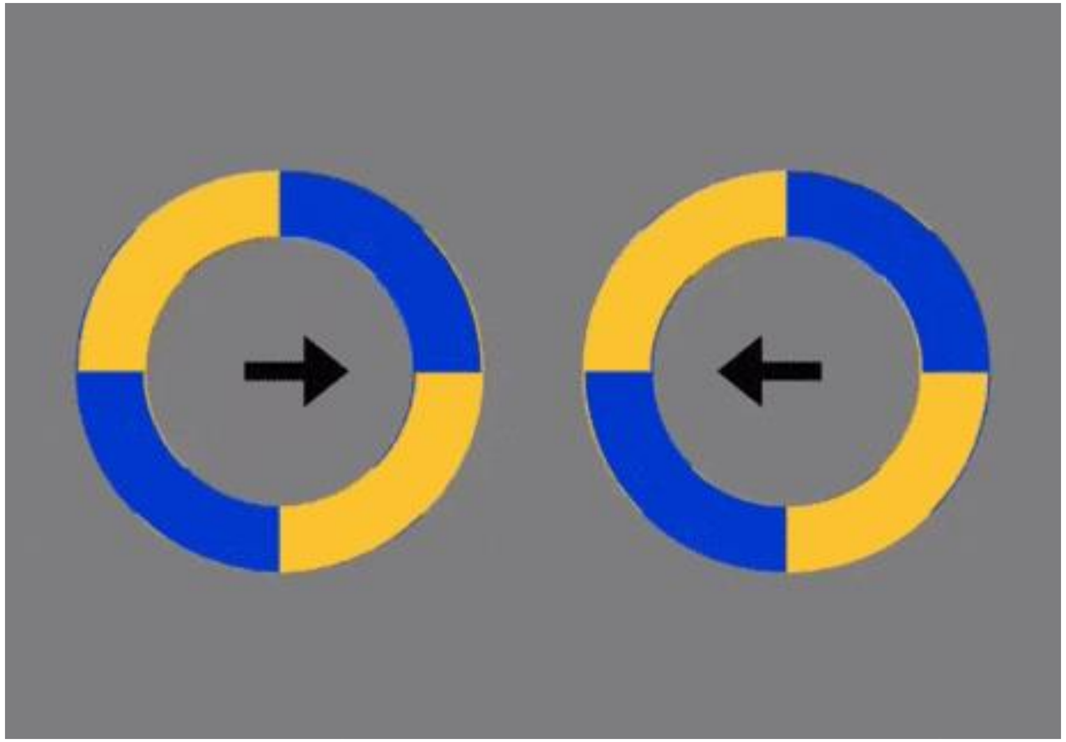
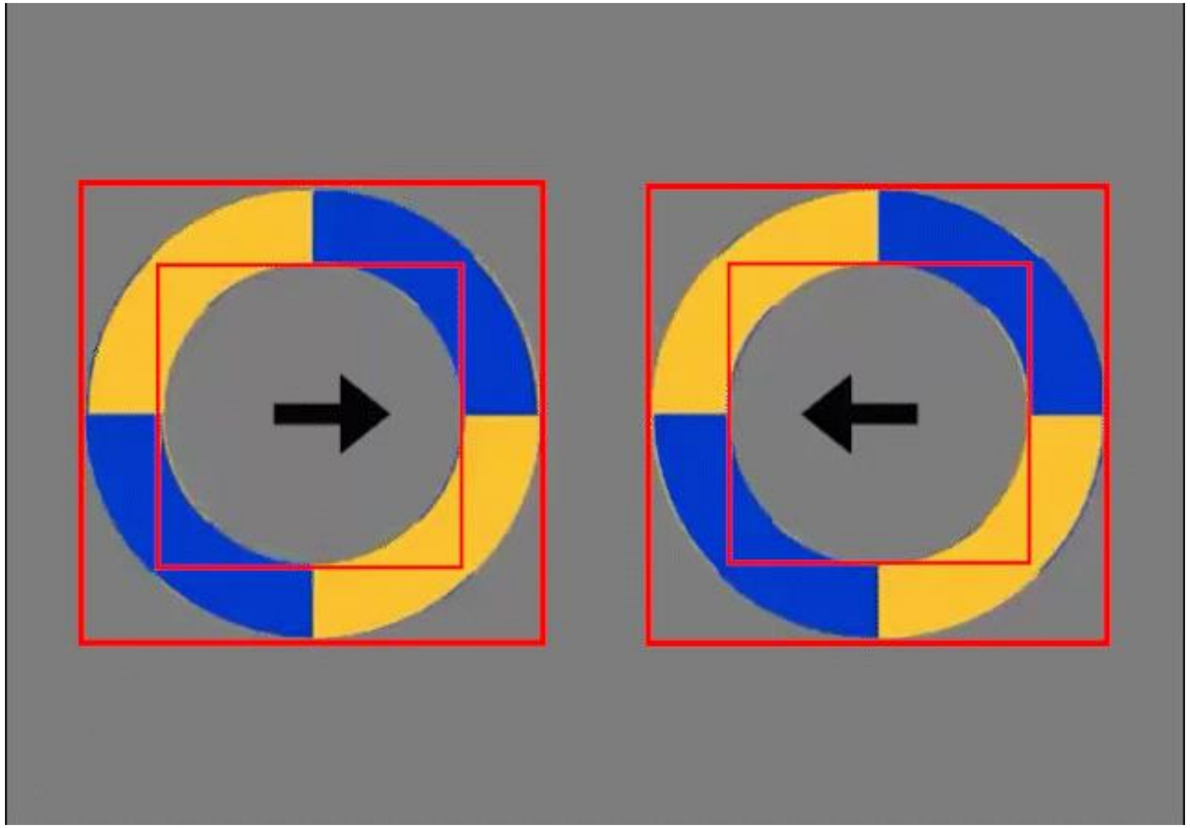
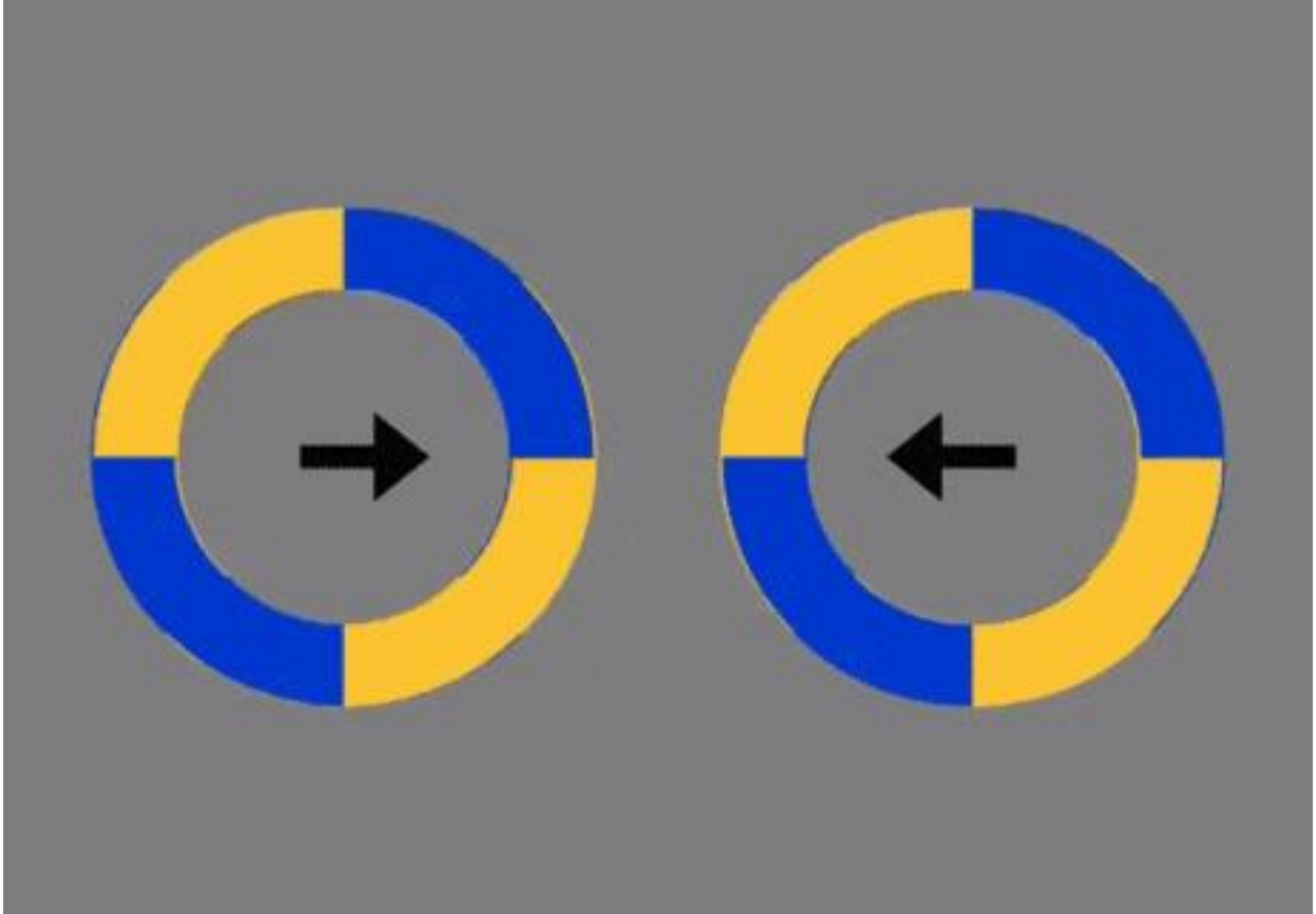
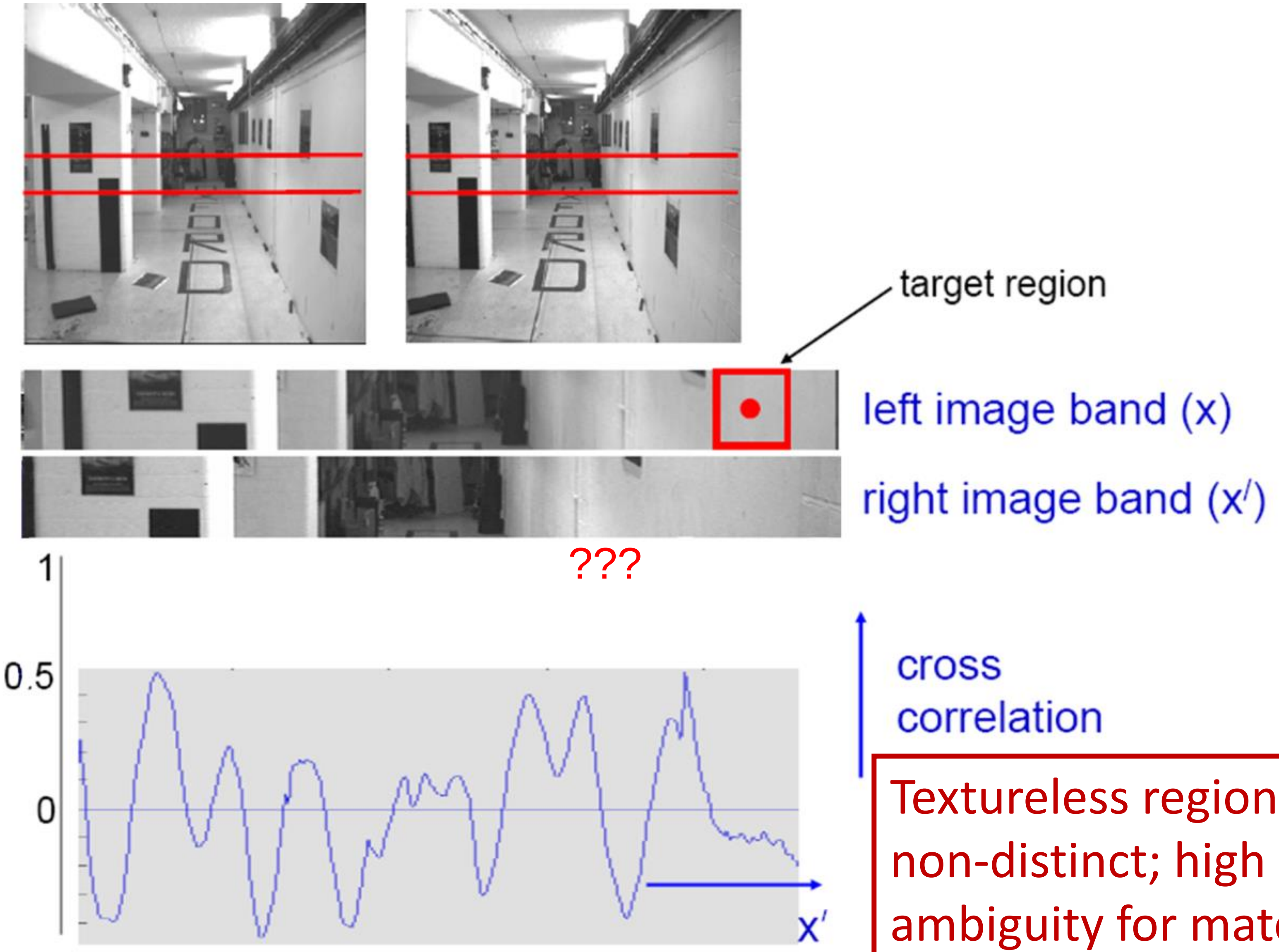




