Motion I

Introduction to Computer Vision
CSE 152
Lecture 17
Announcements

• Assignment 3: Due today at midnight. Hardcopy due date extended to 12 noon tomorrow.

• Today
  – Review of Photometric Stereo
  – Discrete Structure from Motion
  – Continuous motion
Now if BRDF and light source direction/strength are known, then for each image point

1. Image intensity is a function of only the direction of the surface normal.

2. In gradient space, we have $E(x,y) = R(p,q)$ where $E$ is measured, $(p,q)$ is unknown, but form of function $R(p,q)$ is known.
Coordinate system

Normal vector
\[ \mathbf{n} = \frac{\partial s}{\partial x} \times \frac{\partial s}{\partial y} = \left( \frac{\partial z}{\partial x}, \frac{\partial z}{\partial y}, -1 \right) \]

Gradient Space: (p,q)
\[ p = \frac{\partial z}{\partial x}, \quad q = \frac{\partial z}{\partial y} \]
Two Light Sources
Two reflectance maps

Third image would disambiguate match
Plastic Baby Doll: Normal Field
Recovering the surface $z(x,y)$

Many methods: Simplest approach

1. From estimate $n = (n_x, n_y, n_z)$, $p=n_x/n_z$, $q=n_y/n_z$
2. Integrate $p=\frac{dz}{dx}$ along a row $(x,0)$ to get $z(x,0)$
3. Then integrate $q=\frac{dz}{dy}$ along each column starting with value of first row to get $z(x,y)$
Integrability

If \( f(x,y) \) is the height function, we expect that

\[
\frac{\partial}{\partial y} \frac{\partial z}{\partial x} = \frac{\partial}{\partial x} \frac{\partial z}{\partial y}
\]

In terms of estimated gradient space \((p,q)\), this means:

\[
\frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}
\]

But since \( p \) and \( q \) were estimated independently at each point as intersection of curves on three reflectance maps, equality is not going to exactly hold.
At image location \((u,v)\), the intensity of a pixel \(x(u,v)\) is:

\[
e(u,v) = [a(u,v) \hat{n}(u,v)] \cdot [s_0 \hat{s}]
\]

\[
= b(u,v) \cdot s
\]

where

- \(a(u,v)\) is the albedo of the surface projecting to \((u,v)\).
- \(\hat{n}(u,v)\) is the direction of the surface normal.
- \(s_0\) is the light source intensity.
- \(s\) is the direction to the light source.
Important Special Case: Lambertian Photometric Stereo

- If the light sources $s_1$, $s_2$, and $s_3$ are known, then we can recover $b$ from as few as three images. (Photometric Stereo: Silver 80, Woodham 81).

\[
\begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} = b^T \begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}
\]

- i.e., we measure $e_1$, $e_2$, and $e_3$ and we know $s_1$, $s_2$, and $s_3$. We can then solve for $b$ by solving a linear system.

\[
b^T = \begin{bmatrix} e_1 & e_2 & e_3 \end{bmatrix} \begin{bmatrix} s_1 & s_2 & s_3 \end{bmatrix}^{-1}
\]

- Normal is: $n = b/|b|$, albedo is: $|b|$
Bas-Relief Ambiguity

Light Sources ($s_1$, $s_2$, $s_3$) are unknown
III. Photometric Stereo with unknown lighting and Lambertian surfaces
How do you construct subspace?

\[
\begin{bmatrix}
e_1 & e_2 & e_3
\end{bmatrix} = B^T \begin{bmatrix}
s_1 & s_2 & s_3
\end{bmatrix}
\]

- Given three or more images \(e_1 \ldots e_n\), estimate \(B\) and \(s_i\).
- How? Given images in form of \(E = [e_1 \ e_2 \ldots]\), Compute SVD(\(E\)) and \(B^*\) is \(n\) by 3 matrix formed by first 3 singular values.
Matrix Decompositions

• Definition: The factorization of a matrix $M$ into two or more matrices $M_1, M_2, \ldots, M_n$, such that $M = M_1 M_2 \ldots M_n$.

