Calibrated Stereo (Part 2) and Feature Matching

Computer Vision I
CSE 252A
Lecture 8
Announcements

• Assignment 1 is due today, 11:59 PM
• Assignment 2 will be released today
  – Due Nov 3, 11:59 PM
• Reading:
  – Szeliski
    • Sections 12.3, 12.5, 12.6
Stereo Vision Outline

• Offline
  – Calibration of stereo cameras

• Online
  1. Acquire stereo images
  2. Epipolar rectify stereo images
  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
Establish correspondences
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 most similar brightness.

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
end

$O(n\text{rows} * n\text{cols} * \text{disparities} * \text{winx} * \text{winy})$
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>
<tbody>
<tr>
<td>Normalized Cross-Correlation (NCC)</td>
<td>[ \frac{\sum_{u,v} (I_1(u,v) - \bar{I}_1)(I_2(u + d, v) - \bar{I}<em>2)}{\sqrt{\sum</em>{u,v} (I_1(u,v) - \bar{I}<em>1)^2 \cdot \sum</em>{u,v} (I_2(u + d, v) - \bar{I}_2)^2}} ]</td>
</tr>
<tr>
<td>Sum of Squared Differences (SSD)</td>
<td>[ \sum_{u,v} (I_1(u,v) - I_2(u + d, v))^2 ]</td>
</tr>
<tr>
<td>Normalized SSD</td>
<td>[ \sum_{u,v} \left( \frac{(I_1(u,v) - \bar{I}<em>1)}{\sqrt{\sum</em>{u,v} (I_1(u,v) - \bar{I}_1)^2}} - \frac{(I_2(u + d, v) - \bar{I}<em>2)}{\sqrt{\sum</em>{u,v} (I_2(u + d, v) - \bar{I}_2)^2}} \right)^2 ]</td>
</tr>
<tr>
<td>Sum of Absolute Differences (SAD)</td>
<td>[ \sum_{u,v}</td>
</tr>
<tr>
<td>Zero Mean SAD</td>
<td>[ \sum_{u,v}</td>
</tr>
<tr>
<td>Rank</td>
<td>[ \hat{I}<em>k(u,v) = \sum</em>{m,n} I_k(m,n) &lt; I_k(u,v) ]</td>
</tr>
<tr>
<td></td>
<td>[ \sum_{u,v}</td>
</tr>
<tr>
<td>Census</td>
<td>[ \hat{I}<em>k(u,v) = \text{BITSTRING}</em>{m,n} (I_k(m,n) &lt; I_k(u,v)) ]</td>
</tr>
<tr>
<td></td>
<td>[ \sum_{u,v} \text{HAMMING}(I_1(u,v), I_2(u + d, v)) ]</td>
</tr>
</tbody>
</table>
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
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?

$W(P_l)$  

$W(P_r)$
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)

(Cox et al. CVGIP’96; Koch’96; Falkenhagen´97;
Van Meerbergen,Vergauwen,Pollefeys,VanGool IJCV‘02)
Estimate depth
Reconstruction: General 3D case

Given two image measurements $x$ and $x'$, estimate scene point

$$\hat{x} = P\hat{X}$$

$$\hat{x}' = P'\hat{X}$$

Estimate $\hat{X}$ that minimizes $d(x, \hat{x})^2 + d(x', \hat{x}')^2$
Binocular Stereo: Estimating Depth
2-D world with 1-D image plane

Two measurements: $X_L, X_R$
Two unknowns: $X, Z$

Constants:
- Baseline: $d$
- Focal length: $f$

Disparity: $(X_L - X_R)$

\[
X = \frac{d \cdot X_L}{(X_L - X_R)}
\]

\[
Z = \frac{d \cdot f}{(X_L - X_R)}
\]

(Adapted from Hager)
Faraway points – small disparity
Infinitely far, zero disparity

