

# Motion and Optical Flow

EE/CSE 576

Linda Shapiro

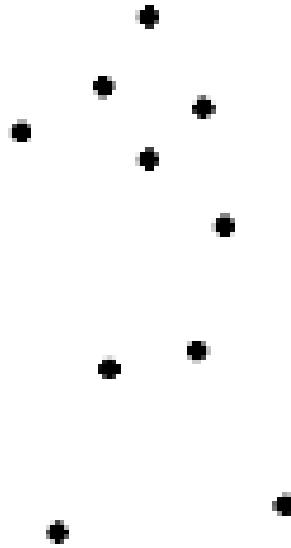
# We live in a moving world

- Perceiving, understanding and predicting motion is an important part of our daily lives



# Motion and perceptual organization

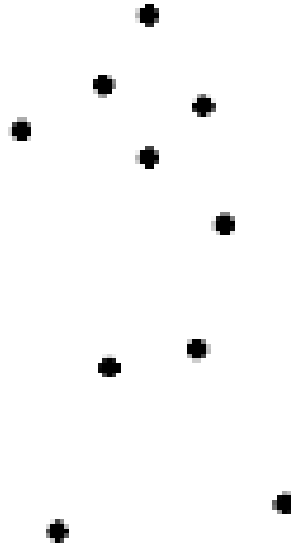
- Even “impoverished” motion data can evoke a strong percept



G. Johansson, “Visual Perception of Biological Motion and a Model For Its Analysis”, *Perception and Psychophysics* 14, 201-211, 1973.

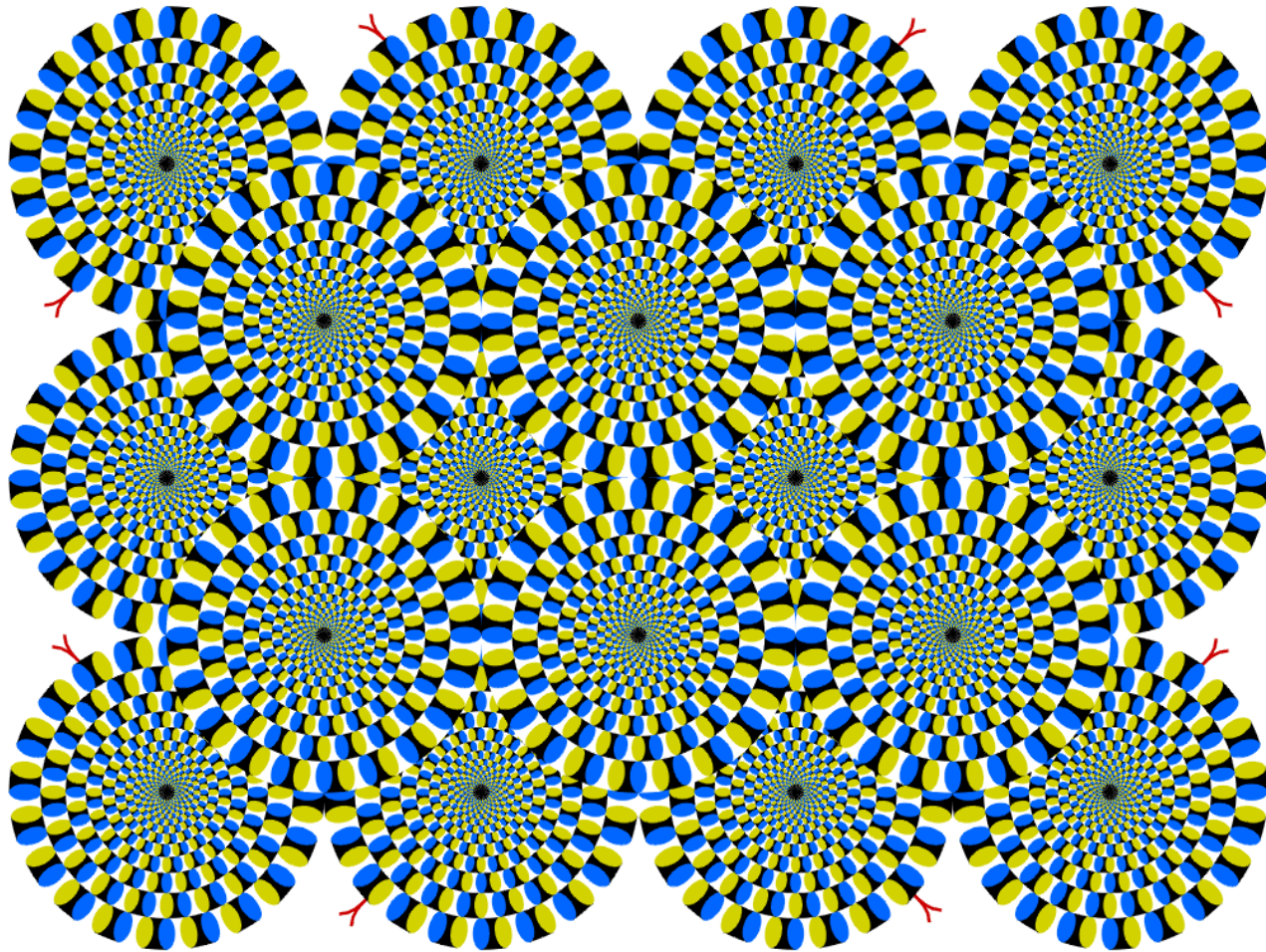
# Motion and perceptual organization

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



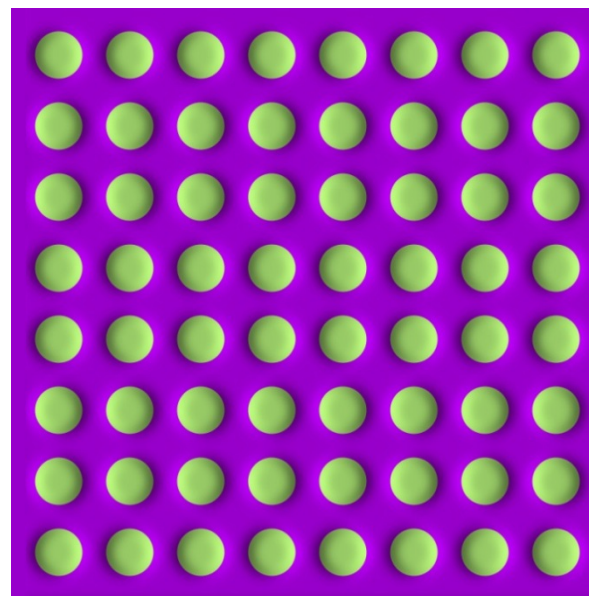
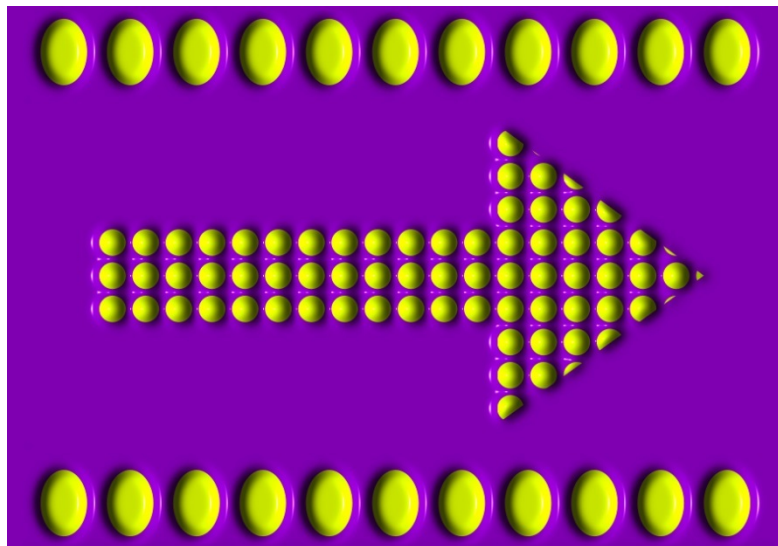
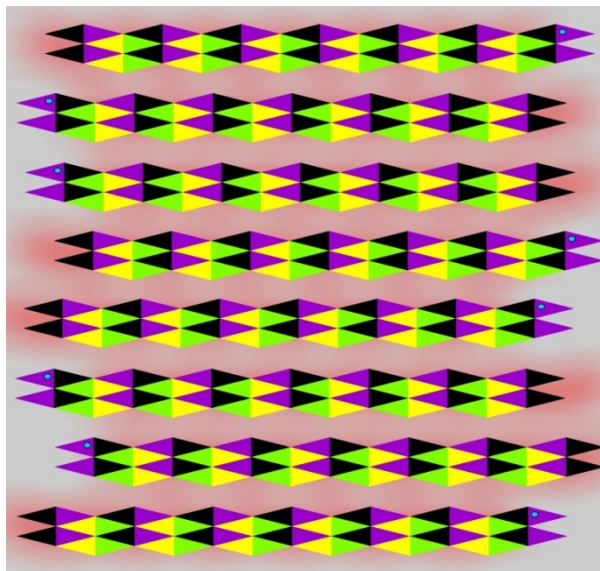
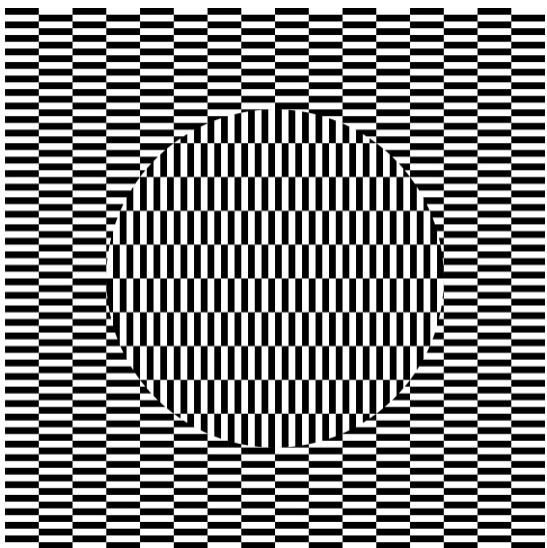
G. Johansson, “Visual Perception of Biological Motion and a Model For Its Analysis”, *Perception and Psychophysics* 14, 201-211, 1973.

