CSE 152: Computer Vision

Hao Su

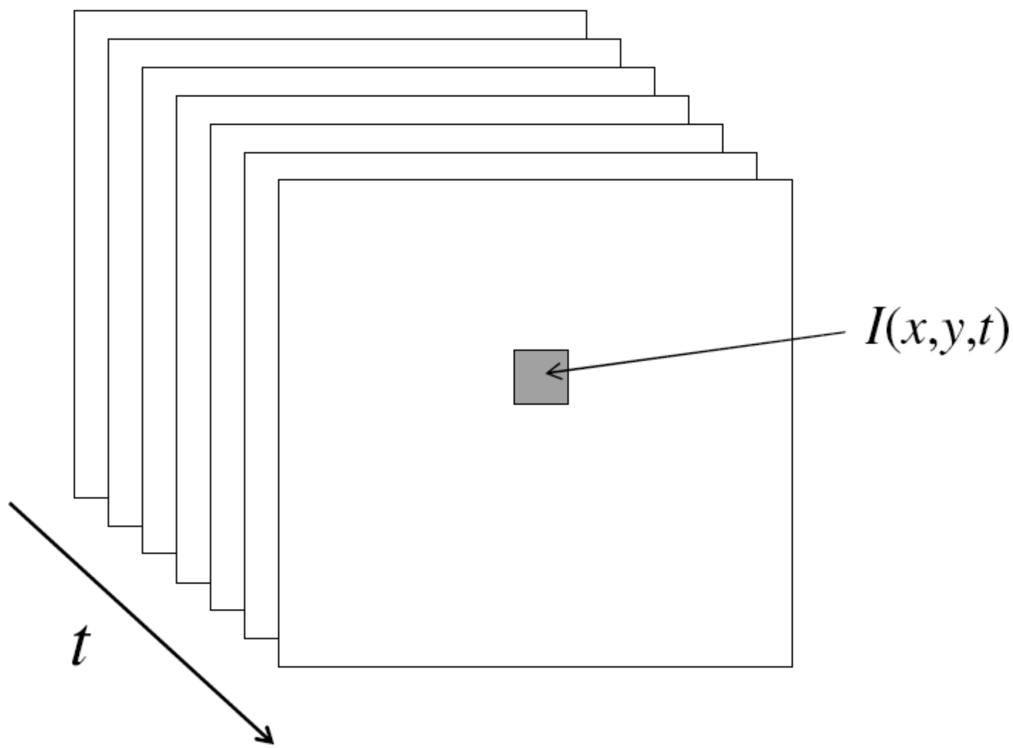
Lecture 17: Motion Estimation

Goal of Video Understanding

- Given an input video, obtain an understanding
- Involves:
  - Objects
  - Humans
  - Actions/events

Video

- A video is a sequence of frames captured over time
- A ‘function’ of space (x, y) and time (t)



Motion Cue is Important

Even “impoverished” motion data from videos can evoke a strong percept

Motion Cue is Important

Even “impoverished” motion data from videos can evoke a strong percept

Motion Cue is Important

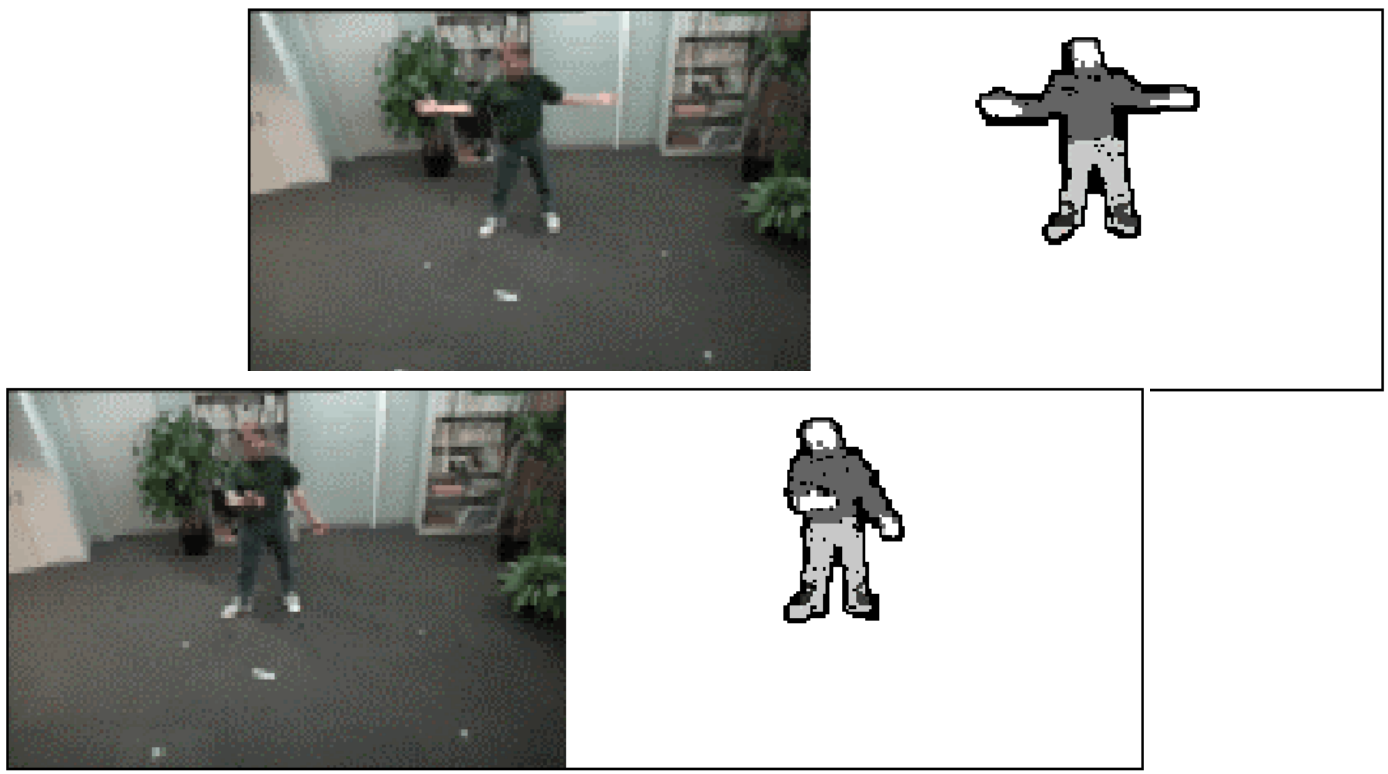
Animation from:
Heider, F. & Simmel, M. (1944).
An experimental study of apparent behavior.
American Journal of Psychology, 57, 243-259.