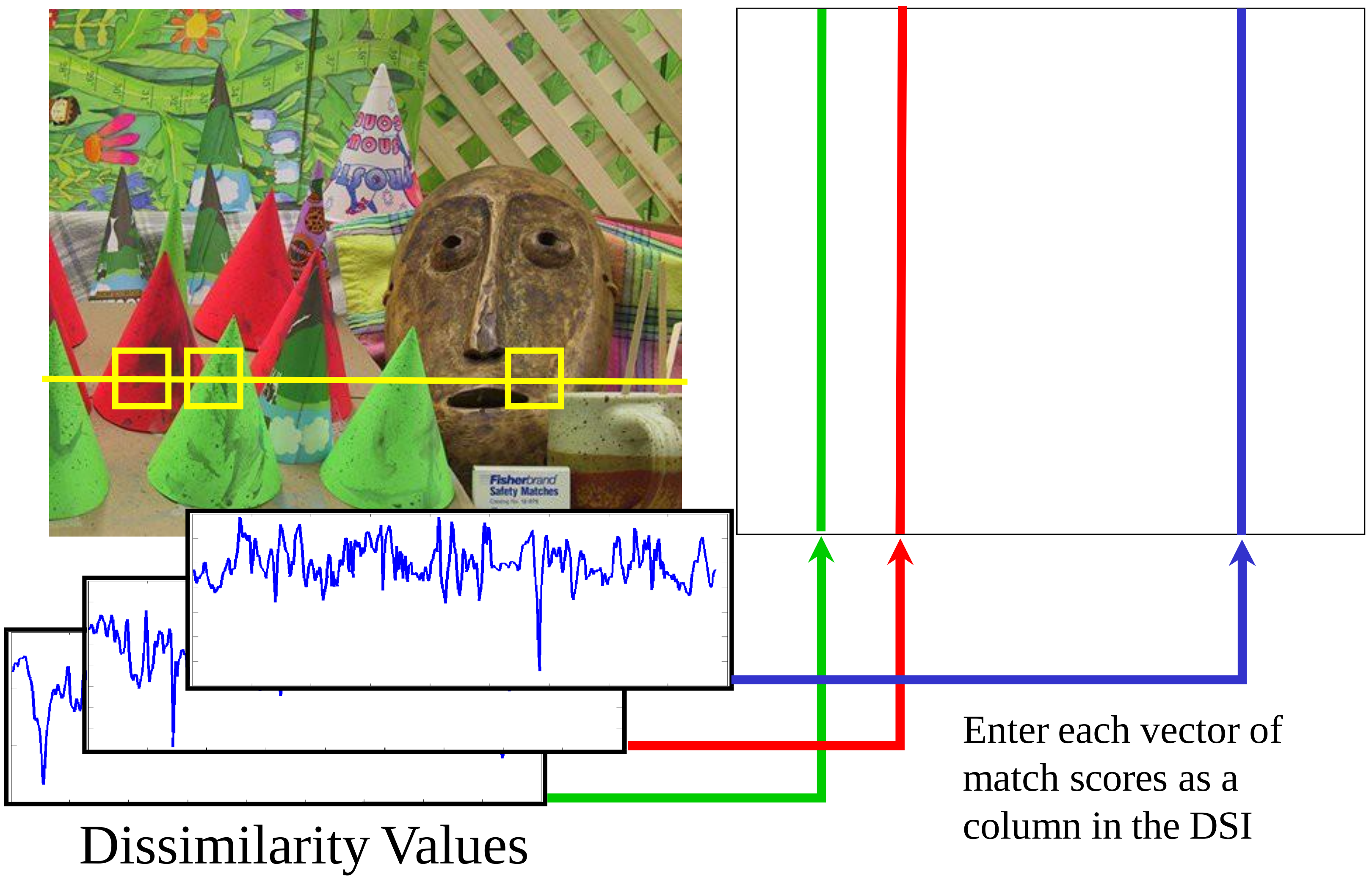


Review: last lecture on stereo matching



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

Review: last lecture on stereo matching



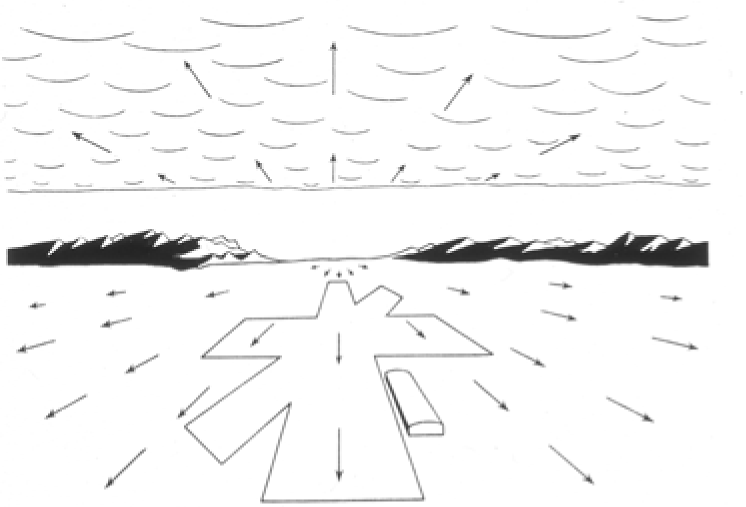
- Fight back against ambiguity by considering matches jointly with a preference for smooth disparities
  - Scanline stereo uses dynamic programming to optimize each scanline (1d smoothness)
  - Graph cut formulations include 2d smoothness
  - Semi-global matches is a greedy optimization that considers 2d smoothness. Fast but still pretty accurate.

Review: last lecture on stereo matching

- Or simply remove the ambiguity by projecting structured light patterns onto the scene
  - Simplifies the correspondence problem
  - Replaces one camera in the stereo system with a projector

Computer Vision

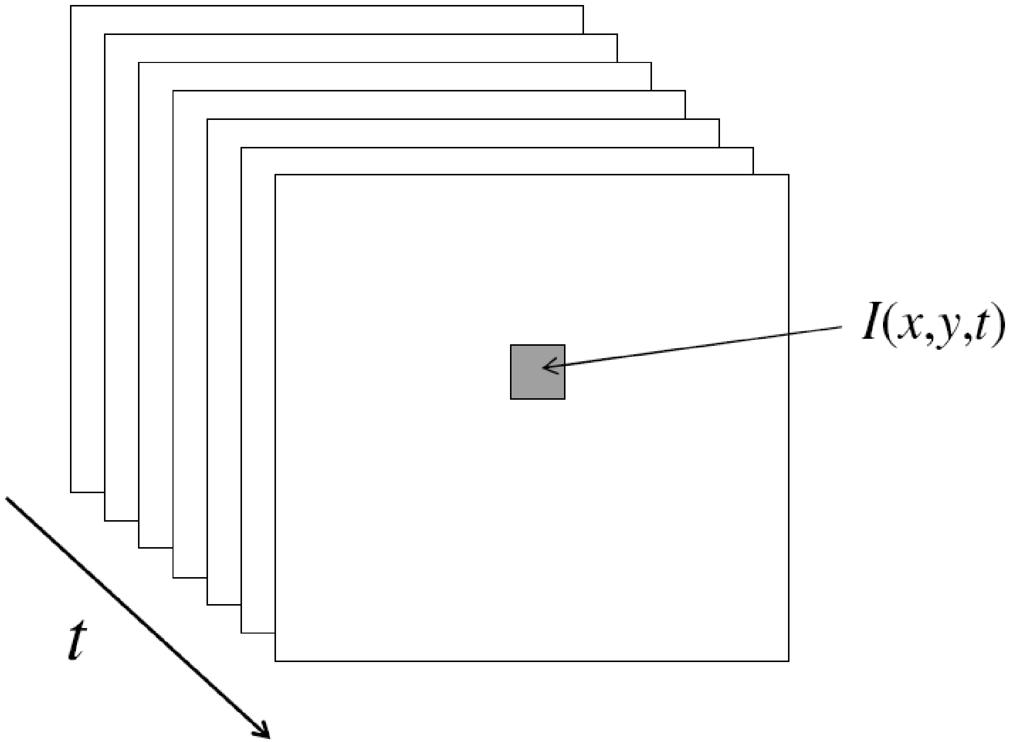
Motion and Optical Flow



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)

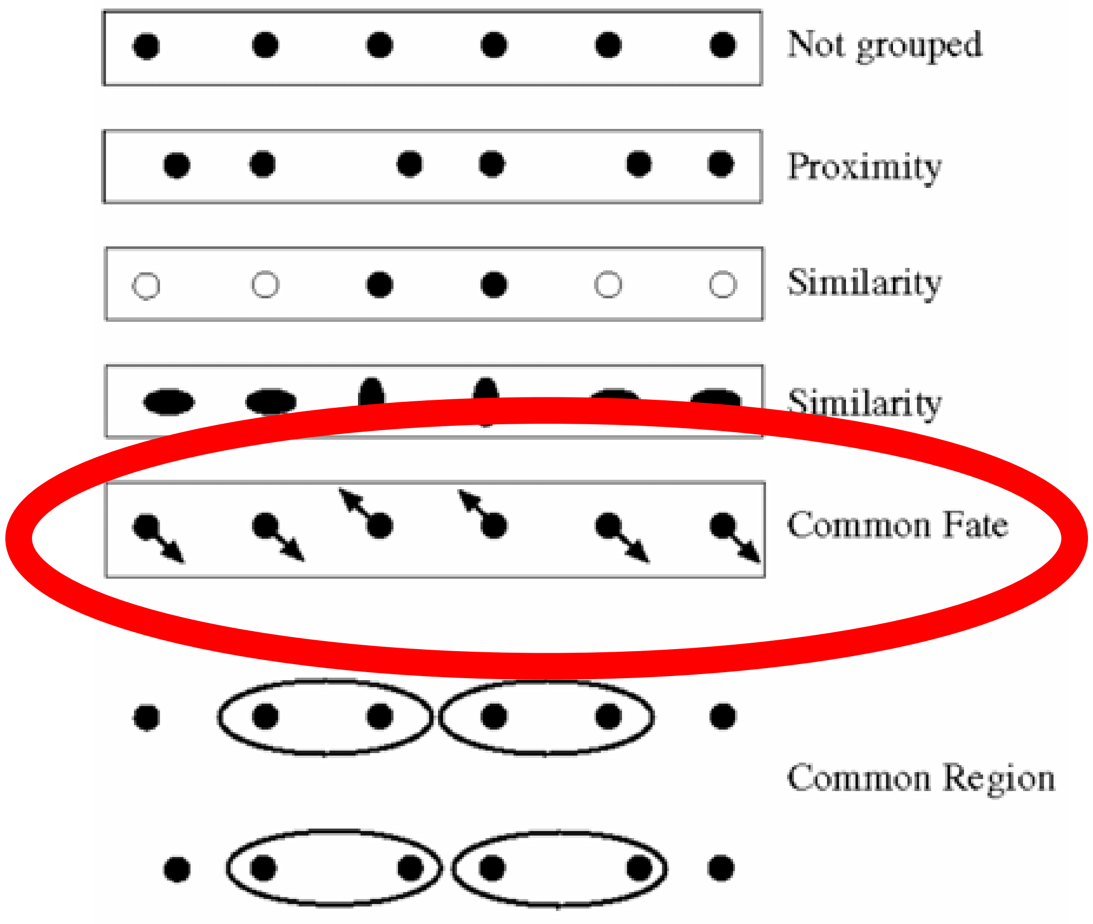


Motion and perceptual organization

Gestalt psychology (Max Wertheimer, 1880-1943)

Motion and perceptual organization

- Sometimes, motion is the only cue



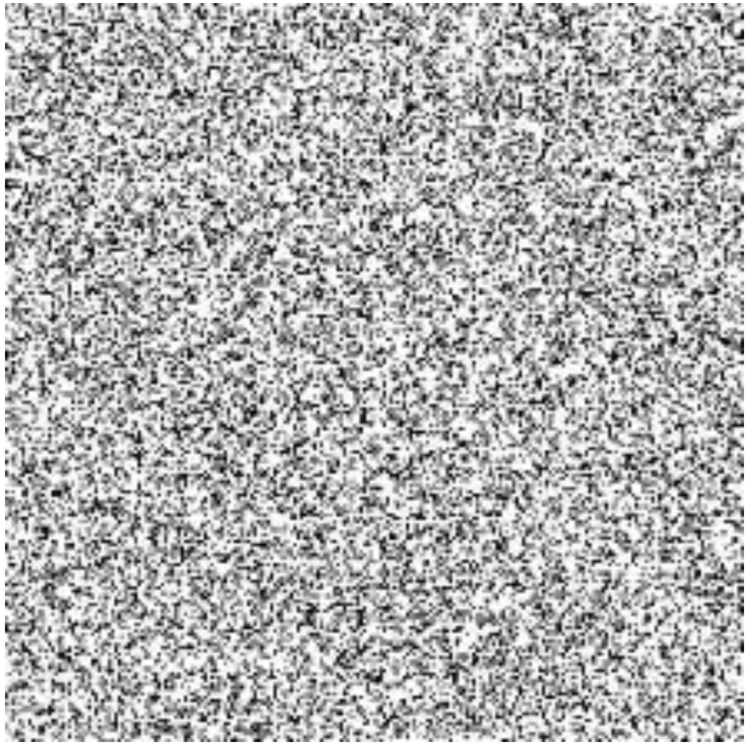
Gestalt psychology (Max Wertheimer, 1880-1943)

