Computer Vision
CSCI 4420

Instructor: Faisal Qureshi

Slides Credit: These slides borrow heavily from on-line resources, particularly from other similar computer vision courses at Brown University, the University of Washington and Stanford.
Feature Tracking
Previously

• Recognition
  – Bag of words models
  – Interest points
  – Instance level recognition
  – Sliding window detectors
  – Geometry recognition / context
This class: 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

- Many problems, such as structure from motion require matching points
- If motion is small, tracking is an easy way to get them
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

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

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

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

**Difference over frames**

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

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

Hence,

\[I_x \cdot u + I_y \cdot v + I_t \approx 0 \quad \rightarrow \nabla I \cdot [u \ v]^T + I_t = 0\]
How does this make sense?

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

• What do the static image gradients have to do with motion estimation?
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:

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

\[
A \ d = b
\]

25x2  2x1  25x1
Matching patches across images

- Overconstrained linear system

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

Least squares solution for \(d\) given by

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

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

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

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

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

\[
\begin{bmatrix}
\sum I_xI_x & \sum I_xI_y \\
\sum I_xI_y & \sum I_yI_y
\end{bmatrix}
\begin{bmatrix}
u \\
v
\end{bmatrix} = -
\begin{bmatrix}
\sum I_xI_t \\
\sum I_yI_t
\end{bmatrix}
\]

\[A^TA\]

\[A^Tb\]

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

- \(A^TA\) should be invertible
- \(A^TA\) should not be too small due to noise
  - eigenvalues \(\lambda_1\) and \(\lambda_2\) of \(A^TA\) should not be too small
- \(A^TA\) 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
$M = A^T A$ is the second moment matrix!
(Harris corner detector...)

$$A^T A = \left[ \begin{array}{cc} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{array} \right] = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \ I_y] = \sum \nabla I (\nabla I)^T$$

- Eigenvectors and eigenvalues of $A^T A$ relate to edge direction and magnitude
  - The eigenvector associated with the larger eigenvalue points in the direction of fastest intensity change
  - The other eigenvector is orthogonal to it
Low-texture region

\[ \sum \nabla I (\nabla I)^T \]

- gradients have small magnitude
- small \( \lambda_1 \), small \( \lambda_2 \)
Edge

\[
\sum \nabla I (\nabla I)^T
\]

- gradients very large or very small
- large \( \lambda_1 \), small \( \lambda_2 \)
High-texture region

\[
\sum \nabla I (\nabla I)^T
\]

- gradients are different, large magnitudes
- large \( \lambda_1 \), large \( \lambda_2 \)
The aperture problem resolved
The aperture problem resolved
Dealing with larger movements: Iterative refinement

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

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

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

3. Shift window by \((u, v)\): \(x' = x' + u; y' = y' + v;\)

4. Recalculate \(I_t\)

5. Repeat steps 2-4 until small change
   - Use interpolation for subpixel values
Dealing with larger movements: coarse-to-fine registration

Gaussian pyramid of image 1 (t)

run iterative L-K

upsample

run iterative L-K

Gaussian pyramid of image 2 (t+1)
Shi-Tomasi feature tracker

• Find good features using eigenvalues of second-moment matrix (e.g., Harris detector or threshold on the smallest eigenvalue)
  – Key idea: “good” features to track are the ones whose motion can be estimated reliably

• Track from frame to frame with Lucas-Kanade
  – This amounts to assuming a translation model for frame-to-frame feature movement

• Check consistency of tracks by affine registration to the first observed instance of the feature
  – Affine model is more accurate for larger displacements
  – Comparing to the first frame helps to minimize drift

Tracking example

Figure 1: Three frame details from Woody Allen’s *Manhattan*. The details are from the 1st, 11th, and 21st frames of a subsequence from the movie.

Figure 2: The traffic sign windows from frames 1, 6, 11, 16, 21 as tracked (top), and warped by the computed deformation matrices (bottom).