Courtesy of:
Department of Psychology,
University of Kansas, Lawrence

Experimental study of apparent behavior.
Fritz Heider & Marianne Simmel. 1944

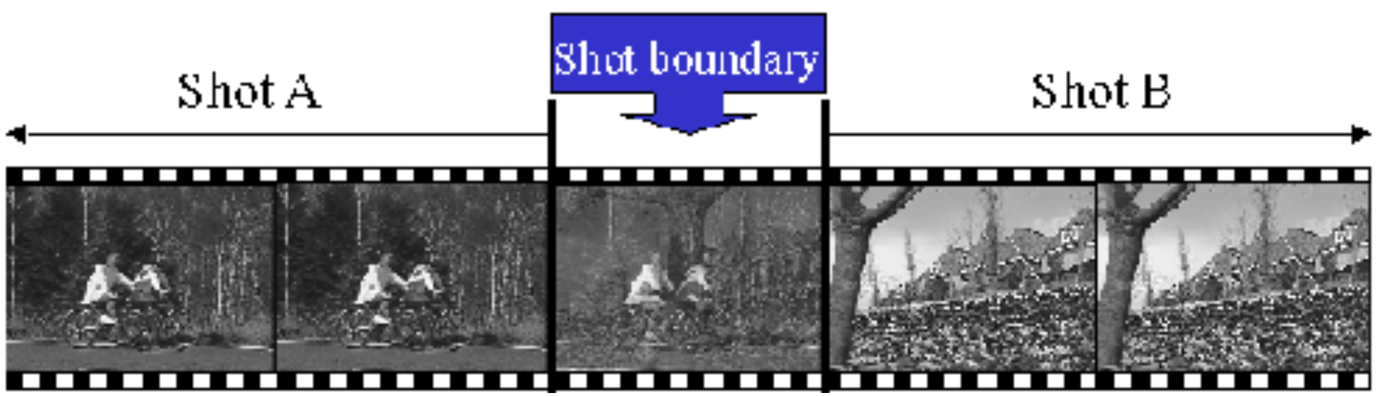
Motion Applications: Segmentation of video

- Background subtraction
  - Goal: separate the static *background* from the moving *foreground*



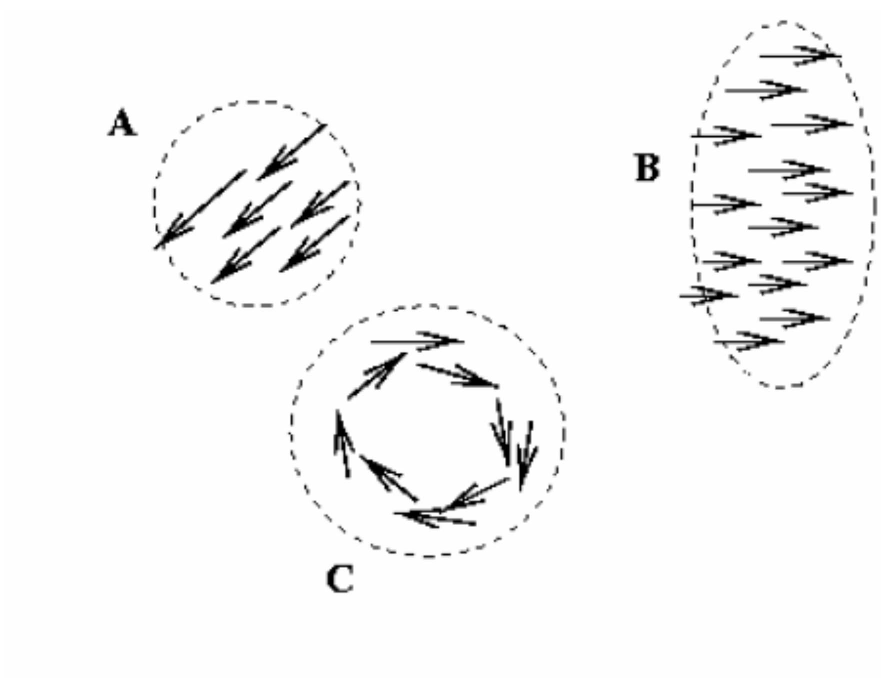
Motion Applications: Segmentation of video

- Background subtraction
- Shot boundary detection in edited video
  - Goal: segment video into shots for summarization and browsing (each shot can be represented by a single keyframe in a user interface)



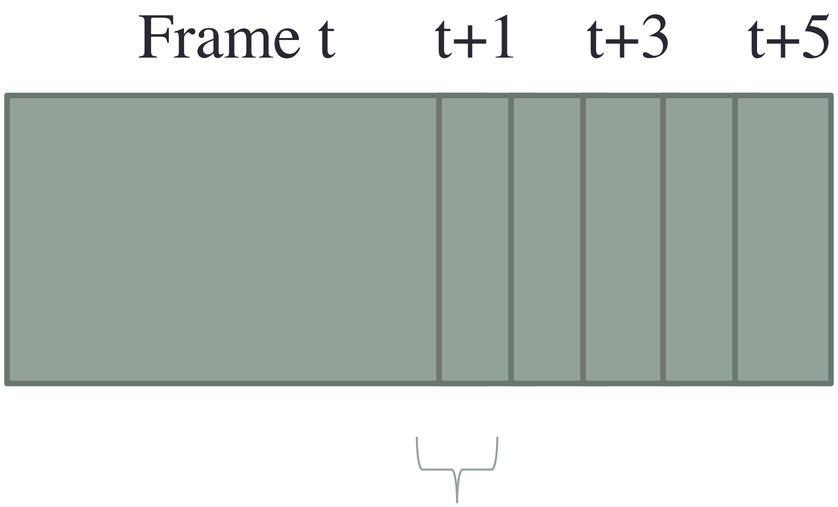
Motion Applications: Segmentation of video