Motion and perceptual organization

- Sometimes, motion is the only cue

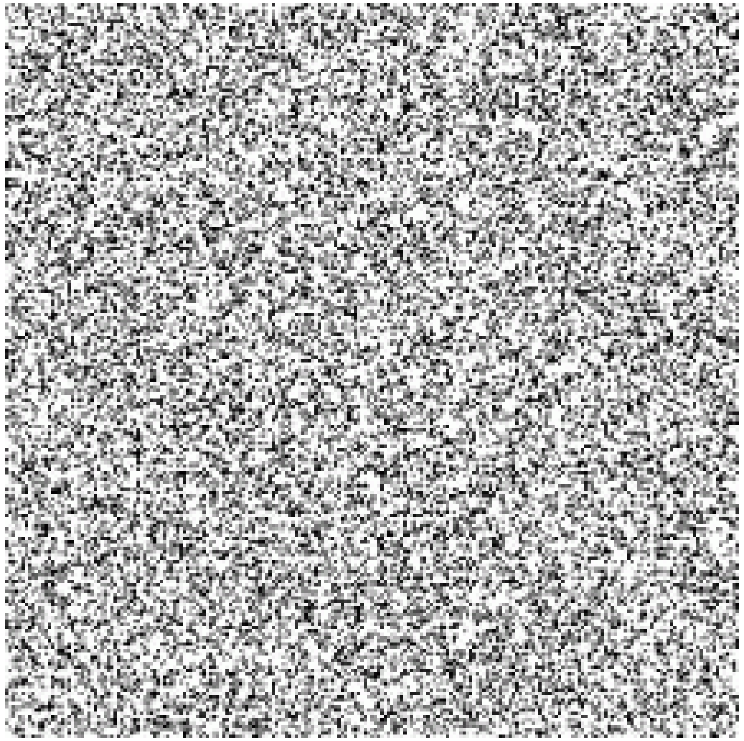
Motion and perceptual organization

- Sometimes, motion is the only cue



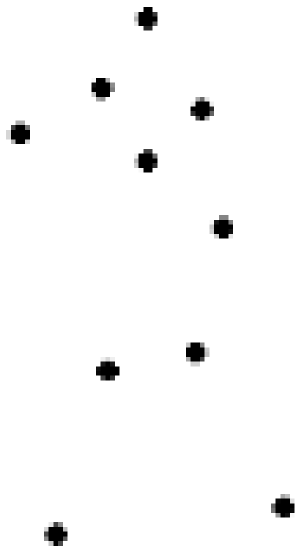
Motion and perceptual organization

- Sometimes, motion is the only cue



Motion and perceptual organization

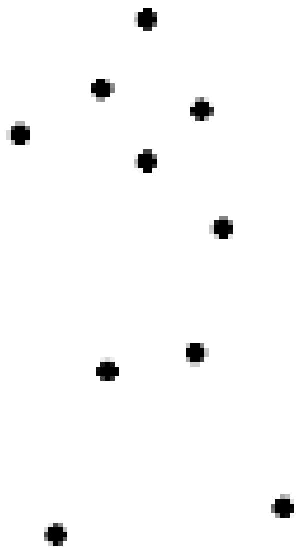
- Even “impoverished” motion data can evoke a strong percept



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.*

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

Q. 3: What kind of person is the circle?

Does not like fighting, is frightened, afraid, fearful, cowardly, shy, timid, meek, not too sure of herself, goes where *t* goes, a follower, not much personality of her own, less initiative and nerve, relies for protection on *t*, helpless, dependent. N = 27 (75%)

Girl, woman, female, feminine. N = 22 (61%)

Shrewd, intelligent, clever, smart. N = 5 (14%)

Courageous, resistent, has courage. N = 4 (11%)

Weak. N = 3 (8%)

Opportunist, looks after own good, teasing, curious, playful, good natured, more gentle, very refined, nervous, retiring, beautiful, loyal, affectionate, coming to aid when necessary. (One *S* each)

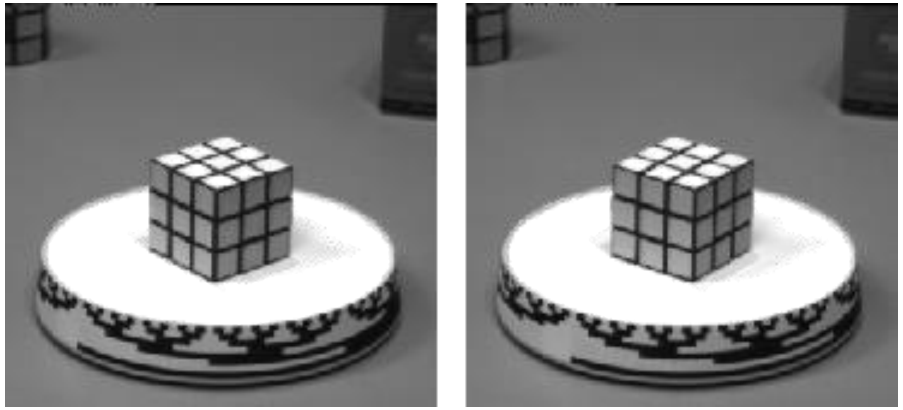
The withdrawing of *c* during the fight and the fact that *c* never hits *T* accounts for the description of it as 'afraid, meek," etc. Some *S*s obviously make *c* at least partly responsible for the ruse played on *T* and call *c* clever.

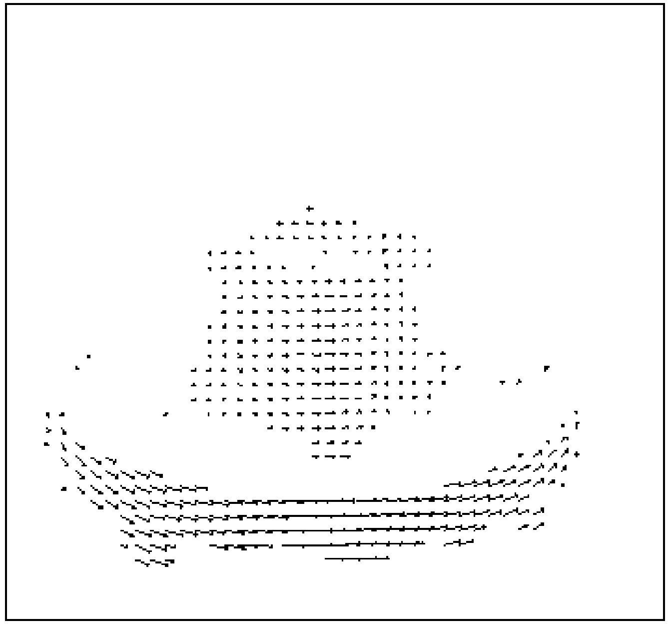
Q. 5: Why did the circle go into the house?

For protection, afraid to watch fight, frightened by fighting, to get out of the way of the fight, scared, tried to hide, for shelter against *T*, to escape villain *T*, afraid of what *T* might do to *t*. N = 33 (92%)

Motion estimation: Optical flow

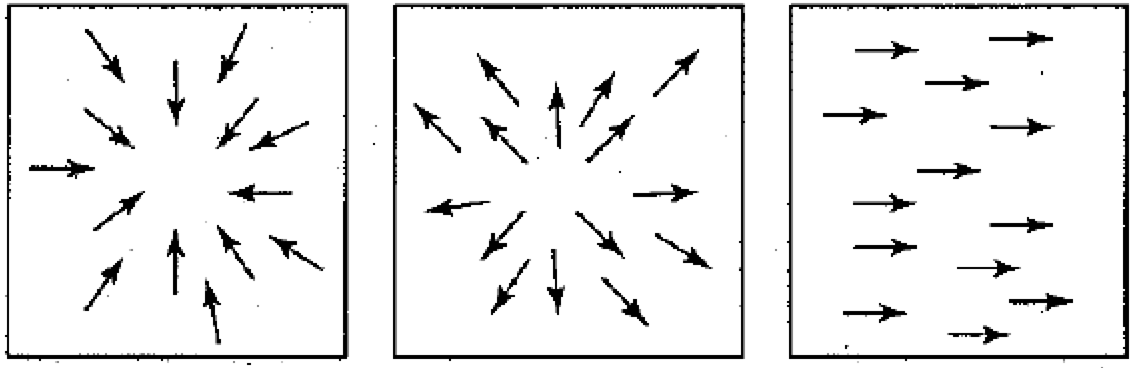
Optical flow is the **apparent** motion of objects or surfaces





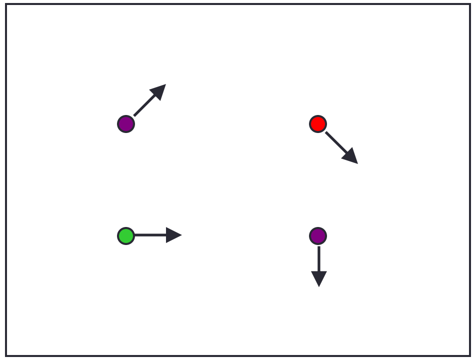
The term “scene flow” is used to describe 3d motion estimation

Motion field + camera motion

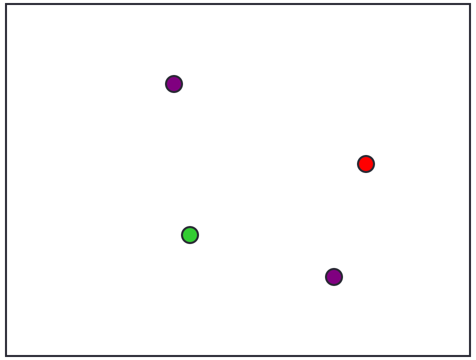


Zoom out **Zoom in** **Pan right to left**

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

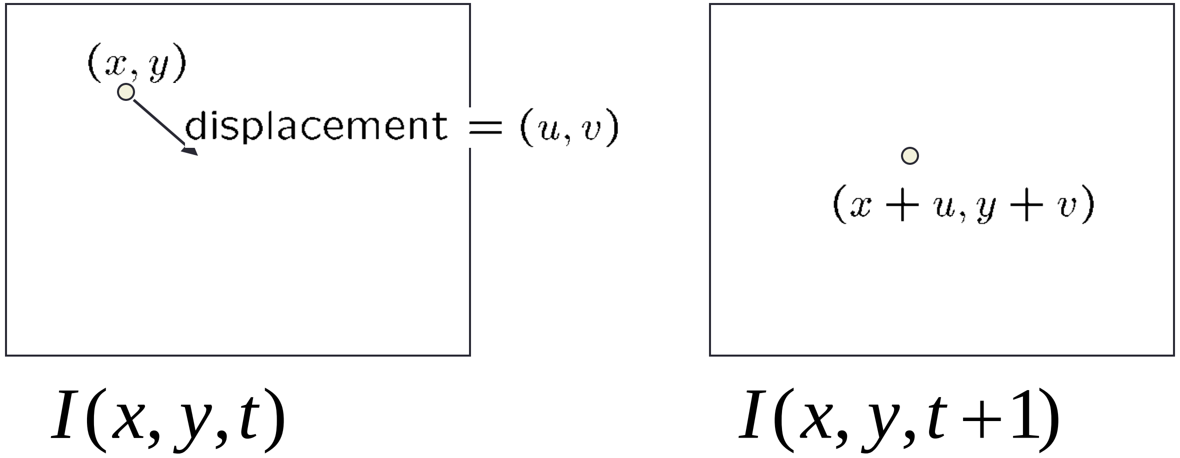
- Solve pixel correspondence problem
  - given a pixel in I(x,y,t), look for nearby pixels of the same color in *I(x,y,t+1)*

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)

