



# Reminder: last lecture on stereo matching

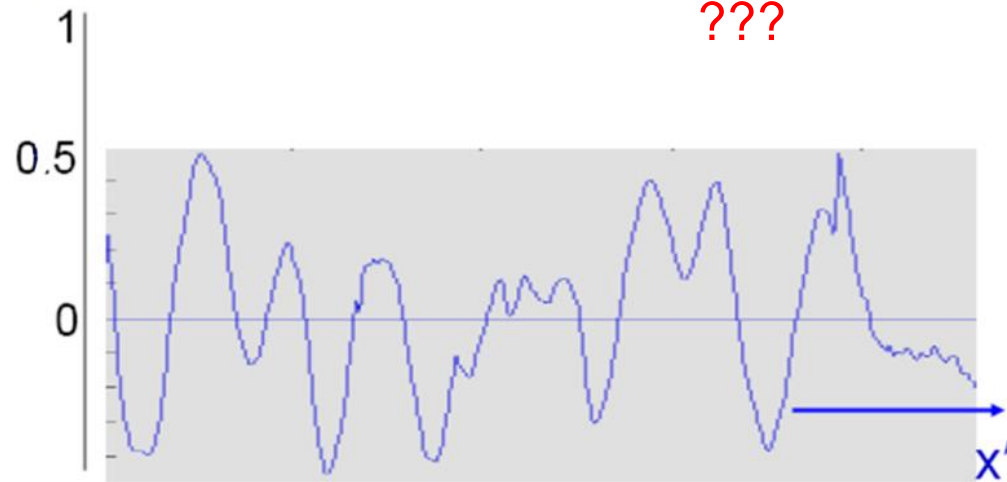


target region



left image band (x)

right image band (x')



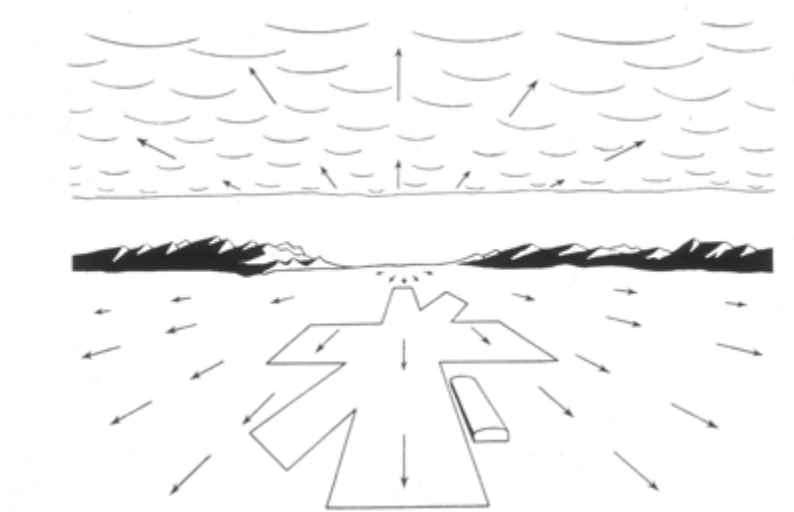
cross  
correlation

Textureless regions are  
non-distinct; high  
ambiguity for matches.

# Computer Vision

## *Motion and Optical Flow*

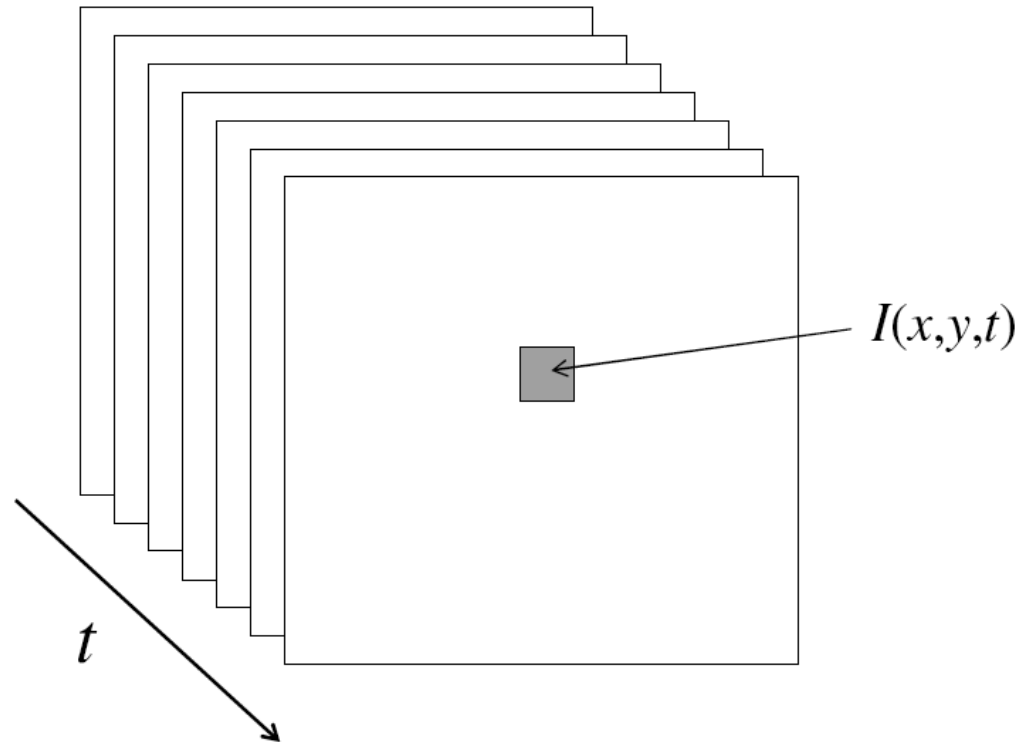
---



Many slides adapted from S. Seitz, R. Szeliski, M. Pollefeys, K. Grauman and others...

# Video

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



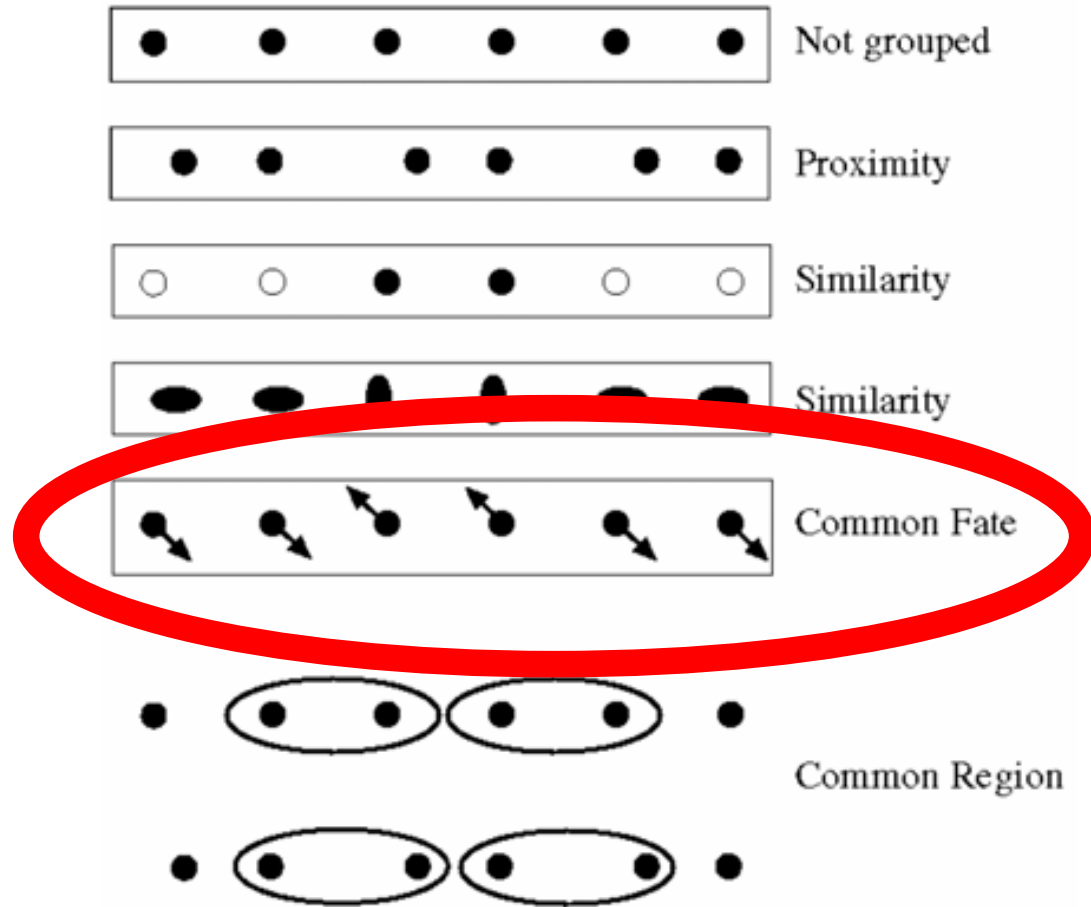
# Motion and perceptual organization



Gestalt psychology  
(Max Wertheimer,  
1880-1943)

# Motion and perceptual organization

- Sometimes, motion is the only cue



Gestalt psychology  
(Max Wertheimer,  
1880-1943)

