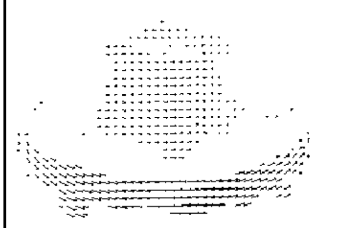


Motion and optical flow

Thursday, Nov 19

Kristen Grauman

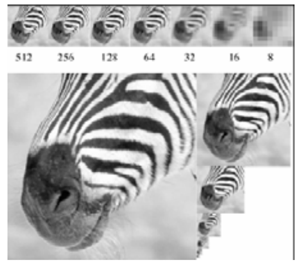
UT-Austin

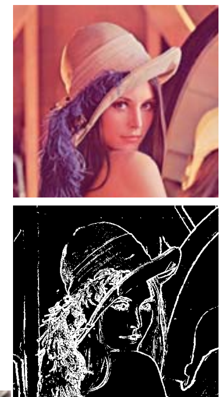
Many slides adapted from S. Seitz, R. Szeliski, M. Pollefeys, S. Lazebnik

Today

- Introduction to motion
- Optical flow

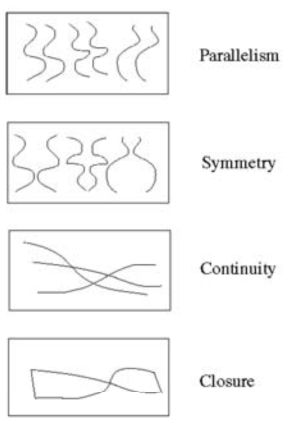
So far: Features and filters

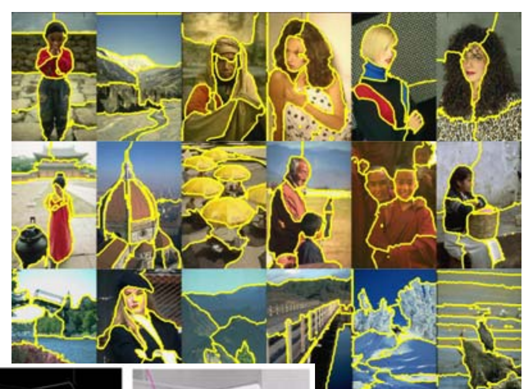




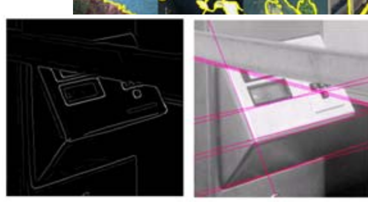
Transforming and describing images; textures and colors

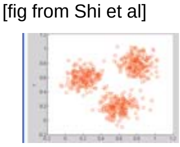
So far: Grouping





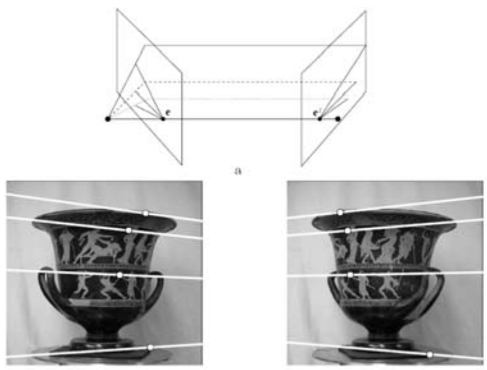
[fig from Shi et al]



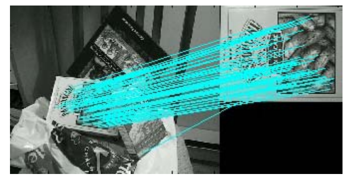


Clustering, segmentation, fitting; what parts belong together?