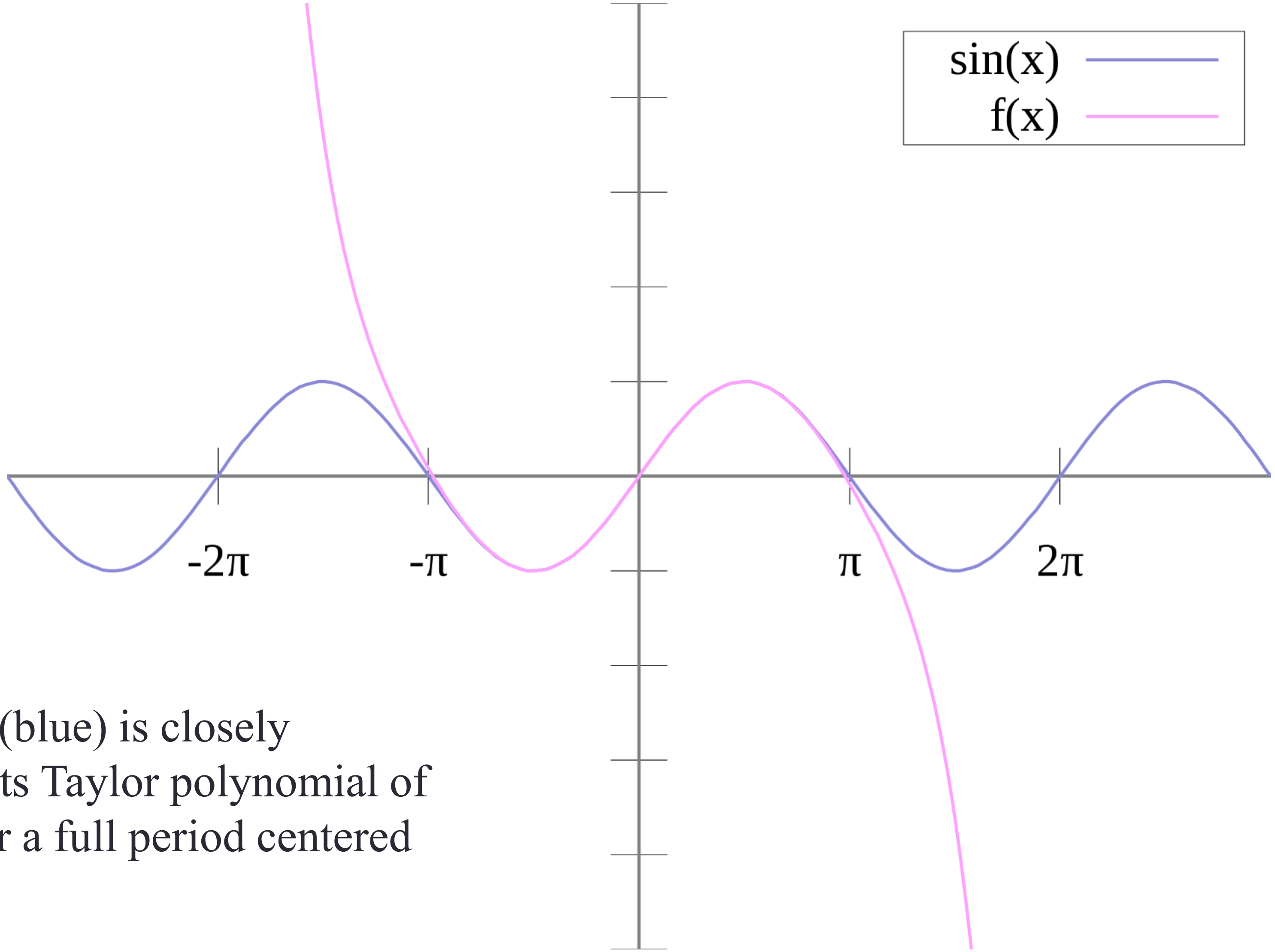


- Let's look at these constraints more closely
  - brightness constancy constraint (equation)

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

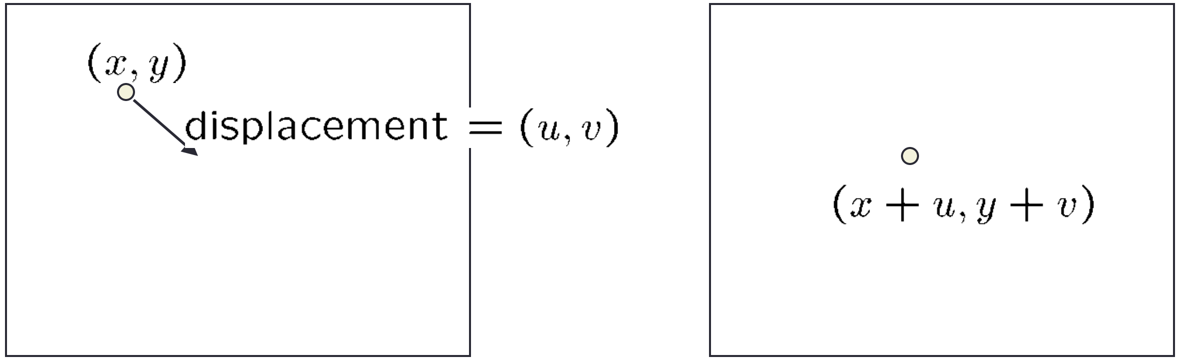
  - small motion: (u and v are less than 1 pixel, or smooth)
    Taylor series expansion of the spatial changes of I:

Taylor Series Reminder



The sine function (blue) is closely approximated by its Taylor polynomial of degree 7 (pink) for a full period centered at the origin.

Optical flow constraints (grayscale images)



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

- 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 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

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

(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?

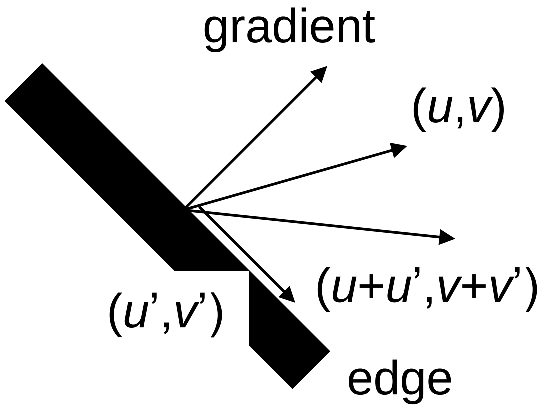
$$0 = I_t + \nabla I \cdot < u, v > \quad \text{or} \quad I_x u + I_y v + I_t = 0$$

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

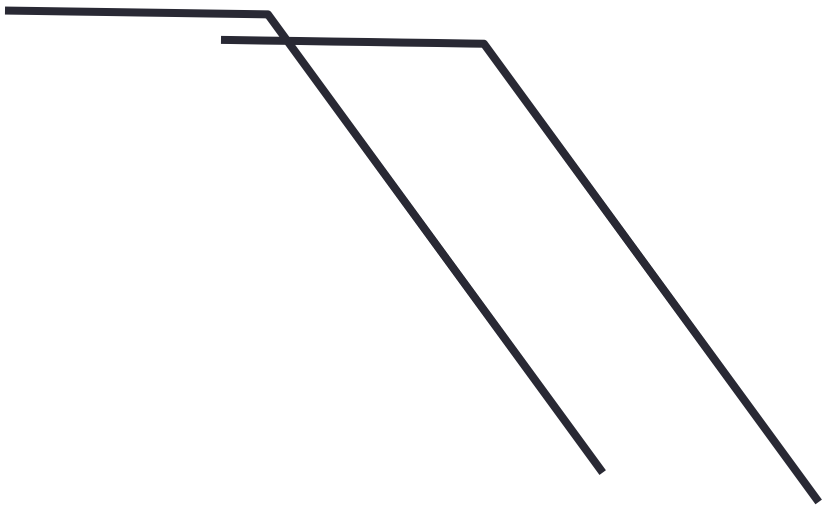
The component of the motion perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

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

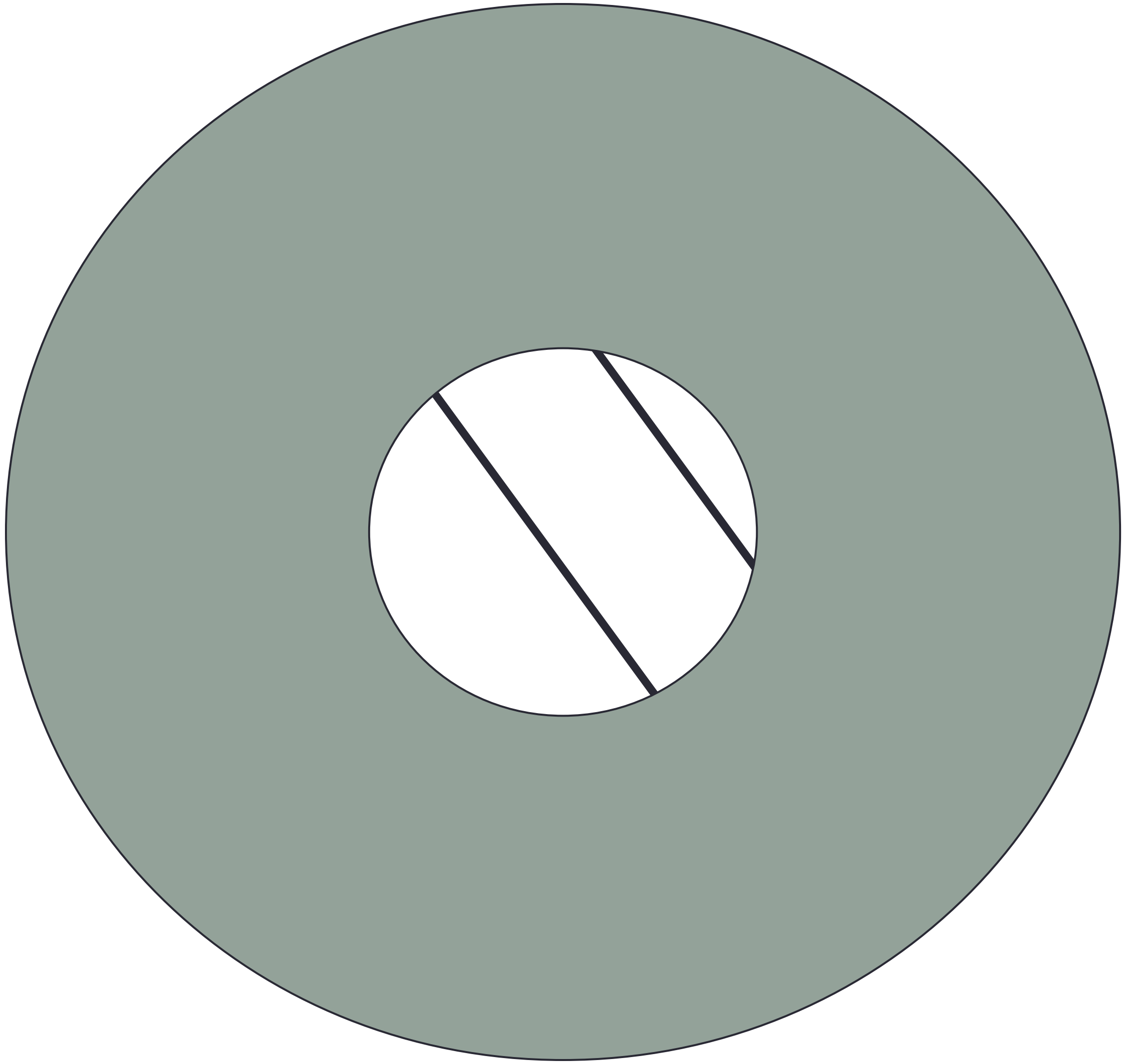
$$\nabla \mathrm{I} \cdot [\mathrm{u'} \ \ \mathrm{v'}]^{\mathrm{T}} = 0$$



