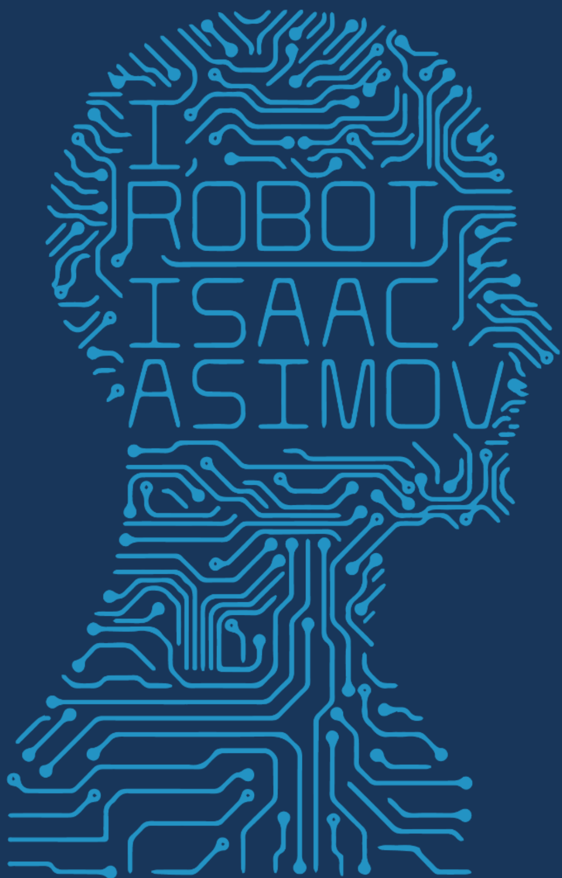
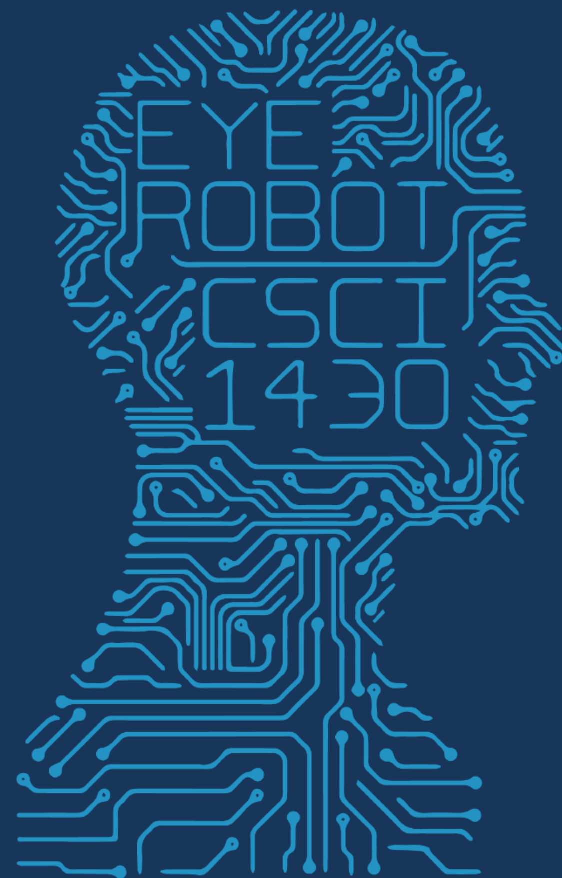