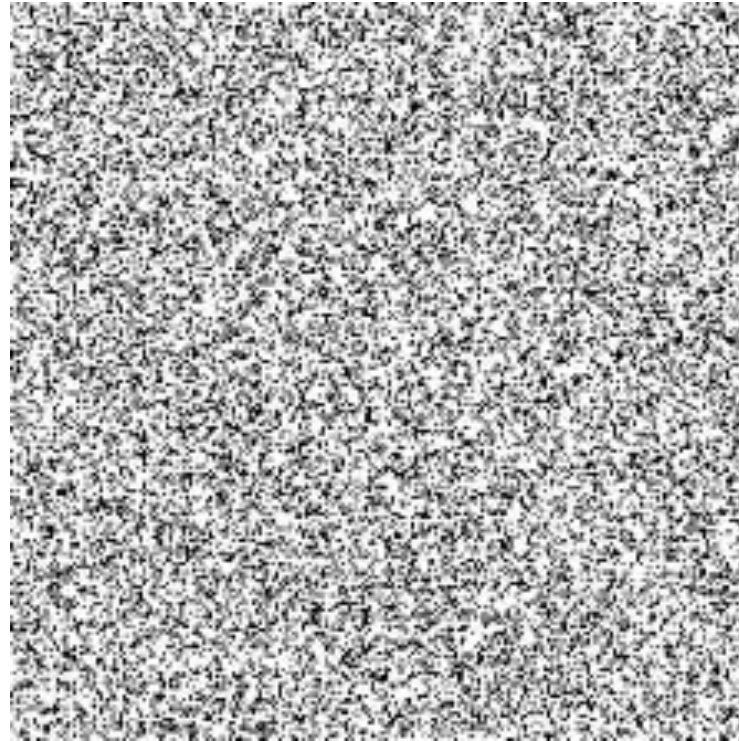


# Motion and perceptual organization

- Sometimes, motion is the only cue

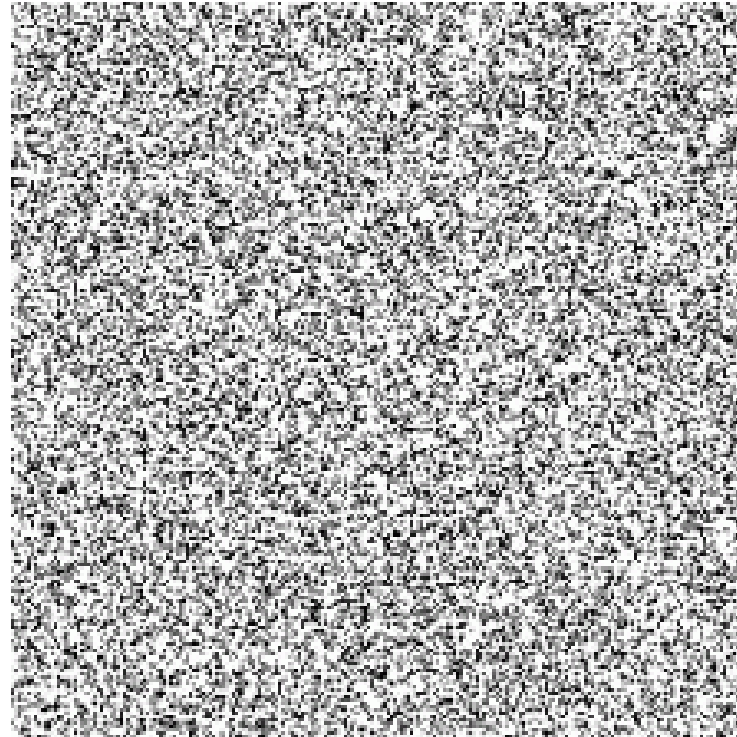
# Motion and perceptual organization

- Sometimes, motion is the only cue



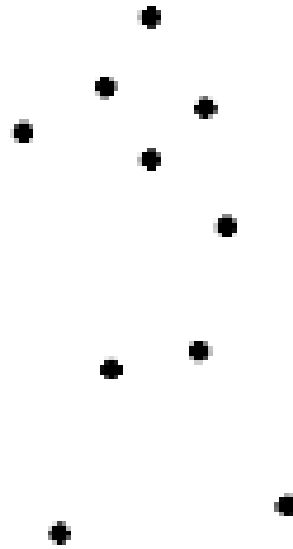
# Motion and perceptual organization

- Sometimes, motion is the only cue



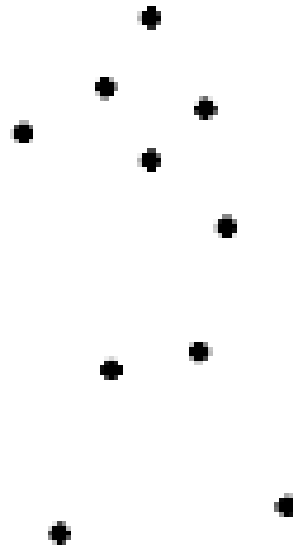
# Motion and perceptual organization

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



# Motion and perceptual organization

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



# Motion and perceptual organization

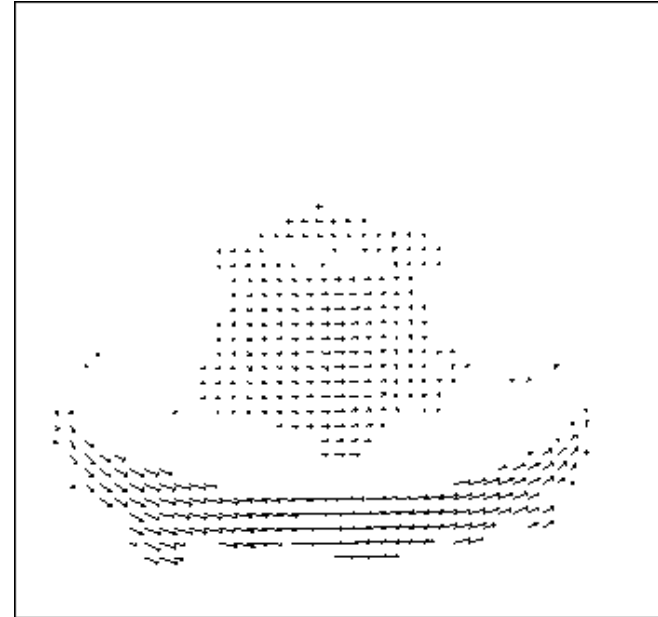
Animation from:  
Heider, F. & Simmel, M. (1944).  
An experimental study of apparent behavior.  
*American Journal of Psychology*, 57, 243-259.

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

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

# Motion estimation: Optical flow

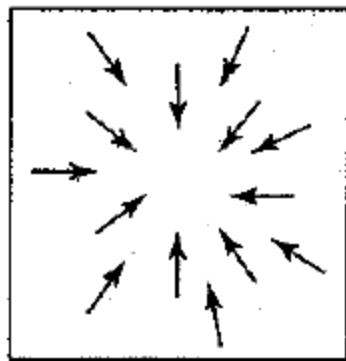
*Optical flow* is the **apparent** motion of objects or surfaces



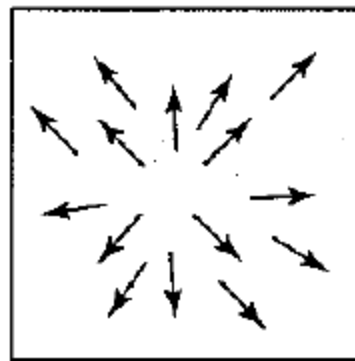
Will start by estimating motion of each pixel separately  
Then will consider motion of entire image

The term “scene flow” is used to describe 3d motion estimation

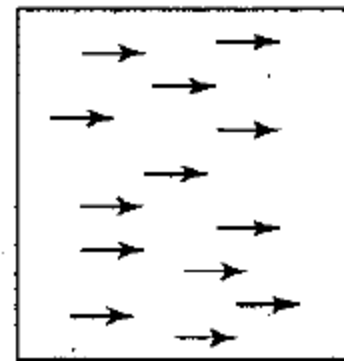
## Motion field + camera motion



**Zoom out**



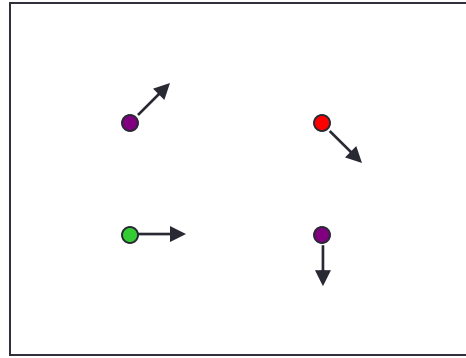
**Zoom in**



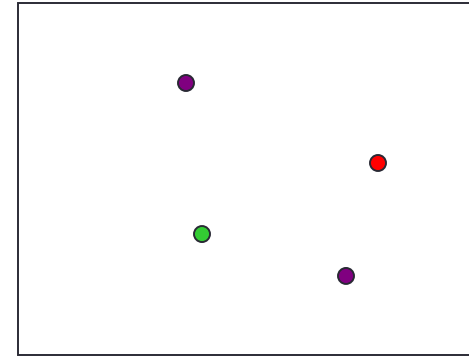
**Pan right to  
left**



# Problem definition: optical flow



$I(x, y, t)$



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

How to estimate pixel motion from image  $I(x, y, t)$  to  $I(x, y, t + 1)$  ?