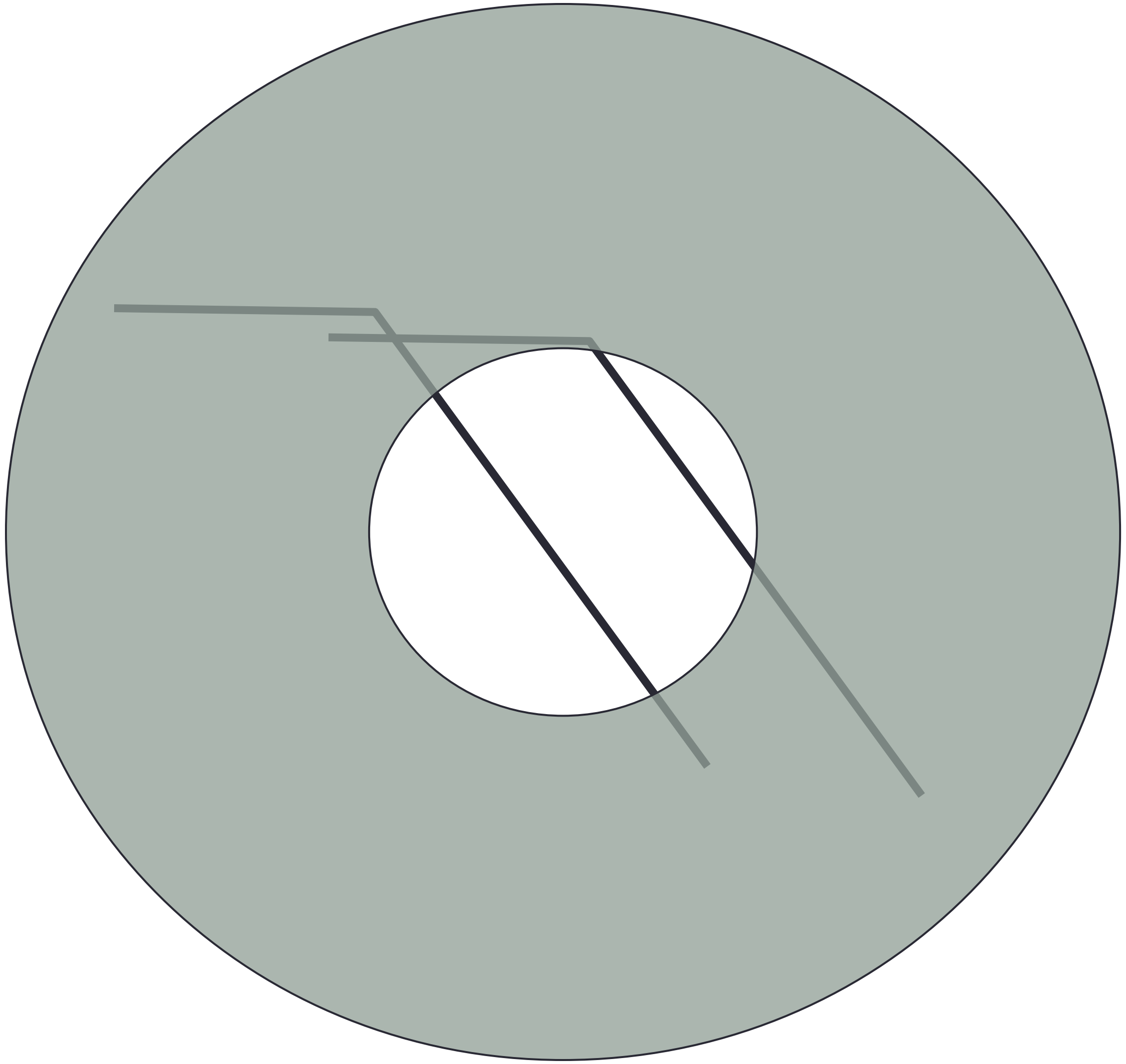
Aperture problem



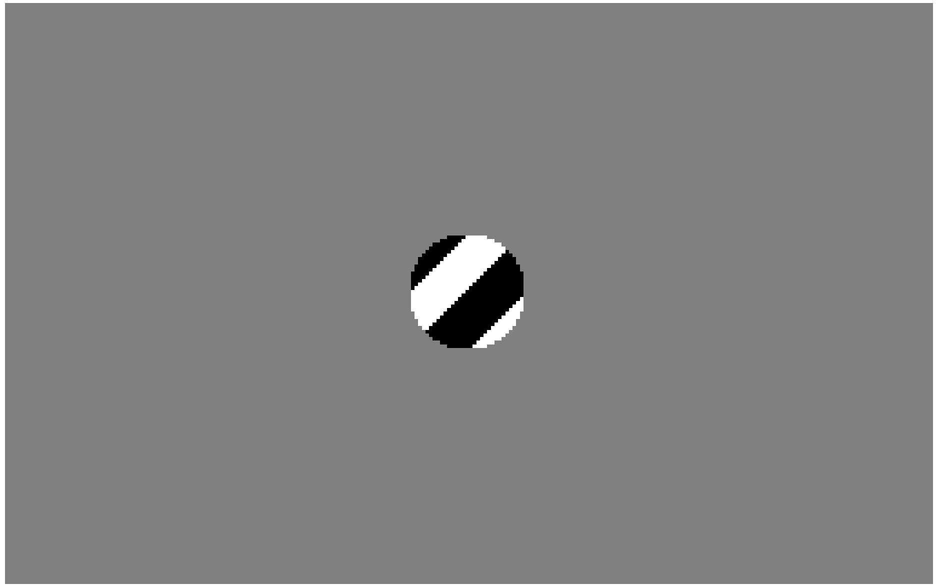
Aperture problem



Aperture problem

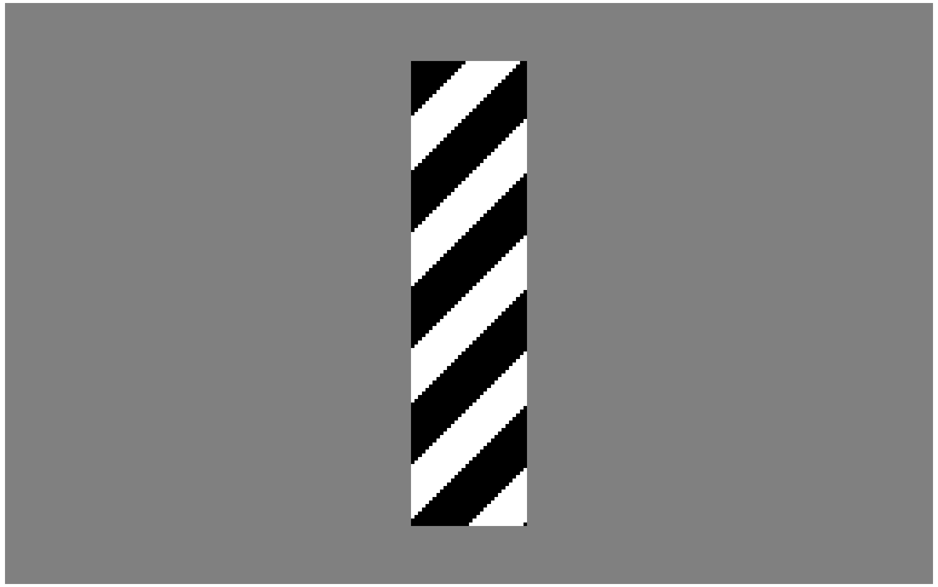


The barber pole illusion



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:

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

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

Matching patches across images

- Overconstrained linear system

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\text{x}2}{A} \; \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b}$$

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

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

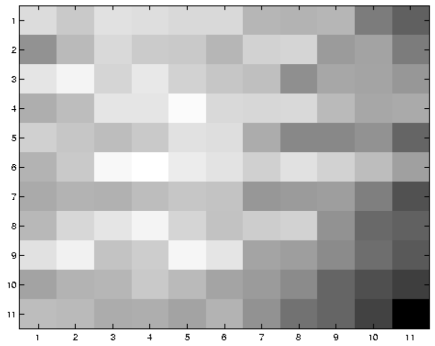
When is this solvable? I.e., what are good points to track?

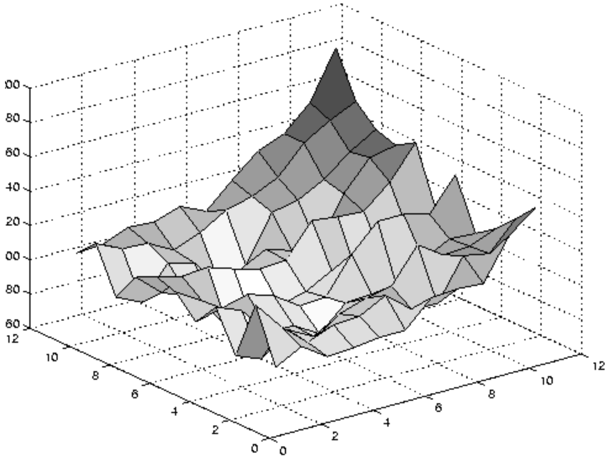
- **A^TA** should be invertible
- **A^TA** should not be too small due to noise
  - eigenvalues λ_1 and λ_2 of **A^TA** should not be too small
- **A^TA** should be well-conditioned
  - λ_1 / λ_2 should not be too large (λ_1 = larger eigenvalue)

Does this remind you of anything?