Nearby points – large disparity

\[ Z = \frac{df}{x_L - x_R} \]
More on stereo …

The Middleburry Stereo Vision Research Page
http://cat.middlebury.edu/stereo/

Recommended reading

D. Scharstein and R. Szeliski.  

Some Challenges & Problems

• Photometric issues:
  – specularities
  – strongly non-Lambertian BRDFs

• Surface structure
  – lack of texture
  – repeating texture within horopter bracket

• Geometric ambiguities
  – as surfaces turn away, difficult to get accurate reconstruction
    (affine approximate can help)
  – at the occluding contour, likelihood of good match but incorrect reconstruction
Variations on Binocular Stereo

1. Trinocular Stereopsis
2. Helmholtz Reciprocity Stereopsis
Trinocular Epipolar Constraints
Helmholtz reciprocity

\[ \rho(\theta_{\text{in}}, \phi_{\text{in}} ; \theta_{\text{out}}, \phi_{\text{out}}) = \rho(\theta_{\text{out}}, \phi_{\text{out}} ; \theta_{\text{in}}, \phi_{\text{in}}) \]

[Helmholtz, 1910], [Minnaert, 1941], [Nicodemus et al, 1977]
Point Source Illumination

\[ \mathbf{i}_l = \rho(\hat{\mathbf{v}}_r, \hat{\mathbf{v}}_l) \frac{\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_r}{|\mathbf{o}_r - \mathbf{p}|^2} \]

\[ \mathbf{i}_r = \rho(\hat{\mathbf{v}}_l, \hat{\mathbf{v}}_r) \frac{\hat{\mathbf{n}} \cdot \hat{\mathbf{v}}_l}{|\mathbf{o}_l - \mathbf{p}|^2} \]
Matching Constraint

\[
\left( i_l \frac{\hat{v}_l}{|o_l - p|^2} - i_r \frac{\hat{v}_r}{|o_r - p|^2} \right) \cdot \hat{n} = 0
\]

measured \quad \hat{w}_l \quad \hat{w}_r

computed from geometric calibration
Using Multiple Helmholtz Stereo Pairs

- Multiple views (at least three pairs) yield a matrix constraint equation
- Matrix must be Rank 2
- Search for depth where rank constraint is satisfied

\[ \begin{pmatrix}
  i_{l1} \mathbf{w}_{l1}^T - i_{r1} \mathbf{w}_{r1}^T \\
  i_{l2} \mathbf{w}_{l2}^T - i_{r2} \mathbf{w}_{r2}^T \\
  \vdots
\end{pmatrix} \Rightarrow \hat{n} = 0 \]
Finding the Normal at each point

Additionally, the surface normal $\hat{n}$ must lie in the null space of the matrix

$$\begin{pmatrix}
i_{l1} \mathbf{w}_{l1}^T - i_{r1} \mathbf{w}_{r1}^T \\
i_{l2} \mathbf{w}_{l2}^T - i_{r2} \mathbf{w}_{r2}^T \\
\vdots
\end{pmatrix} \hat{n} = 0$$
Bulldog: Disparity
Bulldog: Normal Field
Reciprocal Images: Typical Dataset
Reciprocal Images: Typical Dataset

Conventional Stereo
- Constant brightness
- No structure in textureless regions
Reciprocal Images: Typical Dataset

Conventional Stereo
• Constant brightness
• No structure in textureless regions

Photometric Stereo
• Needs reflectance model
• No direct depth estimates
Reciprocal Images: Typical Dataset

**Conventional Stereo**
- Constant brightness
- No structure in textureless regions

**Photometric Stereo**
- Needs reflectance model
- No direct depth estimates

**Helmholtz Stereo**
- No assumed reflectance
- Gives depth and surface normals
Metric Reconstruction
Next Lecture

• Uncalibrated stereo and feature extraction
• Reading:
  – Szeliski
    • Sections 11.3.3, 3.5.3, 7.1.1, and 7.1.2