# Seeing motion from a static picture?



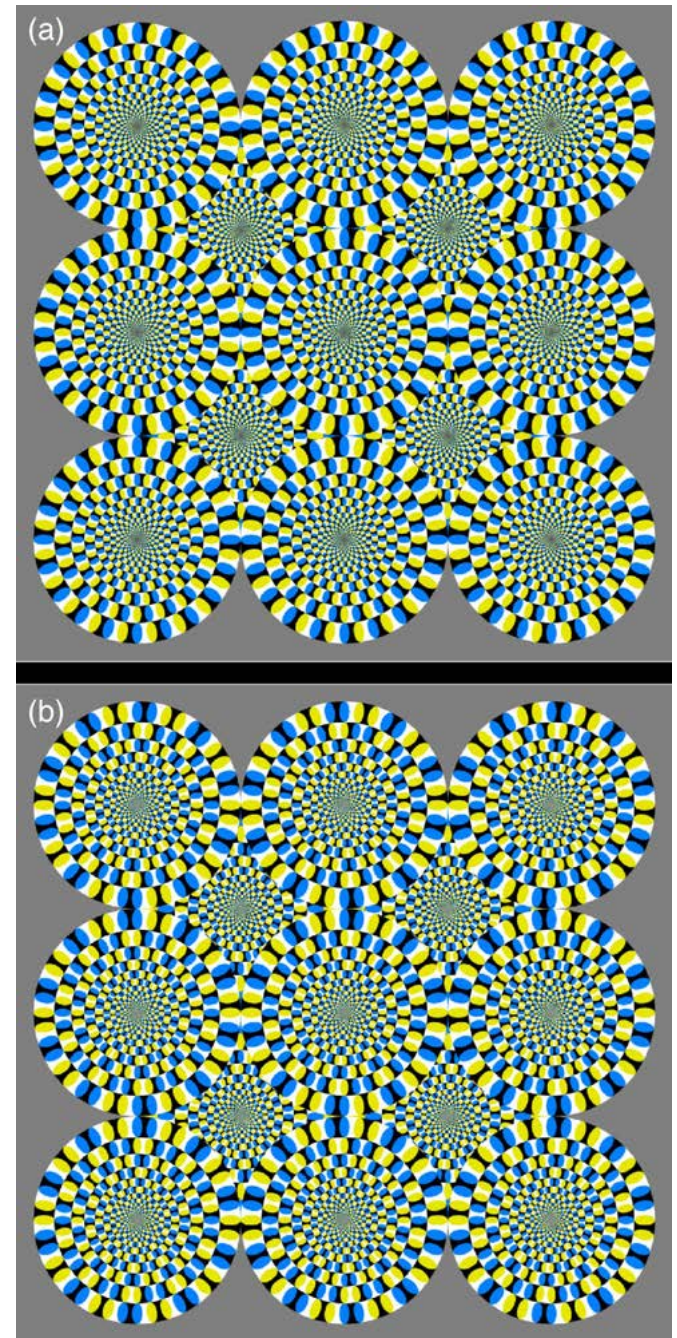


# More examples



# How is this possible?

- The true mechanism is yet 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



# 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

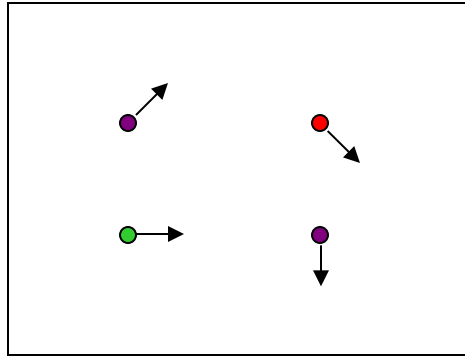
B. Lucas and T. Kanade. [An iterative image registration technique with an application to stereo vision](#). In *Proceedings of the International Joint Conference on Artificial Intelligence*, pp. 674–679, 1981.

# Feature tracking

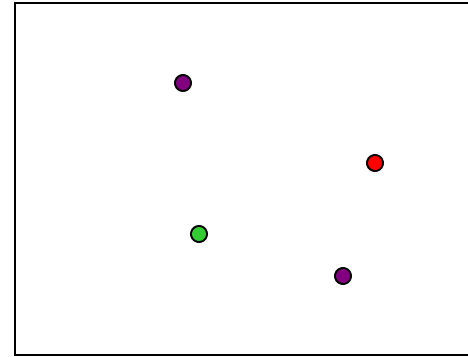
- Challenges
  - Figure out which features can be tracked
  - Efficiently track across frames
  - Some points may change appearance over time (e.g., due to rotation, moving into shadows, etc.)
  - Drift: small errors can accumulate as appearance model is updated
  - Points may appear or disappear: need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



# Feature tracking



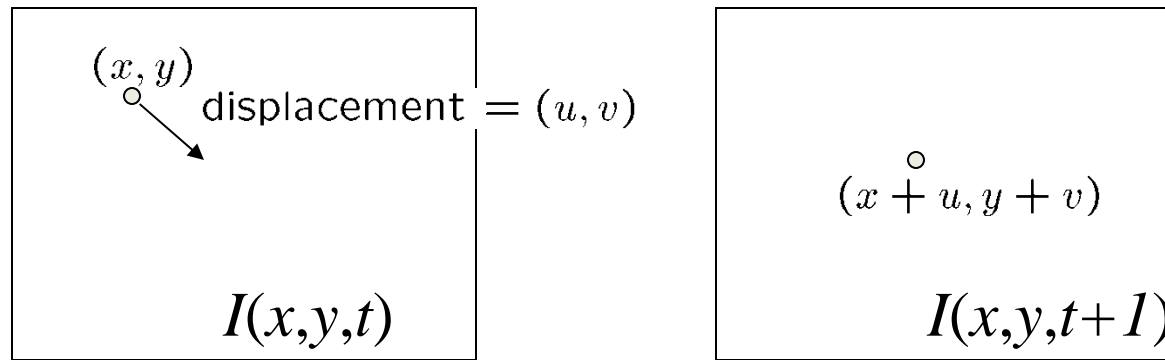
$I(x,y,t)$



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

- 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



- 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:

$$I(x + u, y + v, t + 1) \approx I(x, y, t) + \overset{\text{Image derivative along x}}{I_x} \cdot u + I_y \cdot v + \overset{\text{Difference over frames}}{I_t}$$

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

So:  $I_x \cdot u + I_y \cdot v + I_t \approx 0 \rightarrow \nabla I \cdot [u \ v]^T + I_t = 0$

# The brightness constancy constraint

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

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

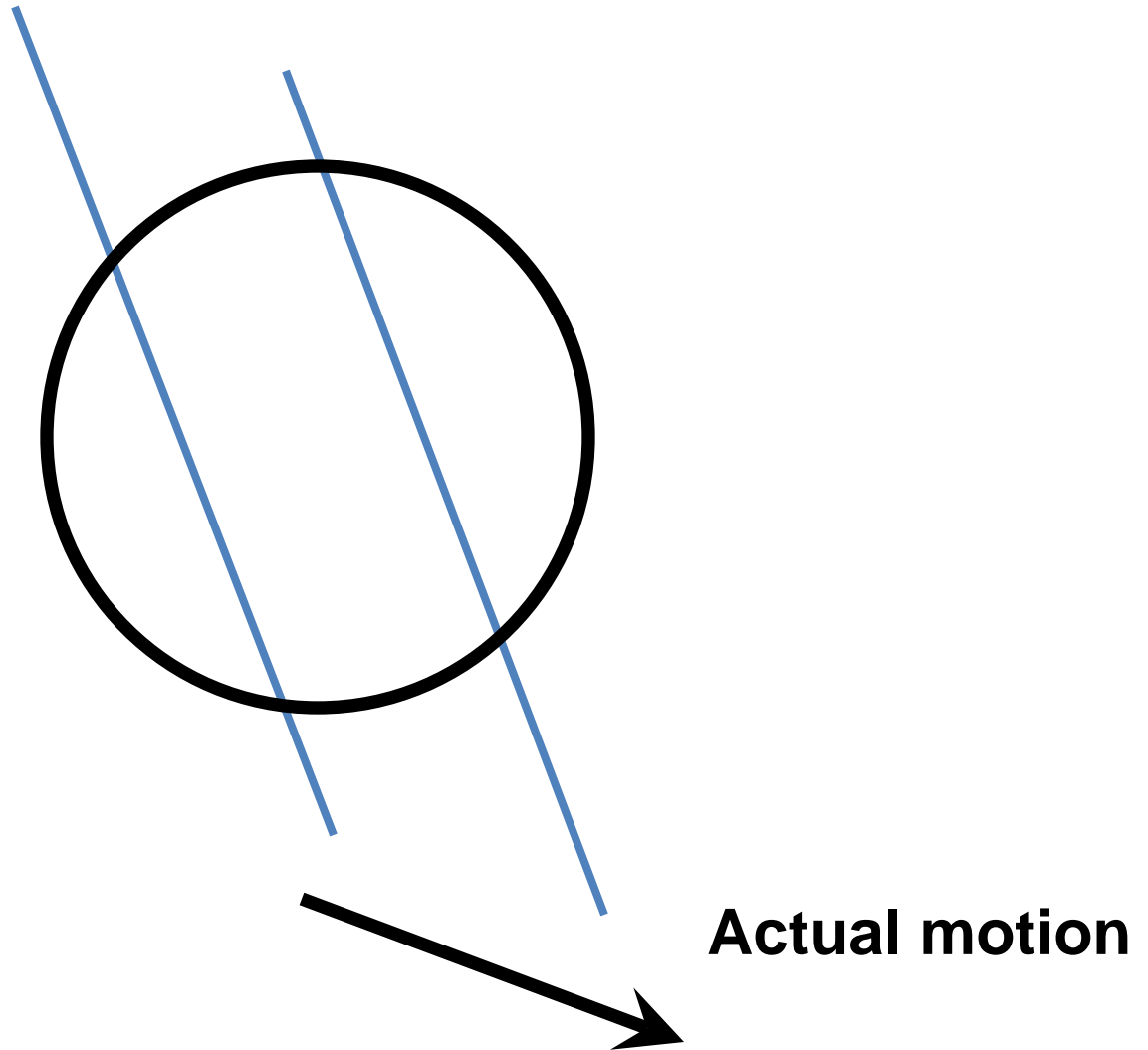
- How many equations and unknowns per pixel?
  - One equation (this is a scalar equation!), two unknowns  $(u, v)$

