Motion and optical flow

Thursday, Nov 19
Kristen Grauman
UT-Austin

Many slides adapted from S. Seitz, R. Szeliski, M. Pollefeys, S. Lazebnik

Today

• Introduction to motion
• Optical flow
So far: Features and filters

Transforming and describing images; textures and colors

So far: Grouping

Parallelism
Symmetry
Continuity
Closure

Clustering, segmentation, fitting; what parts belong together?
So far: Multiple views

Multi-view geometry and matching, stereo

So far: Recognition and learning

Shape matching, recognizing objects and categories, learning techniques
Finally: Motion and tracking

Tracking objects, video analysis, low level motion

Video

- A video is a sequence of frames captured over time
- Now our image data is a function of space (x, y) and time (t)
Applications of segmentation to video

• Background subtraction
  • A static camera is observing a scene
  • Goal: separate the static background from the moving foreground

Applications of segmentation to video

• Background subtraction
• Shot boundary detection
  • Commercial video is usually composed of shots or sequences showing the same objects or scene
  • Goal: segment video into shots for summarization and browsing (each shot can be represented by a single keyframe in a user interface)
  • Difference from background subtraction: the camera is not necessarily stationary
Applications of segmentation to video

- Background subtraction
- Shot boundary detection
  - For each frame
    - Compute the distance between the current frame and the previous one
      » Pixel-by-pixel differences
      » Differences of color histograms
      » Block comparison
    - If the distance is greater than some threshold, classify the frame as a shot boundary

- Motion segmentation
  - Segment the video into multiple coherently moving objects
Motion and perceptual organization

- Sometimes, motion is the only cue

Motion and perceptual organization

- Sometimes, motion is foremost cue
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 dynamical models
- Recognizing events and activities
- Improving video quality (motion stabilization)

Today

- Introduction to motion
- Optical flow
Motion field

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

Motion parallax

http://psych.hanover.edu/KRANTZ/MotionParallax/MotionParallax.html
Motion field + camera motion

Figure 1.2: Two images taken from a helicopter flying through a canyon and the computed optical flow field.

Figure from Michael Black, Ph.D. Thesis

Length of flow vectors inversely proportional to depth Z of 3d point

points closer to the camera move more quickly across the image plane

Motion field + camera motion

Zoom out  Zoom in  Pan right to left
Motion estimation techniques

• Direct methods
  • Directly recover image motion at each pixel from spatio-temporal image brightness variations
  • Dense motion fields, but sensitive to appearance variations
  • Suitable for video and when image motion is small

• Feature-based methods
  • Extract visual features (corners, textured areas) and track them over multiple frames
  • Sparse motion fields, but more robust tracking
  • Suitable when image motion is large (10s of pixels)

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
Apparent motion ≠ motion field

Problem definition: optical flow

How to estimate pixel motion from image H to image I?
- Solve pixel correspondence problem
  - given a pixel in H, look for nearby pixels of the same color in I

Key assumptions
- **color constancy**: a point in H looks the same in I
  - For grayscale images, this is brightness constancy
- **small motion**: points do not move very far

This is called the **optical flow** problem
Brightness constancy

Let’s look at these constraints more closely

- brightness constancy: Q: what’s the equation?
  \[ H(x, y) = I(x + u, y + v) \]

- small motion: (u and v are less than 1 pixel)
  \[ I(x+u, y+v) = I(x, y) + \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v + \text{higher order terms} \]
  \[ \approx I(x, y) + \frac{\partial I}{\partial x} u + \frac{\partial I}{\partial y} v \]

Optical flow constraints (grayscale images)
Optical flow equation

Combining these two equations

\[ 0 = I(x + u, y + v) - H(x, y) \]
\[ \approx I(x, y) + I_x u + I_y v - H(x, y) \]
\[ \approx (I(x, y) - H(x, y)) + I_x u + I_y v \]
\[ \approx I_t + I_x u + I_y v \]
\[ \approx I_t + \nabla I \cdot [u \ v] \]

shorthand: \( I_x = \frac{\partial I}{\partial x} \)

Optical flow equation

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

Q: how many unknowns and equations per pixel?