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



Motion Applications: Mosaicing for Panoramas

Left to right sweep of video camera



Compare small
overlap for efficiency

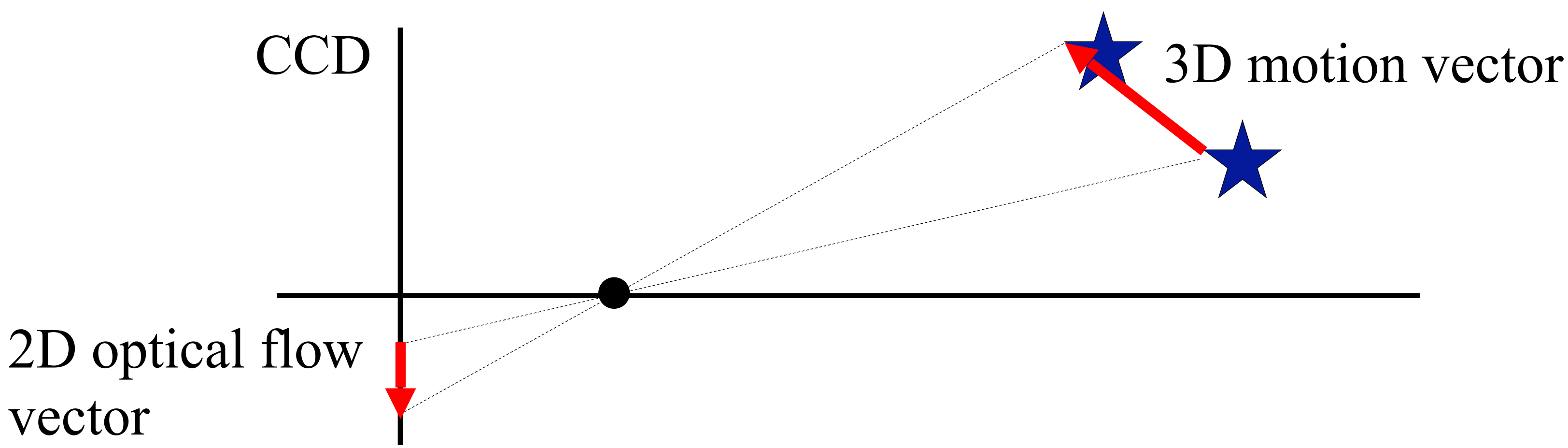
Motion Applications: Mosaicing for Panoramas

Agenda

- **Optical Flow Definition**
- Optical Flow Estimation

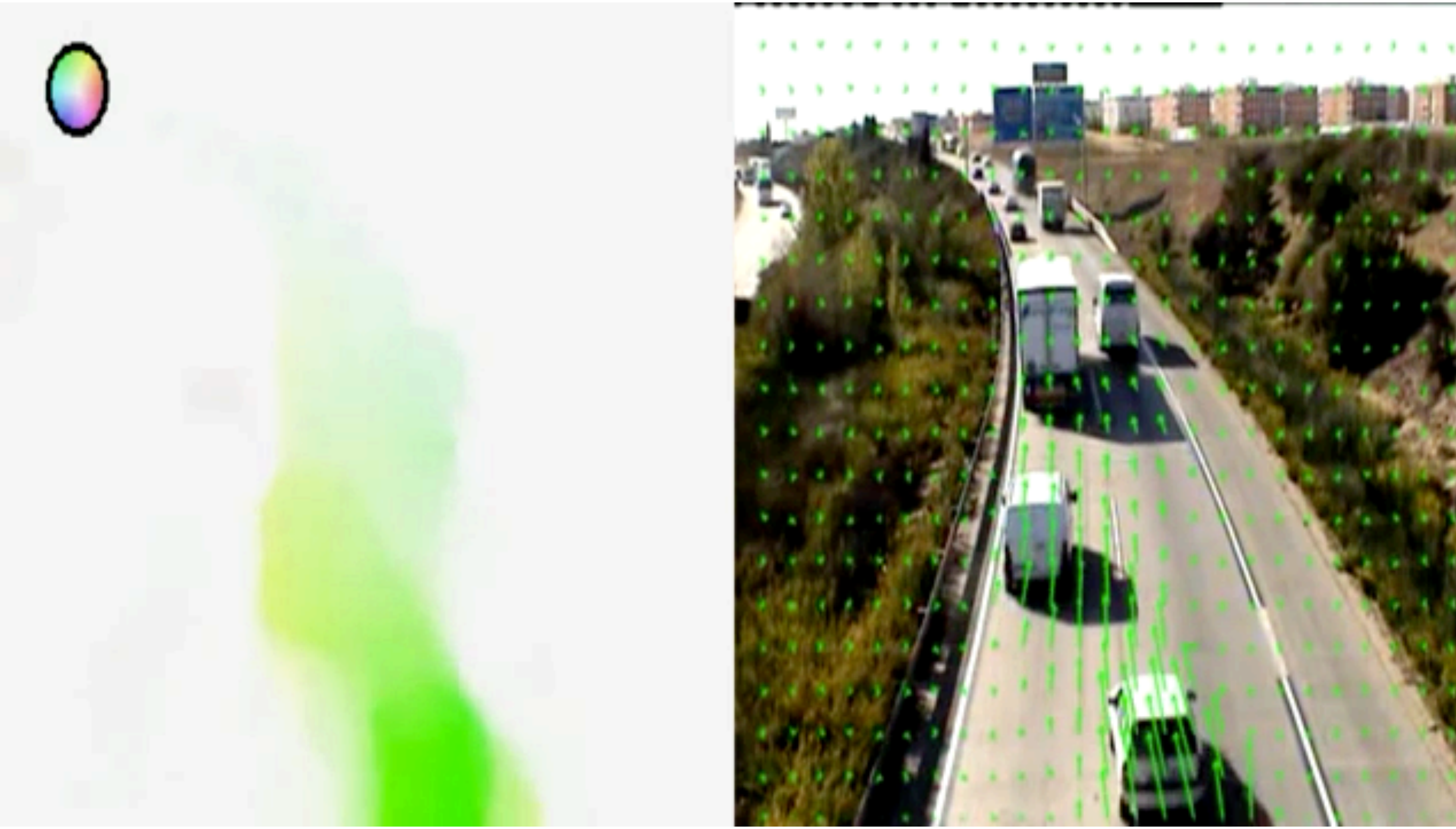
Motion Field & Optical Flow Field

- Motion Field = Real world 3D motion
- Optical Flow Field = Projection of the motion field onto the 2d image