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

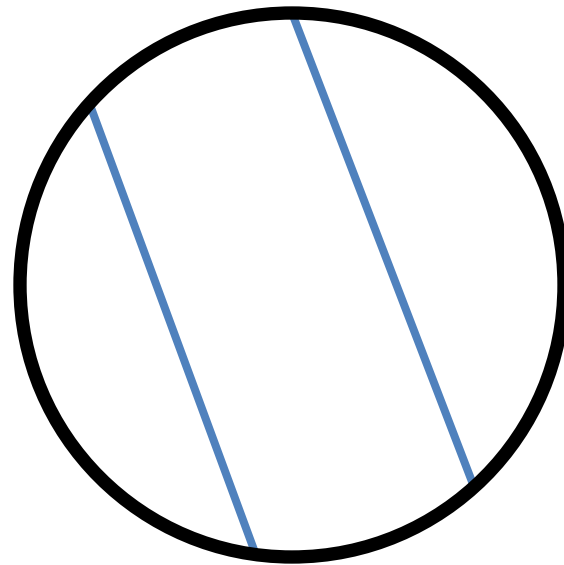
If  $(u, v)$  satisfies the equation,  
so does  $(u+u', v+v')$  if

$$\nabla I \cdot [u' \ v']^T = 0$$

# The aperture problem



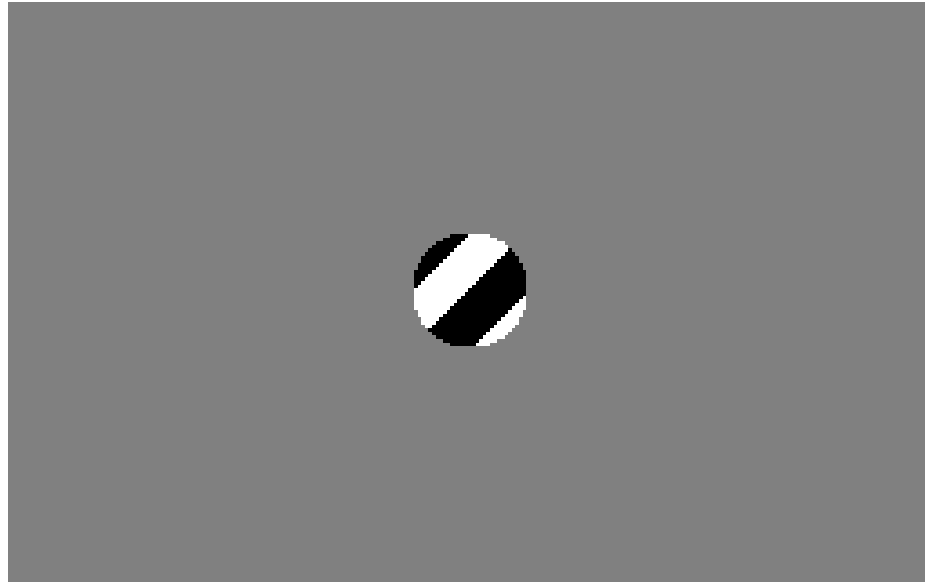
# The aperture problem



**Perceived motion**

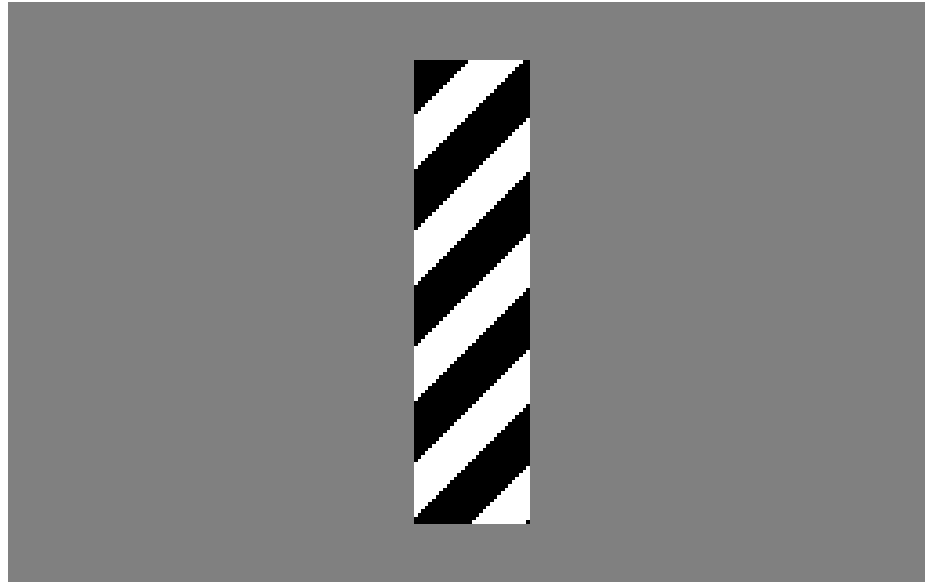


# The barber pole illusion



[http://en.wikipedia.org/wiki/Barberpole\\_illusion](http://en.wikipedia.org/wiki/Barberpole_illusion)

# The barber pole illusion



[http://en.wikipedia.org/wiki/Barberpole\\_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:

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \quad \begin{matrix} A & d = b \\ 25 \times 2 & 2 \times 1 & 25 \times 1 \end{matrix}$$

# Matching patches across images

- Overconstrained linear system

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \quad \begin{matrix} A & d = b \\ 25 \times 2 & 2 \times 1 & 25 \times 1 \end{matrix}$$

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

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

$A^T A$   $A^T b$

The summations are over all pixels in the  $K \times K$  window



# Conditions for solvability

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

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

$A^T A$   $A^T b$

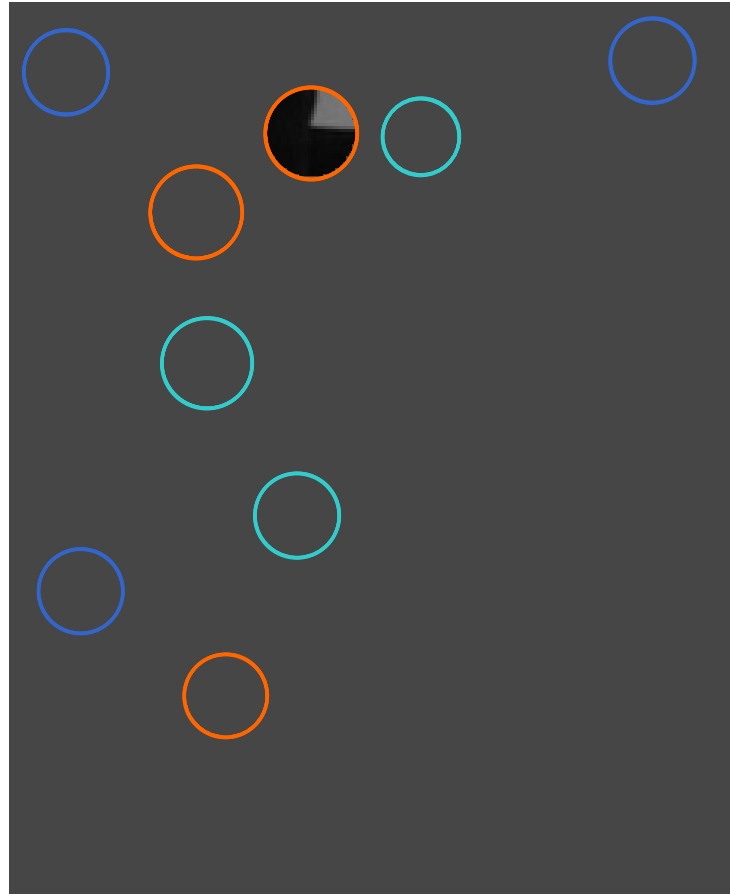
When is this solvable? I.e., what are good points to track?

- $A^T A$  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

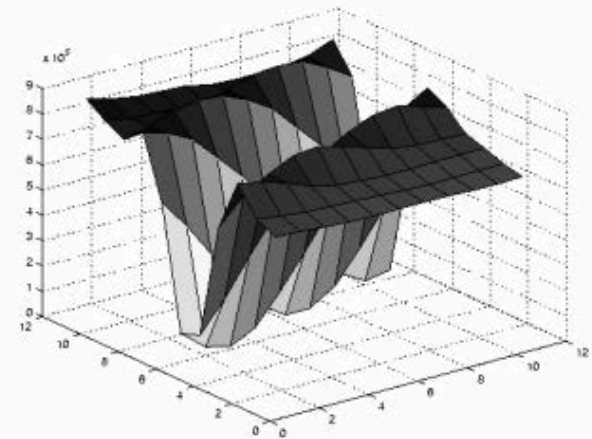
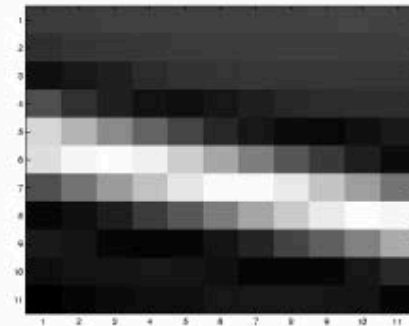