1950

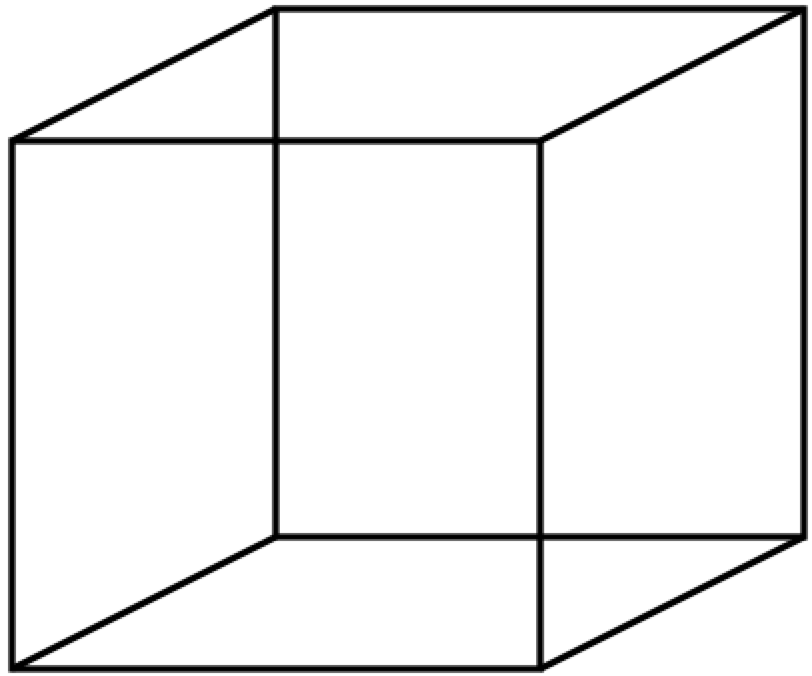
Future Vision



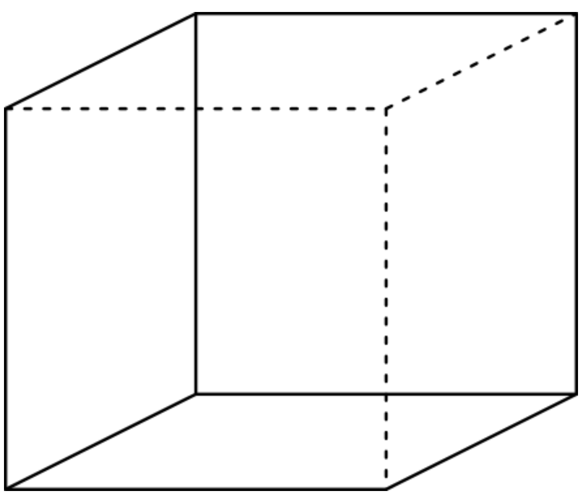
2017 MWF 1pm 368

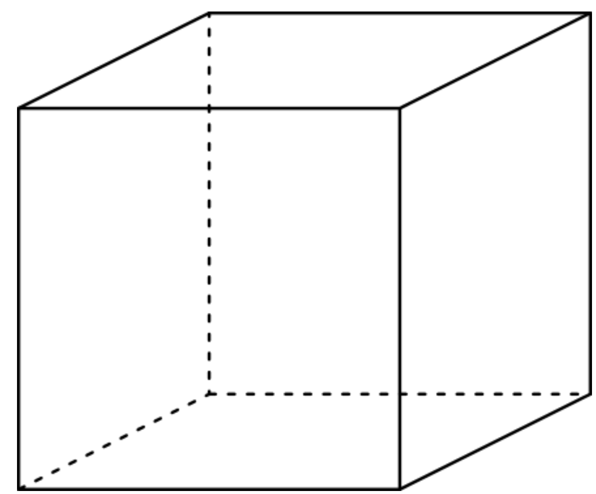
Computer Vision

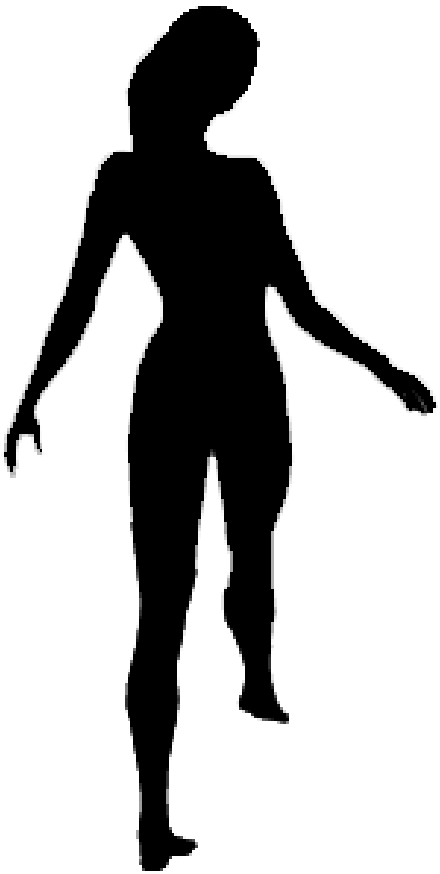
Multi-stable Perception



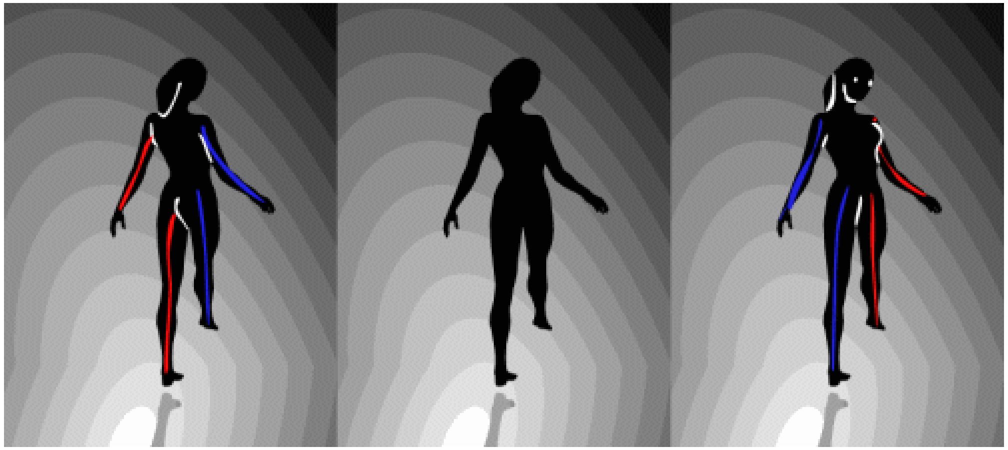
Necker Cube



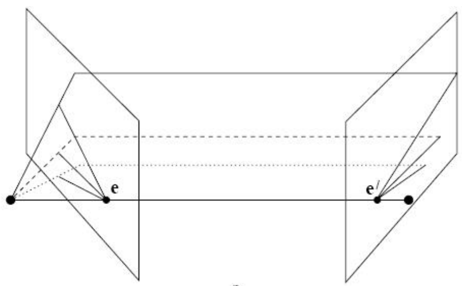




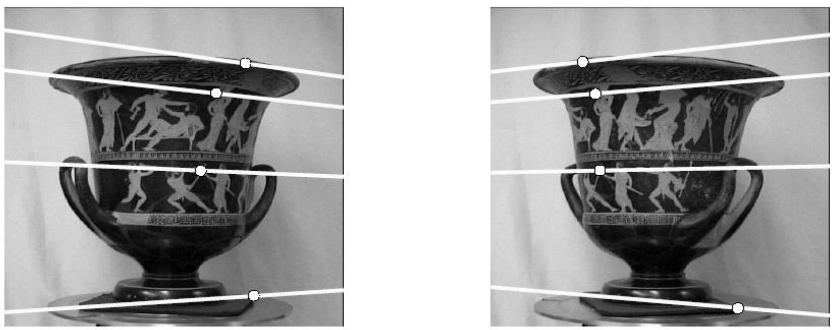