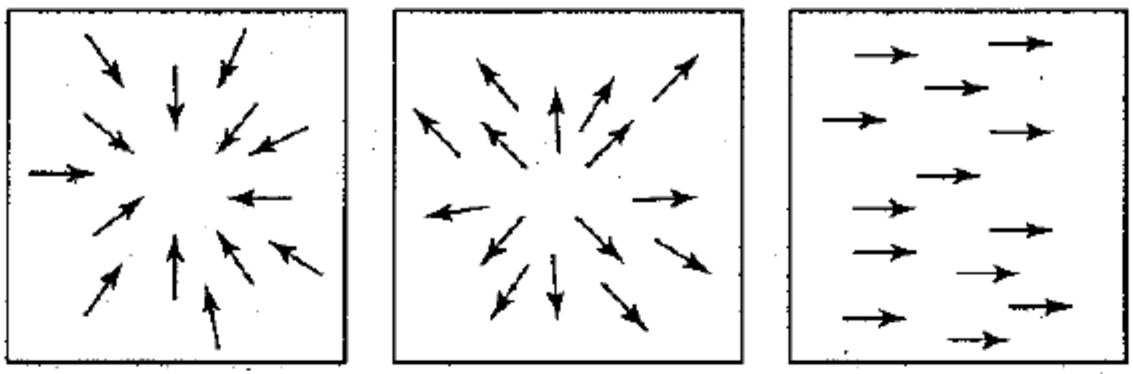
$$\vec{u} = (u, v)$$

Slide adapted from Savarese.



https://www.youtube.com/watch?v=5rR_9YIcg_s

Motion Field + Camera Motion



Zoom out Zoom in Pan right to left

Slide adapted from Savarese and Grauman.

Motion Field + Camera Motion

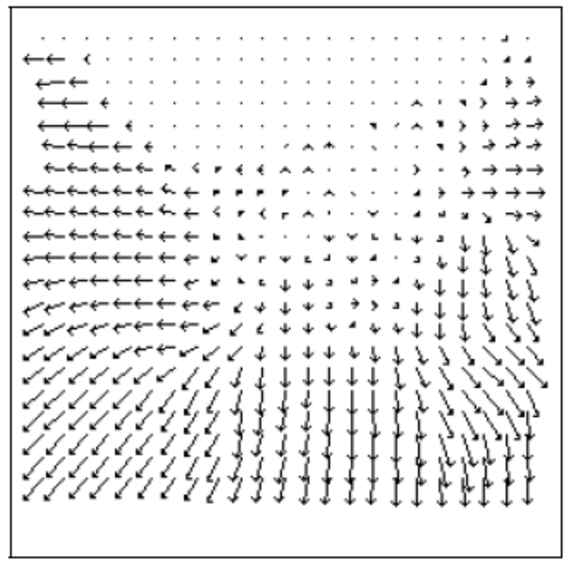


Figure 1.2: Two images taken from a helicopter flying through a canyon and the computed optical flow field.

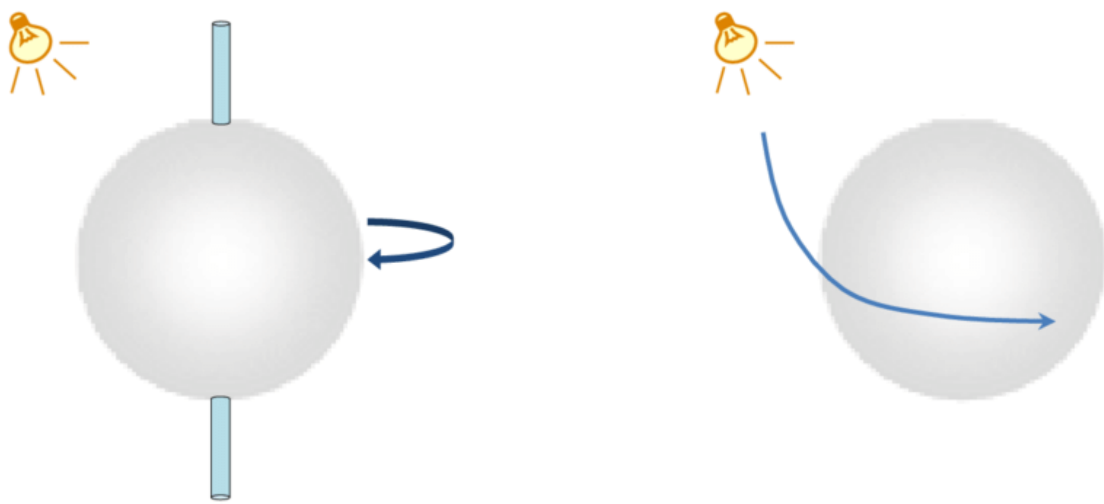
Length of flow vectors inversely proportional to depth Z of 3d point

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

Figure from Michael Black, Ph.D. Thesis

Apparent motion

- Optical flow differs from actual motion field:
  - (a) intensity remains constant, so that no motion is perceived;
  - (b) no object motion exists, however moving light source produces shading changes.



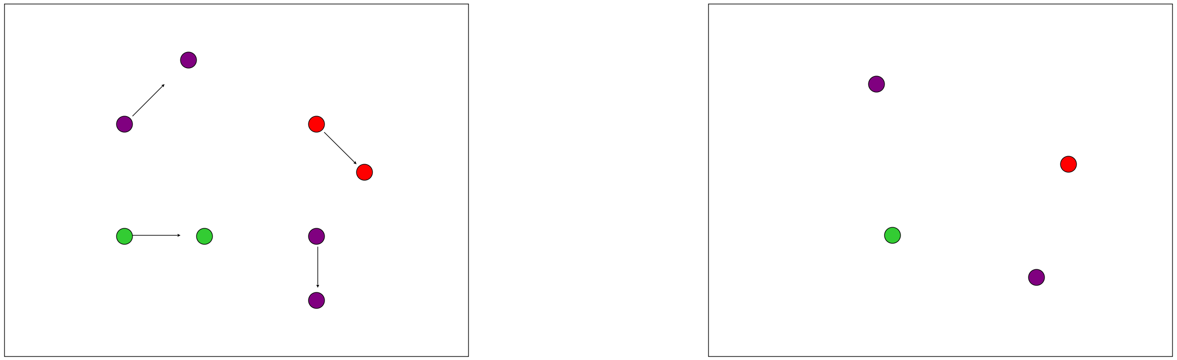
(a) (b)