Corners

Lines

Flat regions

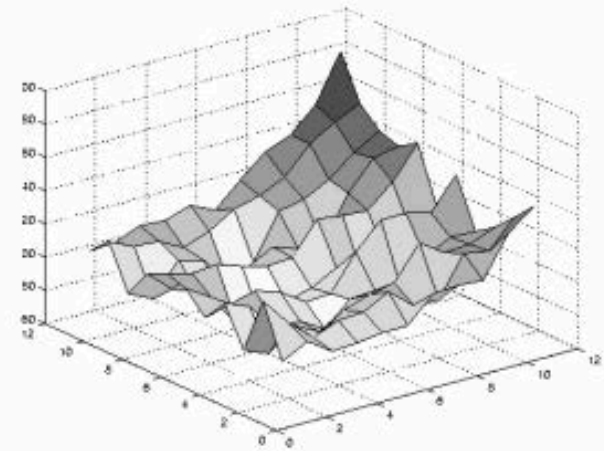
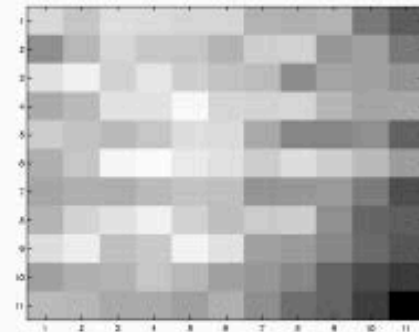
# Edge



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

- large gradients, all the same
- large  $\lambda_1$ , small  $\lambda_2$

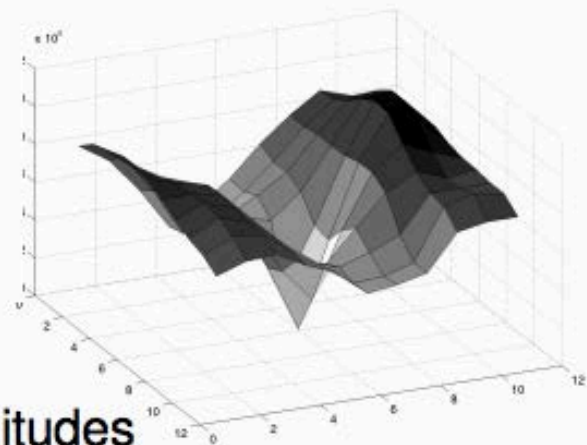
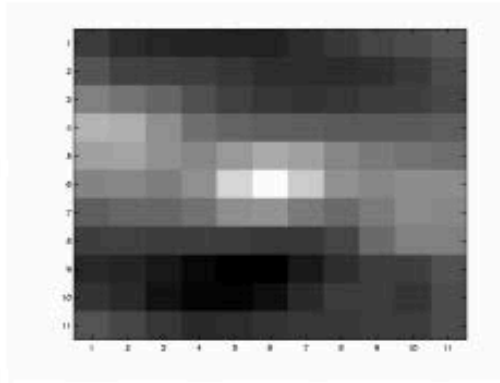
# Low Texture Region



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

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

# High Texture Region



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

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

When our assumptions are violated

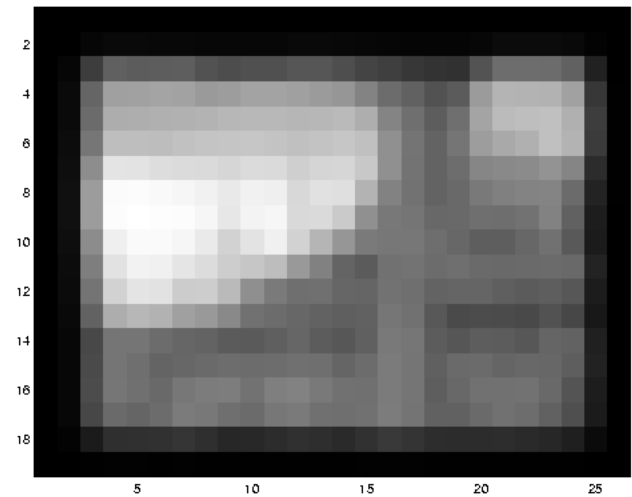
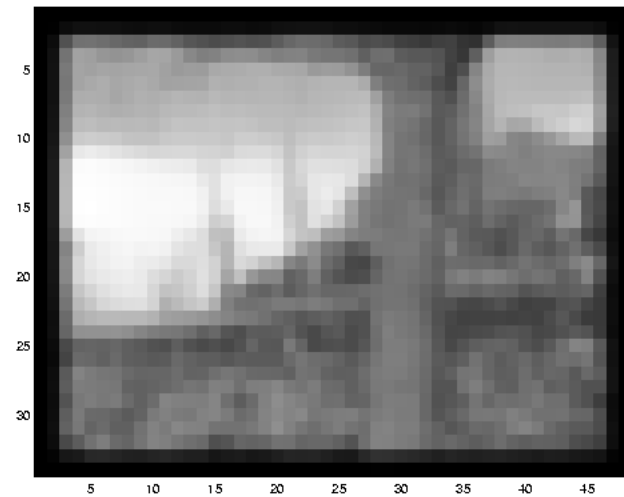
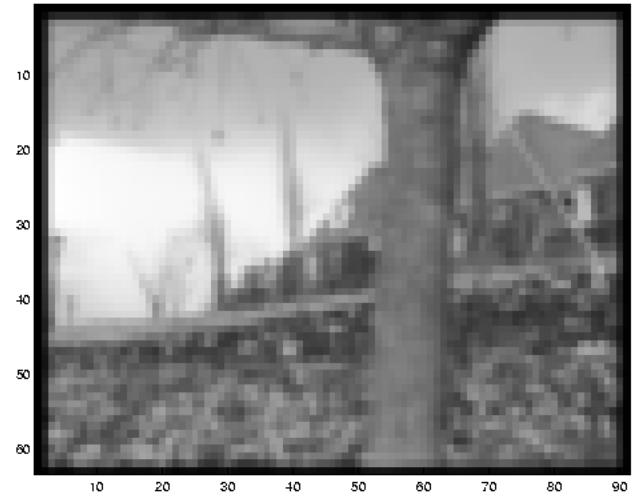
- Brightness constancy is **not** satisfied
- The motion is **not** small
- A point does **not** move like its neighbors
  - window size is too large
  - what is the ideal window size?

# Revisiting the small motion assumption

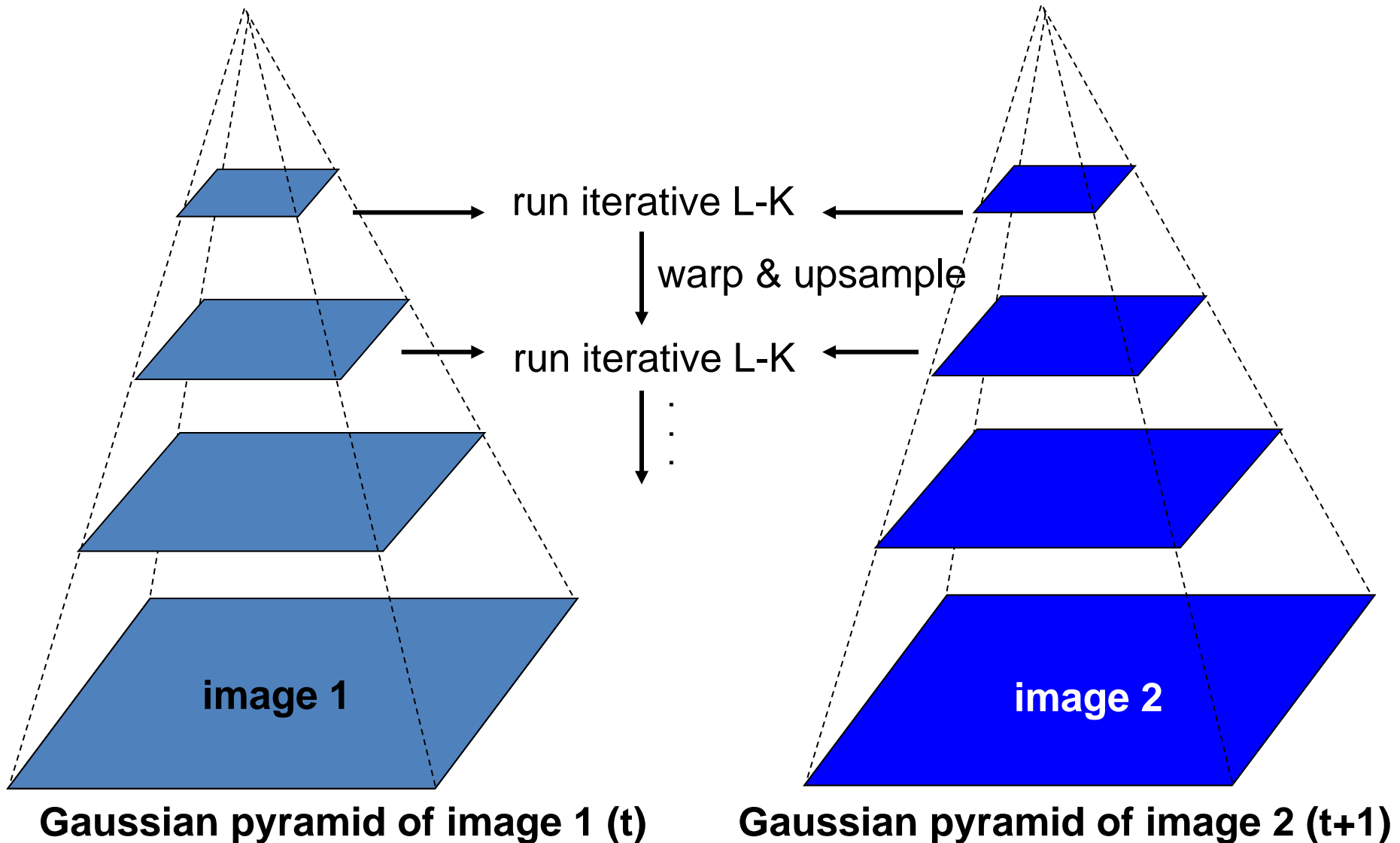


- Is this motion small enough?
  - Probably not—it's much larger than one pixel ( $2^{\text{nd}}$  order terms dominate)
  - How might we solve this problem?

