Motion and Optical Flow

Slides from Ce Liu, Steve Seitz, Larry Zitnick, Ali Farhadi
We live in a moving world

• Perceiving, understanding and predicting motion is an important part of our daily lives
Motion and perceptual organization

• Even “impoverished” motion data can evoke a strong percept

Motion and perceptual organization

- Even “impoverished” motion data can evoke a strong percept

Seeing motion from a static picture?

http://www.ritsumei.ac.jp/~akitaoka/index-e.html
More examples
How is this possible?

- The true mechanism is to be revealed
- FMRI data suggest that illusion is related to some component of eye movements
- We don’t expect computer vision to “see” motion from these stimuli, yet
What do you see?
In fact, ...
The cause of motion

• Three factors in imaging process
  – Light
  – Object
  – Camera

• Varying either of them causes motion
  – Static camera, moving objects (surveillance)
  – Moving camera, static scene (3D capture)
  – Moving camera, moving scene (sports, movie)
  – Static camera, moving objects, moving light (time lapse)
Motion scenarios (priors)

Static camera, moving scene

Moving camera, static scene

Moving camera, moving scene

Static camera, moving scene, moving light
We still don’t touch these areas
How can we recover motion?
Recovering motion

• Feature-tracking
  – Extract visual features (corners, textured areas) and “track” them over multiple frames

• Optical flow
  – Recover image motion at each pixel from spatio-temporal image brightness variations (optical flow)

Two problems, one registration method

Feature tracking

• Challenges
  – Figure out which features can be tracked
  – Efficiently track across frames
  – Some points may change appearance over time (e.g., due to rotation, moving into shadows, etc.)
  – Drift: small errors can accumulate as appearance model is updated
  – Points may appear or disappear: need to be able to add/delete tracked points
Feature tracking

- Given two subsequent frames, estimate the point translation

- Key assumptions of Lucas-Kanade Tracker
  - **Brightness constancy**: projection of the same point looks the same in every frame
  - **Small motion**: points do not move very far
  - **Spatial coherence**: points move like their neighbors
The brightness constancy constraint

\[
(x, y) \quad \text{displacement} = (u, v)
\]

\[
I(x, y, t) \quad I(x, y, t+1)
\]

\[
(x^\circ, y + v)
\]

- **Brightness Constancy Equation:**

\[
I(x, y, t) = I(x + u, y + v, t + 1)
\]

Take Taylor expansion of \(I(x+u, y+v, t+1)\) at \((x,y,t)\) to linearize the right side:

Image derivative along \(x\) \quad Difference over frames

\[
I(x+u, y+v, t+1) \approx I(x, y, t) + I_x \cdot u + I_y \cdot v + I_t
\]

\[
I(x+u, y+v, t+1) - I(x, y, t) = +I_x \cdot u + I_y \cdot v + I_t
\]

So:

\[
I_x \cdot u + I_y \cdot v + I_t \approx 0 \quad \rightarrow \nabla I \cdot [u \ v]^T + I_t = 0
\]
The brightness constancy constraint

Can we use this equation to recover image motion \((u,v)\) at each pixel?

\[
\nabla I \cdot [u \ v]^T + I_t = 0
\]

• How many equations and unknowns per pixel?
  • One equation (this is a scalar equation!), two unknowns \((u,v)\)

The component of the motion perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

If \((u, v)\) satisfies the equation, so does \((u' + u, v' + v')\) if

\[
\nabla I \cdot [u' \ v']^T = 0
\]
The aperture problem

Actual motion
The aperture problem

Perceived motion
The barber pole illusion

http://en.wikipedia.org/wiki/Barberpole_illusion
The barber pole illusion

http://en.wikipedia.org/wiki/Barberpole_illusion
Solving the ambiguity...


- How to get more equations for a pixel?
- **Spatial coherence constraint**
- Assume the pixel’s neighbors have the same \((u,v)\)
  - If we use a 5x5 window, that gives us 25 equations per pixel

\[
0 = I_t(p_i) + \nabla I(p_i) \cdot [u \ v]
\]

\[
\begin{bmatrix}
I_x(p_1) & I_y(p_1) \\
I_x(p_2) & I_y(p_2) \\
\vdots & \vdots \\
I_x(p_{25}) & I_y(p_{25})
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix}
= -
\begin{bmatrix}
I_t(p_1) \\
I_t(p_2) \\
\vdots \\
I_t(p_{25})
\end{bmatrix}
\]
Solving the ambiguity...