Criteria for Harris corner detector

Low texture region



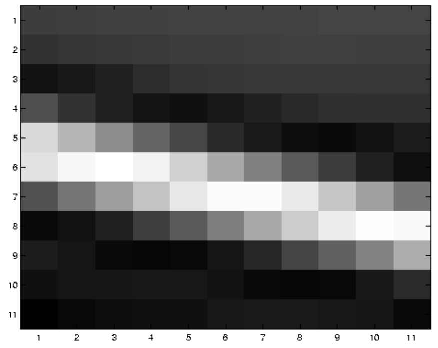


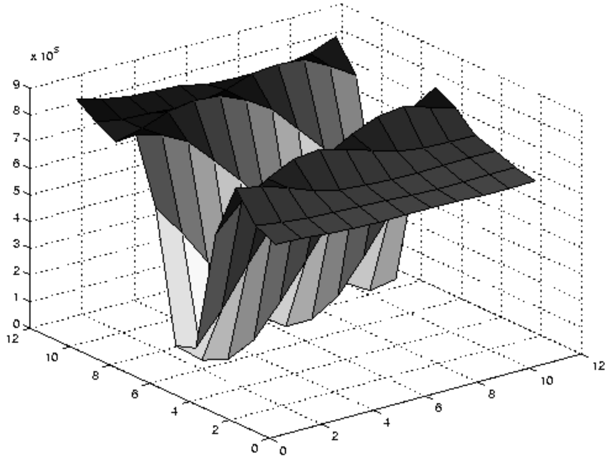
$$\sum \nabla I (\nabla I)^T$$

– gradients have small magnitude

– small λ_1, small λ_2

Edge



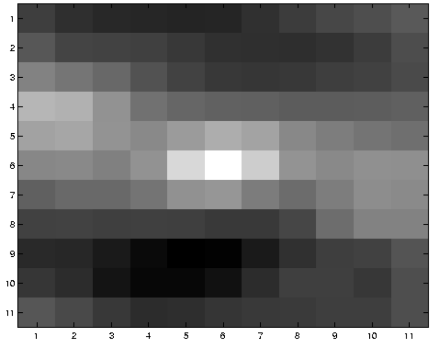


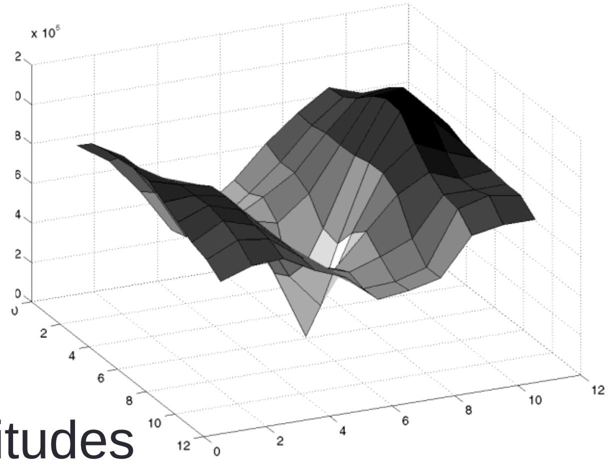
$$\sum \nabla I (\nabla I)^T$$

– large gradients, all the same

– large λ_1, small λ_2

High textured region



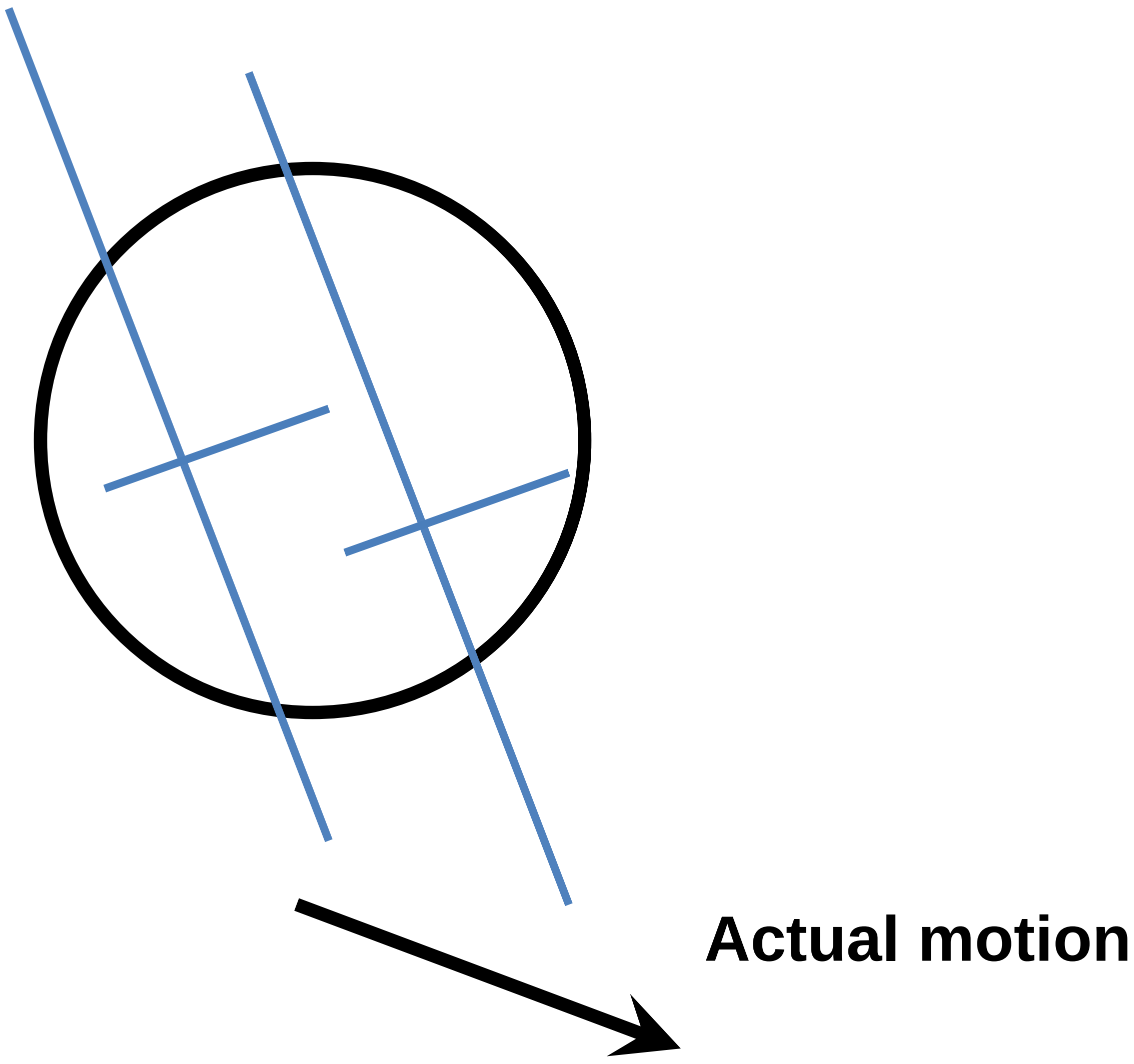


$$\sum \nabla I (\nabla I)^T$$

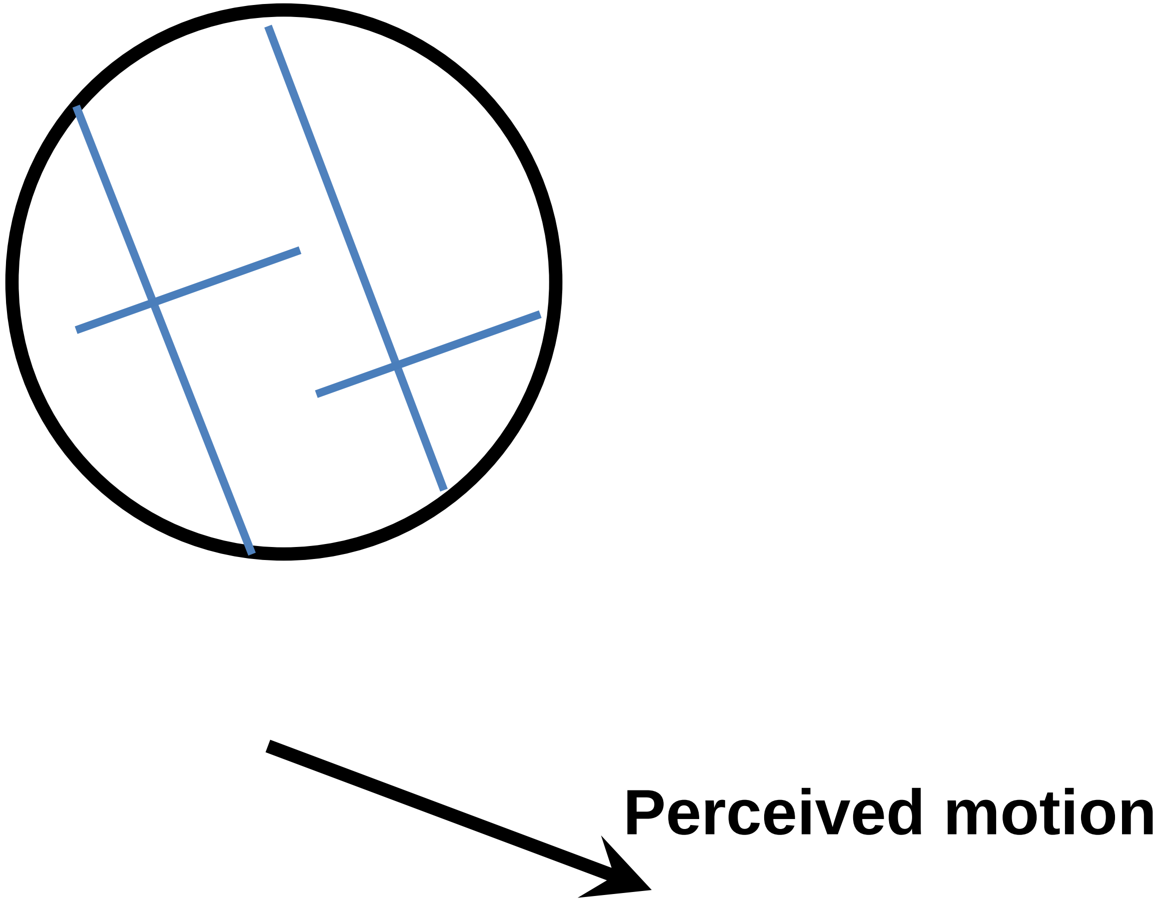
– gradients are different, large magnitudes

– large λ_1, large λ_2

The aperture problem resolved



The aperture problem resolved



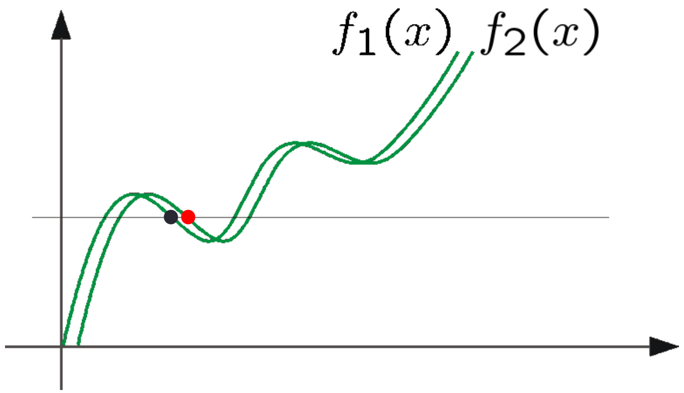
Revisiting the small motion assumption

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

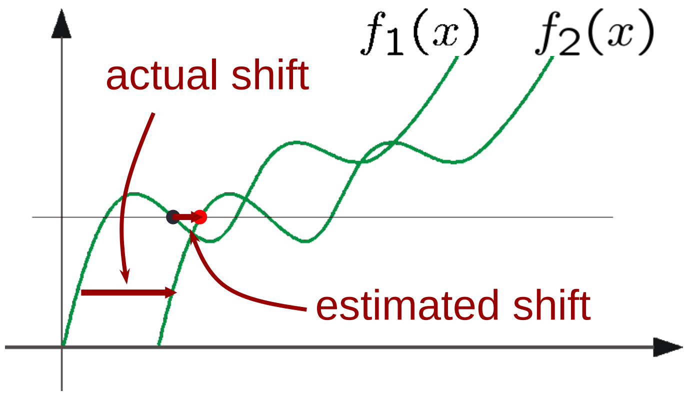
Optical Flow: Aliasing

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?



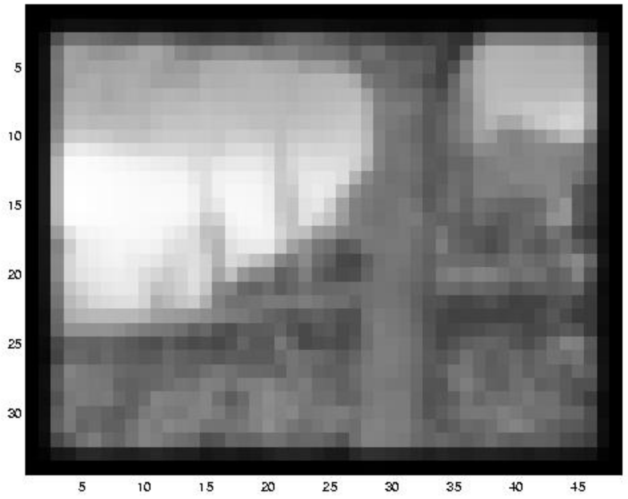
nearest match is correct
(no aliasing)

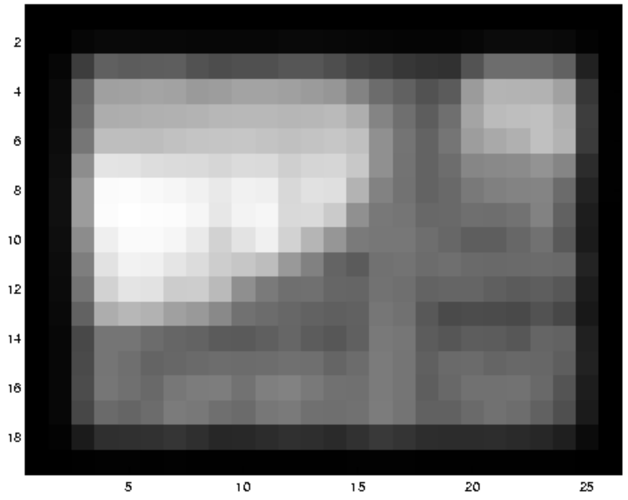


nearest match is incorrect
(aliasing)

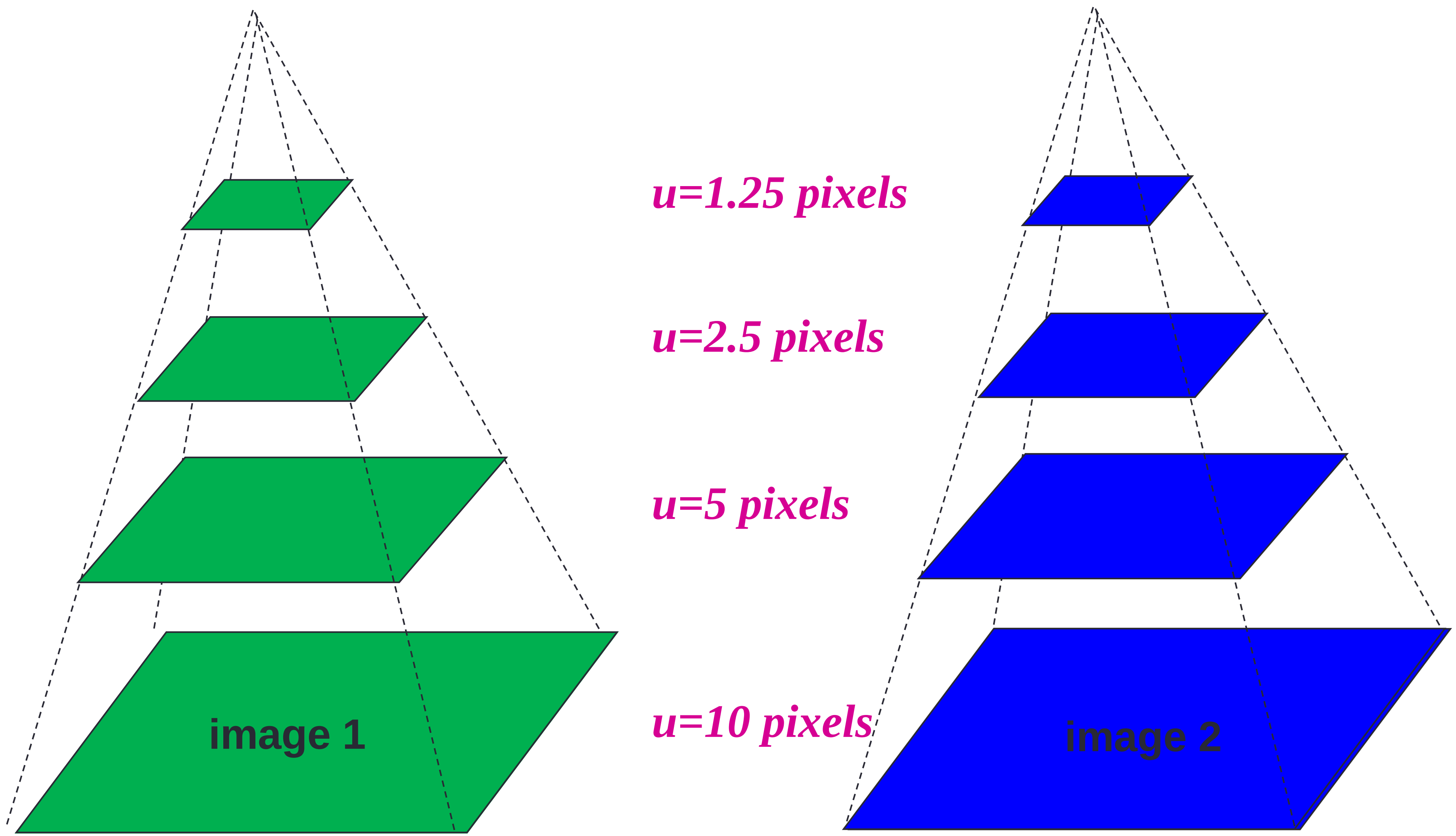
To overcome aliasing: coarse-to-fine estimation.

Reduce the resolution!