# Reduce the resolution!



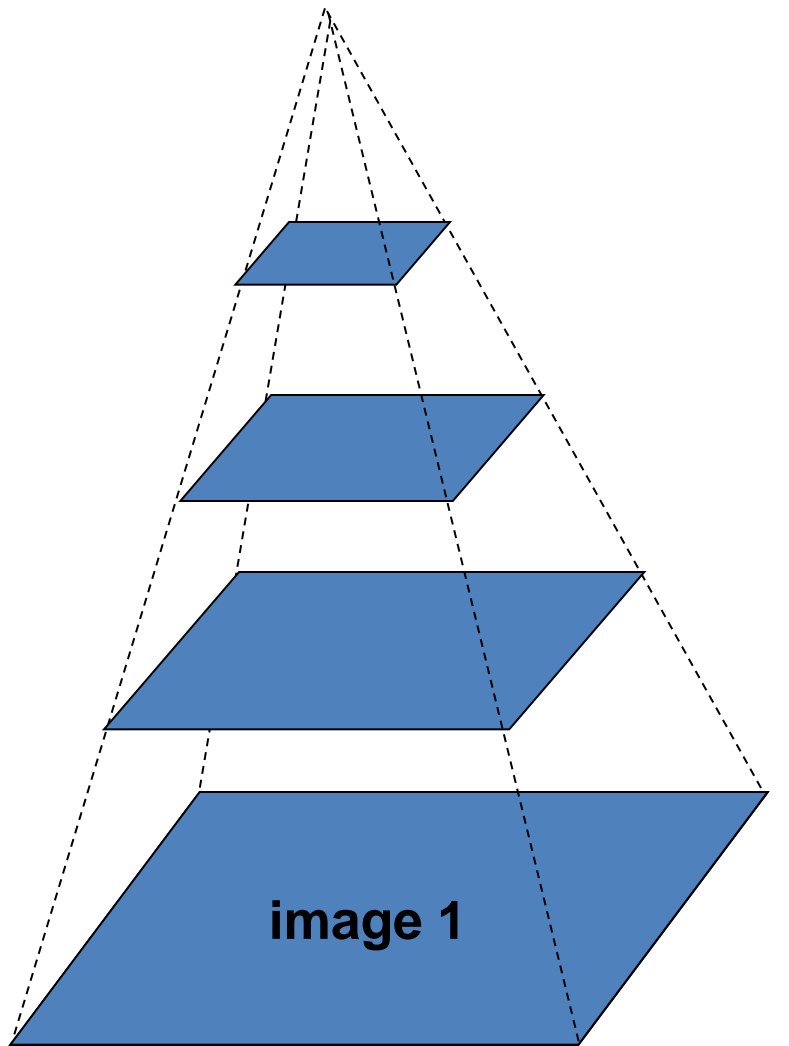
# Coarse-to-fine optical flow estimation



# 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



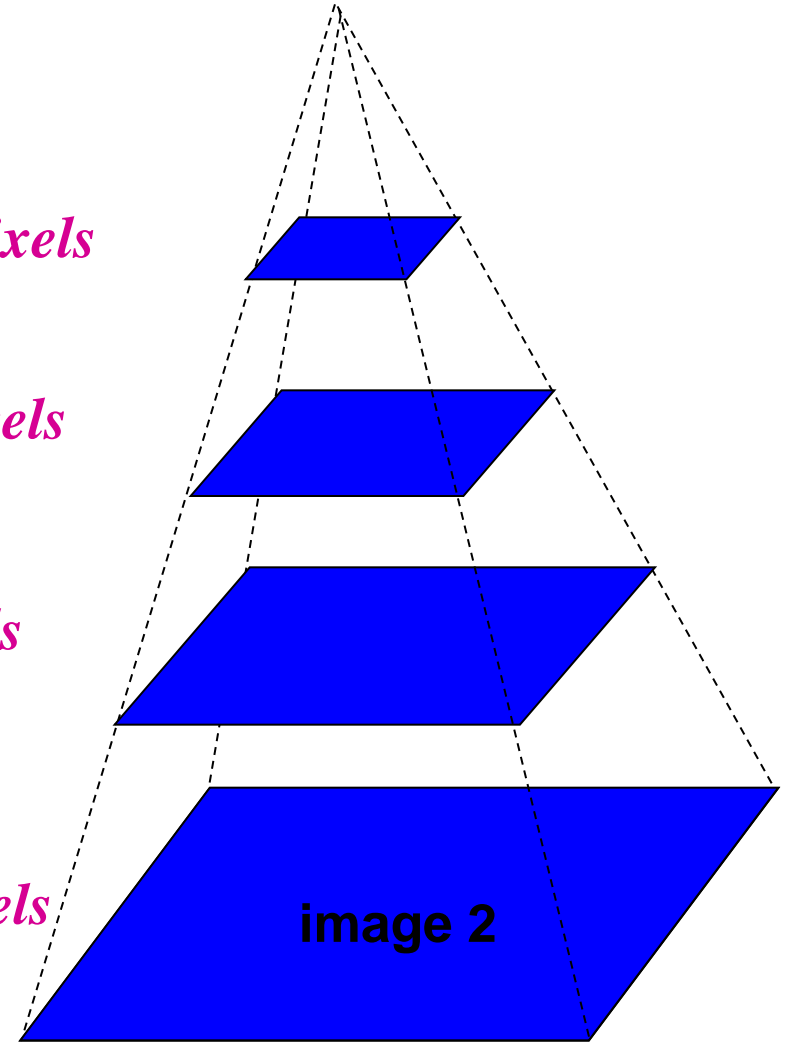
**Gaussian pyramid of image 1**

*$u=1.25$  pixels*

*$u=2.5$  pixels*

*$u=5$  pixels*

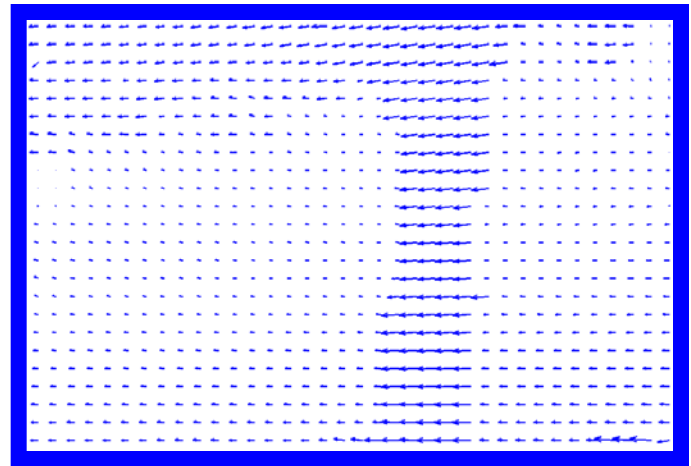
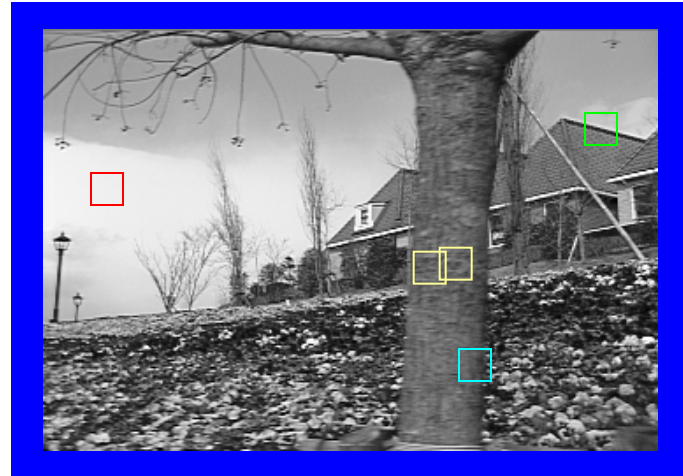
*$u=10$  pixels*



**Gaussian pyramid of image 2**

# The Flower Garden Video

What should the  
optical flow be?