So far: Multiple views

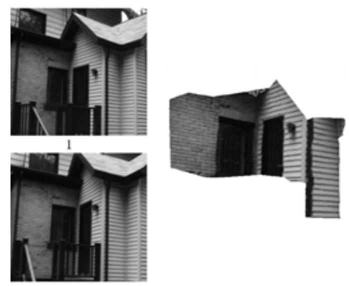


Hartley and Zisserman



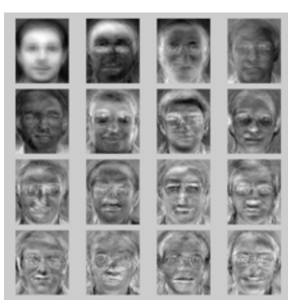
Lowe

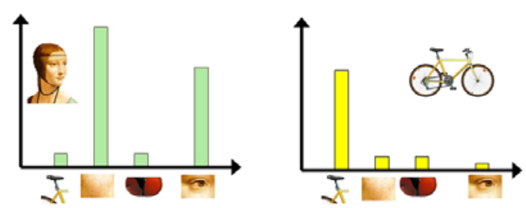
Multi-view geometry and matching, stereo



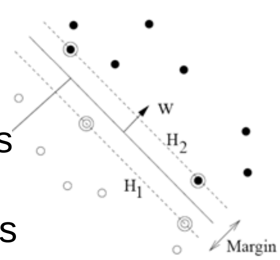
Tomasi and Kanade

So far: Recognition and learning



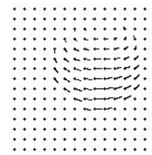


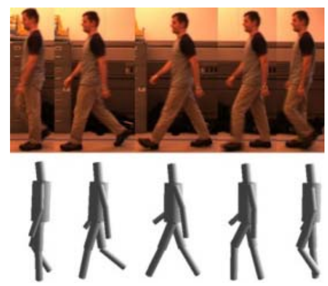
Shape matching, recognizing objects and categories, learning techniques