• Many decompositions exist…
  – $QR$ Decomposition
  – $LU$ Decomposition
  – $LDU$ Decomposition
  – Etc.
Singular Value Decomposition

Excellent ref: ‘Matrix Computations,” Golub, Van Loan

- Any $m$ by $n$ matrix $A$ may be factored such that
  \[ A = U \Sigma V^T \]
  \[ [m \times n] = [m \times m][m \times n][n \times n] \]

- $U$: $m$ by $m$, orthogonal matrix
  - Columns of $U$ are the eigenvectors of $AA^T$

- $V$: $n$ by $n$, orthogonal matrix,
  - columns are the eigenvectors of $A^TA$

- $\Sigma$: $m$ by $n$, diagonal with non-negative entries ($\sigma_1, \sigma_2, \ldots, \sigma_s$) with $s=\min(m,n)$ are called the singular values
  - Singular values are the square roots of eigenvalues of both $AA^T$ and $A^TA$

- Result of SVD algorithm: $\sigma_1 \geq \sigma_2 \geq \ldots \geq \sigma_s$
Applying SVD to Photometric stereo

• The images are formed by
\[
\begin{bmatrix}
e_1 & e_2 & e_3 & \ldots & e_n
\end{bmatrix} = \mathbf{B}^T \begin{bmatrix}
s_1 & s_2 & s_3 & \ldots & s_n
\end{bmatrix}
\]
E = \mathbf{B}^T \mathbf{S}

• So, svd(E) = \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T where \mathbf{U} is N by n, \mathbf{\Sigma} is n by n, and \mathbf{V}^T is n by N.

• Without noise, we expect 3 non-zero singular values, and so \( \mathbf{U} \mathbf{\Sigma} \mathbf{V}^T = \mathbf{U'} \mathbf{\Sigma'} \mathbf{V'}^T \)
where \( \mathbf{U'} \) is N by 3, \( \mathbf{\Sigma'} \) is 3 by 3, and \( \mathbf{V'}^T \) is 3 by n.

• In particular \( \mathbf{B} = \mathbf{U'} \mathbf{A} \) where \( \mathbf{A} \) is some 3x3 matrix.
Do Ambiguities Exist? Yes

• Is B unique?

• For any \( A \in \text{GL}(3), \ B^* = BA \) also a solution

• For any image of \( B \) produced with light source \( S \), the same image can be produced by lighting \( B^* \) with \( S^* = A^{-1}S \) because

\[
X = B^*S^* = B AA^{-1}S = BS
\]

• When we estimate B using SVD, the rows are NOT generally normal * albedo.
Surface Integrability

In general, $B^*$ does not have a corresponding surface.

Linear transformations of the surface normals in general do not produce an integrable normal field.
GBR Transformation

Only **Generalized Bas-Relief** transformations satisfy the integrability constraint:

\[
A = G^T = \begin{bmatrix}
\lambda & 0 & -\mu \\
0 & \lambda & -\nu \\
0 & 0 & 1
\end{bmatrix}^T
\]

\[
\bar{f}(x, y) = \lambda f(x, y) + \mu x + \nu y
\]
Generalized Bas-Relief Transformations

Objects differing by a GBR have the same illumination cone.

Without knowledge of light source location, one can only recover surfaces up to GBR transformations.
Uncalibrated photometric stereo

1. Take n images as input, perform SVD to compute $B^*$.

2. Find some $A$ such that $B^*A$ is close to integrable.

3. Integrate resulting gradient field to obtain height function $f^*(x,y)$.

Comments:

– $f^*(x,y)$ differs from $f(x,y)$ by a GBR.
– Can use specularities to resolve GBR for non-Lambertian surface.
Structure-from-Motion (SFM)

Goal: Take as input two or more images or video w/o any information on camera position/motion, and estimate camera position and 3-D structure of scene.

Two Approaches
1. Discrete motion (wide baseline)
   1. Orthographic (affine) vs. Perspective
   2. Two view vs. Multi-view
   3. Calibrated vs. Uncalibrated
2. Continuous (Infinitesimal) motion
Discrete Motion: Some Counting

Consider $M$ images of $N$ points, how many unknowns?

1. Affix coordinate system to location of first camera location: $(M-1)*6$ Unknowns
2. 3-D Structure: $3*N$ Unknowns
3. Can only recover structure and motion up to scale. Why?

Total number of unknowns: $(M-1)*6+3*N-1$

Total number of measurements: $2*M*N$

Solution is possible when $(M-1)*6+3*N-1 \leq 2*M*N$
Epipolar Constraint: Calibrated Case

\[ \overrightarrow{O_p} \cdot [\overrightarrow{OO'} \times \overrightarrow{O'p'}] = 0 \]
\[ \mathbf{p} \cdot [\mathbf{t} \times (\mathcal{R}\mathbf{p'})] = 0 \]

where

\[ \mathbf{p} = (u, v, 1)^T \]
\[ \mathbf{p'} = (u', v', 1)^T \]
\[ \mathcal{M} = (\mathbf{I} \quad \mathbf{0}) \]
\[ \mathcal{M}' = (\mathcal{R}^T, -\mathcal{R}^T\mathbf{t}) \]