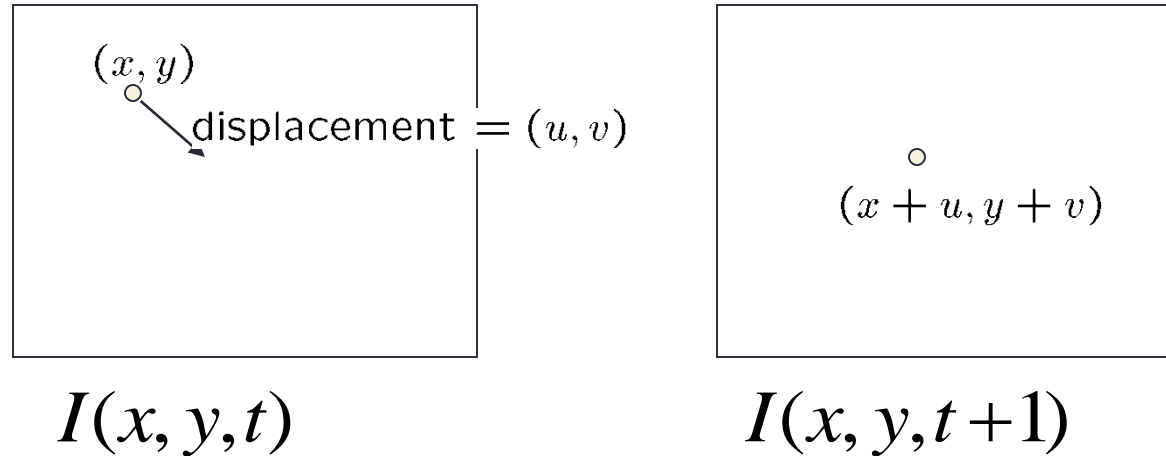
- Solve pixel correspondence problem
  - given a pixel in  $I(x, y, t)$ , look for nearby pixels of the same color in  $I(x, y, t + 1)$

Key assumptions

- **color constancy**: a point in  $I(x, y, t)$  looks the same in  $I(x, y, t + 1)$ 
  - For grayscale images, this is brightness constancy
- **small motion**: points do not move very far

This is called the optical flow problem

# Optical flow constraints (grayscale images)



- 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 smooth)

Taylor series expansion of the spatial changes 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

$$\begin{aligned} 0 &= I(x + u, y + v, t + 1) - I(x, y, t) \\ &\approx I(x, y, t + 1) + I_x u + I_y v - I(x, y, t) \end{aligned}$$

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

# Optical flow equation

- Combining these two equations

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

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

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

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

$$\approx I_t + I_x u + I_y v$$

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

# 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*

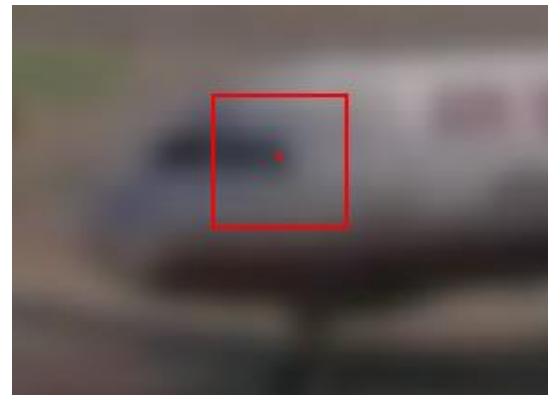
$$I_x u + I_y v + I_t = 0$$

# 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?



If I told you  
 $I_t$  is -5  
 $I_x$  is 2.5  
 $I_y$  is 0

What was  
the pixel  
shift  $(u,v)$ ?

# The brightness constancy constraint

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

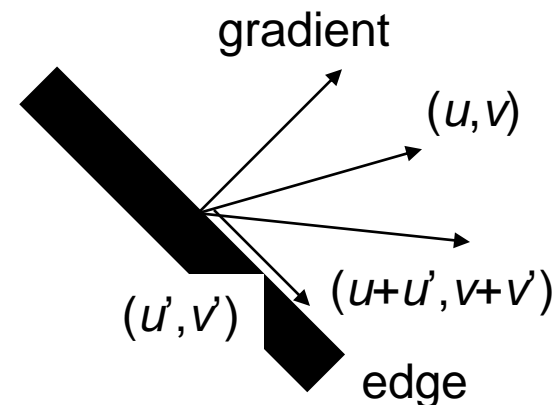
$$0 = I_t + \nabla I \cdot \langle u, v \rangle \quad \text{or} \quad I_x u + I_y v + I_t = 0$$

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

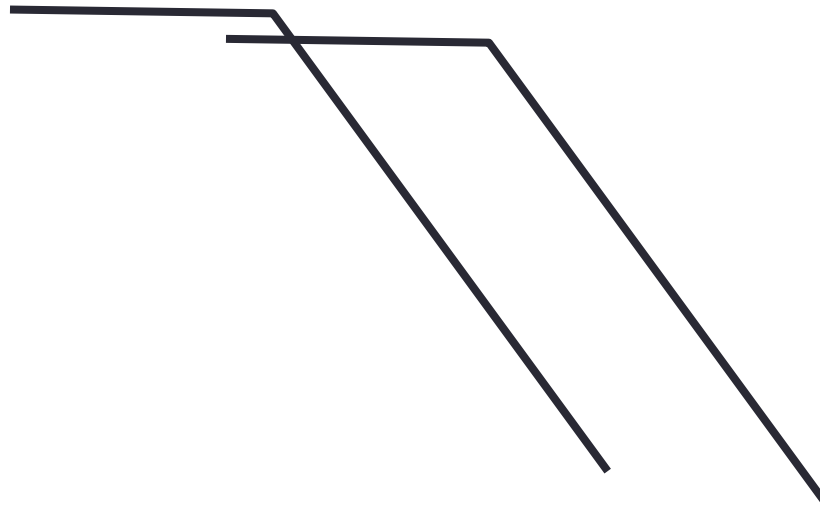
The component of the motion perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