Finally: Motion and tracking

Tracking objects, video analysis, low level motion

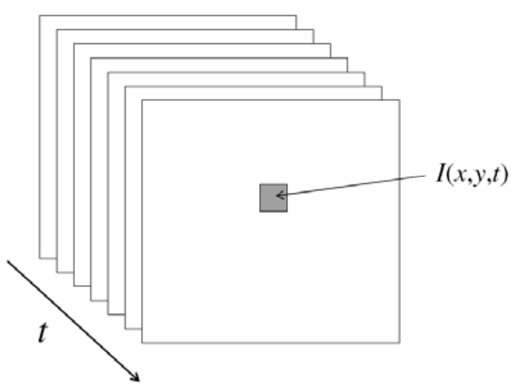




Tomas Izo

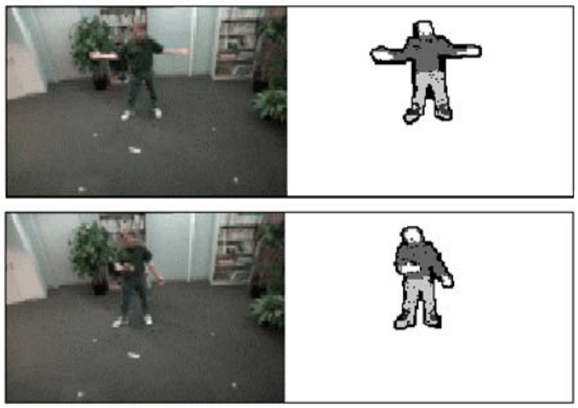
Video

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



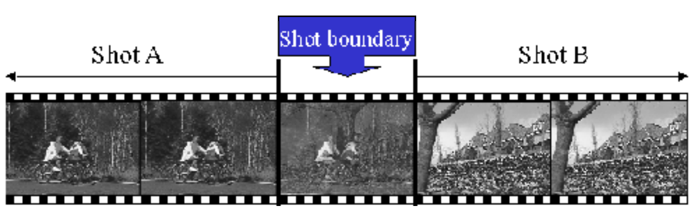
Applications of segmentation to video

- 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

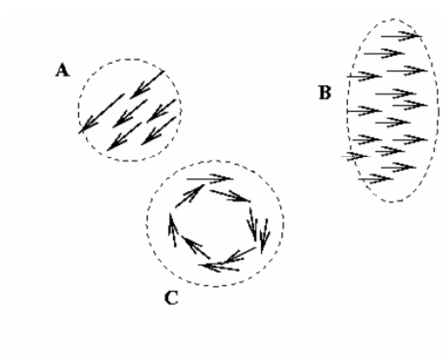


Applications of segmentation to video

- 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

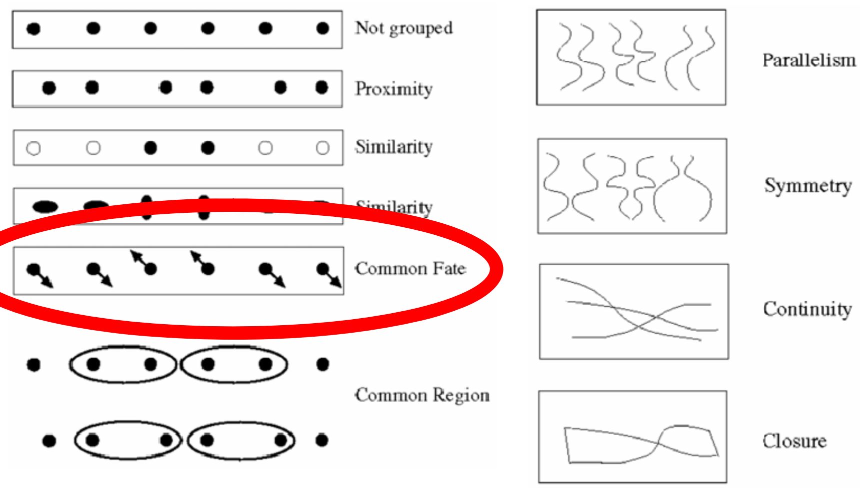
Applications of segmentation to video

- Background subtraction
- Shot boundary detection
- Motion segmentation
  - Segment the video into multiple *coherently* moving objects



Motion and perceptual organization

- Sometimes, motion is the only cue

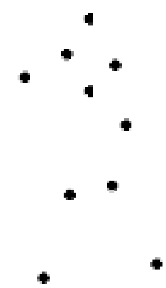


Motion and perceptual organization

- Sometimes, motion is foremost cue

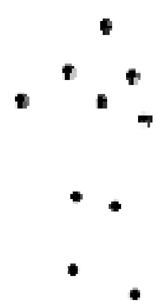
Motion and perceptual organization

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



Motion and perceptual organization

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



Uses of motion

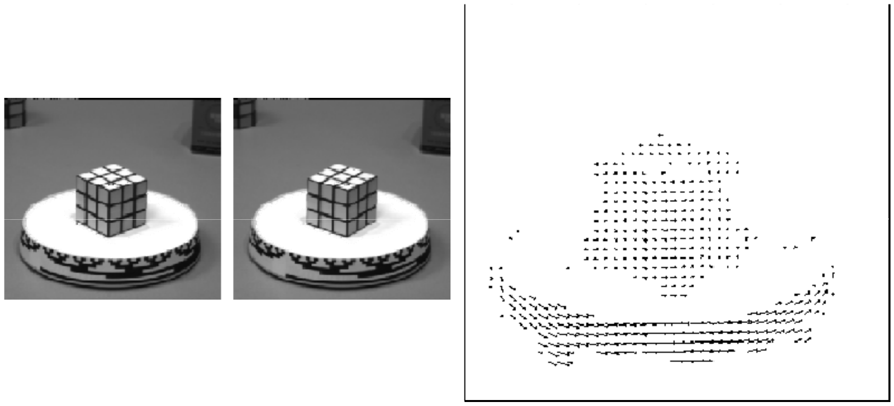
- Estimating 3D structure
- Segmenting objects based on motion cues
- Learning dynamical models
- Recognizing events and activities
- Improving video quality (motion stabilization)