Summary of KLT tracking

• Find a good point to track (harris corner)

• Use intensity second moment matrix and difference across frames to find displacement

• Iterate and use coarse-to-fine search to deal with larger movements

• When creating long tracks, check appearance of registered patch against appearance of initial patch to find points that have drifted
Implementation issues

• Window size
  – Small window more sensitive to noise and may miss larger motions (without pyramid)
  – Large window more likely to cross an occlusion boundary (and it’s slower)
  – 15x15 to 31x31 seems typical

• Weighting the window
  – Common to apply weights so that center matters more (e.g., with Gaussian)
Optical flow

Vector field function of the spatio-temporal image brightness variations

Picture courtesy of Selim Temizer - Learning and Intelligent Systems (LIS) Group, MIT
Motion and perceptual organization

- Sometimes, motion is the only cue
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

Uses of motion

- Estimating 3D structure
- Segmenting objects based on motion cues
- Learning and tracking dynamical models
- Recognizing events and activities
- Improving video quality (motion stabilization)
Motion field

- The motion field is the projection of the 3D scene motion into the image.

What would the motion field of a non-rotating ball moving towards the camera look like?
Optical flow

• Definition: optical flow is the *apparent* motion of brightness patterns in the image

• Ideally, optical flow would be the same as the motion field

• Have to be careful: apparent motion can be caused by lighting changes without any actual motion
  – Think of a uniform rotating sphere under fixed lighting vs. a stationary sphere under moving illumination
Lucas-Kanade Optical Flow

• Same as Lucas-Kanade feature tracking, but for each pixel
  – As we saw, works better for textured pixels

• Operations can be done one frame at a time, rather than pixel by pixel
  – Efficient
Multi-resolution Lucas Kanade Algorithm

- Compute ‘simple’ LK at highest level
- At level $i$
  - Take flow $u_{i-1}, v_{i-1}$ from level $i-1$
  - Bilinear interpolate it to create $u_i^*, v_i^*$ matrices of twice resolution for level $i$
  - Multiply $u_i^*, v_i^*$ by 2
  - Compute $f_i$ from a block displaced by $u_i^*(x,y), v_i^*(x,y)$
  - Apply LK to get $u_i'(x,y), v_i'(x,y)$ (the correction in flow)
  - Add corrections $u_i' v_i'$, i.e. $u_i = u_i^* + u_i'$, $v_i = v_i^* + v_i'$. 
Iterative Refinement

• Iterative Lukas-Kanade Algorithm
  1. Estimate displacement at each pixel by solving Lucas-Kanade equations
  2. Warp $I(t)$ towards $I(t+1)$ using the estimated flow field
     - Basically, just interpolation
  3. Repeat until convergence

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003
Coarse-to-fine optical flow estimation

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

Gaussian pyramid of image 2 (t+1)
Coarse-to-fine optical flow estimation

Gaussian pyramid of image 1
- $u=1.25$ pixels
- $u=2.5$ pixels
- $u=5$ pixels
- $u=10$ pixels

Gaussian pyramid of image 2
Example
Multi-resolution registration

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003
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
Errors in Lucas-Kanade

• The motion is large
  – Possible Fix: Keypoint matching

• A point does not move like its neighbors
  – Possible Fix: Region-based matching

• Brightness constancy does not hold
  – Possible Fix: Gradient constancy
State-of-the-art optical flow

Start with something similar to Lucas-Kanade
+ gradient constancy
+ energy minimization with smoothing term
+ region matching
+ keypoint matching (long-range)

Large displacement optical flow, Brox et al., CVPR 2009
Stereo vs. Optical Flow

- Similar dense matching procedures
- Why don’t we typically use epipolar constraints for optical flow?

Summary

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

• Key ideas
  – By assuming brightness constancy, truncated Taylor expansion leads to simple and fast patch matching across frames
  – Coarse-to-fine registration