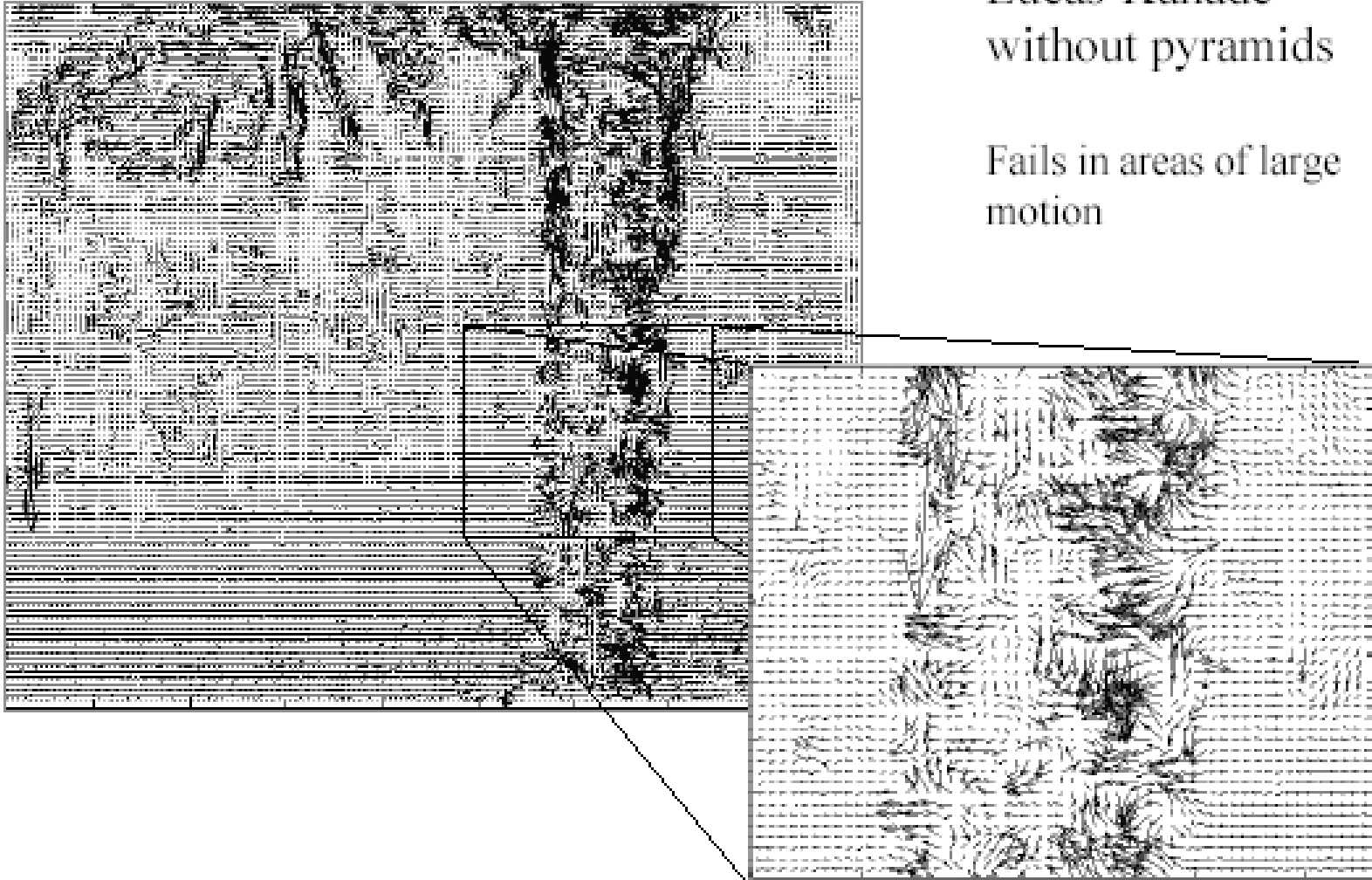




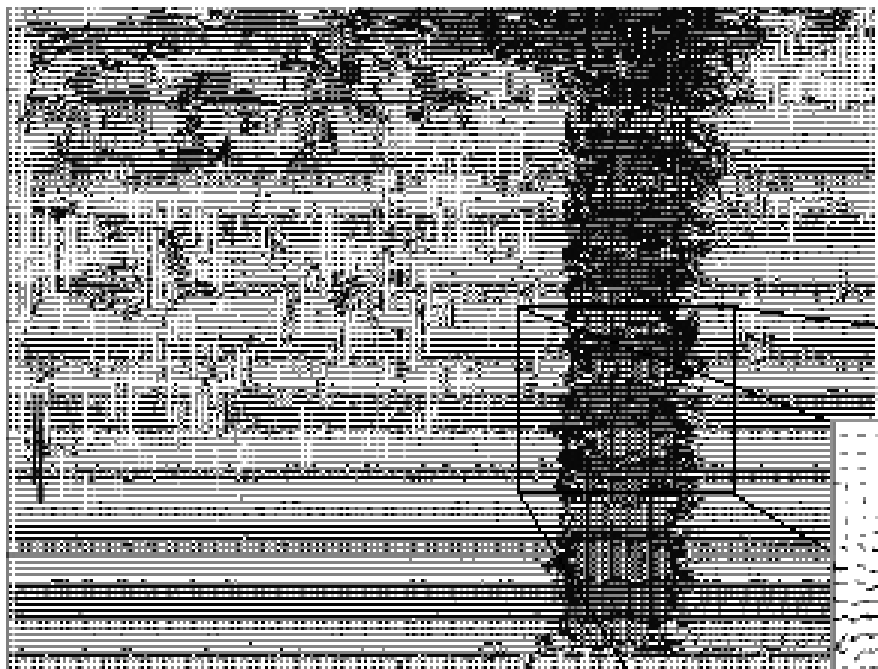
# Optical Flow Results

Lucas-Kanade  
without pyramids

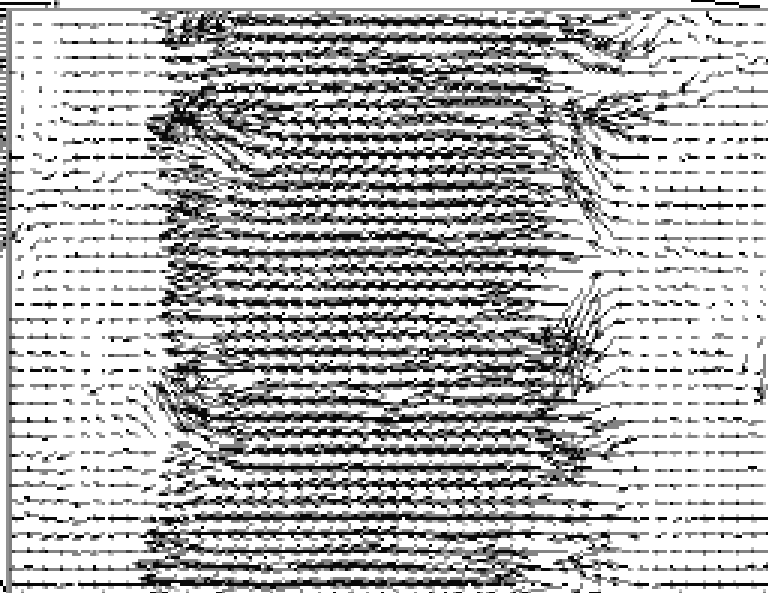
Fails in areas of large  
motion



# Optical Flow Results



Lucas-Kanade with Pyramids





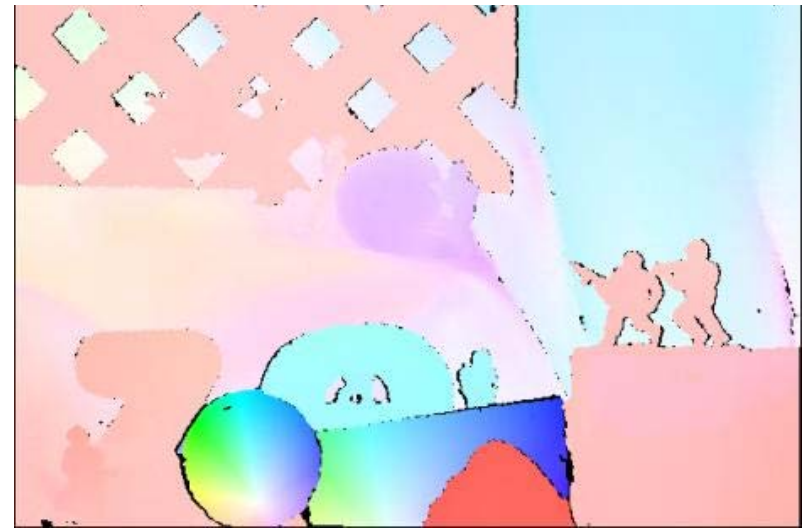
# Flow quality evaluation





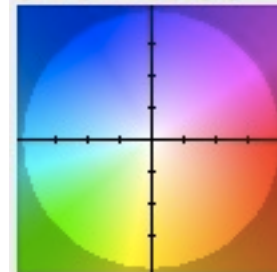
# Flow quality evaluation

- Middlebury flow page
  - <http://vision.middlebury.edu/flow/>



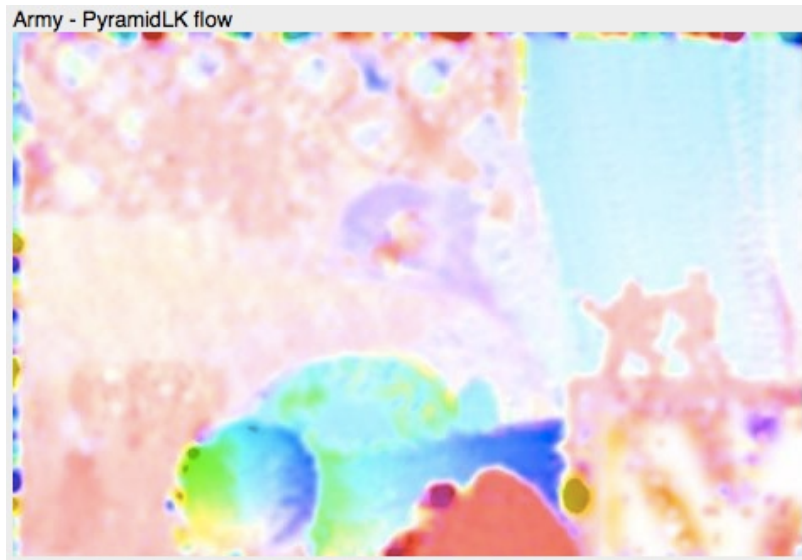
Ground Truth

Color encoding  
of flow vectors

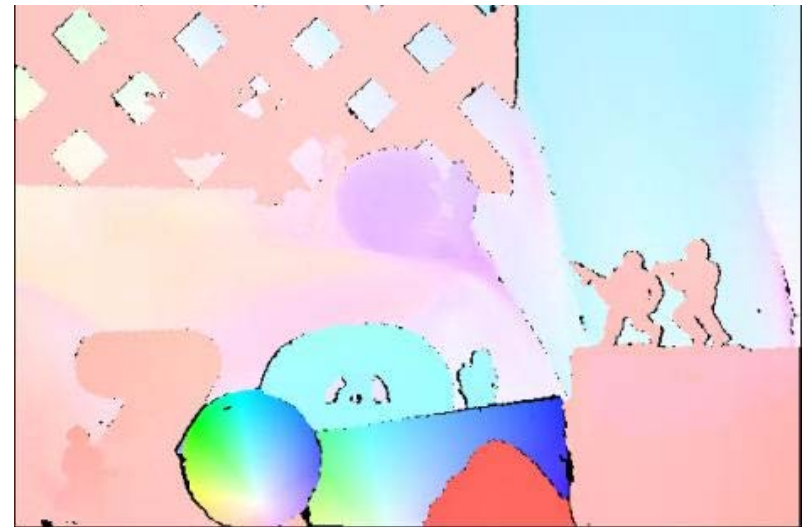


# Flow quality evaluation

- Middlebury flow page
  - <http://vision.middlebury.edu/flow/>

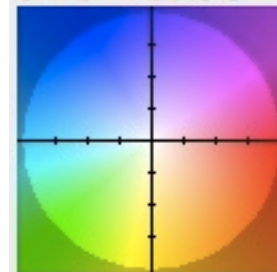


Lucas-Kanade flow



Ground Truth

Color encoding  
of flow vectors

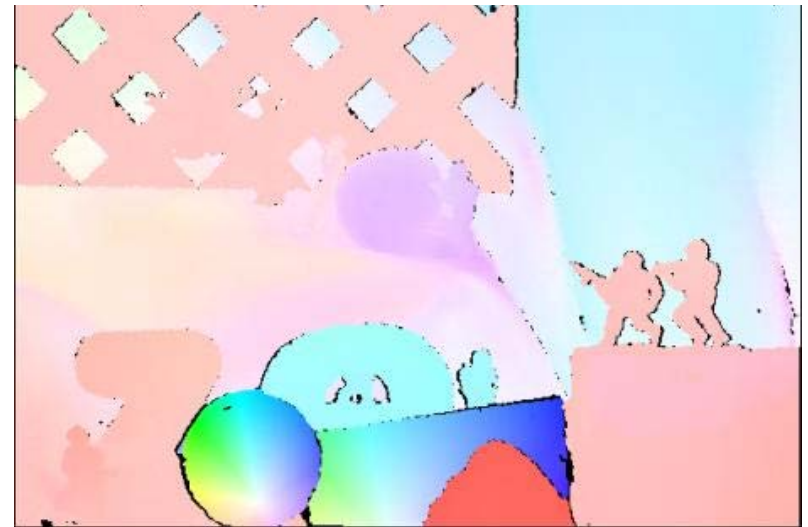


# Flow quality evaluation

- Middlebury flow page
  - <http://vision.middlebury.edu/flow/>

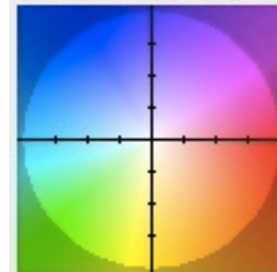


Best-in-class alg



Ground Truth

Color encoding  
of flow vectors



# Video stabilization





# Video denoising

Original



Denoised



# Video super resolution

Low-Res



# Robust Visual Motion Analysis: Piecewise-Smooth Optical Flow

Ming Ye

Electrical Engineering  