Today

- Introduction to motion
- Optical flow

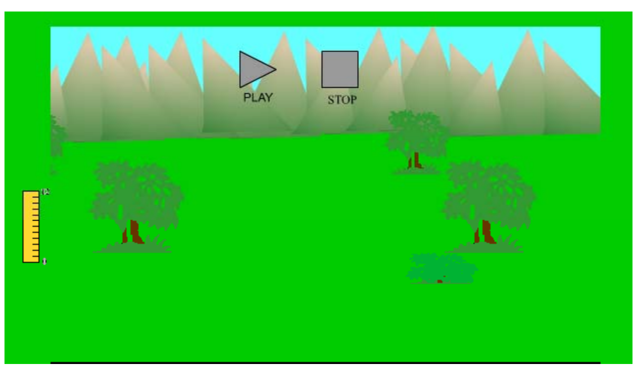
Motion field

- 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

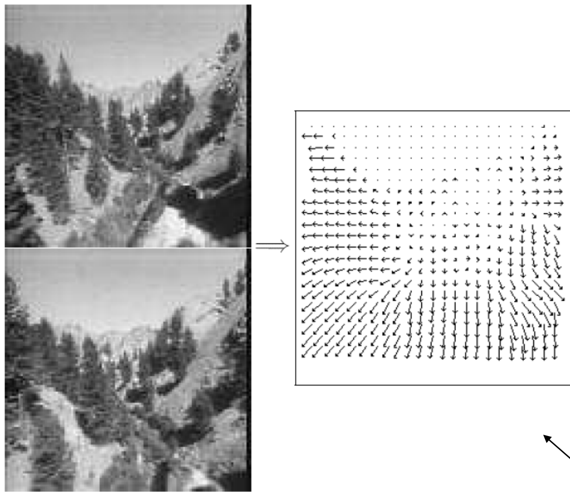


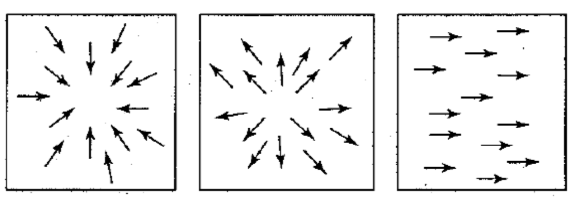
Figure 1.2: Two images taken from a helicopter flying through a canyon and the computed optical flow field.

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

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

Figure from Michael Black, Ph.D. Thesis

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

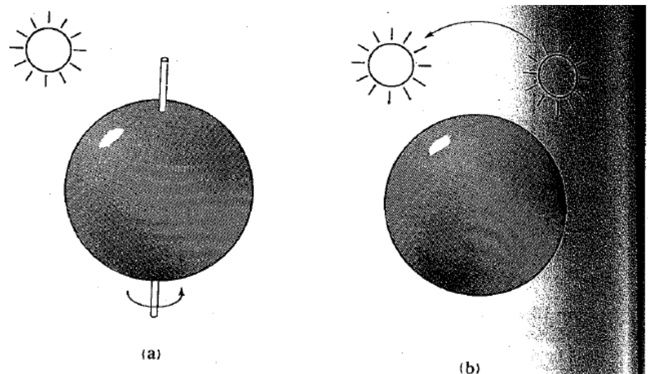
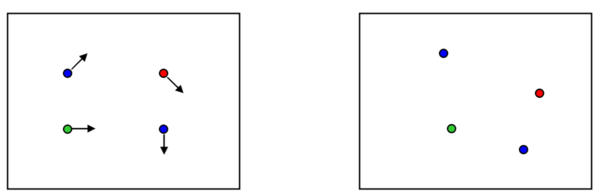


Figure 12-2. The optical flow is not always equal to the motion field. In (a) a smooth sphere is rotating under constant illumination—the image does not change, yet the motion field is nonzero. In (b) a fixed sphere is illuminated by a moving source—the shading in the image changes, yet the motion field is zero.