Spinning dancer illusion, Nobuyuki Kayahara



Multiple view geometry



a



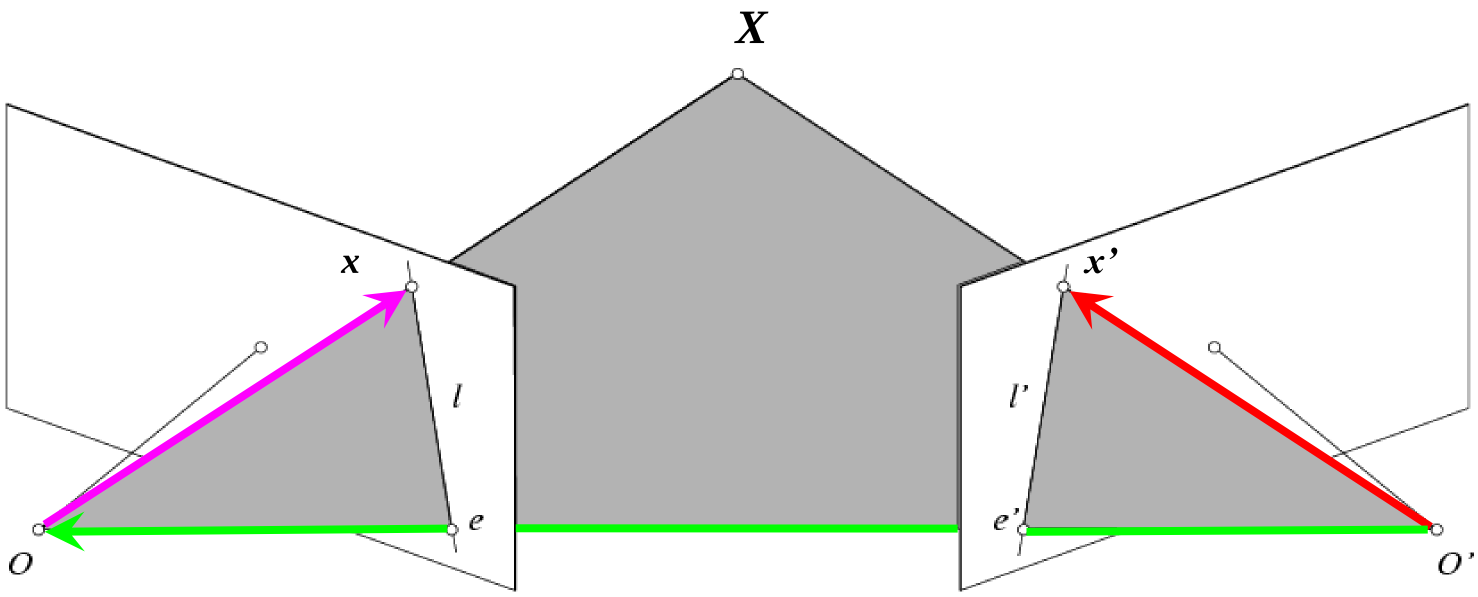
Hartley and Zisserman

Stereo vision

Epipolar geometry

Depth map extraction

Essential matrix



$$\hat{x} \cdot [t \times (R\hat{x}')] = 0 \quad \Rightarrow \quad \hat{x}^T E \hat{x}' = 0 \quad \text{with} \quad E = [t]_\times R$$