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

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

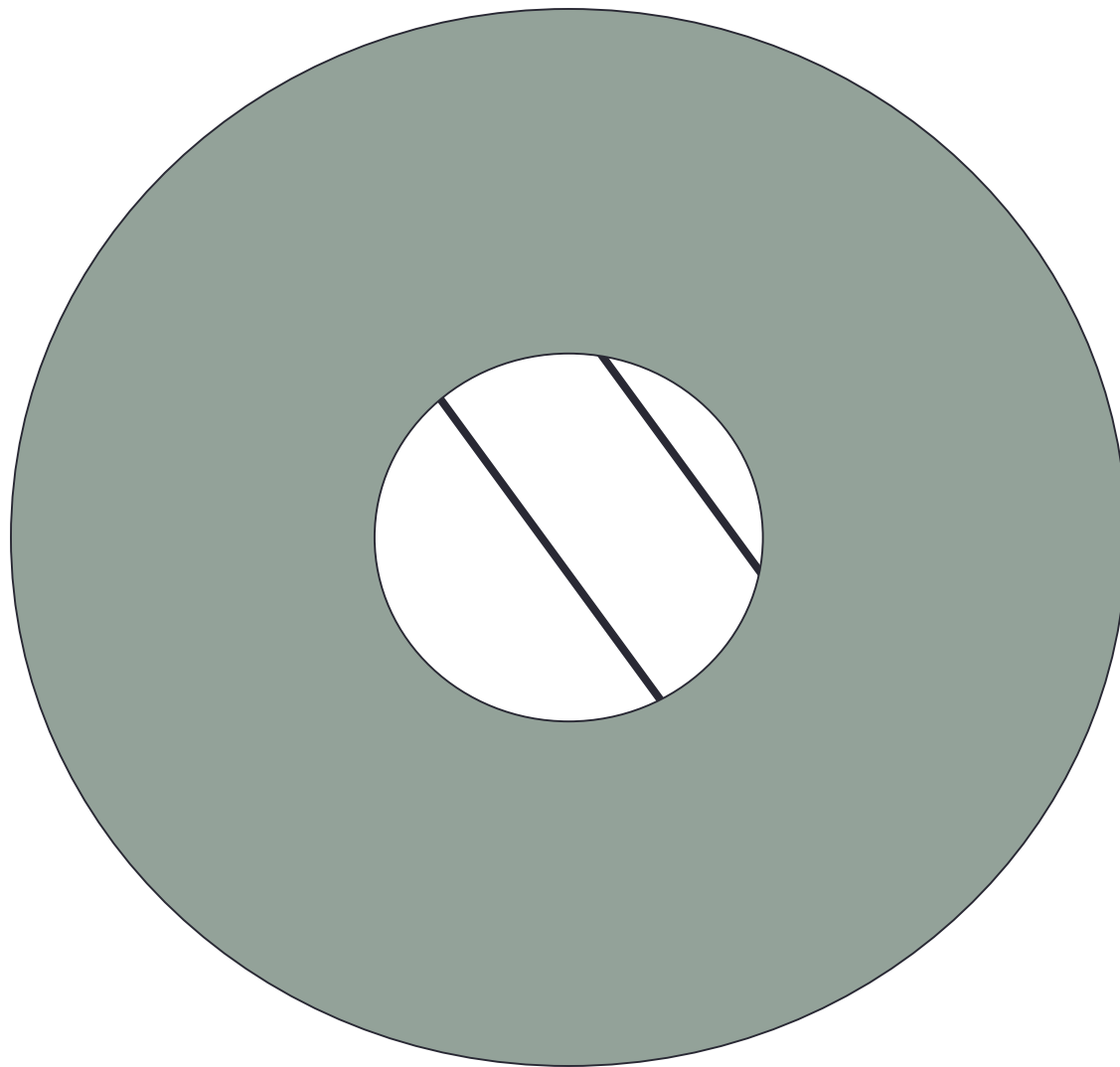


# Aperture problem

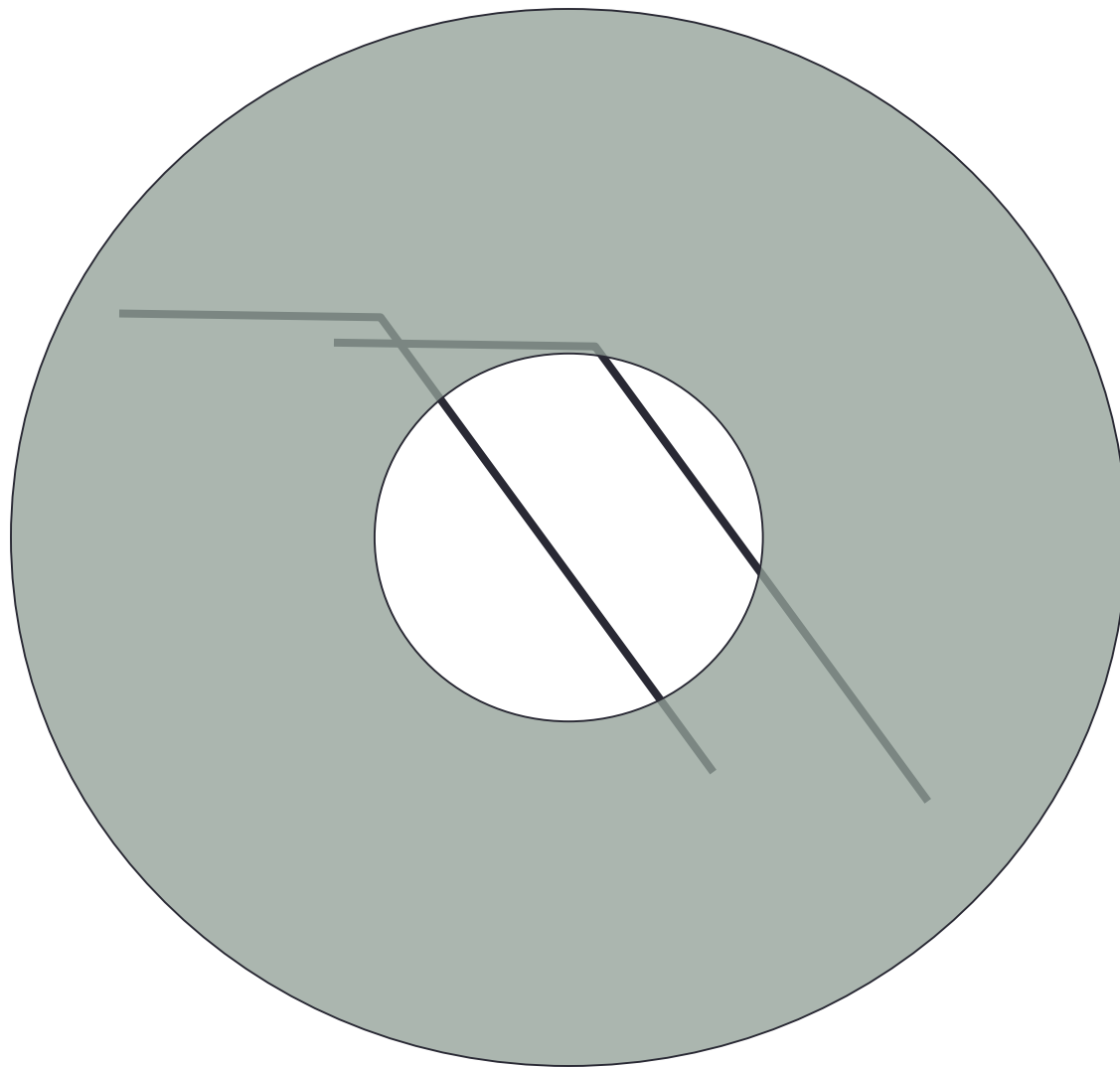




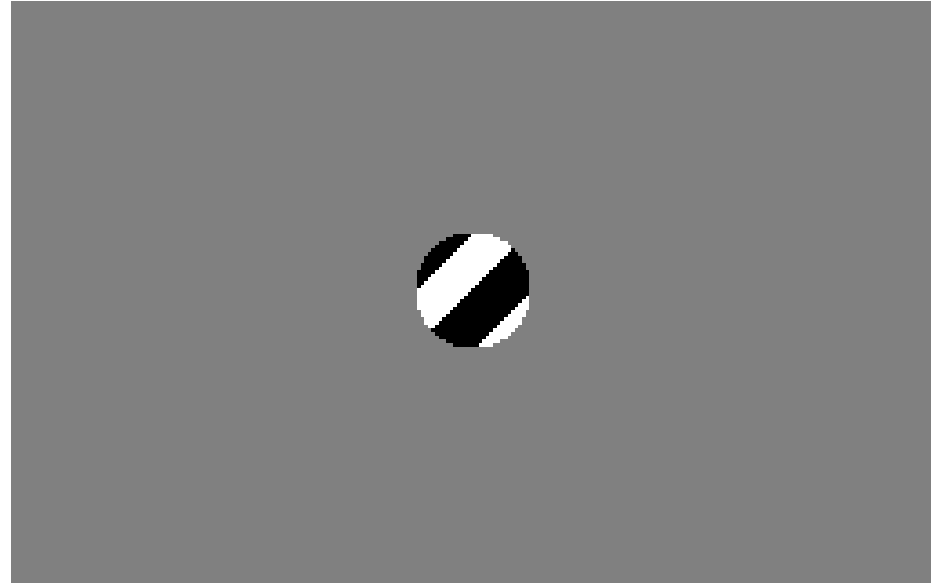
# Aperture problem



# Aperture problem

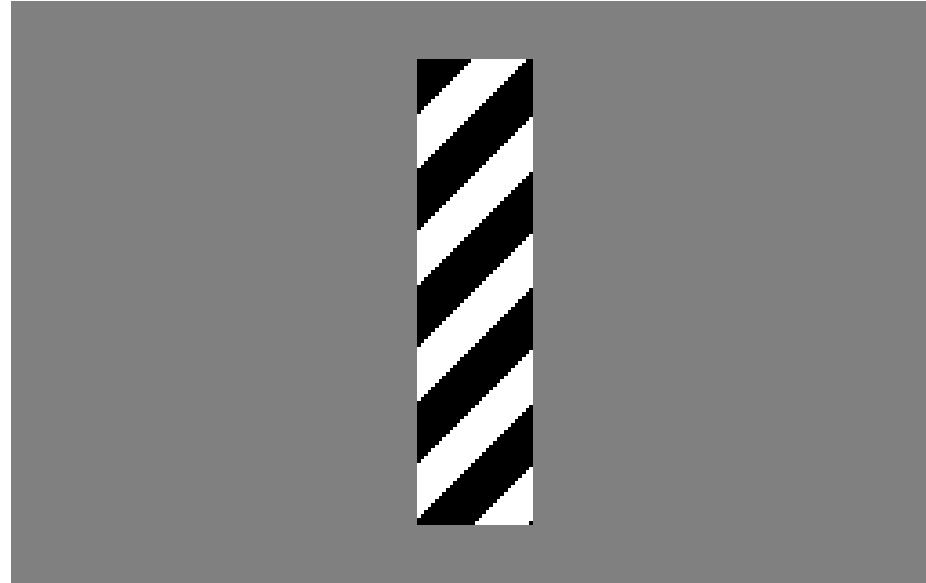


# The barber pole illusion



[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?