Figure from Horn book

Problem definition: optical flow



$H(x, y)$ $I(x, y)$

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

Slide credit: Steve Seitz

Brightness constancy

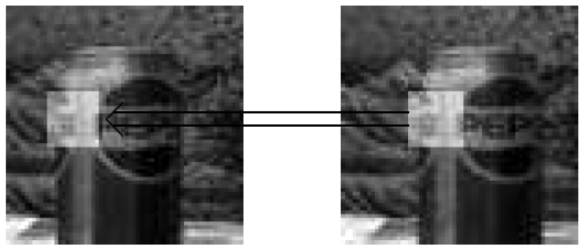
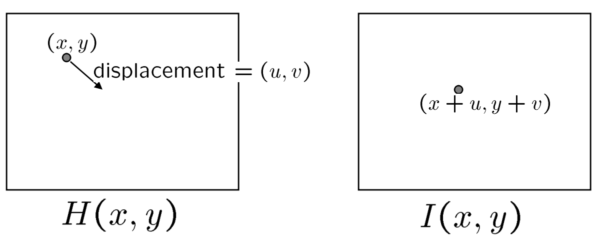


Figure 1.5: Data conservation assumption. The highlighted region in the right image looks roughly the same as the region in the left image, despite the fact that it has moved.

Figure by Michael Black

Optical flow constraints (grayscale images)



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)

$$I(x+u,y+v) = I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order terms}$$
$$\approx I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

Slide credit: Steve Seitz

Optical flow equation

Combining these two equations

shorthand: $I_x = \frac{\partial I}{\partial x}$

$$
\begin{aligned}
0 &= I(x+u, y+v) - H(x,y) \\
&\approx I(x,y) + I_x u + I_y v - H(x,y) \\
&\approx (I(x,y) - H(x,y)) + I_x u + I_y v \\
&\approx I_t + I_x u + I_y v \\
&\approx I_t + \nabla I \cdot [u \;\; v]
\end{aligned}
$$

Slide credit: Steve Seitz

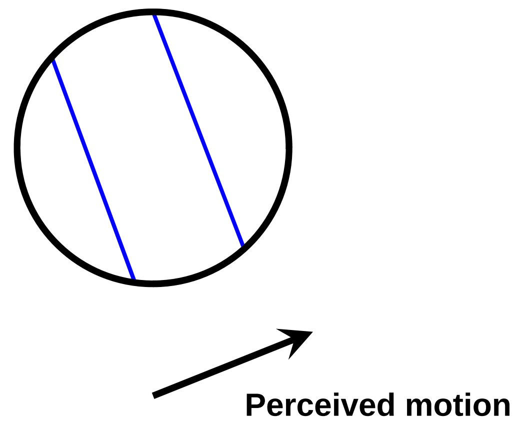
Optical flow equation

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

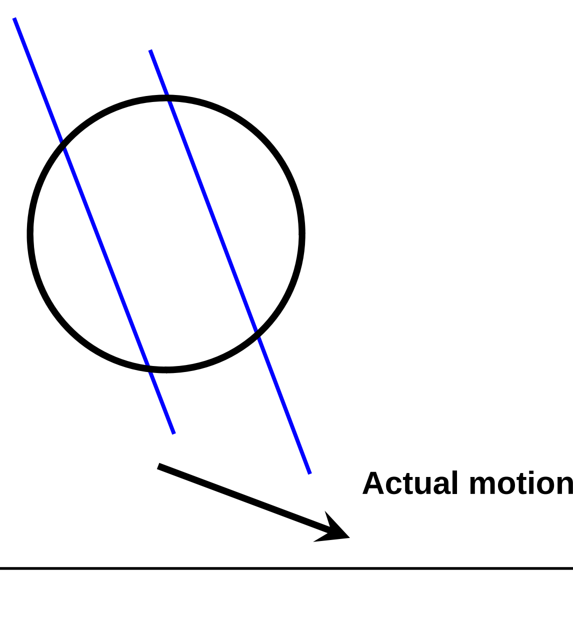
Q: how many unknowns and equations per pixel?