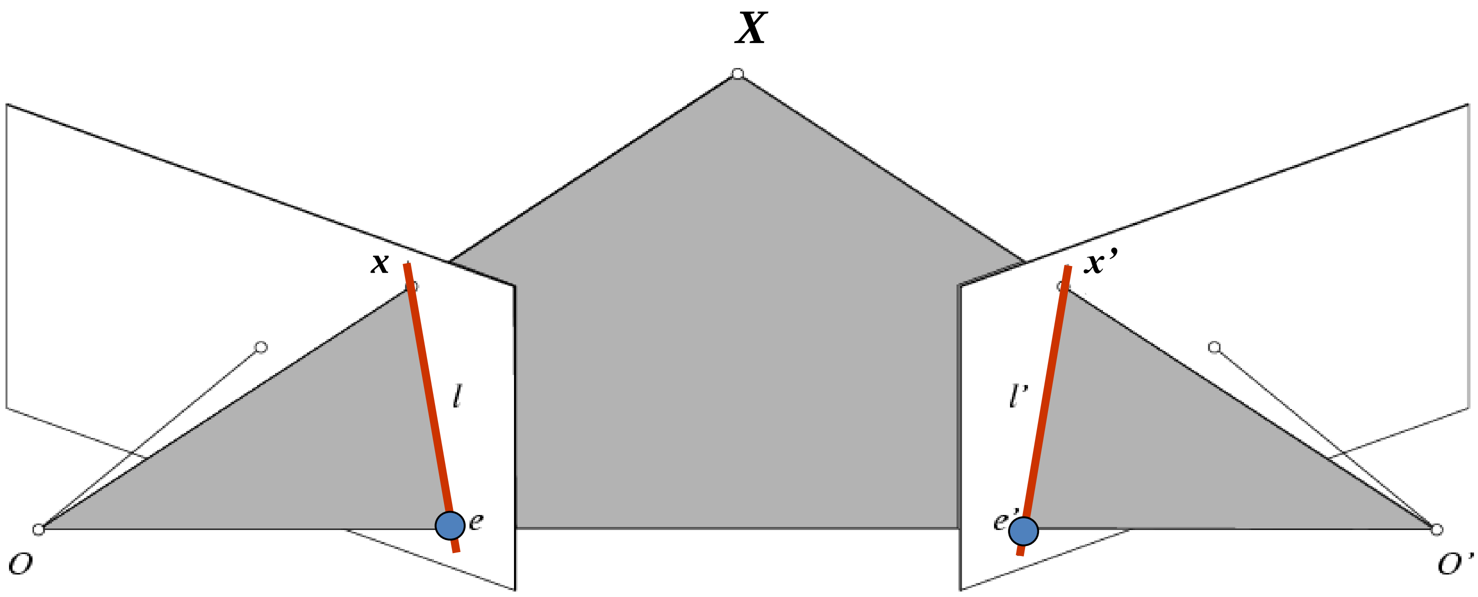
E is a 3x3 matrix which relates corresponding pairs of normalized homogeneous image points across pairs of images – for *K* calibrated cameras.

Essential Matrix
(Longuet-Higgins, 1981)

Estimates relative position/orientation.