University of Washington

# Estimating Piecewise-Smooth Optical Flow with Global Matching and Graduated Optimization

## *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.**

**Uses least median of squares instead of regular least squares.**

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

**Minimizes a global energy function  $E = \sum (E_B(V_i) + E_S(V_i))$  where  $E_B$  is the brightness difference and  $E_S$  is the smoothness at flow vector  $V_i$**

- **Step 3. Global matching that minimizes energy by a greedy approach.**

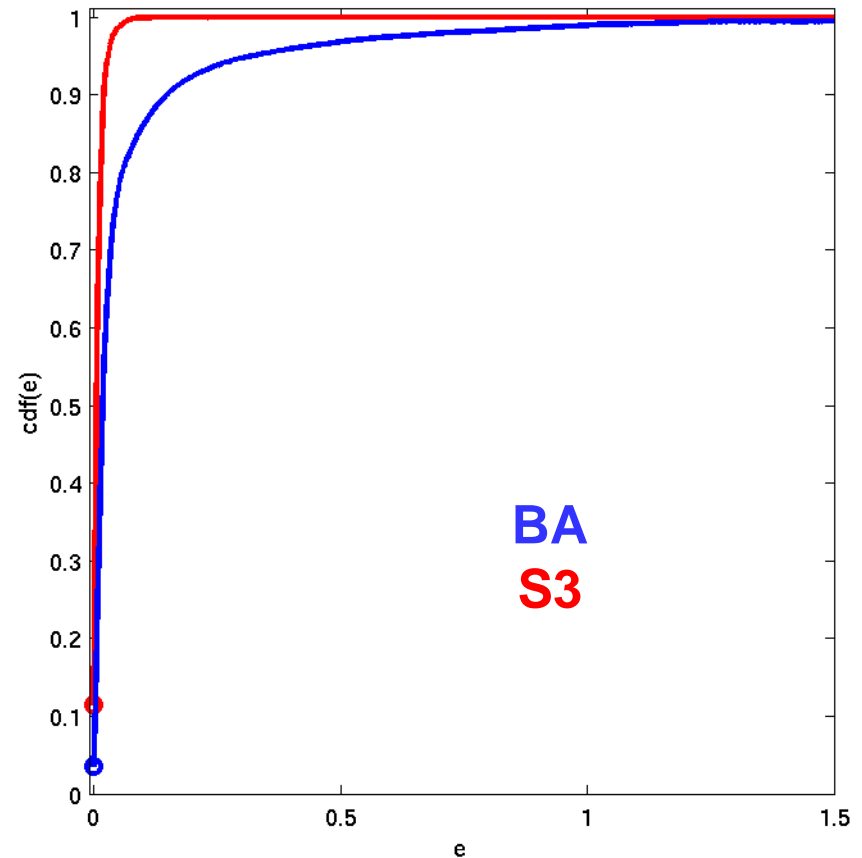
**Visits each pixel and updates it to be consistent with neighbors, iteratively.**

# TT: Translating Tree



150x150 (Barron 94)

	$e_{\angle}(\text{°})$	$e_{ \bullet }(\text{pix})$	$\bar{e}(\text{pix})$
<b>BA</b>	<b>2.60</b>	<b>0.128</b>	<b>0.0724</b>
<b>S3</b>	<b>0.248</b>	<b>0.0167</b>	<b>0.00984</b>



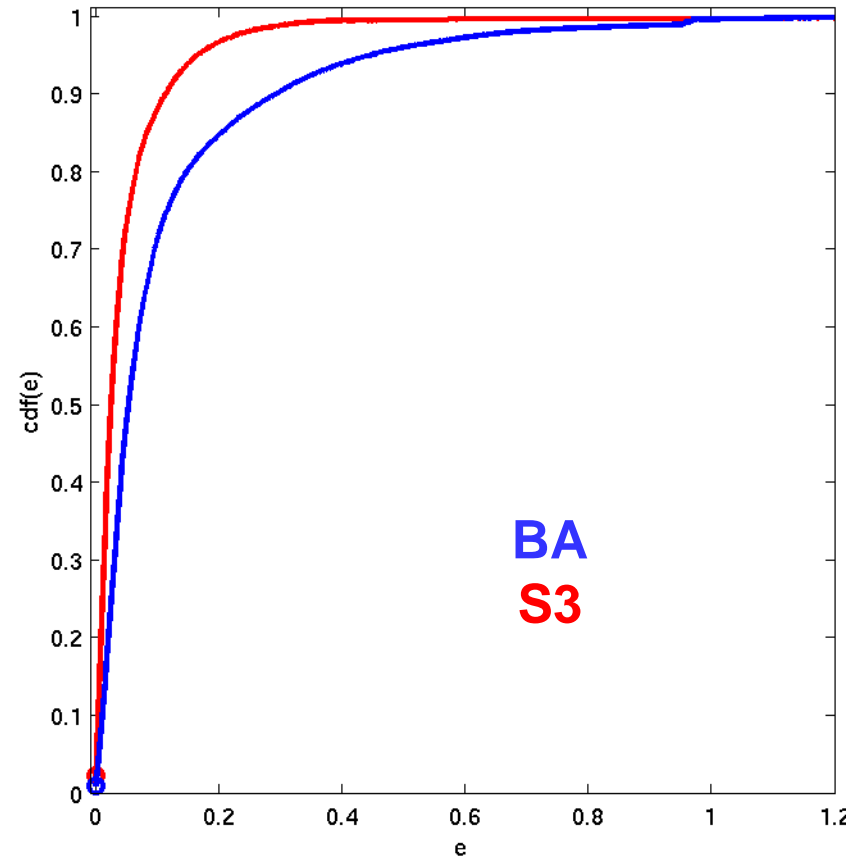
**e: error in pixels, cdf: cumulative distribution function for all pixels**

# DT: Diverging Tree

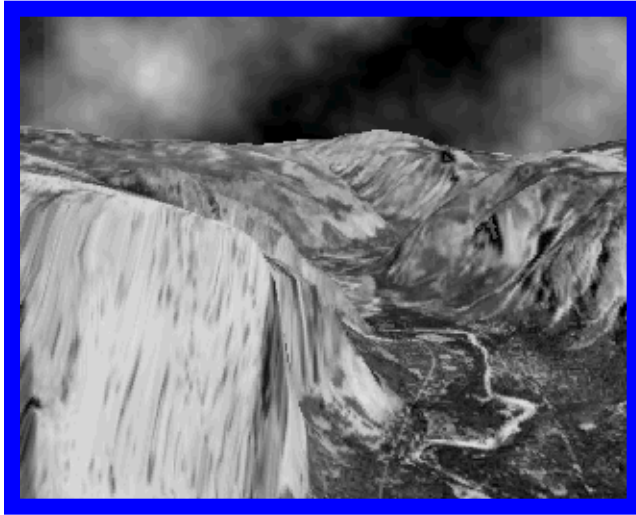


150x150 (Barron 94)

	$e_{\angle}(\text{°})$	$e_{ \bullet }(\text{pix})$	$\bar{e}(\text{pix})$
<b>BA</b>	<b>6.36</b>	<b>0.182</b>	<b>0.114</b>
<b>S3</b>	<b>2.60</b>	<b>0.0813</b>	<b>0.0507</b>