- Least squares problem:

\[
\begin{bmatrix}
I_x(p_1) & I_y(p_1) \\
I_x(p_2) & I_y(p_2) \\
\vdots & \vdots \\
I_x(p_{25}) & I_y(p_{25}) \\
\end{bmatrix}
\begin{bmatrix}
u \\
v \\
\end{bmatrix} = -
\begin{bmatrix}
I_t(p_1) \\
I_t(p_2) \\
\vdots \\
I_t(p_{25}) \\
\end{bmatrix}
\]

\[
A \begin{bmatrix}
d \\
\end{bmatrix} = b
\]

25x2 2x1 25x1
Matching patches across images

- Overconstrained linear system

\[
\begin{bmatrix}
I_x(p_1) & I_y(p_1) \\
I_x(p_2) & I_y(p_2) \\
\vdots & \vdots \\
I_x(p_{25}) & I_y(p_{25})
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix}
= -
\begin{bmatrix}
I_t(p_1) \\
I_t(p_2) \\
\vdots \\
I_t(p_{25})
\end{bmatrix}
\]

\[
A \quad d = b
25x2 \quad 2x1 \quad 25x1
\]

Least squares solution for \(d\) given by

\[
(A^T A) \quad d = A^T b
\]

\[
\begin{bmatrix}
\sum I_x I_x & \sum I_x I_y \\
\sum I_x I_y & \sum I_y I_y
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix}
= -
\begin{bmatrix}
\sum I_x I_t \\
\sum I_y I_t
\end{bmatrix}
\]

\[
A^T A \\
A^T b
\]

The summations are over all pixels in the K x K window
Conditions for solvability

Optimal \((u, v)\) satisfies Lucas-Kanade equation

\[
\begin{bmatrix}
\sum I_x I_x & \sum I_x I_y \\
\sum I_x I_y & \sum I_y I_y 
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix} =
-\begin{bmatrix}
\sum I_x I_t \\
\sum I_y I_t
\end{bmatrix}
\]

\[A^T A\]

\[A^T b\]

When is this solvable? I.e., what are good points to track?

- \(A^T A\) should be invertible
- \(A^T A\) should not be too small due to noise
  - eigenvalues \(\lambda_1\) and \(\lambda_2\) of \(A^T A\) should not be too small
- \(A^T A\) should be well-conditioned
  - \(\lambda_1 / \lambda_2\) should not be too large (\(\lambda_1 = \) larger eigenvalue)

Does this remind you of anything?

Criteria for Harris corner detector
Aperture problem

Corners

Lines

Flat regions
$$\sum \nabla I (\nabla I)^T$$
- large gradients, all the same
- large $\lambda_1$, small $\lambda_2$
Low Texture Region

\[ \sum \nabla I (\nabla I)^T \]
- gradients have small magnitude
- small \( \lambda_1 \), small \( \lambda_2 \)
\[ \sum \nabla I (\nabla I)^T \]

- gradients are different, large magnitudes
- large \( \lambda_1 \), large \( \lambda_2 \)
Errors in Lukas-Kanade

• What are the potential causes of errors in this procedure?
  – Suppose $A^TA$ is easily invertible
  – Suppose there is not much noise in the image

When our assumptions are violated

• Brightness constancy is not satisfied
• The motion is not small
• A point does not move like its neighbors
  – window size is too large
  – what is the ideal window size?
Dealing with larger movements: Iterative refinement

1. Initialize \((x', y') = (x, y)\)

2. Compute \((u, v)\) by

   \[
   \begin{bmatrix}
   \sum I_x I_x & \sum I_x I_y \\
   \sum I_x I_y & \sum I_y I_y \\
   \end{bmatrix}
   \begin{bmatrix}
   u \\
   v \\
   \end{bmatrix}
   = -
   \begin{bmatrix}
   \sum I_x I_t \\
   \sum I_y I_t \\
   \end{bmatrix}
   \]

   2\textsuperscript{nd} moment matrix for feature patch in first image

3. Shift window by \((u, v)\):

   \[
   x' = x' + u; \quad y' = y' + v; \\
   \]

4. Recalculate \(I_t\)

5. Repeat steps 2-4 until small change

   - Use interpolation for subpixel values
Revisiting the small motion assumption

- Is this motion small enough?
  - Probably not—it’s much larger than one pixel (2\text{nd} order terms dominate)
  - How might we solve this problem?
Reduce the resolution!
Coarse-to-fine optical flow estimation

Gaussian pyramid of image 1 (t)  run iterative L-K  warp & upsample

Gaussian pyramid of image 2 (t+1)  run iterative L-K  ...

image 1

image 2
A Few Details

• **Top Level**
  – Apply L-K to get a flow field representing the flow from the first frame to the second frame.
  – Apply this flow field to warp the first frame toward the second frame.
  – Rerun L-K on the new warped image to get a flow field from it to the second frame.
  – Repeat till convergence.

