Shape From X
(Stereo, Photometric Stereo, and Structure from Motion)

Lecture 12

Shape From X

• Recovery of 3D (shape) from one or two (2D images).
Shape From X

- Stereo
- Motion
- Shading
- Photometric Stereo
- Texture
- Contours
- Silhouettes
- Defocus

Copyright Mubarak Shah 2003

Shape from Stereo

(a) (b) (c)
Stereo

\[ \frac{Z + f}{Z} = \frac{x_1 + x_2 + B}{B}, \quad Z = \frac{fB}{x_1 + x_2}, \]

\[ x_1 + x_2 = \text{disparity} = d \]

Stereo Pairs and Depth Maps  
(from Szeliski’s book)
Rectification

Correlation Based Stereo Methods

• Once disparity is available compute depth using

\[ Z = \frac{fB}{d} \]
Correspondence using Search

Criterion function:

\[ Z = \frac{fB}{d} \]

Correlation Based Stereo Methods

- Disparity map can be constructed based on a correlation measure

\[
\begin{align*}
\text{SSD} &= \sum \sum (I_{\text{left}} - I_{\text{right}}) \quad \text{Sum of squares difference} \\
\text{AD} &= \sum \sum |I_{\text{left}} - I_{\text{right}}| \quad \text{Absolute difference} \\
\text{CC} &= \sum \sum I_{\text{left}} I_{\text{right}} \quad \text{Cross correlation} \\
\text{NC} &= \frac{\sum \sum (I_{\text{left}} I_{\text{right}})}{\sqrt{\sum \sum I_{\text{left}}^2} \sqrt{\sum \sum I_{\text{right}}^2}} \quad \text{Normalized Correlation} \\
\text{MC} &= \frac{1}{64\sigma_{i,i+1}^2} \sum \sum (I_{i+1} - \mu_{i+1})(I_i - \mu_i) \quad \text{Mutual Correlation}
\end{align*}
\]

Alper Yilmaz, Mubarak shah, Fall 2011 UCF
Correlation

- Similarity/Dissimilarity Measures
  - Sum of Squares Difference (SSD)
  - Normalized Correlation
  - Mutual Correlation
  - Mutual information 
    \[ I(x, y) = \sum \sum p(x, y) \log \frac{p(x, y)}{p_1(x)p_2(y)} \]

- Use
  - Gray levels
  - Laplacian of Gaussian
  - Gradient magnitude

Block Matching

- Can be used for
  - computing MPEG motion vectors
  - Optical flow
  - Stereo
  - Image matching
Block Matching (MPEG)

• For each 8X8 block, centered around pixel \((x,y)\) in frame \(k\), \(B_k\)
  – Obtain 16X16 block in frame \(k-1\), centered around \((x,y)\), \(B_{k-1}\)
  – Compute Sum of Squares Differences (SSD) between 8X8 block, \(B_k\) and all possible 8X8 blocks in \(B_{k-1}\)
  – The 8X8 block in \(B_{k-1}\) centered around \((x',y')\), which gives the least SSD is the match
  – The displacement vector (optical flow) is given by \(u=x-x'; \ v=y-y'\)

Sum of Squares Differences (SSD)

\[
(u(x,y),v(x,y)) = \arg \min_{u,v=-4...4} \sum_{i=0}^{7} \sum_{j=0}^{7} \left( f_k(x+i,y+j) - f_{k-1}(x+i+u,y+j+v) \right)^2
\]

\[
u = \arg \min_{u=-4...4} \sum_{i} \left( f_k(x_i) - f_{k-1}(x_i+u) \right)^2
\]
Minimum Absolute Difference (MAD)

\[(u(x, y), v(x, y)) = \arg \min_{u, v = -4, \ldots, 4} \sum_{i=0}^{7} \sum_{j=0}^{7} \| f_k(x + i, y + j) - f_{k-i}(x + i + u, y + j + v) \| \]