Source: Silvio Savarese

Agenda

- Optical Flow Definition
- **Optical Flow Estimation**

Estimating Optical Flow

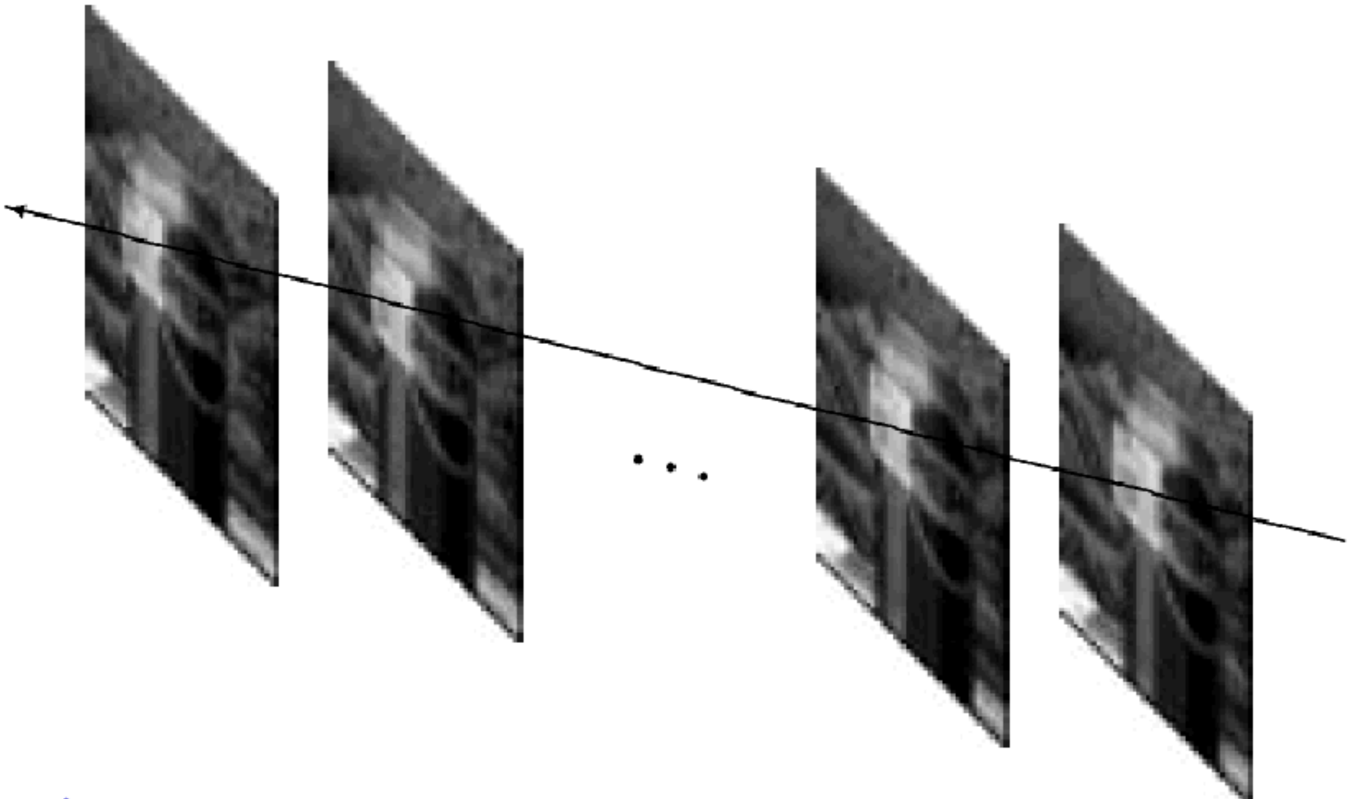


$I(x,y,t-1)$ $I(x,y,t)$

- Given two subsequent frames, estimate the apparent motion field u(x,y), v(x,y) between them
- **Key assumptions**
  - 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

Source: Silvio Savarese

Key Assumptions: Small Motions

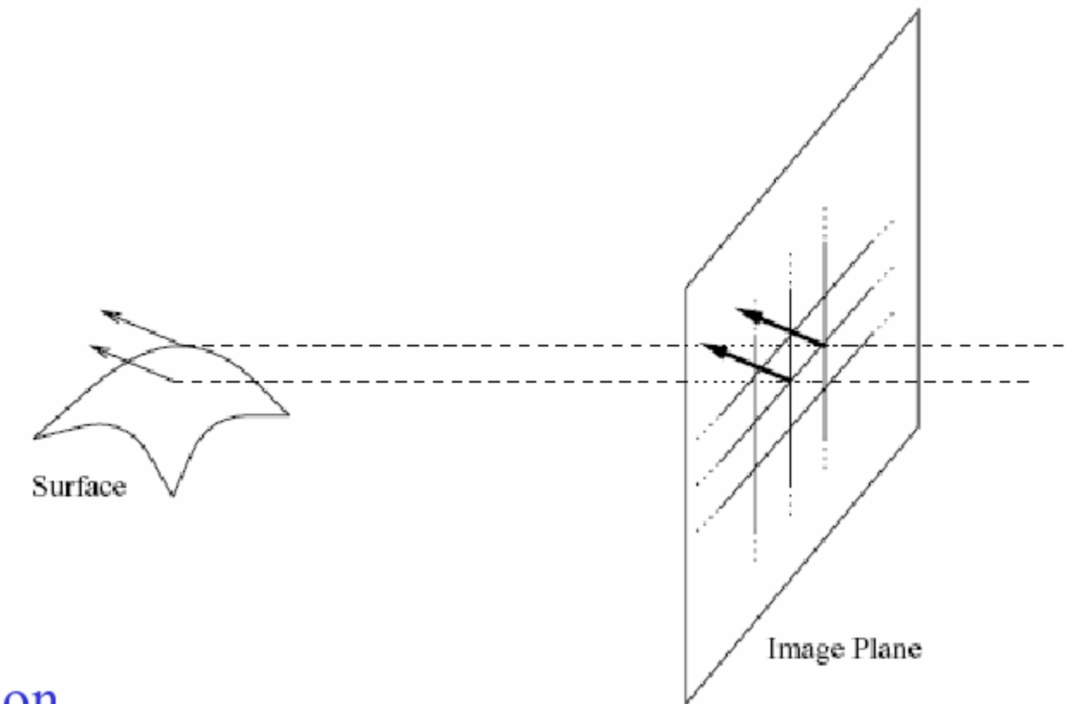


Assumption:

The image motion of a surface patch changes gradually over time.