• **Next Level**
  – Upsample the flow field to the next level as the first guess of the flow at that level.
  – Apply this flow field to warp the first frame toward the second frame.
  – Rerun L-K and warping till convergence as above.

• **Etc.**
Coarse-to-fine optical flow estimation

Gaussian pyramid of image 1

image 1

Gaussian pyramid of image 2

image 2

$u=1.25$ pixels

$u=2.5$ pixels

$u=5$ pixels

$u=10$ pixels
The Flower Garden Video

What should the optical flow be?
Optical Flow Results

Lucas-Kanade without pyramids

Fails in areas of large motion

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003
Optical Flow Results

Lucas-Kanade with Pyramids

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003
Flow quality evaluation
Flow quality evaluation
Flow quality evaluation

• Middlebury flow page
  – http://vision.middlebury.edu/flow/

Ground Truth

Color encoding of flow vectors
Flow quality evaluation

- Middlebury flow page
  - [http://vision.middlebury.edu/flow/](http://vision.middlebury.edu/flow/)

Lucas-Kanade flow

Ground Truth

Color encoding of flow vectors
Flow quality evaluation

- Middlebury flow page
  - [http://vision.middlebury.edu/flow/](http://vision.middlebury.edu/flow/)

**Best-in-class alg**

**Ground Truth**

**Color encoding of flow vectors**
Video stabilization
Video denoising
Video super resolution

Low-Res
Robust Visual Motion Analysis:
Piecewise-Smooth Optical Flow

Ming Ye
Electrical Engineering
University of Washington
Problem Statement:

Assuming only brightness conservation and piecewise-smooth motion, find the optical flow to best describe the intensity change in three frames.
Approach: Matching-Based Global Optimization

• Step 1. Robust local gradient-based method for high-quality initial flow estimate.

• Step 2. Global gradient-based method to improve the flow-field coherence.

• Step 3. Global matching that minimizes energy by a greedy approach.
TT: Translating Tree

150x150 (Barron 94)

|   | $e_\angle (\degree)$ | $e_{|\varepsilon|} (\text{pix})$ | $\bar{e} (\text{pix})$ |
|---|------------------|------------------------|-----------------|
| BA | 2.60             | 0.128                   | 0.0724          |
| S3 | 0.248            | 0.0167                  | 0.00984         |

e: error in pixels, cdf: cumulative distribution function for all pixels
DT: Diverging Tree

150x150 (Barron 94)

<table>
<thead>
<tr>
<th></th>
<th>$e_\angle$ (°)</th>
<th>$e_{\text{pix}}$</th>
<th>$\bar{e}_{\text{pix}}$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>6.36</td>
<td>0.182</td>
<td>0.114</td>
</tr>
<tr>
<td>S3</td>
<td>2.60</td>
<td>0.0813</td>
<td>0.0507</td>
</tr>
</tbody>
</table>
YOS: Yosemite Fly-Through

316x252 (Barron, cloud excluded)

<table>
<thead>
<tr>
<th></th>
<th>$e_\angle(\degree)$</th>
<th>$e_{|}(\text{pix})$</th>
<th>$\bar{e}(\text{pix})$</th>
</tr>
</thead>
<tbody>
<tr>
<td>BA</td>
<td>2.71</td>
<td>0.185</td>
<td>0.118</td>
</tr>
<tr>
<td>S3</td>
<td>1.92</td>
<td>0.120</td>
<td>0.0776</td>
</tr>
</tbody>
</table>
TAXI: Hamburg Taxi

256x190, (Barron 94)
max speed 3.0 pix/frame

LMS

BA

Ours

Error map

Smoothness error
Traffic

512x512 (Nagel)
max speed: 6.0 pix/frame

Ours

Error map

Smoothness error
FG: Flower Garden

360x240 (Black)
Max speed: 7pix/frame

BA
LMS

Ours
Error map
Smoothness error
Summary

• Major contributions from Lucas, Tomasi, Kanade
  – Tracking feature points
  – Optical flow
  – Stereo
  – Structure from motion

• Key ideas
  – By assuming brightness constancy, truncated Taylor expansion leads to simple and fast patch matching across frames
  – Coarse-to-fine registration
  – Global approach by former EE student Ming Ye