Intuitively, what does this ambiguity mean?

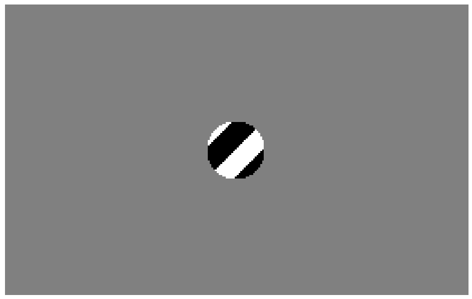
The aperture problem



The aperture problem

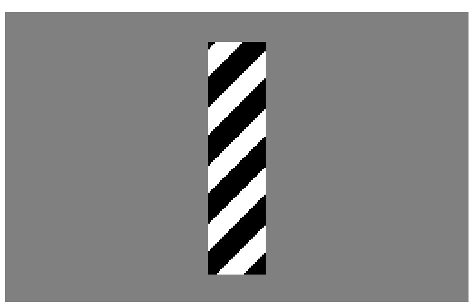


The barber pole illusion



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

The barber pole 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

$$0 = I_t(\mathbf{p_i}) + \nabla I(\mathbf{p_i}) \cdot [u \;\; v]$$

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

$$\underset{25\times2}{A} \;\; \underset{2\times1}{d} = \underset{25\times1}{b}$$

Slide credit: Steve Seitz

Solving the aperture problem

Prob: we have more equations than unknowns

$$\underset{25\times2}{A} \;\; \underset{2\times1}{d} = \underset{25\times1}{b} \longrightarrow \text{minimize } \|Ad - b\|^2$$

Solution: solve least squares problem

- minimum least squares solution given by solution (in d) of:

$$\underset{2\times2}{(A^T A)} \;\; \underset{2\times1}{d} = \underset{2\times1}{A^T b}$$

$$\underbrace{\boxed{\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix}}}_{A^T A} \begin{bmatrix} u \\ v \end{bmatrix} = - \underbrace{\begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}}_{A^T b}$$

- 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

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

$$A^T A \qquad\qquad\qquad A^T b$$

When is this solvable?

- **$A^T A$** should be invertible
- **$A^T A$** should not be too small
  - eigenvalues λ_1 and λ_2 of **$A^T A$** should not be too small
- **$A^T A$** should be well-conditioned
  - λ_1 / λ_2 should not be too large (λ_1 = larger eigenvalue)