* Slide from Michael Black, CS143 2003

Key Assumptions: Spatial Coherence

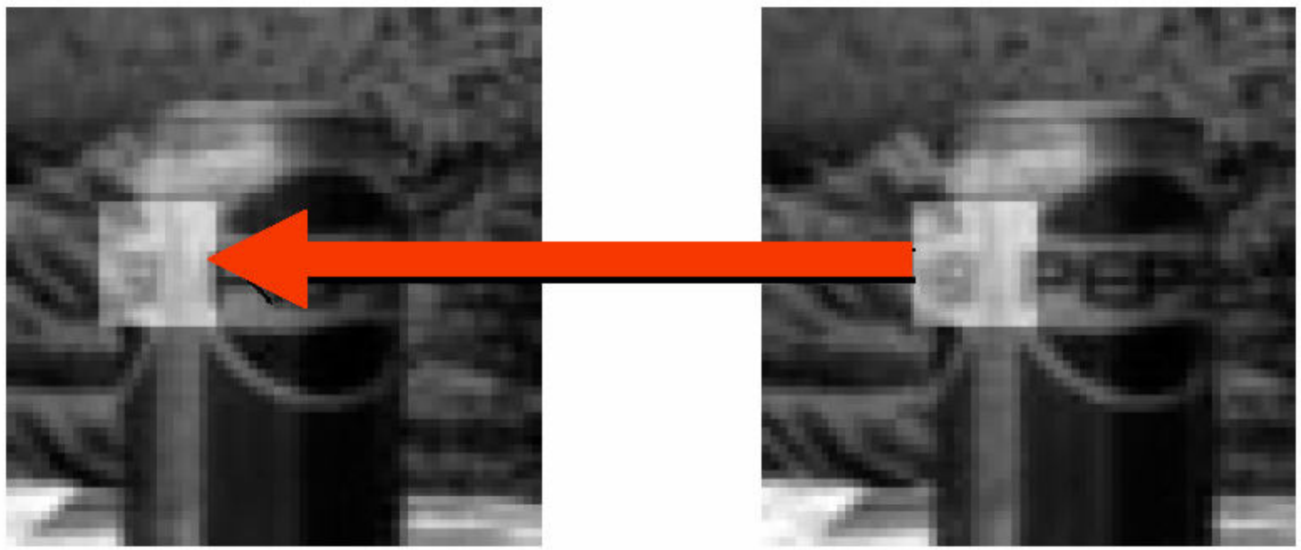


Assumption

* Neighboring points in the scene typically belong to the same surface and hence typically have similar motions.
* Since they also project to nearby points in the image, we expect spatial coherence in image flow.

* Slide from Michael Black, CS143 2003

Key Assumptions: Brightness Constancy



Assumption

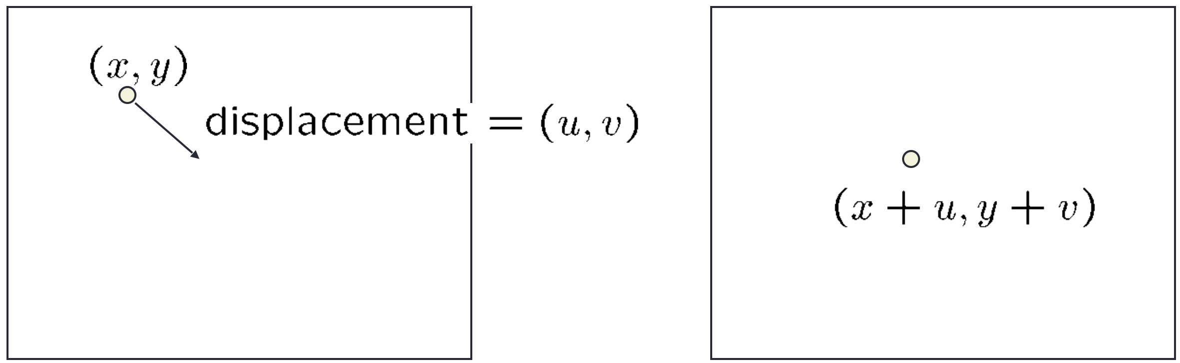
Image measurements (e.g. brightness) in a small region remain the same although their location may change.

$$I(x+u, y+v, t+1) = I(x, y, t)$$

(assumption)

* Slide from Michael Black, CS143 2003

Optical Flow Constraints (grayscale images)



$$I(x, y, t) \qquad\qquad I(x, y, t+1)$$

- Let's look at these constraints more closely
  - Brightness constancy constraint (equation)

$$I(x, y, t) = I(x+u, y+v, t+1)$$

  - Small motion: (u and v are less than 1 pixel, or smoothly varying)
    Taylor series expansion of I:

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

$$\approx I(x, y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

Optical Flow Equation

- Combining these two equations

$$0 = I(x+u, y+v, t+1) - I(x, y, t)$$

(Short hand: $I_x = \frac{\partial I}{\partial x}$

for t **or** $t+1$)

Optical Flow Equation

- Combining these two equations

$$
\begin{aligned}
0 &= I(x+u, y+v, t+1) - I(x, y, t) \\
&\approx I(x, y, t+1) + I_x u + I_y v - I(x, y, t) \\
&\approx [I(x, y, t+1) - I(x, y, t)] + I_x u + I_y v \\
&\approx I_t + I_x u + I_y v \\
&\approx I_t + \nabla I \cdot < u, v >
\end{aligned}
$$

(Short hand: $I_x = \frac{\partial I}{\partial x}$ for t **or** $t+1$)

Optical Flow Equation

- Combining these two equations

$$
\begin{aligned}
0 &= I(x+u, y+v, t+1) - I(x,y,t) \\
&\approx I(x,y,t+1) + I_x u + I_y v - I(x,y,t) \\
&\approx [I(x,y,t+1) - I(x,y,t)] + I_x u + I_y v \\
&\approx I_t + I_x u + I_y v \\
&\approx I_t + \nabla I \cdot \langle u, v \rangle
\end{aligned}
$$