Note: $[t]_\times$ is matrix representation of cross product

Fundamental matrix for uncalibrated cases

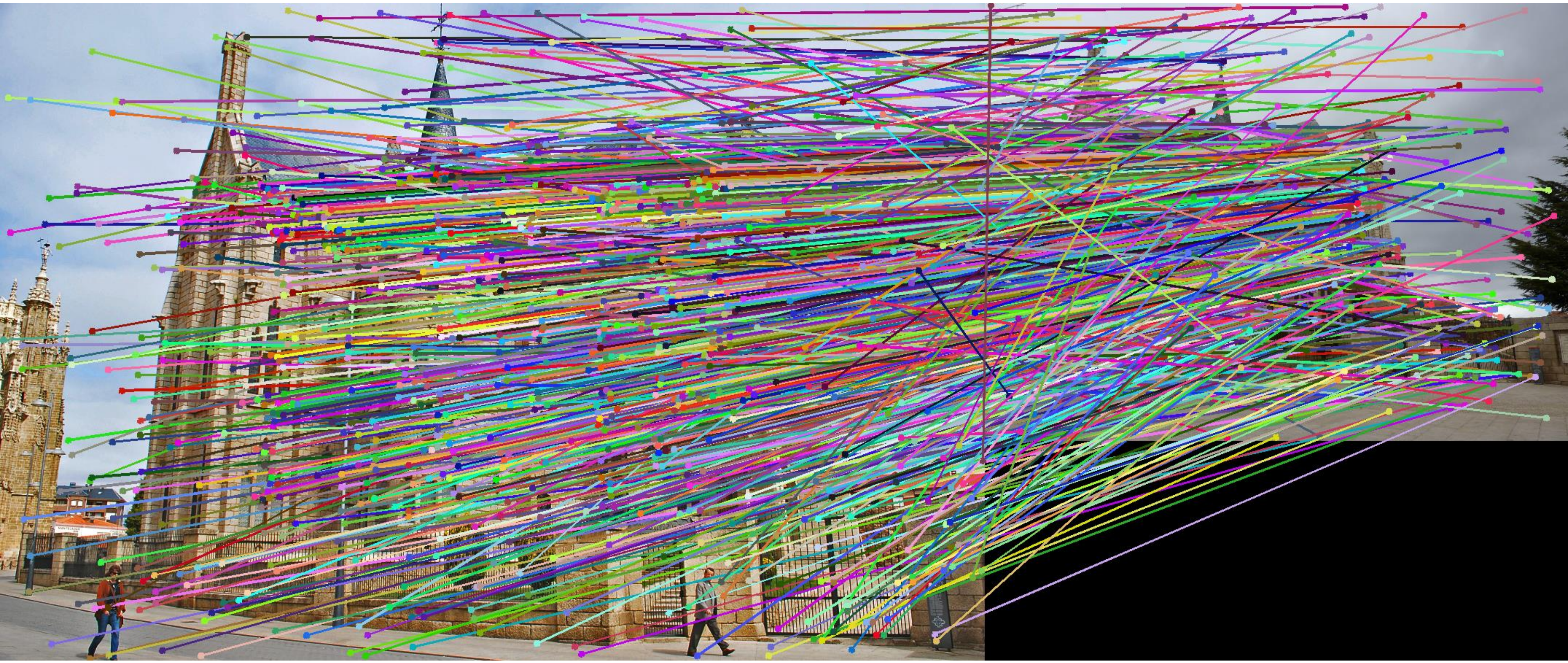


$$x^T F x' = 0 \quad \text{with} \quad F = K^{-T} E K'^{-1}$$

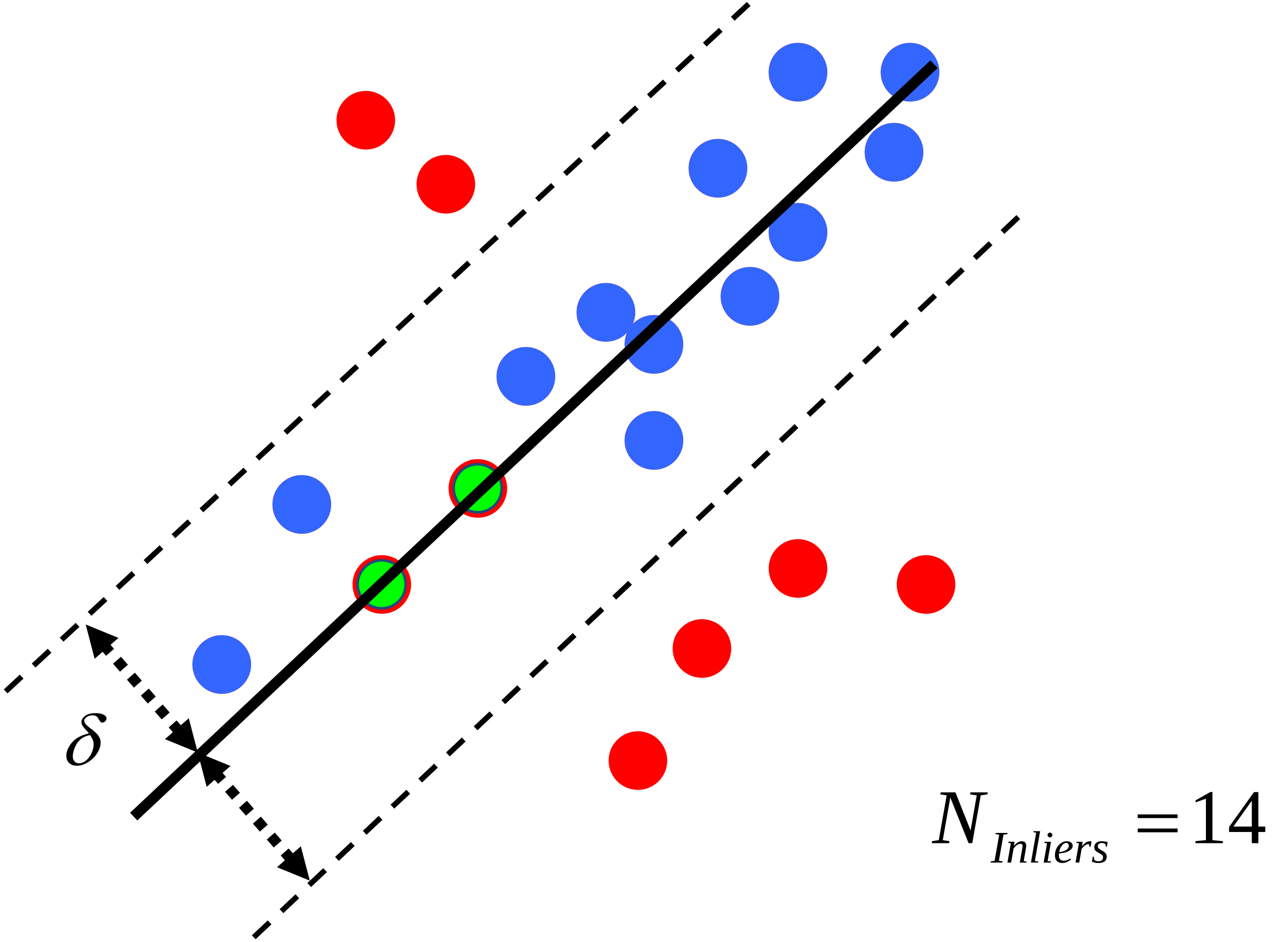
- *F x'* = 0 is the epipolar line *l* associated with *x'*
- $F^T x$ = 0 is the epipolar line *l'* associated with *x*

- *F* is singular (rank two): det(F)=0
- *F e'* = 0   and $F^T e = 0$ (nullspaces of F = e'; nullspace of F^T = e')
- *F* has seven degrees of freedom: 9 entries but defined up to scale, det(F)=0

VLFeat's 800 most confident matches among 10,000+ local features.