Intuitively, what does this ambiguity mean?
The aperture problem

Perceived motion

The aperture problem

Actual motion
The barber pole illusion

http://en.wikipedia.org/wiki/Barberpole_illusion

http://www.sandlotscience.com/Ambiguous/Barberpole_Illusion.htm
Solving the aperture problem (grayscale image)

- How to get more equations for a pixel?
- **Spatial coherence constraint:** pretend the pixel's neighbors have the same (u,v)
  - If we use a 5x5 window, that gives us 25 equations per pixel
    \[ 0 = I_t(p_i) + \nabla I(p_i) \cdot [u \ v] \]
    
    \[
    \begin{bmatrix}
    I_x(p_1) & I_y(p_1) \\
    I_x(p_2) & I_y(p_2) \\
    \vdots & \vdots \\
    I_x(p_{25}) & I_y(p_{25})
    \end{bmatrix}
    \begin{bmatrix}
    u \\
    v
    \end{bmatrix}
    =
    \begin{bmatrix}
    I_t(p_1) \\
    I_t(p_2) \\
    \vdots \\
    I_t(p_{25})
    \end{bmatrix}
    \]

\[
A d = b
\]

\[25 \times 2 \quad 2 \times 1 \quad 25 \times 1\]

Slide credit: Steve Seitz

---

Solving the aperture problem

Prob: we have more equations than unknowns

\[
A \ 25 \times 2 \ 2 \times 1 \ 25 \times 1 \quad d = b \rightarrow \text{minimize } ||Ad - b||^2
\]

Solution: solve least squares problem

- minimum least squares solution given by solution (in d) of:
  
  \[
  (A^T A) \ 2 \times 2 \ 2 \times 1 \quad d = A^T b
  \]

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

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

- The summations are over all pixels in the K x K window
- This technique was first proposed by Lucas & Kanade (1981)

Slide credit: Steve Seitz
Conditions for solvability

When is this solvable?

- $\mathbf{A}^T \mathbf{A}$ should be invertible
- $\mathbf{A}^T \mathbf{A}$ should not be too small
  - eigenvalues $\lambda_1$ and $\lambda_2$ of $\mathbf{A}^T \mathbf{A}$ should not be too small
- $\mathbf{A}^T \mathbf{A}$ should be well-conditioned
  - $\lambda_1 / \lambda_2$ should not be too large ($\lambda_1$ = larger eigenvalue)

Edge

- gradients very large or very small
- large $\lambda_1$, small $\lambda_2$
Low-texture region

- gradients have small magnitude
- small $\lambda_1$, small $\lambda_2$

High-texture region

- gradients are different, large magnitudes
- large $\lambda_1$, large $\lambda_2$
### Motion vs. Stereo: Similarities

- Both involve solving
  - Correspondence: disparities, motion vectors
  - Reconstruction

### Motion vs. Stereo: Differences

- **Motion:**
  - Uses velocity: consecutive frames must be close to get good approximate time derivative
  - 3d movement between camera and scene not necessarily single 3d rigid transformation
- **Whereas with stereo:**
  - Could have any disparity value
  - View pair separated by a single 3d transformation
Using optical flow: recognizing facial expressions

Recognizing Human Facial Expression (1994)
by Yaser Yacoob, Larry S. Davis
Using optical flow: action recognition at a distance

- Features = optical flow within a region of interest
- Classifier = nearest neighbors

Challenge: low-res data, not going to be able to track each limb.

[Efros, Berg, Mori, & Malik 2003]
http://graphics.cs.cmu.edu/people/efros/research/action/

Using optical flow: action recognition at a distance

Correlation-based tracking
Extract person-centered frame window
Using optical flow: action recognition at a distance

Extract optical flow to describe the region’s motion.

Using optical flow: action recognition at a distance

Input Sequence

Matched Frames

Use nearest neighbor classifier to name the actions occurring in new video frames.
Using optical flow: action recognition at a distance

Use nearest neighbor classifier to name the actions occurring in new video frames.

Do as I do: motion retargeting

- Include constraint for similarity within sequence as well as across sequences
Summary

• Motion field: 3d motions projected to 2d images; dependency on depth

• Solving for motion with
  – dense optical flow

• Optical flow
  – Brightness constancy assumption
  – Aperture problem
  – Solution with spatial coherence assumption