(Short hand: $I_x = \frac{\partial I}{\partial x}$ for t **or** $t+1$)

In the limit as u and v go to zero, this becomes exact

$$0 = I_t + \nabla I \cdot \langle u, v \rangle$$

Brightness constancy constraint equation

$$I_x u + I_y v + I_t = 0$$

Filters Used to Find the Derivatives

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \text{first image} \qquad \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \text{first image} \qquad \begin{bmatrix} -1 & -1 \\ -1 & -1 \end{bmatrix} \text{first image}$$

$$\begin{bmatrix} -1 & 1 \\ -1 & 1 \end{bmatrix} \text{second image} \qquad \begin{bmatrix} -1 & -1 \\ 1 & 1 \end{bmatrix} \text{second image} \qquad \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix} \text{second image}$$

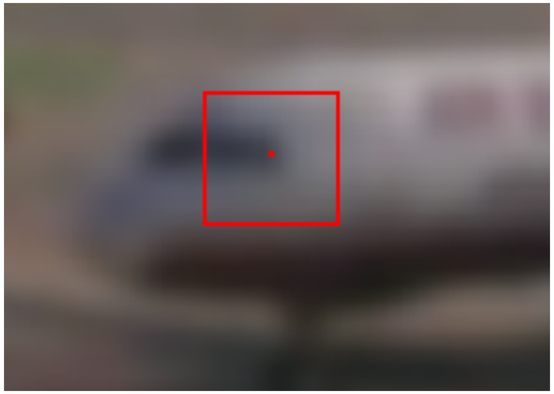
$$I_x \qquad\qquad I_y \qquad\qquad I_t$$

How Does This Make Sense?

Brightness constancy constraint equation

$$I_x u + I_y v + I_t = 0$$

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?

$$0 = I_t + \nabla I \cdot < u, v >$$

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

Source: Silvio Savarese

The Brightness Constancy Constraint

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

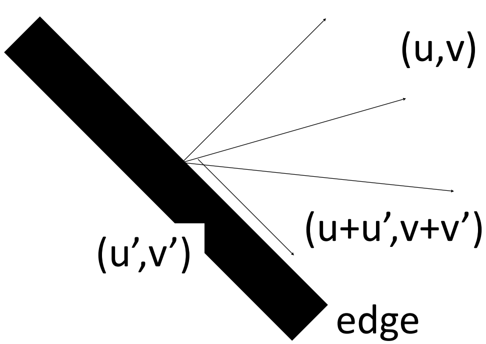
$$0 = I_t + \nabla I \cdot < u, v >$$

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

The component of the flow 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

$$\nabla I \cdot \begin{bmatrix} u' & v' \end{bmatrix}^T = 0$$



Source: Silvio Savarese

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…

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

- 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

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

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

Solving the Ambiguity…

- Least squares problem:

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\times2}{A} \; \underset{2\times1}{d} = \underset{25\times1}{b}$$

Matching Matches Across Images

- Overconstrained linear system

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix}$$

$$\underset{25\text{x}2}{A} \ \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b}$$

Least squares solution for *d* given by $(A^T A)\ d = A^T b$

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

The summations are over all pixels in the K x K window

Conditions for Solvability

Optimal (u, v) satisfies Lucas-Kanade equation

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

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

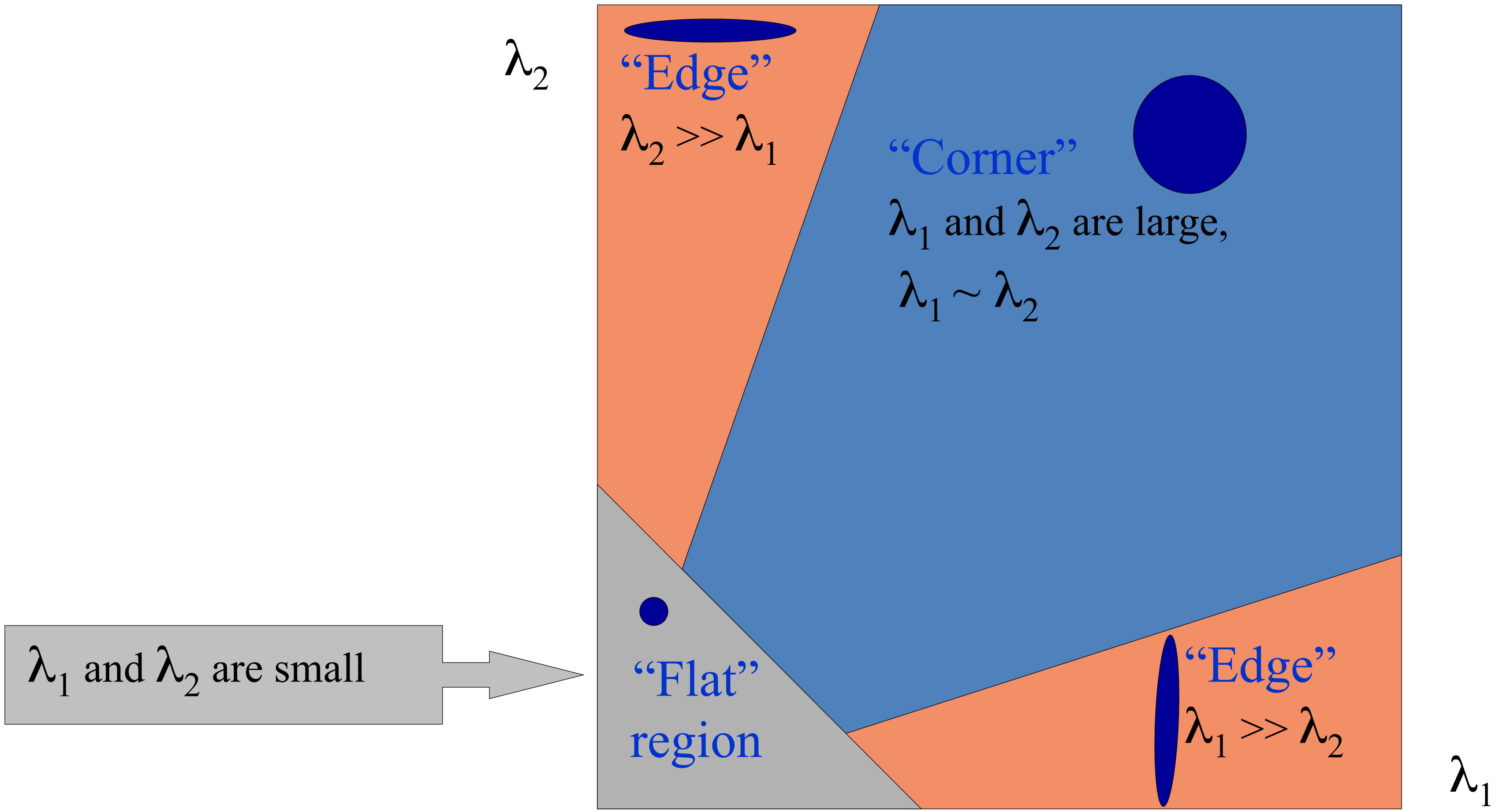
When is this solvable? What are good points to track?