RANSAC

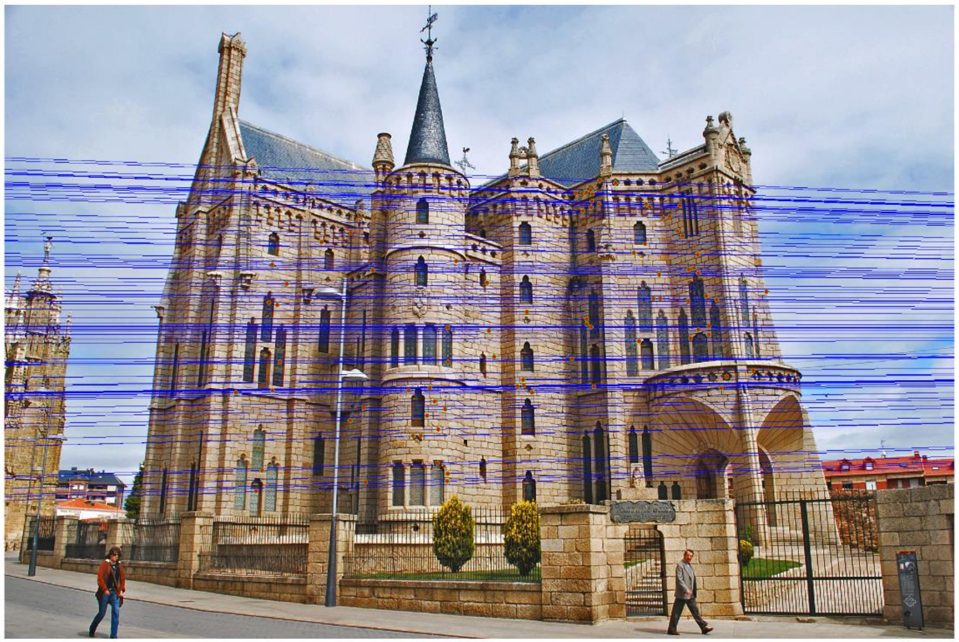


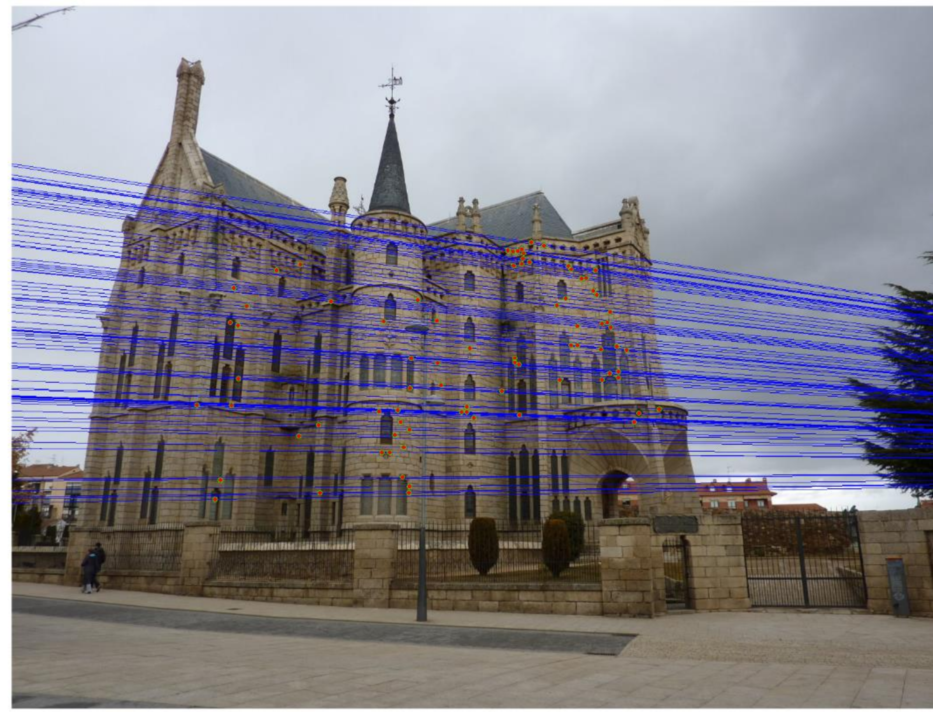
Algorithm:

1. **Sample** (randomly) the number of points required to fit the model (*s*=2)
2. **Solve** for model parameters using samples
3. **Score** by the fraction of inliers within a preset threshold of the model