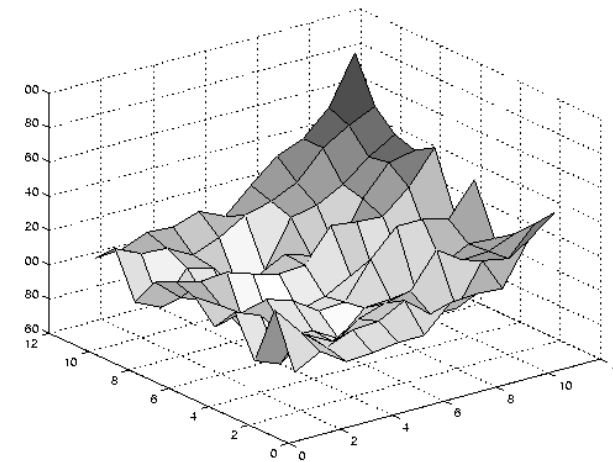
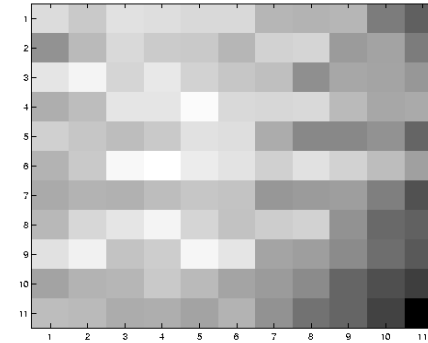
- $A^T A$  should be invertible
- $A^T A$  should not be too small due to noise
  - eigenvalues  $\lambda_1$  and  $\lambda_2$  of  $A^T A$  should not be too small
- $A^T A$  should be well-conditioned
  - $\lambda_1 / \lambda_2$  should not be too large ( $\lambda_1 =$  larger eigenvalue)

Does this remind you of anything?

Criteria for Harris corner detector



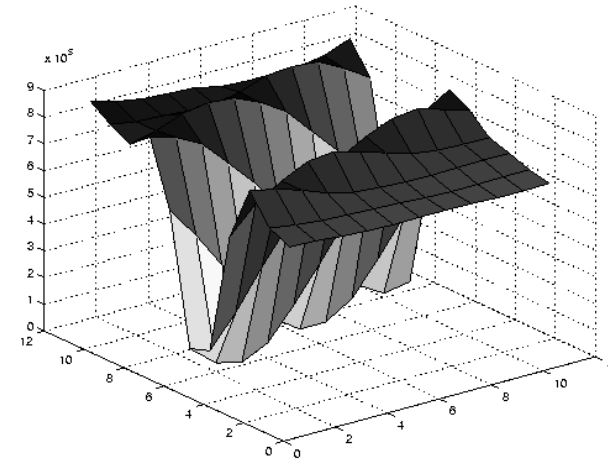
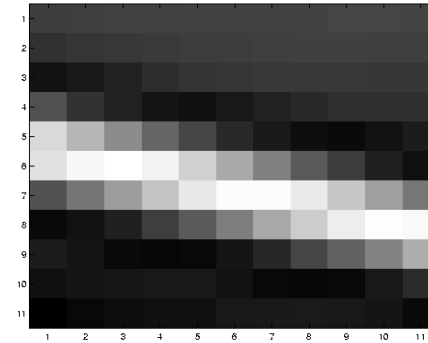
# Low texture region



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

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

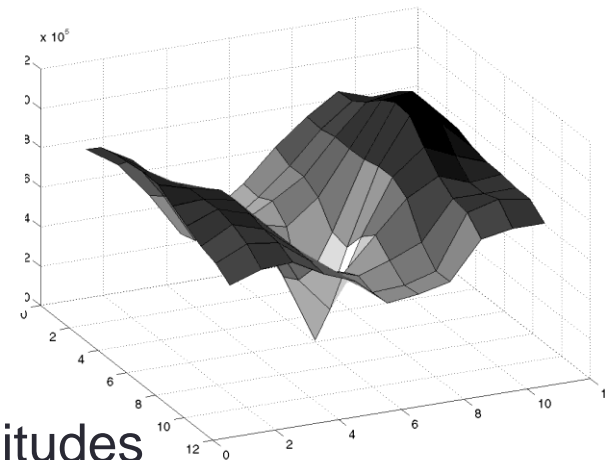
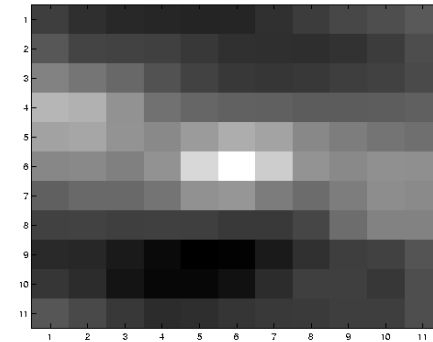
# Edge



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

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

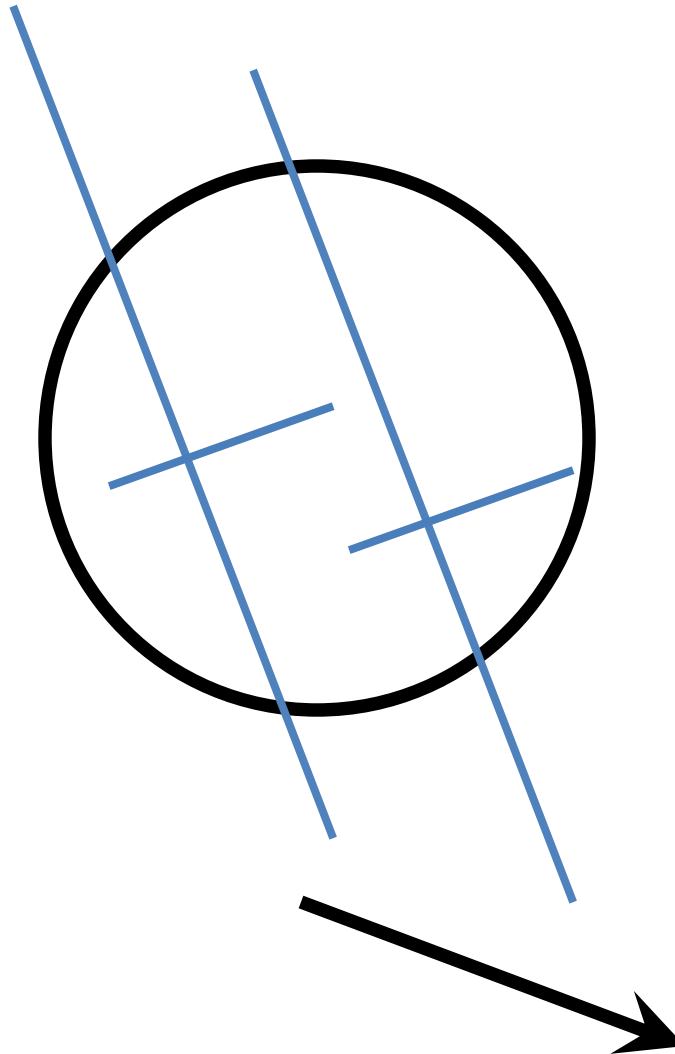
# High textured region



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

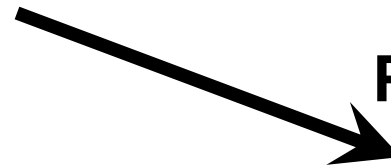
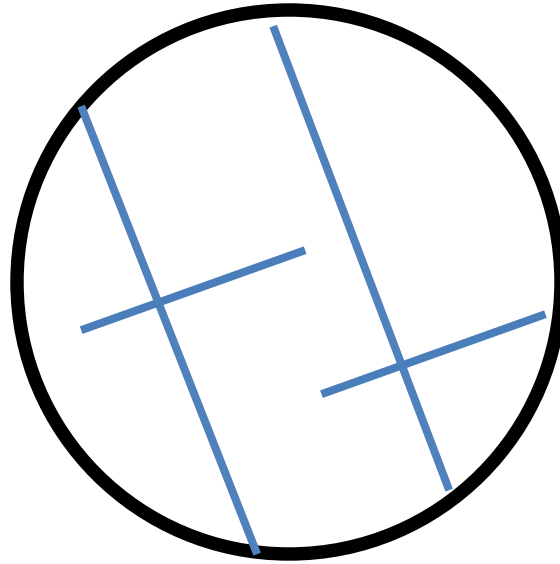
- gradients are different, large magnitudes
- large  $\lambda_1$ , large  $\lambda_2$

# The aperture problem resolved



**Actual motion**

# The aperture problem resolved



**Perceived motion**

# Revisiting the small motion assumption

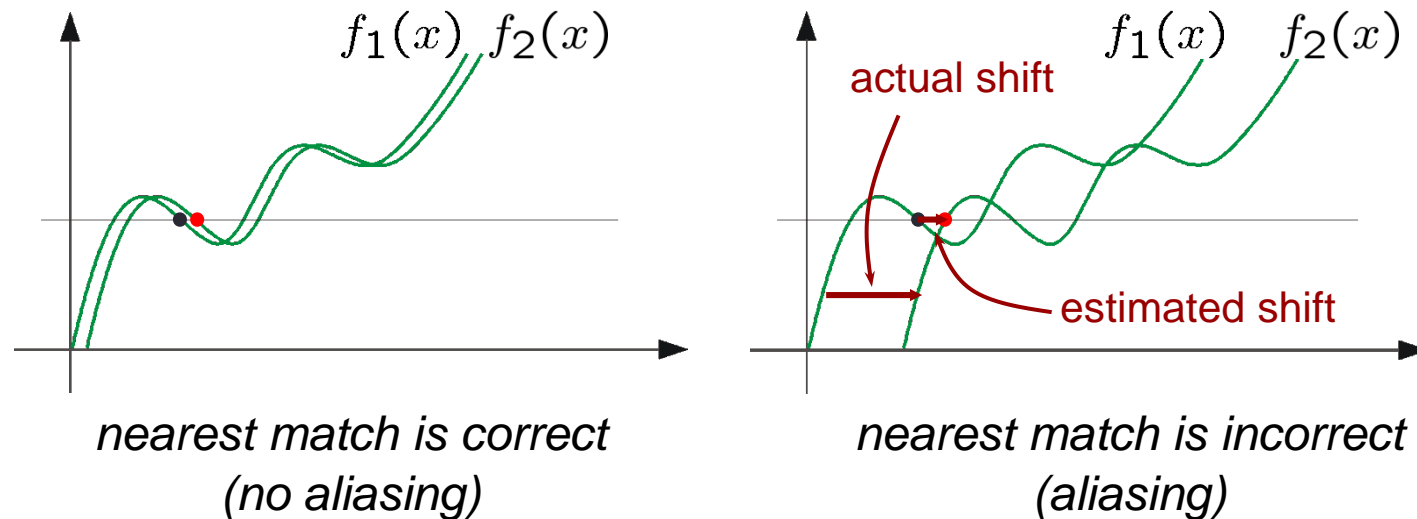


- Is this motion small enough?
  - Probably not—it's much larger than one pixel
  - How might we solve this problem?