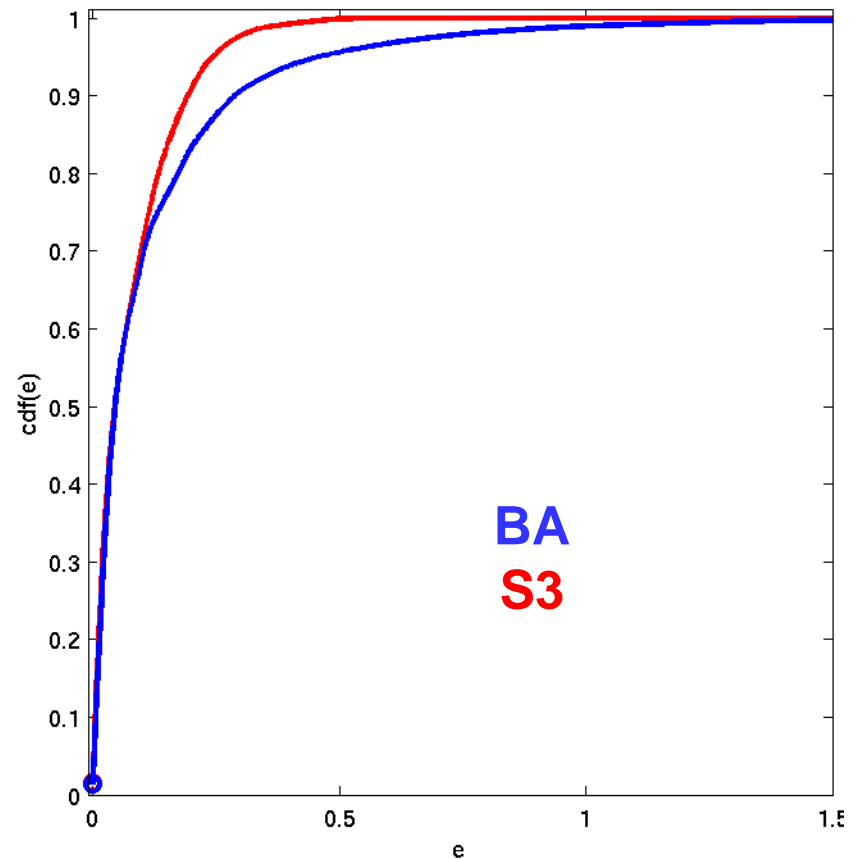


# YOS: Yosemite Fly-Through



316x252 (Barron, cloud excluded)

	$e_{\angle} (^{\circ})$	$e_{ \bullet } (\text{pix})$	$\bar{e} (\text{pix})$
<b>BA</b>	<b>2.71</b>	<b>0.185</b>	<b>0.118</b>
<b>S3</b>	<b>1.92</b>	<b>0.120</b>	<b>0.0776</b>





# TAXI: Hamburg Taxi



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



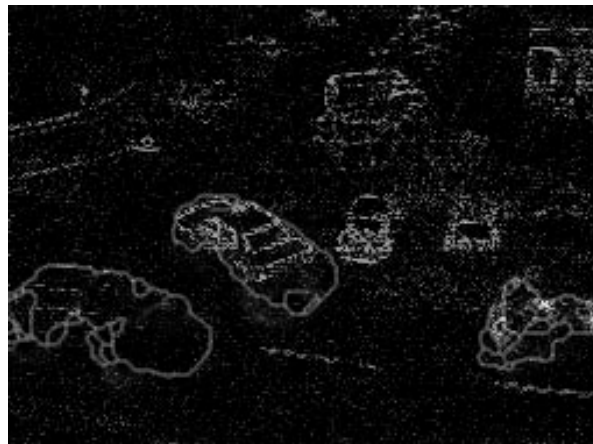
LMS



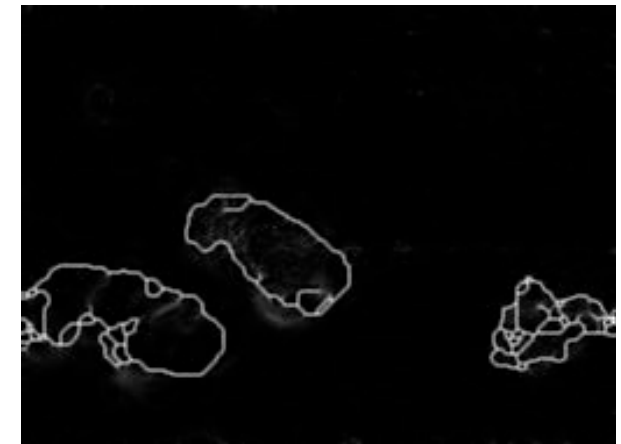
BA



Ours



Error map

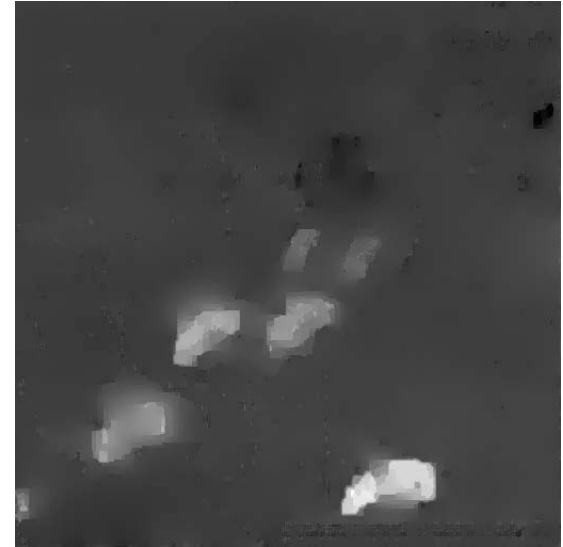


Smoothness error

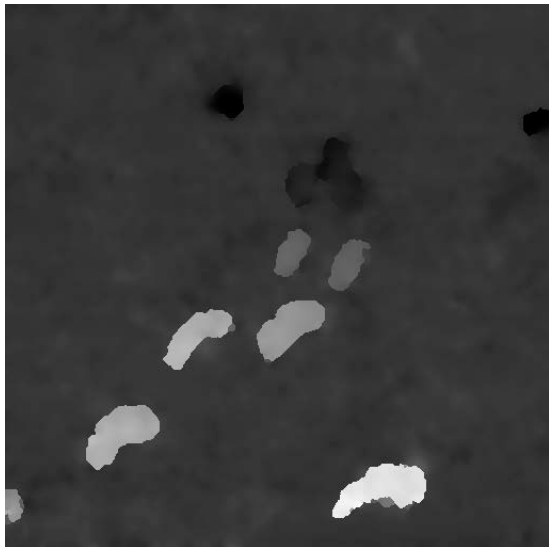
# Traffic



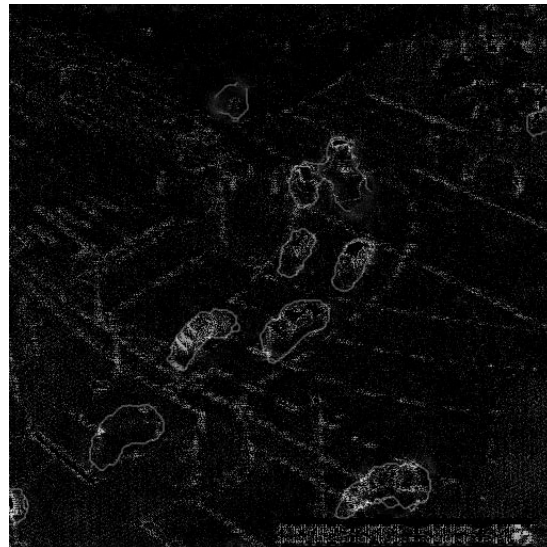
512x512  
(Nagel)  
max speed:  
6.0 pix/frame



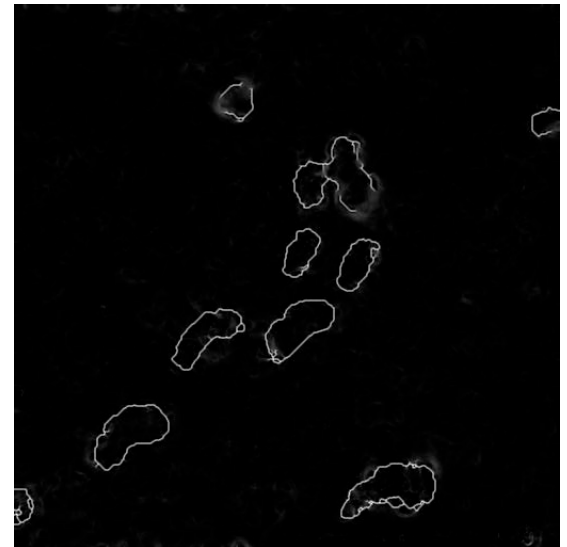
BA



Ours



Error map



Smoothness error

# FG: Flower Garden



**360x240 (Black)**

**Max speed: 7pix/frame**



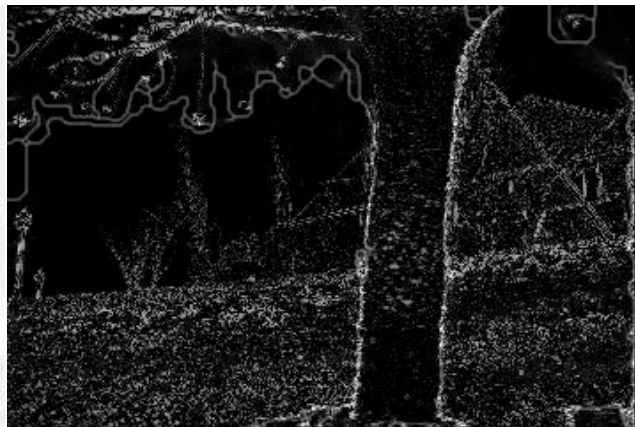
**BA**



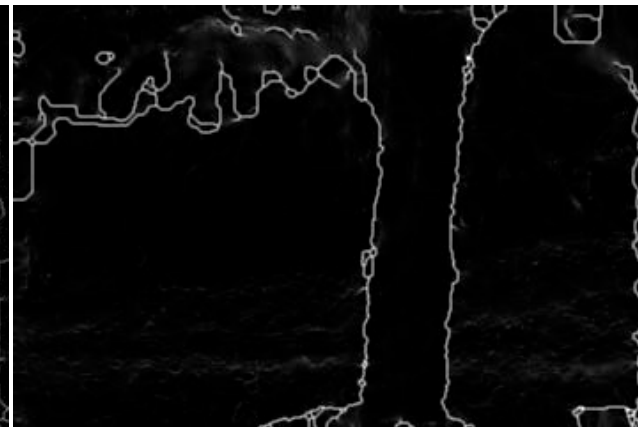
**LMS**



**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