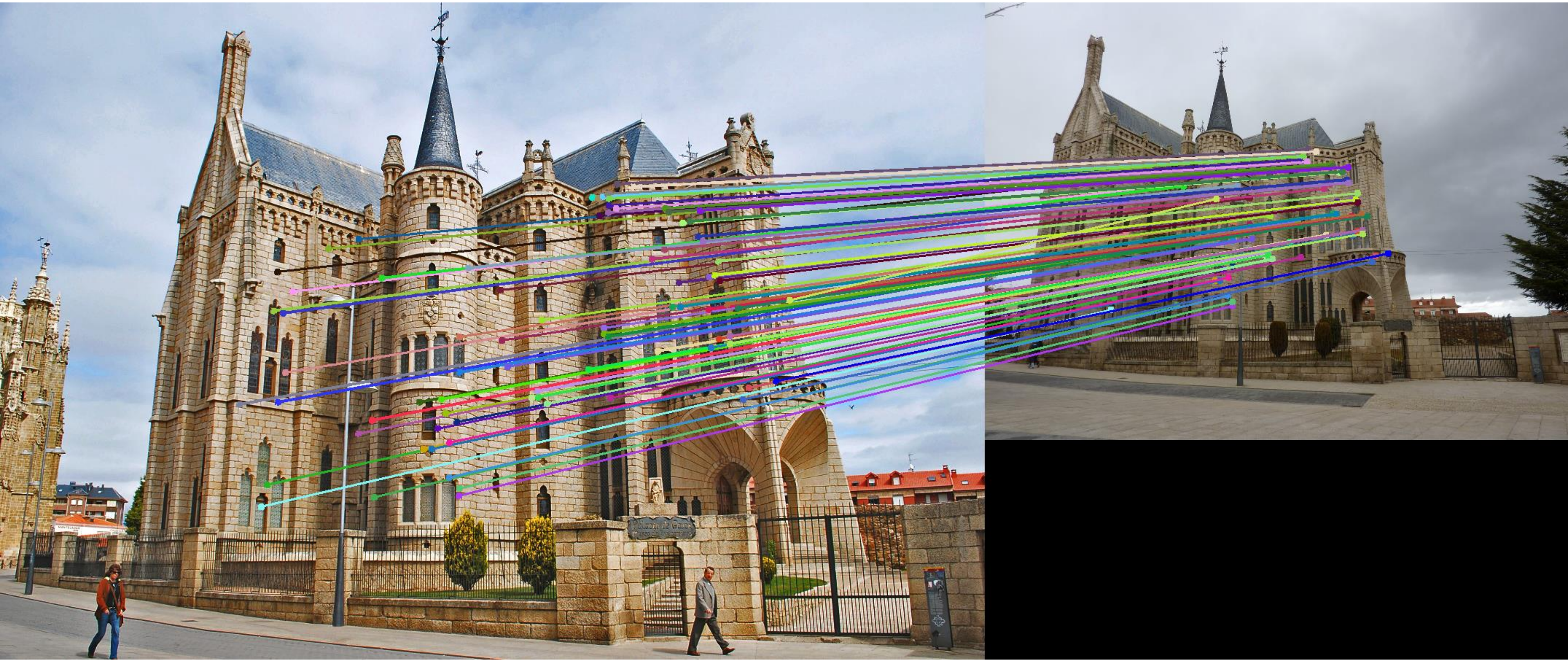
Repeat 1-3 until the best model is found with high confidence

Epipolar lines

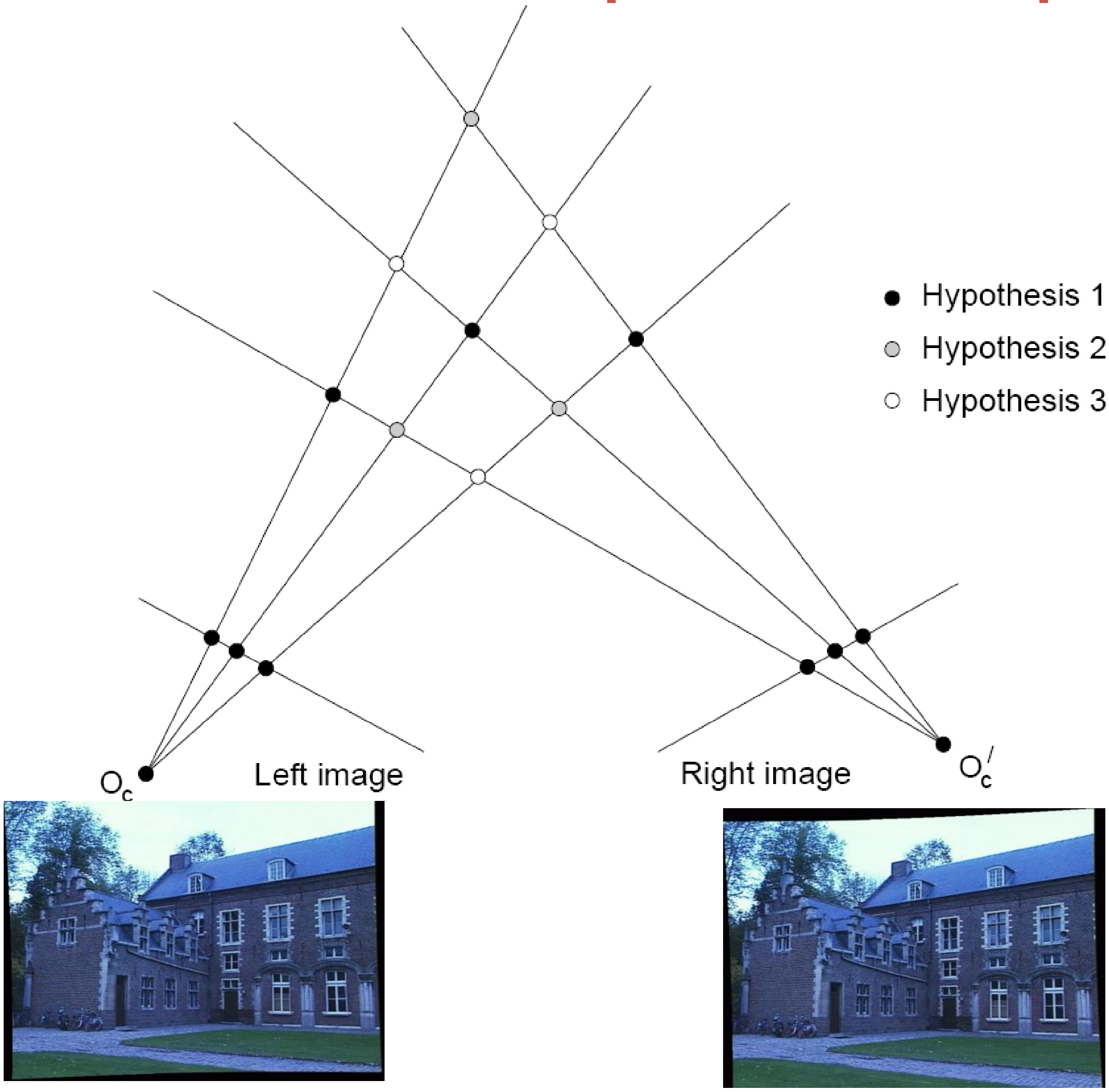




Keep only the matches at are "inliers" with respect to the "best" fundamental matrix