# Optical Flow: Aliasing

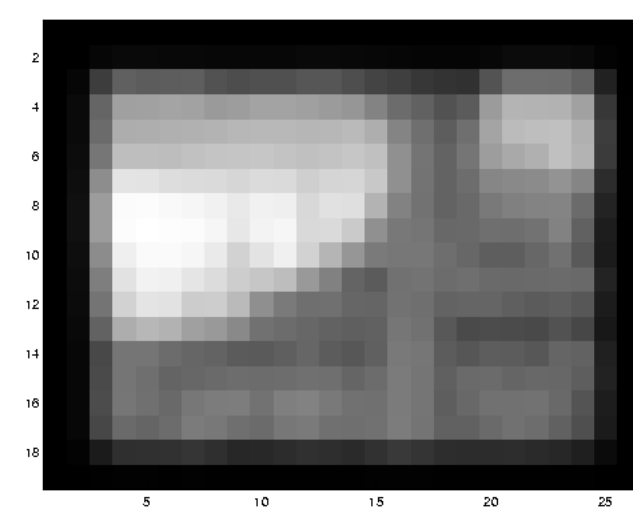
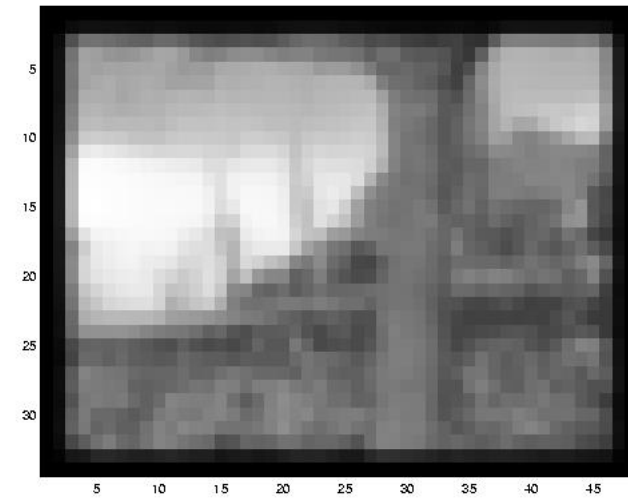
Temporal aliasing causes ambiguities in optical flow because images can have many pixels with the same intensity.

I.e., how do we know which 'correspondence' is correct?



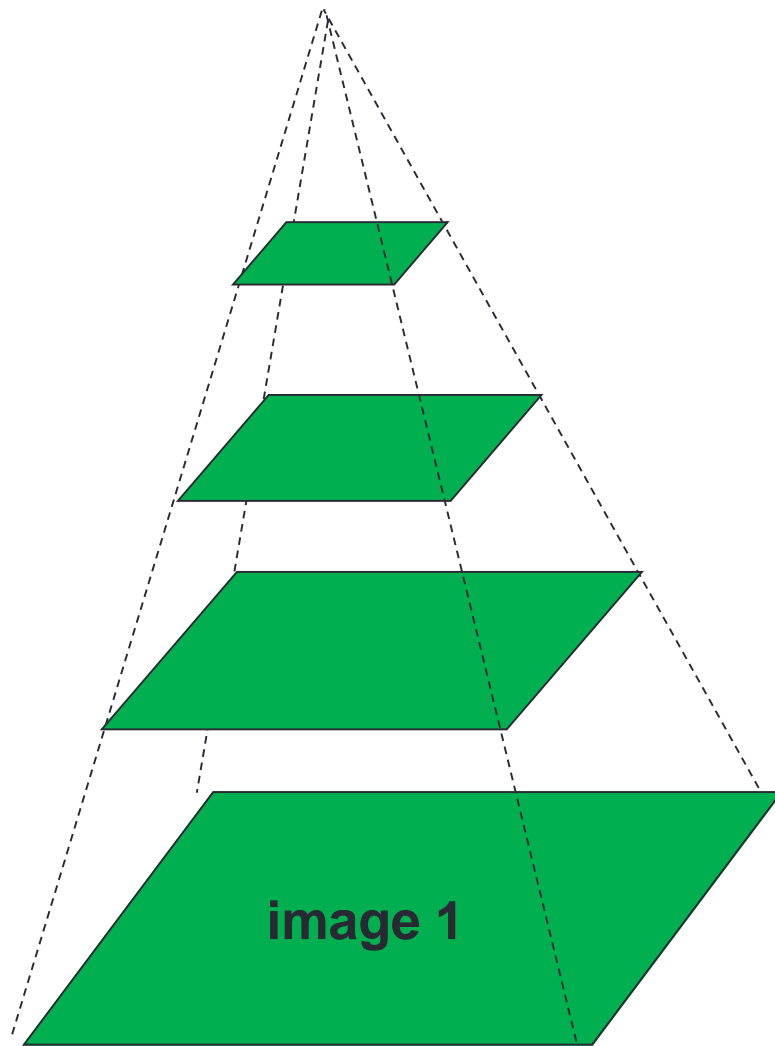
To overcome aliasing: coarse-to-fine estimation.