- **$A^T A$** should be invertible
- **$A^T A$** should not be too small due to noise
  - 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)

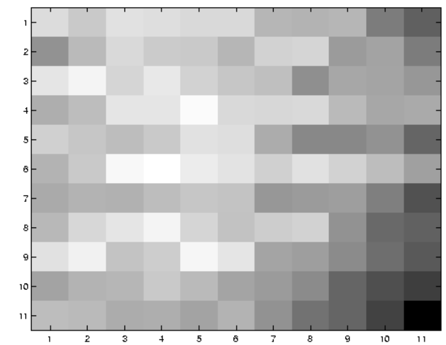
Criteria for Harris corner detector

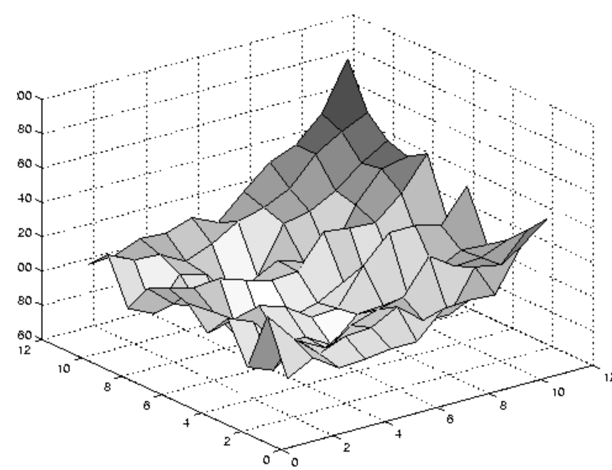
Interpreting the eigenvalues

Classification of image points using eigenvalues of the second moment matrix:



Low Texture Region



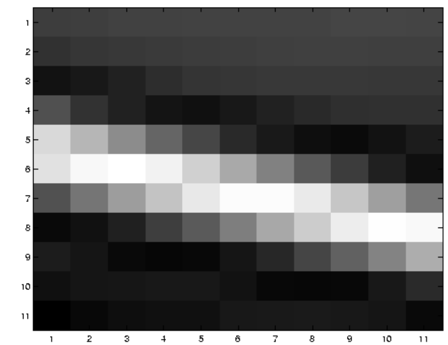


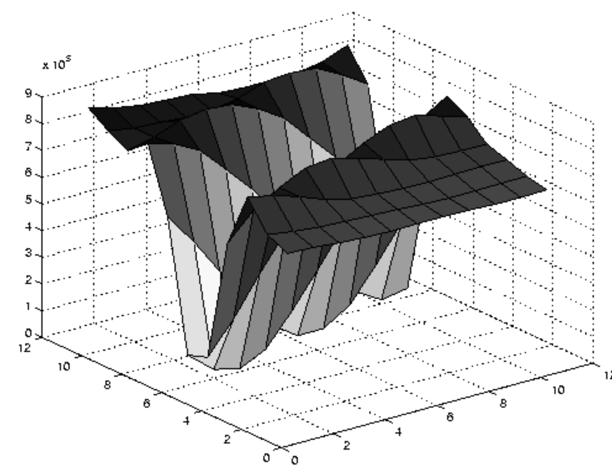
$$\sum \nabla I (\nabla I)^T$$

– gradients have small magnitude

– small λ_1, small λ_2

Edge



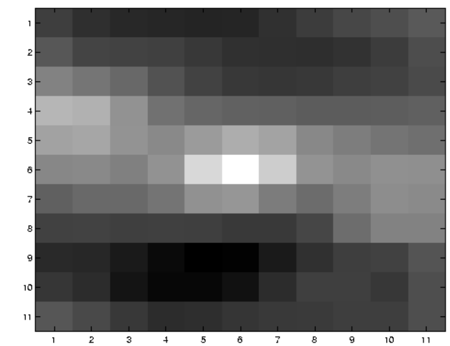


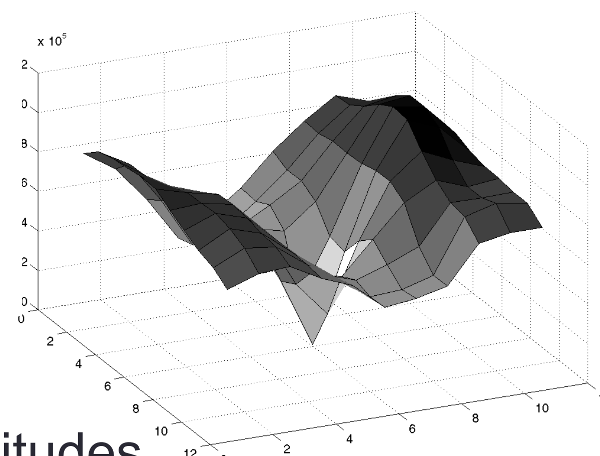
$$\sum \nabla I (\nabla I)^T$$

– large gradients, all the same

– large λ_1, small λ_2

High Textured Region





$$\sum \nabla I (\nabla I)^T$$

– gradients are different, large magnitudes

– large λ_1, large λ_2

Errors in Lukas-Kanade

What are the potential causes of errors in this procedure?

- Suppose $A^T A$ is easily invertible
- Suppose there is not much noise in the image

- When our assumptions are violated (Taylor expansion fails)
  - 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?

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003