Stereo
(Part 2)

Introduction to Computer Vision I
CSE 152A
Lecture 9
Announcements

• Assignment 2 is due Feb 10, 11:59 PM
• Quiz 2 is Feb 12
• Reading:
  – Szeliski
    • Section 12.1.1
• 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

Slanted epipolar lines vs Horizontal, row aligned epipolar lines
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.
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.
Stereo Vision Outline

- Offline
  - Calibrate cameras and determine epipolar geometry

- Online
  1. Acquire stereo images
  2. Rectify images to convenient epipolar geometry
  3. Establish correspondence
  4. Estimate depth
What if stereo geometry is not 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 2D projective transformation (homography).
Rectification

Under perspective projection, the mapping from a plane to a plane is given by a 2D projective transformation (homography)

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

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

Two images

Two homographies

H_L, H_R
Epipolar Rectification

- Create pair of virtual cameras
  - Virtual cameras have the same camera centers as real cameras
  - Both virtual cameras have the same:
    - Camera rotation matrix $R$
    - Camera calibration matrix $K$
- Rectification transformation matrices
  $$H = K_{\text{virtual}} R_{\text{virtual}} R_{\text{real}}^T K_{\text{real}}^{-1}$$
Image pair rectification

Apply projective transformation so that epipolar lines correspond to horizontal scanlines

\[ H \text{ should map epipole } e \text{ to } (1,0,0), \text{ a point at infinity on the } x\text{-axis} \]

\[ H \text{ should minimize image distortion} \]

Note that rectified images are usually not rectangular

\[ \begin{bmatrix} 1 \\ 0 \\ 0 \end{bmatrix} = He \]
Rectification
Given a pair of images, transform both images so that epipolar lines are scan lines.

Input Images
Rectification

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

Rectified Images

epipolar lines run parallel with the x-axis and are aligned between two views (no y disparity)
Rectification
Rectification

- Epipolar lines
Rectification
Polar Rectification

Homography-based Rectification

Polar Rectification

Alternative epipolar rectification method that minimizes pixel distortion
Polar Rectification

Epipoles are in images
(white dot on ball)

Homography-based rectification
is not possible
Features on same epipolar line
Stereo Vision Outline

• Offline
  B – Calibrate cameras and determine epipolar geometry

• Online
  1. Acquire stereo images
  2. Rectify images to convenient epipolar geometry
  3. Establish correspondence
  4. Estimate depth
Two Approaches

1. Feature-Based (sparse)
   - From each image, process “monocular” image to obtain cues (e.g., corners, SIFT features)
   - Establish feature correspondence between the two images

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??
  • There is anecdotal evidence for the latter (camouflage).
  • Random dot stereograms provide an objective answer.
Random Dot Stereograms
Random Dot Stereograms
Stereoscopic 3D
Stereoscopic 3D
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 minimum match cost
    • This will never work, so: match windows

(Seitz)
Finding Correspondences

\[ W(p_1) \]

\[ W(p_r) \]
Correspondence Search Algorithm

For  \( i = 1 : \text{nrows} \)
    for  \( j = 1 : \text{ncols} \)
        \( \text{best}(i,j) = -1 \)
        for  \( k = \text{mindisparity} : \text{maxdisparity} \)
            \( c = \text{Match\_Metric}(I_1(i,j), I_2(i,j+k), \text{winsize}) \)
            if (\( c > \text{best}(i,j) \))
                \( \text{best}(i,j) = c \)
                \( \text{disparities}(i,j) = k \)
        end
    end
end

\( O(n\text{rows} \times n\text{cols} \times \text{disparities} \times \text{winx} \times \text{winy}) \)
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

\[ \overline{W_i} = \frac{1}{n} \sum_{x,y} W_i \], \quad \sigma_{W_i} = \sqrt{\frac{1}{n} \sum_{x,y} (W_i - \overline{W_i})^2} \]
# Match Metric Summary

<table>
<thead>
<tr>
<th>MATCH METRIC</th>
<th>DEFINITION</th>
</tr>
</thead>
</table>
| Normalized Cross-Correlation (NCC) | \[
\sum_{u,v} \frac{\left(I_1(u,v) - \bar{I}_1\right) \cdot \left(I_2(u+d,v) - \bar{I}_2\right)}{\sqrt{\sum_{u,v} (I_1(u,v) - \bar{I}_1)^2 \cdot \sum_{u,v} (I_2(u+d,v) - \bar{I}_2)^2}}
\] |
| Sum of Squared Differences (SSD)    | \[
\sum_{u,v} (I_1(u,v) - I_2(u+d,v))^2
\] |
| Normalized SSD                      | \[
\sum_{u,v} \left(\frac{I_1(u,v) - \bar{I}_1}{\sqrt{\sum_{u,v} (I_1(u,v) - \bar{I}_1)^2}} - \frac{I_2(u+d,v) - \bar{I}_2}{\sqrt{\sum_{u,v} (I_2(u+d,v) - \bar{I}_2)^2}}\right)^2
\] |
| Sum of Absolute Differences (SAD)   | \[
\sum_{u,v} |I_1(u,v) - I_2(u+d,v)|
\] |
| Zero Mean SAD                       | \[
\sum_{u,v} |I_1(u,v) - I_2(u+d,v) - \bar{I}_2|
\] |
| Rank                                | \[
I_k'(u,v) = \sum_{m,n} I_k(m,n) < I_k(u,v)
\sum_{u,v} (I_1'(u,v) - I_2'(u+d,v))
\] |
| Census                              | \[
I_k'(u,v) = \text{BITSTRING}_{m,n} (I_k(m,n) < I_k(u,v))
\sum_{u,v} HAMMING(I_1'(u,v), I_2'(u+d,v))
\] |

These two are actually the same.
Stereo results

– Data from University of Tsukuba

Scene

Ground truth

(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
Boykov et al., Fast Approximate Energy Minimization via Graph Cuts,
International Conference on Computer Vision, September 1999.

Ground truth
(Seitz)
State of the Art Results

Using neural networks

Ground truth

Some Issues

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

Each feature on left epipolar line match one and only one feature on right epipolar line.
Multiple Interpretations

Each feature on left epipolar line match one and only one feature on right epipolar line.
Multiple Interpretations

Each feature on left epipolar line match one and only one feature on right epipolar line.
Multiple Interpretations

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

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

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

With the Greedy Algorithm $W_1$ might match $W_R$, but $W_2$ might also match $W_R$.
Some Issues

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

- Effect of window size

Better results with *adaptive window*

(Seitz)
Some Issues

- Epipolar ordering
- Ambiguity
- Window size
- Window shape
- Lighting
- Half occluded regions
Window Shape and Forshortening
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 $w_l$ and $w_r$.
- But when scene plane is tilted, $w_p$ projects to a quadrilateral in the images.
Some Issues

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

Does the match metric handle matching across differences of brightness?
Some Issues

• Epipolar ordering
• Ambiguity
• 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 that stereo 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

Similarity measure (SSD or NCC)

Optimal path (dynamic programming)

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

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

(From Pollefeys)
(From Pollefeys, CVGIP ’96; Koch ’96; Falkenhagen ’97; Van Meerbergen, Vergauwen, Pollefeys, VanGool, IJCV ’02)
Next Lecture

• Structure from Motion
• Reading:
  – Szeliski
    • Sections 11.3, 11.4, and 11.5