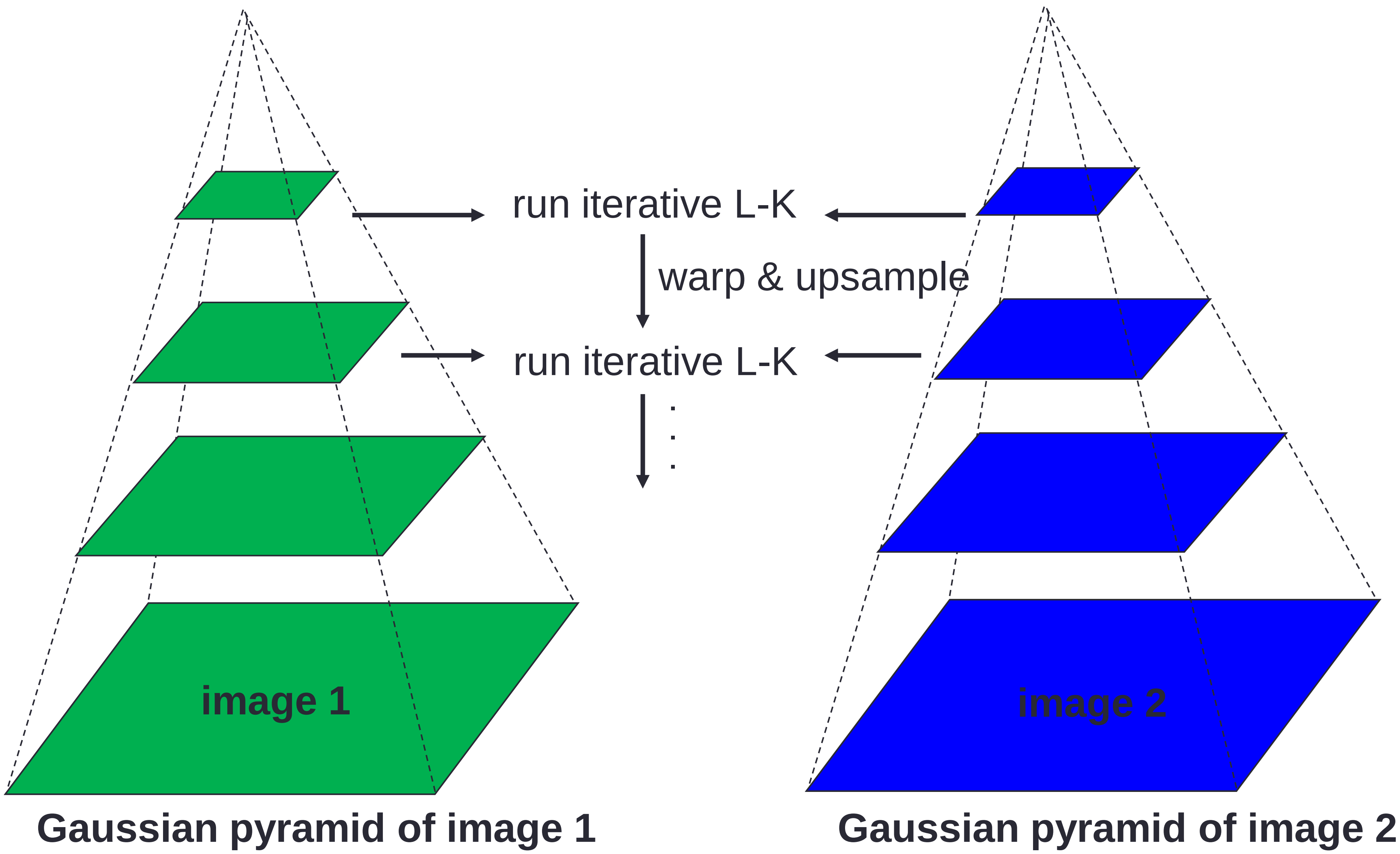
Coarse-to-fine optical flow estimation



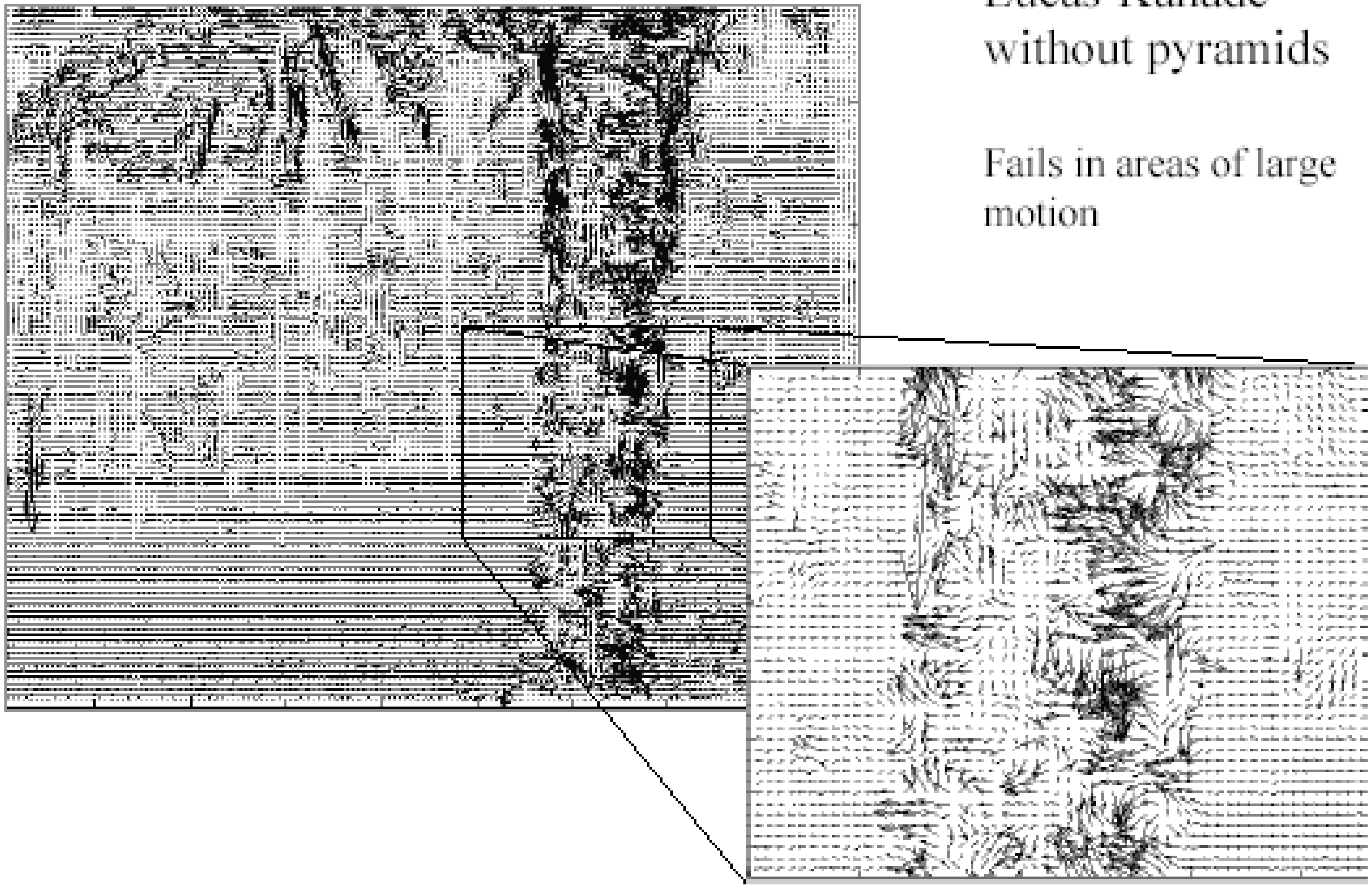
Gaussian pyramid of image 1

Gaussian pyramid of image 2

Coarse-to-fine optical flow estimation



Optical Flow Results

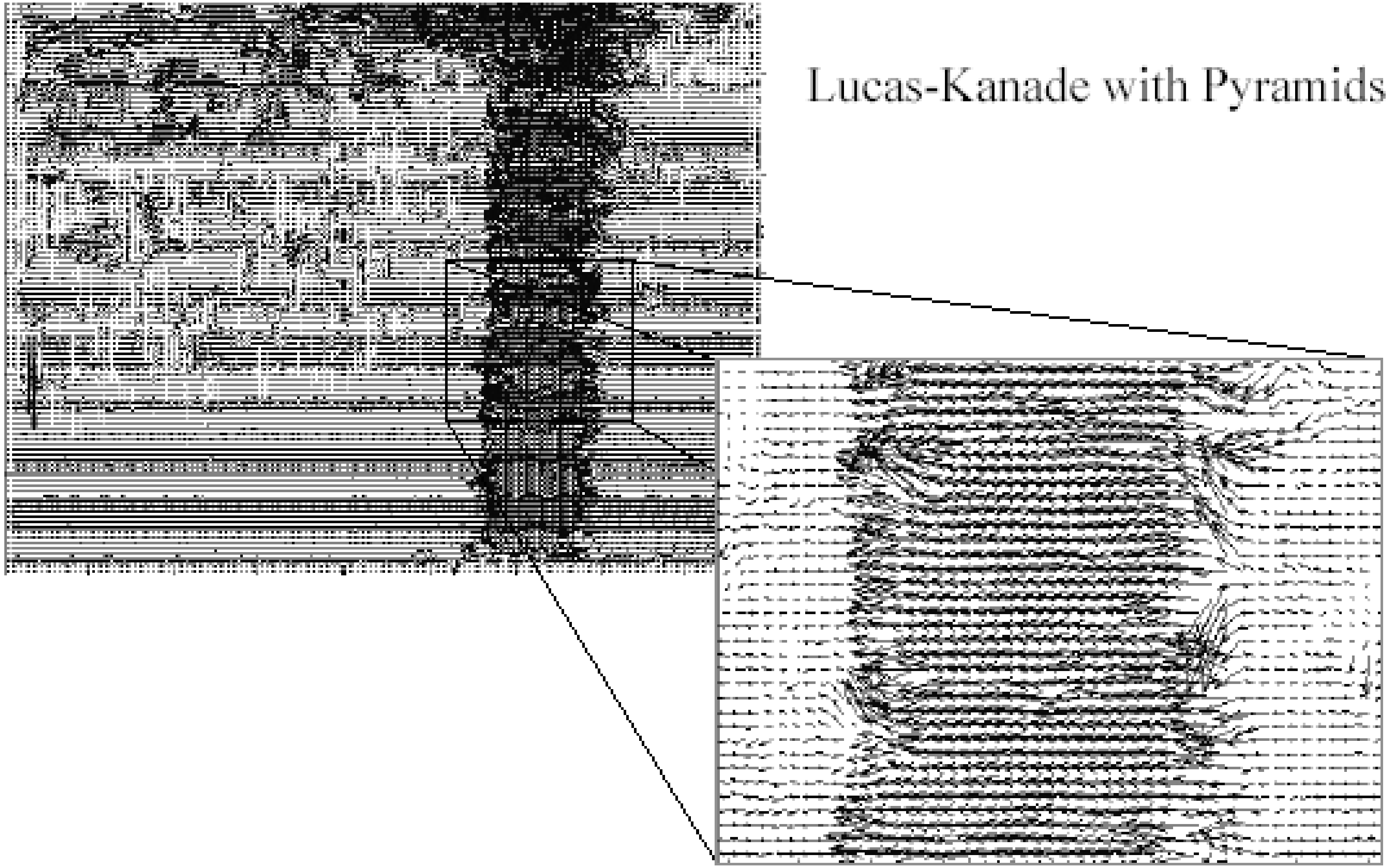


Lucas-Kanade
without pyramids

Fails in areas of large
motion

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003

Optical Flow Results



Lucas-Kanade with Pyramids

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003

State-of-the-art optical flow in 2009

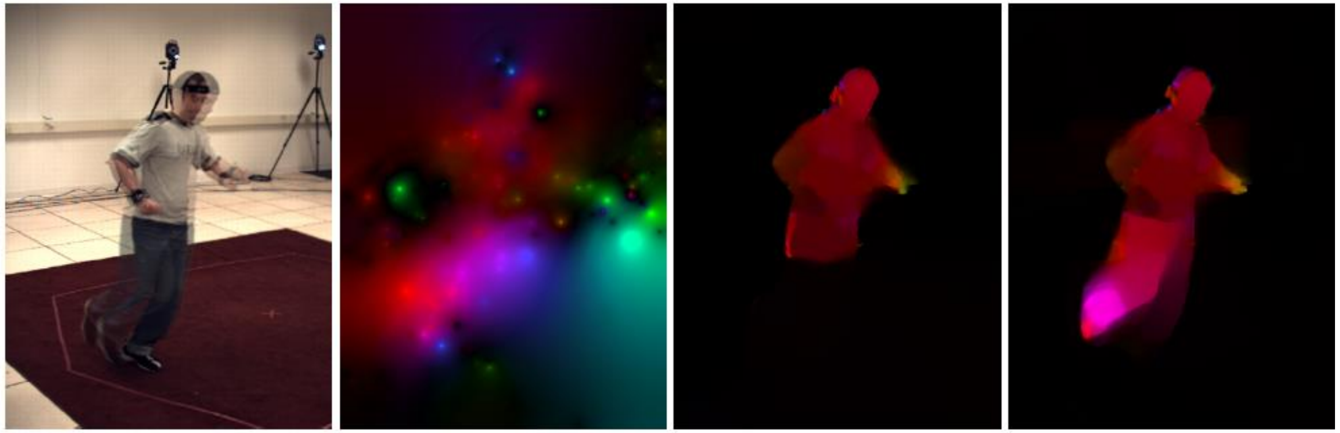
Start with something similar to Lucas-Kanade

+ gradient constancy

+ energy minimization with smoothing term

+ region matching

+ keypoint matching (long-range)