Slide credit: Steve Seitz

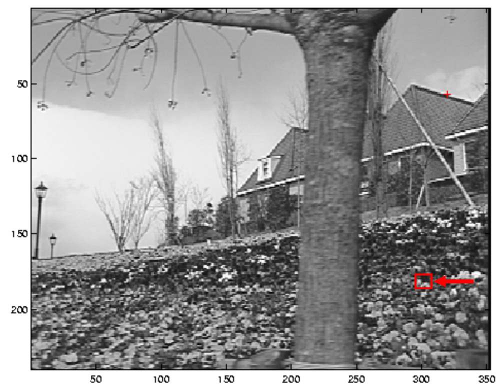
Edge

– gradients very large or very small
– large λ_1, small λ_2

Low-texture region

– gradients have small magnitude
– small λ_1, small λ_2

High-texture region



– gradients are different, large magnitudes
– large λ_1, large λ_2

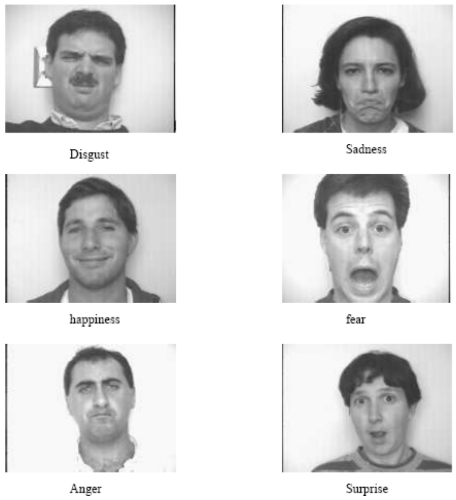
Motion vs. Stereo: Similarities

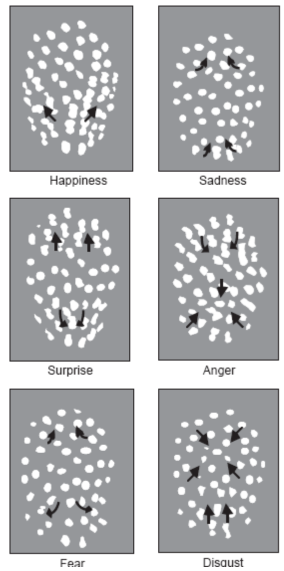
- Both involve solving
  - Correspondence: disparities, motion vectors
  - Reconstruction

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

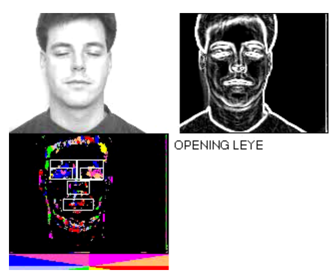
Using optical flow: recognizing facial expressions





Recognizing Human Facial Expression (1994)
by Yaser Yacoob, Larry S. Davis

Using optical flow: recognizing facial expressions



Using optical flow: action recognition at a distance

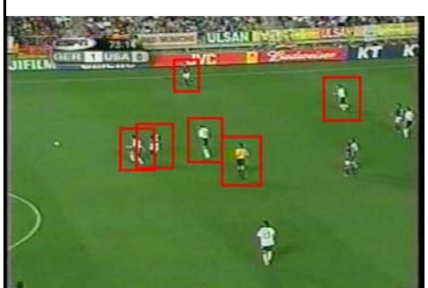
- Features = optical flow within a region of interest
- Classifier = nearest neighbors

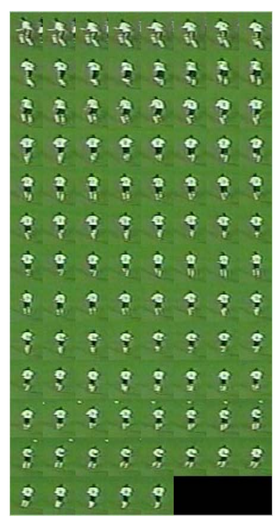
The 30-Pixel Man

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/

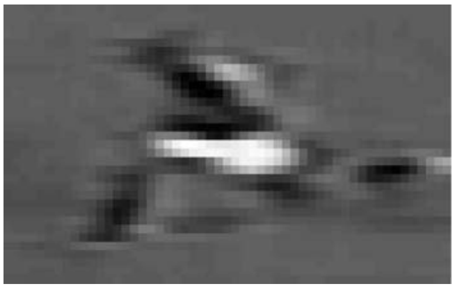
Using optical flow: action recognition at a distance





Correlation-based tracking
Extract person-centered frame window

Using optical flow: action recognition at a distance

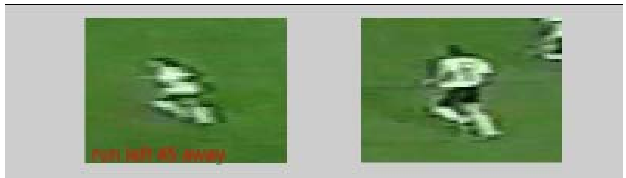


Extract optical flow to describe the region's motion.

Using optical flow: action recognition at a distance

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

Using optical flow: action recognition at a distance



Input Sequence

Matched NN Frame

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