### Stereo results

Data from University of Tsukuba

![Scene](image1.png)  ![Ground truth](image2.png)

(Seitz)
Results with greedy algorithm and correlation match metric

Window-based matching (best window size)
Ground truth
(Seitz)

Results with better method

Using global optimization
Ground truth
(Seitz)

Boykov et al., *Fast Approximate Energy Minimization via Graph Cuts*,
Some Issues

- Ambiguity
- Epipolar ordering
- Window size
- Window shape
- Lighting
- Half occluded regions

Ambiguity

It’s a coin toss whether $W_L$ will match $W_1$ or $W_2$
Some Issues

- Ambiguity
- **Epipolar ordering**
- Window size
- Window shape
- Lighting
- Half occluded regions
A challenge: Multiple Interpretations

Each feature on left epipolar line matches one and only one feature on right epipolar line.

Multiple Interpretations

Each feature on left epipolar line matches one and only one feature on right epipolar line.
Each feature on left epipolar line matches one and only one feature on right epipolar line.
Some Issues

- Ambiguity
- Epipolar ordering
- **Window size**
- Window shape
- Lighting
- Half occluded regions

Window size

Better results with *adaptive window*


(Seitz)
Some Issues

- Ambiguity
- Epipolar ordering
- Window size
- **Window shape**
- Lighting
- Half occluded regions

Window Shape and Forshortening

Hmmm!! Does window matching really work? Here the window is around corresponding points, but the windows don’t look the same.
Window Shape: Fronto-parallel Configuration

- When scene plane is parallel to the image planes, a square $w_p$ in the scene projects to squares in the images $w_i$ and $w_r$.
- But when scene plane is tilted, $w_p$ projects to a quadrilateral in the images.

Some Issues

- Ambiguity
- Epipolar ordering
- Window size
- Window shape
- Lighting
- Half occluded regions
Lighting Conditions (Photometric Variations)

Does the match metric handle matching across differences of brightness?

Some Issues

- Ambiguity
- Epipolar ordering
- Window size
- Window shape
- Lighting
- Half occluded regions
Half occluded regions

- Half occluded regions are visible in one camera, but not in the other
- They can be a cue for a depth change

Summary of Stereo Constraints

<table>
<thead>
<tr>
<th>CONSTRAINT</th>
<th>BRIEF DESCRIPTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-D Epipolar Search</td>
<td>Arbitrary images of the same scene may be rectified based on epipolar geometry such as epipolar matches lie along one-dimensional scanlines. This reduces the computational complexity and also reduces the likelihood of false matches.</td>
</tr>
<tr>
<td>Monotonic Ordering</td>
<td>Points along an epipolar scanline appear in the same order in both stereo images, assuming that all objects in the scene are approximately the same distance from the cameras.</td>
</tr>
<tr>
<td>Image Brightness Constancy</td>
<td>Assuming Lambertian surfaces, the brightness of corresponding points in stereo images are the same.</td>
</tr>
<tr>
<td>Match Uniqueness</td>
<td>For every point in one stereo image, there is at most one corresponding point in the other image.</td>
</tr>
<tr>
<td>Disparity Continuity</td>
<td>Disparities vary smoothly (i.e. disparity gradient is small) over most of the image. This assumption is violated at object boundaries.</td>
</tr>
<tr>
<td>Disparity Limit</td>
<td>The search space may be reduced significantly by limiting the disparity range, reducing both computational complexity and the likelihood of false matches.</td>
</tr>
<tr>
<td>Fronto-Parallel Surfaces</td>
<td>The implicit assumption made by area-based matching is that objects have fronto-parallel surfaces (i.e. depth is constant within the region of local support). This assumption is violated by sloping and creased surfaces.</td>
</tr>
<tr>
<td>Feature Similarity</td>
<td>Corresponding features must be similar (e.g. edges must have roughly the same length and orientation).</td>
</tr>
<tr>
<td>Structural Grouping</td>
<td>Corresponding feature groupings and their connectivity must be consistent.</td>
</tr>
</tbody>
</table>

(From G. Hager)
Stereo matching

Constraints
- epipolar
- ordering
- uniqueness
- disparity limit
- disparity gradient limit

Trade-off
- Matching cost (data)
- Discontinuities (prior)

Variations on Binocular Stereo

1. **Trinocular Stereopsis**
2. Multiview stereo
3. Helmholtz Reciprocity Stereopsis
Trinocular Epipolar Constraints

These constraints are not independent!

\[
\begin{align*}
    p_1^T e_{12} p_2 &= 0 \\
    p_2^T e_{23} p_3 &= 0 \\
    p_3^T e_{31} p_1 &= 0 \\
    e_{31}^T e_{12} e_{32} &= e_{12}^T e_{23} e_{13} = e_{23}^T e_{31} e_{21} = 0
\end{align*}
\]