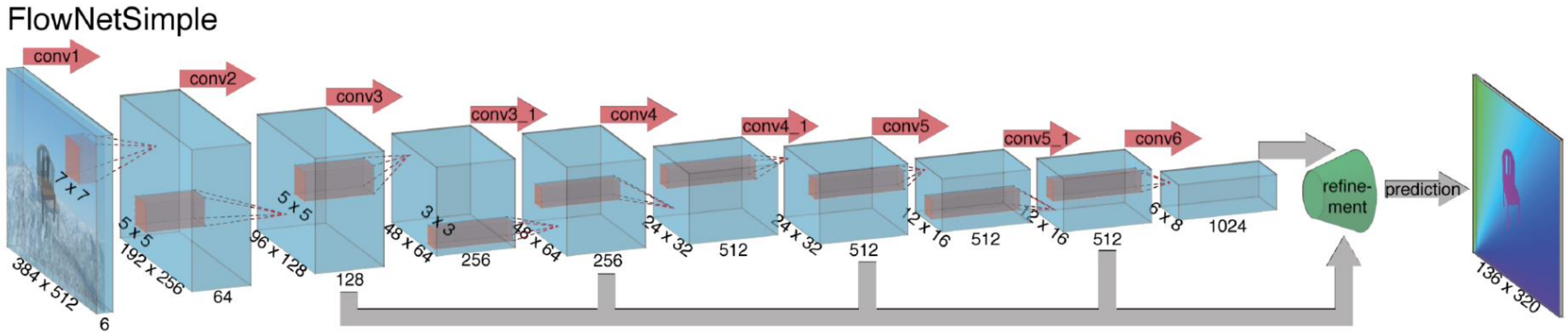


Region-based +Pixel-based +Keypoint-based

Large displacement optical flow, Brox et al., CVPR 2009

State-of-the-art optical flow in 2015

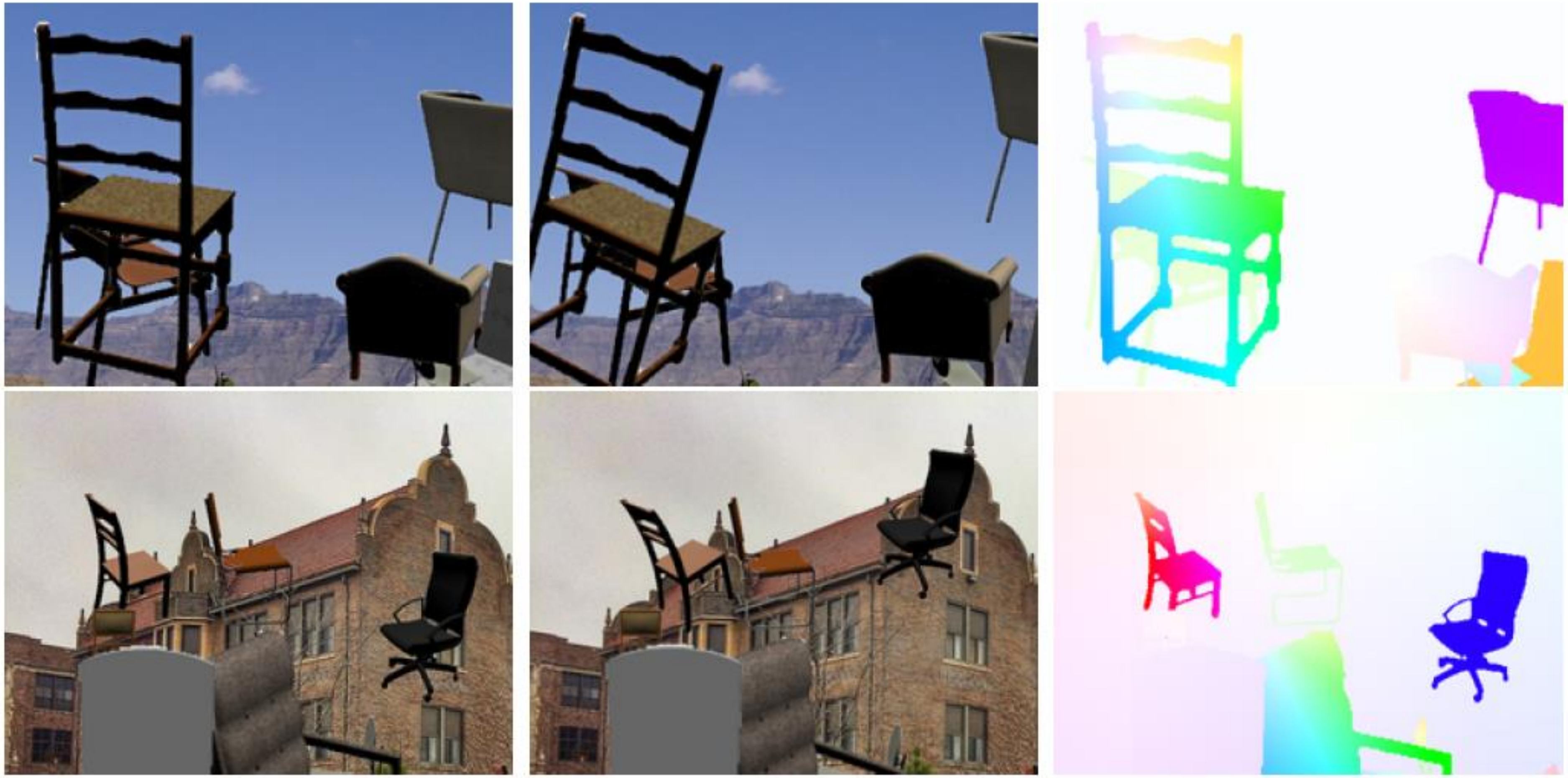
Deep convolutional network which accepts a pair of input frames and upsamples the estimated flow back to input resolution. Very fast because of deep network, near the state-of-the-art in terms of end-point-error.



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

Deep optical flow, 2015

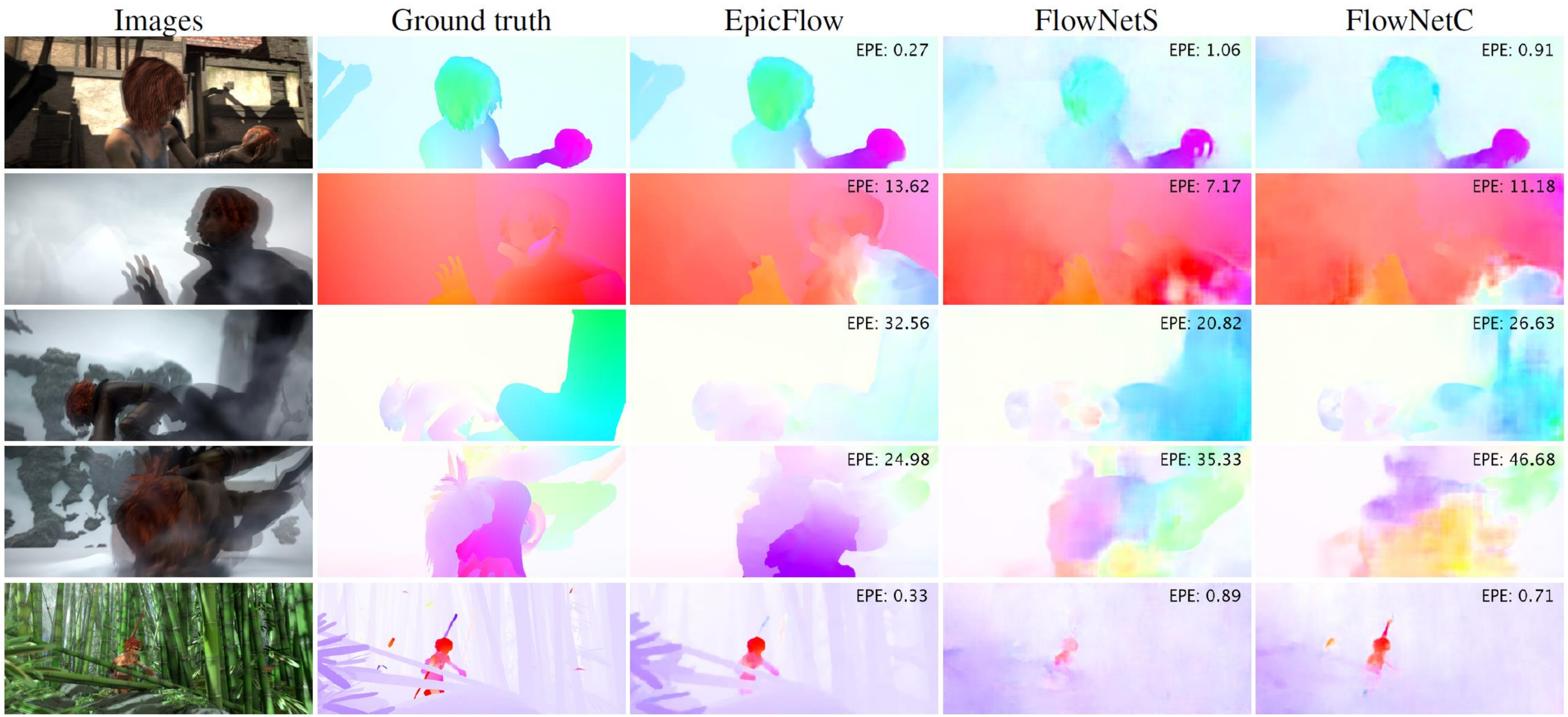
Synthetic Training data



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

Deep optical flow, 2015

Results on Sintel

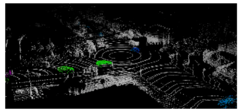


Fischer et al. 2015. https://arxiv.org/abs/1504.06852

Wrap up: 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

Scene Flow: 3D scene motion



Argoverse 2.0 Self-Supervised Scene Flow

★ 6

Toggle Participation

Organized by: argoai-argoverse

Published

Starts on: Mar 29, 2023 8:00:00 PM EST (GMT - 4:00)

Ends on: May 31, 2099 7:59:59 PM EST (GMT - 4:00)

Add Tags or Domain

Overview | Evaluation | Phases | Submit | My Submissions | All Submissions | Leaderboard | Manage

Leaderboard

Phase: **Open Submission**, Split: **Test Split**

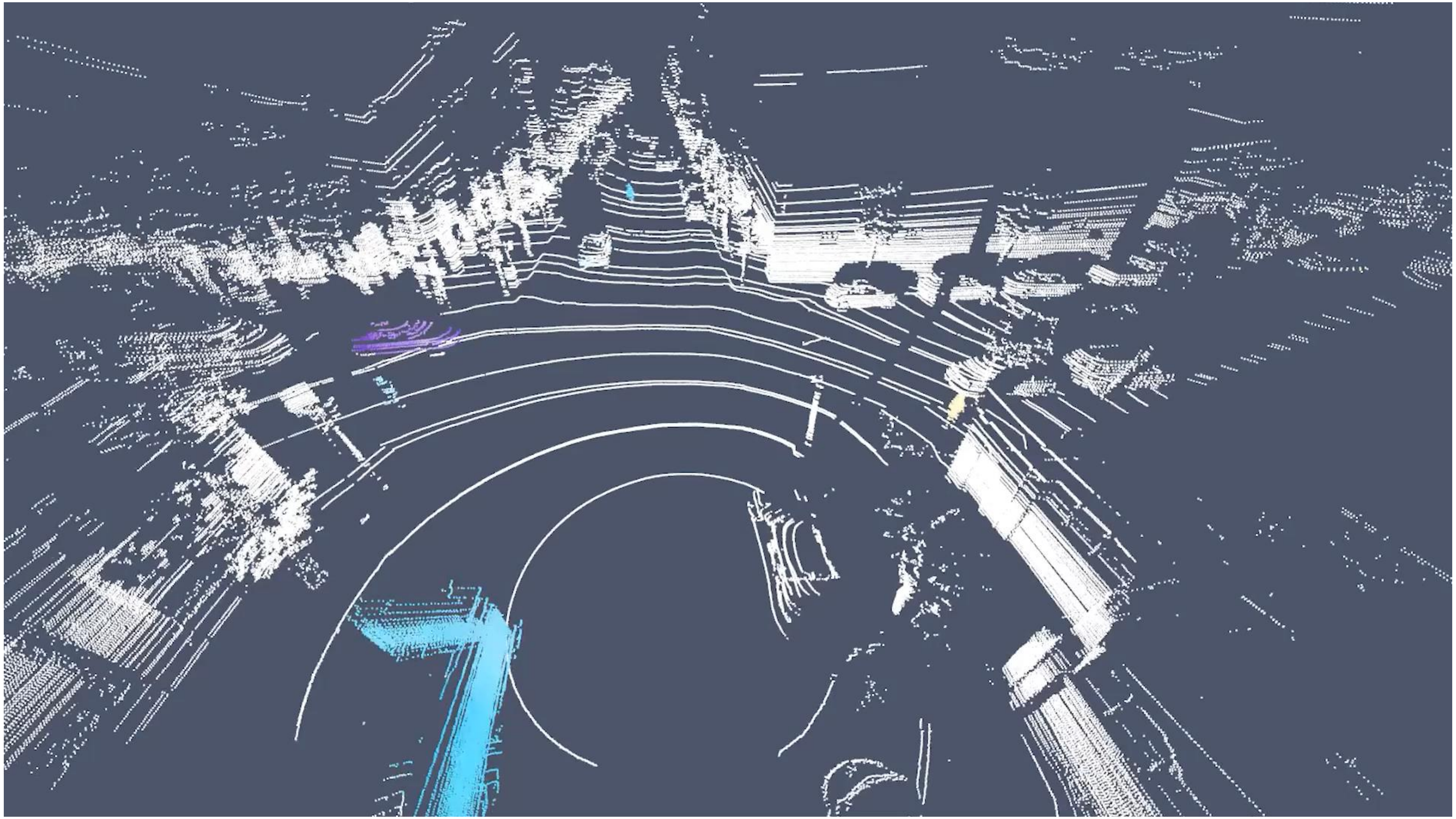
Order by metric

B - Baseline * - Private V - Verified ☐ Include private submissions

Visible Metrics

Rank	Participant team	Dynamic IoU (↑)	EPE 3-Way Average (↓)	EPE/Background /Static (↑)	EPE/Foreground /Dynamic (↑)	EPE/Foreground /Static (↑)	Angle Error/Background /Static (↑)	Angle Error/For /Dynami
1	RPL (Deflow)	0.6289	0.0534	0.0029	0.1340	0.0232	0.0052	0.1920

Scene Flow: 3D scene motion



DeFlow. Qingwen Zhang, Yi Yang, Heng Fang, Ruoyu Geng, and Patric Jensfelt. RPL lab @ KTH

- Sample quiz questions (new window)