Maximum Matching Pixel Count (MPC)

\[T(x, y, u, v) = \begin{cases} 
1 & \text{if } \| f_k(x, y) - f_{k-i}(x + u, y + v) \| \leq t \\
0 & \text{otherwise} 
\end{cases} \]

\[(u(x, y), v(x, y)) = \arg \max_{u, v = -4, \ldots, 4} \sum_{i=0}^{7} \sum_{j=0}^{7} T(x + i, y + j; u, v) \]
Cross Correlation

\[
(u(x, y), v(x, y)) = \arg \max_{u,v=-4...4} \sum_{i=0}^{7} \sum_{j=0}^{7} (f_k(x+i, y+j) f_{k-1}(x+i+u, y+j+v))
\]

Normalized Correlation

\[
(u, v) = \arg \max_{u,v=-4...4} \frac{\sum_{i=0}^{7} \sum_{j=0}^{7} ((f_k(x+i, y+j) - \mu_k)(f_{k-1}(x+i+u, y+j+v) - \mu_{k-1}))}{\left[ \sum_{i=0}^{7} \sum_{j=0}^{7} (f_k(x+i, y+j) - \mu_k)^2 \right]^{1/2} \left[ \sum_{i=0}^{7} \sum_{j=0}^{7} (f_{k-1}(x+i+u, y+j+v) - \mu_{k-1})^2 \right]^{1/2}}
\]

and \(\mu_k\) are the means of patch-1 and patch-2 respectively.
**Mutual Correlation**

\[
(u(x, y), v(x, y)) = \arg \max_{u, v = -4 \cdots 4} \frac{1}{64\sigma_1\sigma_2} \sum_{i=0}^{7} \sum_{j=0}^{7} (f_i(x+i, y+j) - \mu_i)f_j(x+i+u, y+j+v) - \mu_j)
\]

Sigma and mu are standard deviation and mean of patch-1 and patch-2 respectively.

---

**Mutual Information**

\[
U = \arg \max_{u, v = -4 \cdots 4} \sum p(X, Y + U) \log \frac{p(X, Y + U)}{p(X)p(Y + U)}
\]
Barnard’s Stereo Method

- Similar intensity
  - Similar to brightness constraint
- Smoothness of disparity

\[ E = \sum_{i=1}^{1} \sum_{j=1}^{1} \left| I_{left}(x+i, y+j) - I_{right}(x+i + D_x(x, y), y+j) \right| + \lambda \| \nabla D(x, y) \| \]

\[ \nabla D(x, y) = \sum_{i=1}^{1} \sum_{j=1}^{1} |D(x+i, y+j) - D(x, y)| \]

Barnard’s Stereo Method

- Energy can be minimized using brute force search
  - Let max allowed disparity is 10 pixels
  - For 128x128 image for 10 possible levels of disparity
    - There $10^{16384}$ possible disparity values
  - We can select any minimization technique
  - Barnard choose simulated annealing
Simulated Annealing

- Select a random state S (disparities)
- Select a high temperature
  - Select random S’
  - Compute $\Delta E = E(S’) - E(S)$
  - If ($\Delta E < 0$) $S \leftarrow S’$
  - Else
    - $P \leftarrow \exp(-\Delta E / T)$
    - $X \leftarrow \text{random}(0,1)$
      - If $X < P$ then $S \leftarrow S’$
    - If no decrease in several iterations lower T

Examples

- bread
- toy
- apple

Alper Yilmaz, Mubarak Shah, Fall
2011 UCF
Examples

Left Image  Right Image  Depth Map

Stereo results

– Data from University of Tsukuba
– Similar results on other images without ground truth

Scene  Ground truth
Results with window correlation

Window-based matching (best window size)  Ground truth

Results with better method

State of the art method  Ground truth

Applications of Stereo (from Szeliski’s book)

Reading Material

• Fundamental of Computer Vision
  – 6.2.1, 6.2.4 and 6.2.5
• Computer Vision: Algorithms and Applications, Richard Szeliski
  – Chapter 11