# Reduce the resolution!





# Coarse-to-fine optical flow estimation



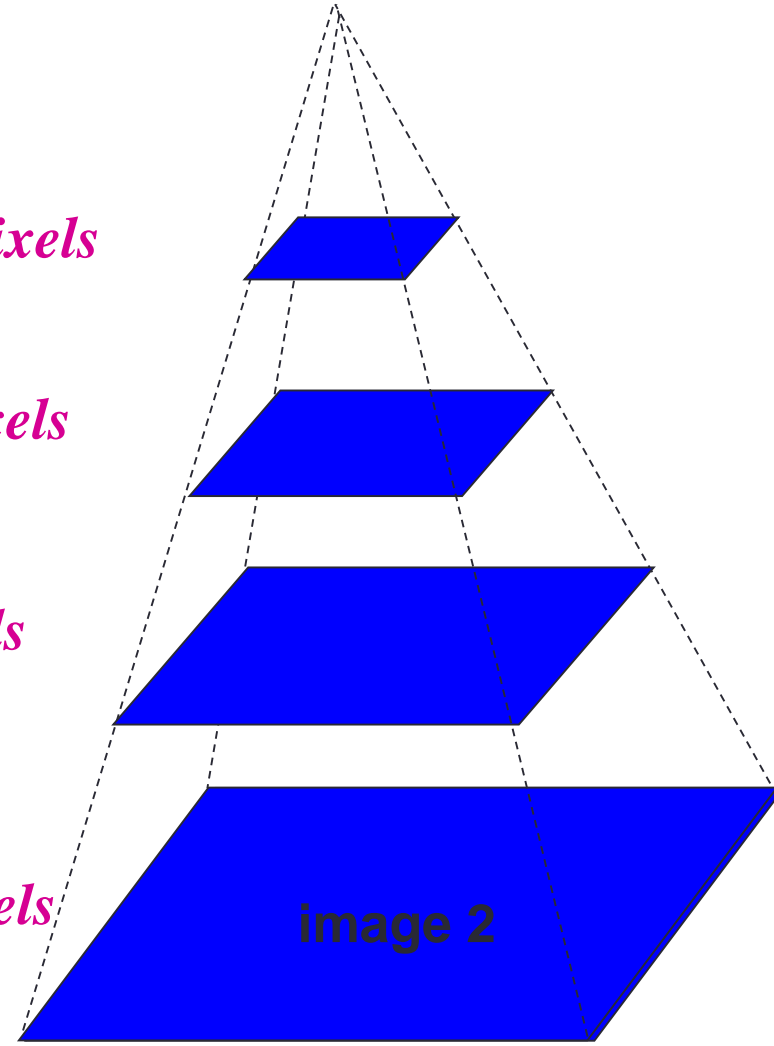
Gaussian pyramid of image 1

*$u=1.25$  pixels*

*$u=2.5$  pixels*

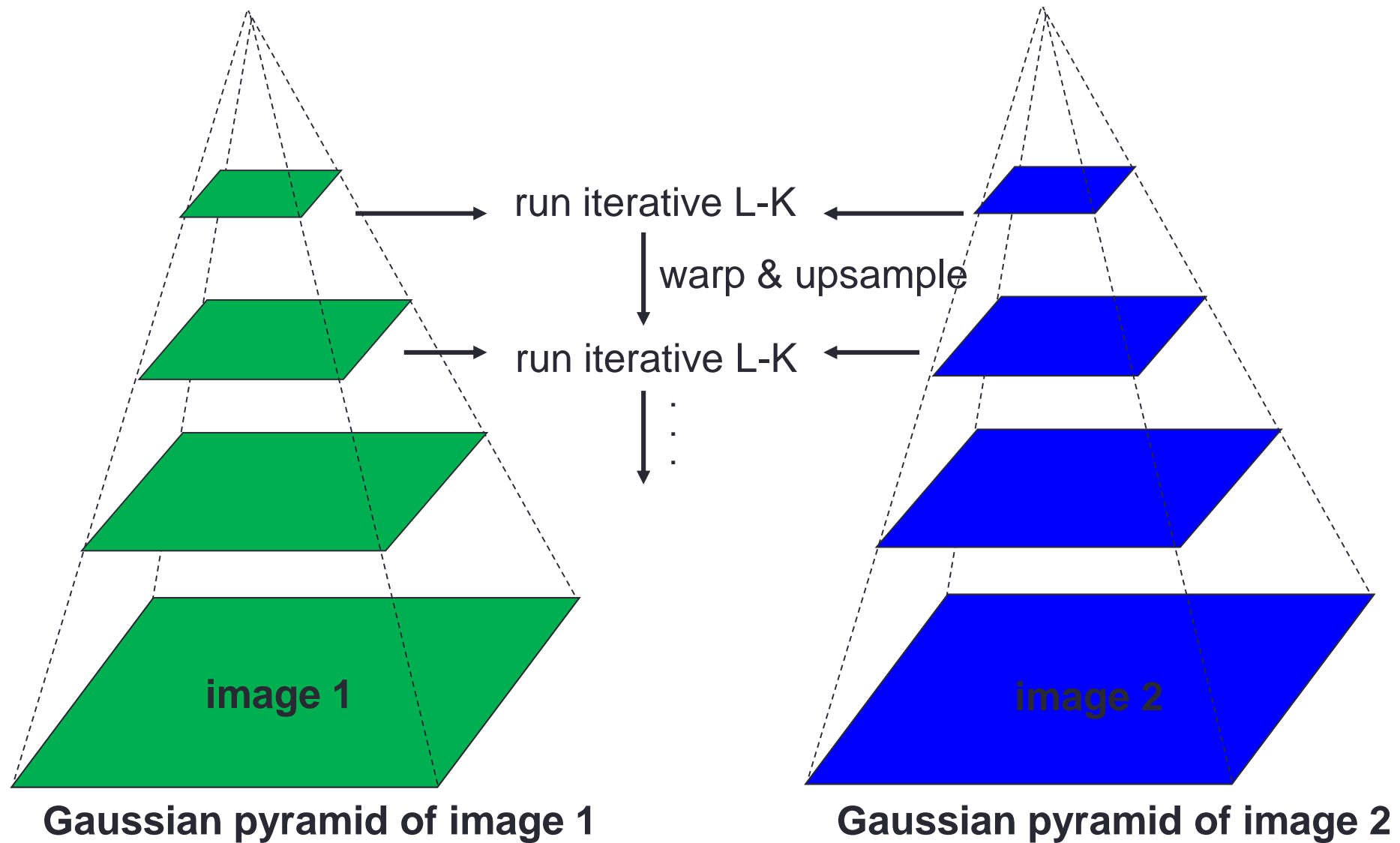
*$u=5$  pixels*

*$u=10$  pixels*

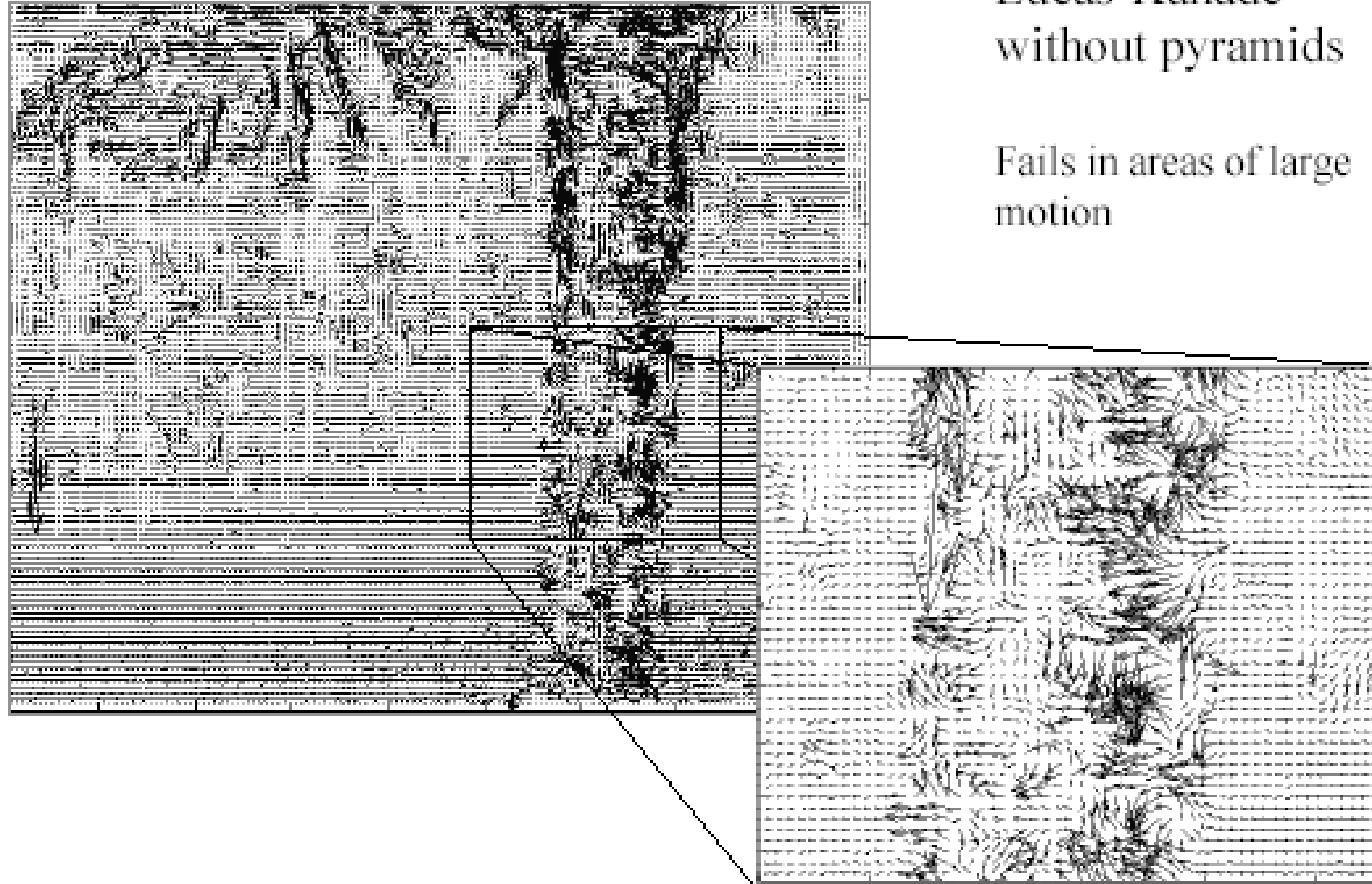


Gaussian pyramid of image 2

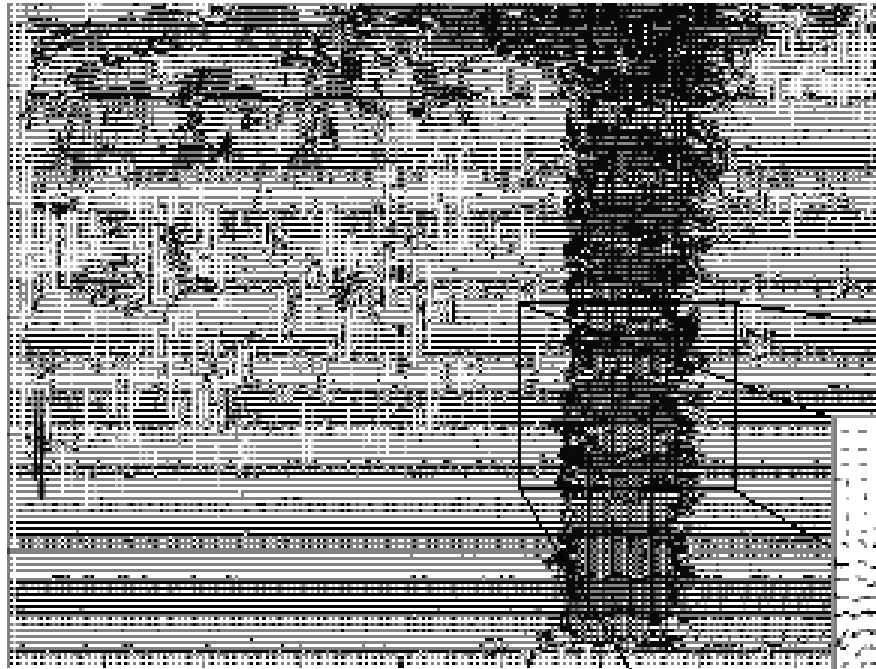
# Coarse-to-fine optical flow estimation