Essential Matrix
(Longuet-Higgins, 1981)

\[ \mathbf{p}^T\mathcal{E}\mathbf{p'} = 0 \]
with
\[ \mathcal{E} = [\mathbf{t}_x]\mathcal{R} \]

where
\[ [\mathbf{t}_x] = \begin{bmatrix} 0 & -t_z & t_y \\ t_z & 0 & -t_x \\ -t_y & t_x & 0 \end{bmatrix} \]
The Eight-Point Algorithm (Longuet-Higgins, 1981)

Let $F$ denote the Essential Matrix. Here

$$
\begin{pmatrix}
(u, v, 1) \\
F_{11} & F_{12} & F_{13} \\
F_{21} & F_{22} & F_{23} \\
F_{31} & F_{32} & F_{33}
\end{pmatrix}
\begin{pmatrix}
u' \\
\end{pmatrix}
= 0
$$

Set $F_{33}$ to 1

$$
\begin{pmatrix}
u' \\
v
\end{pmatrix}
= 0
$$

Solve for $F$

Solve for $R$ and $t$
Sketch of Two View SFM Algorithm

Input: Two images
1. Detect feature points
2. Find 8 matching feature points (easier said than done)
3. Compute the Essential Matrix E using Normalized 8-point Algorithm
4. Computer R and T (recall that E=RS where S is skew symmetric matrix)
5. Perform stereo matching using recovered epipolar geometry expressed via E.
6. Reconstruct 3-D geometry of corresponding points.
Feature points

Select strongest features (e.g. 1000/image)
Finding Corners

Intuition:

• Right at corner, gradient is ill defined.

• Near corner, gradient has two different values.
Detecting Feature points
(e.g. Harris & Stephens' 88; Shi & Tomasi' 94)

Find points that differ as much as possible from all neighboring points

$$SSD \approx \Delta^T M \Delta$$

$$M = \int \int_W \begin{bmatrix} \frac{\partial I}{\partial x} \\ \frac{\partial I}{\partial y} \end{bmatrix} \begin{bmatrix} \frac{\partial I}{\partial x} & \frac{\partial I}{\partial y} \end{bmatrix} w(x, y) dx dy$$

$M$ should have large eigenvalues

Feature = local maxima of $F(\lambda_1, \lambda_2)$
Feature matching

Evaluate normalized cross correlation (or sum of squared differences) for all features with similar coordinates

\[ (x', y') \in [x - \frac{w}{10}, x + \frac{w}{10}] \times [y - \frac{h}{10}, y + \frac{h}{10}] \]

Keep mutual best matches

Still many wrong matches!
Comments

- **Greedy Algorithm:**
  - Given feature in one image, find best match in second image irrespective of other matches.
  - OK for small motions, little rotation, small search window

- **Otherwise**
  - Must compare descriptor over rotation
  - Can’t consider $O(n^8)$ potential pairings (way too many), so
    - Manual correspondence (e.g., façade, photogrametry).
    - use random sampling (RANSAC)
    - More descriptive features (line segments, larger regions, color).
    - Use video sequence to track, but perform SFM w/ first and last image.
Continuous Motion

- Consider a video camera moving continuously along a trajectory (rotating & translating).
- How do points in the image move?
- What does that tell us about the 3-D motion & scene structure?
Motion

Some problems of motion

1. Correspondence: Where have elements of the image moved between image frames
2. Reconstruction: Given correspondence, what is 3-D geometry of scene
3. Segmentation: What are regions of image corresponding to different moving objects
4. Tracking: Where have objects moved in the image? related to correspondence and segmentation.

Variations:

– Small motion (video),
– Wide-baseline (multi-view)
Motion

“When objects move at equal speed, those more remote seem to move more slowly.”
- Euclid, 300 BC
Simplest Idea for video processing
Image Differences

- Given image $I(u,v,t)$ and $I(u,v, t+\delta t)$, compute $I(u,v, t+\delta t) - I(u,v,t)$.

- This is partial derivative: $\frac{\partial I}{\partial t}$

- At object boundaries, $|\frac{\partial I}{\partial t}|$ is large and is a cue for segmentation

- Doesn’t tell which way stuff is moving
Background Subtraction

• Gather image $I(x,y,t_0)$ of background without objects of interest (perhaps computed over average over many images).

• At time $t$, pixels where $|I(x,y,t)-I(x,y,t_0)| > \tau$ are labeled as coming from foreground objects.
The Motion Field
Where in the image did a point move?

Down and left
The Motion Field
What causes a motion field?

1. Camera moves (translates, rotates)
2. Object’s in scene move rigidly
3. Objects articulate (pliers, humans, animals)
4. Objects bend and deform (fish)
5. Blowing smoke, clouds