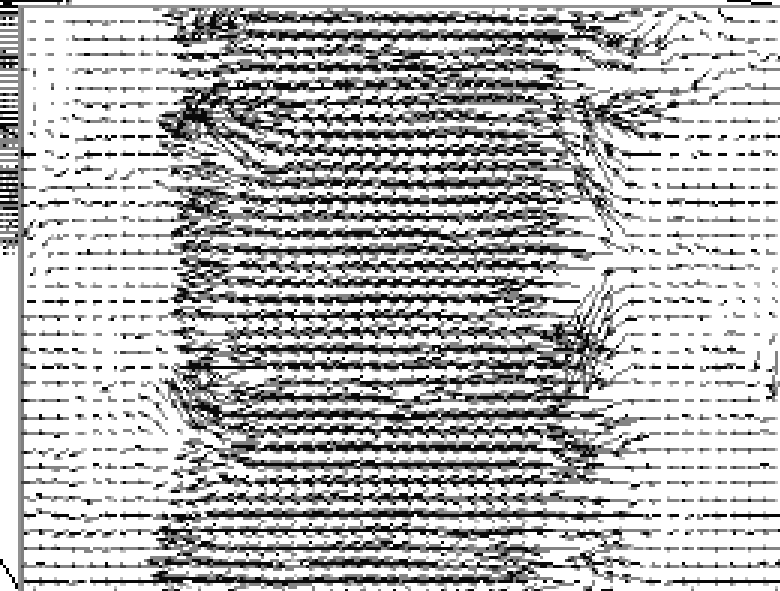
# Optical Flow Results



# Optical Flow Results



Lucas-Kanade with Pyramids



# State-of-the-art optical flow in 2009

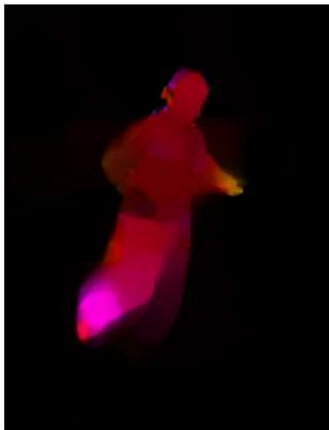
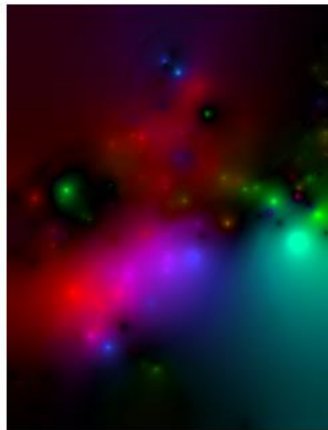
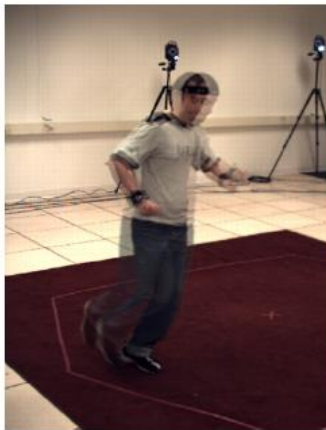
Start with something similar to Lucas-Kanade

+ gradient constancy

+ energy minimization with smoothing term

+ region matching

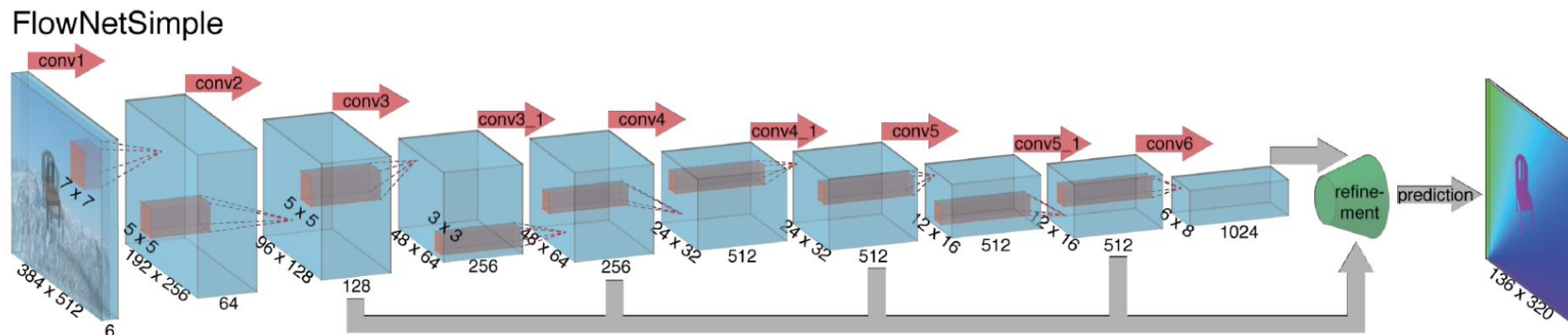
+ keypoint matching (long-range)



Region-based +Pixel-based +Keypoint-based

# State-of-the-art optical flow in 2015

Deep convolutional network which accepts a pair of input frames and upsamples the estimated flow back to input resolution. Very fast because of deep network, near the state-of-the-art in terms of end-point-error.



# Deep optical flow, 2015

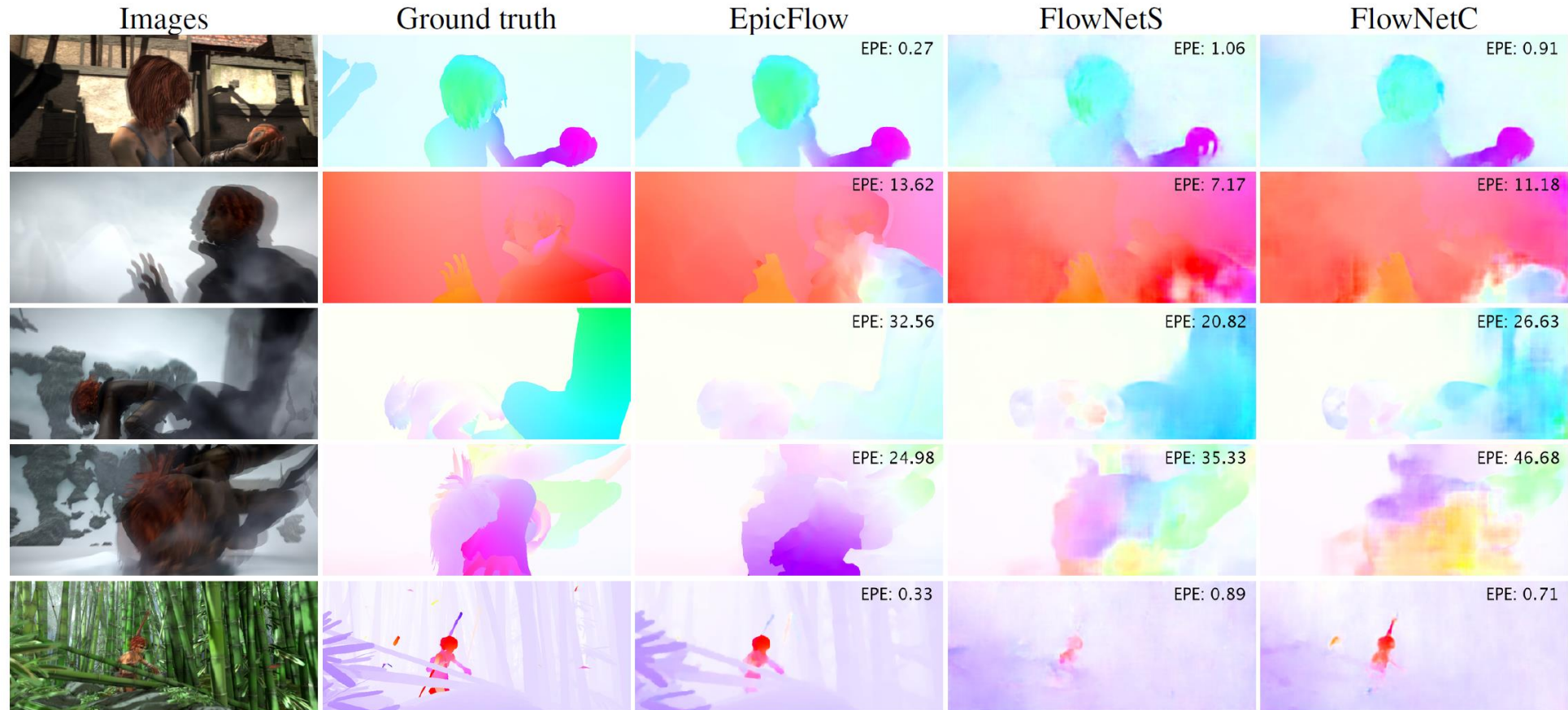
Synthetic Training data



Fischer et al. 2015. <https://arxiv.org/abs/1504.06852>

# Deep optical flow, 2015

## Results on Sintel





# 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