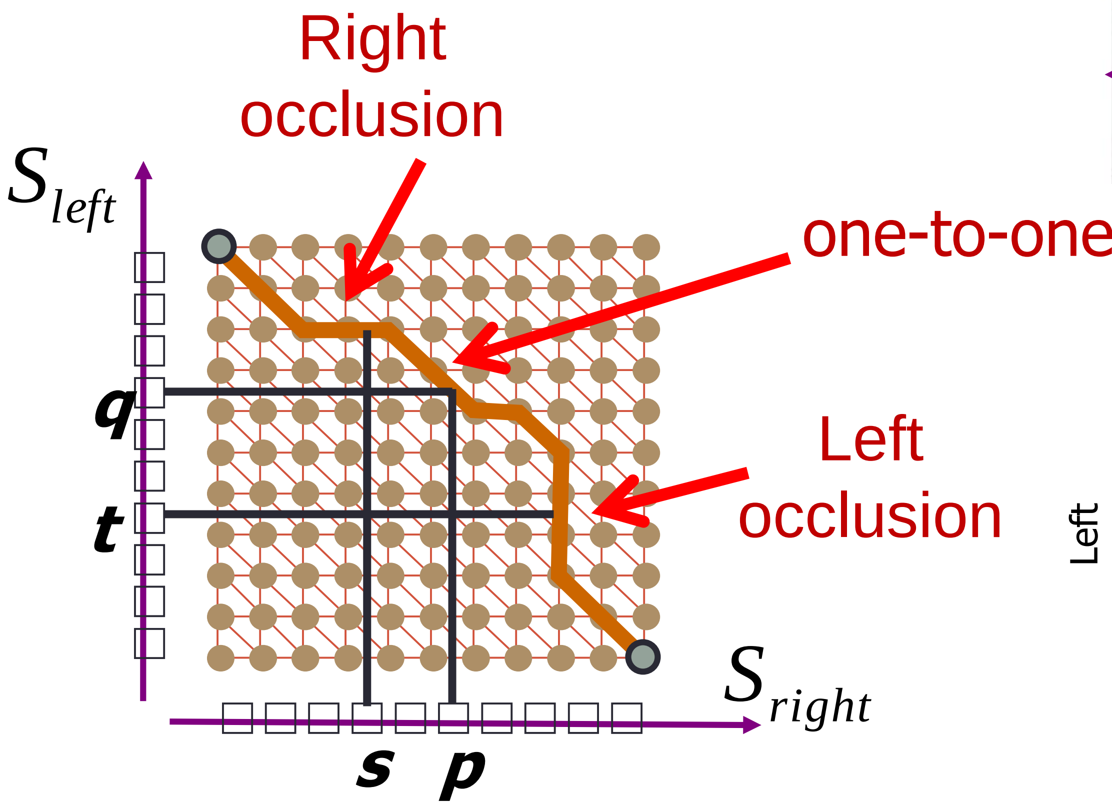
Dense correspondence problem

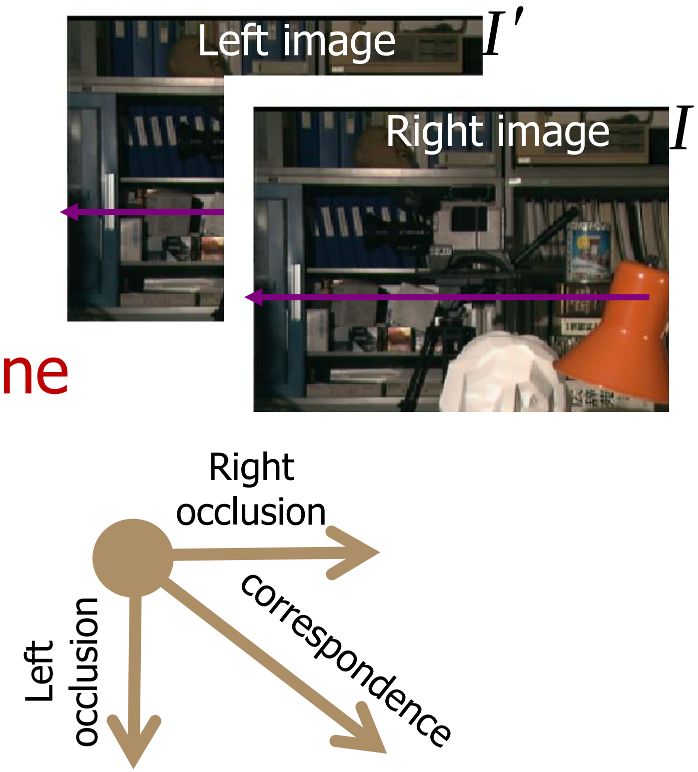


Multiple match hypotheses satisfy epipolar constraint, but which is correct?

Figure from Gee & Cipolla 1999

“Shortest paths” for scan-line stereo





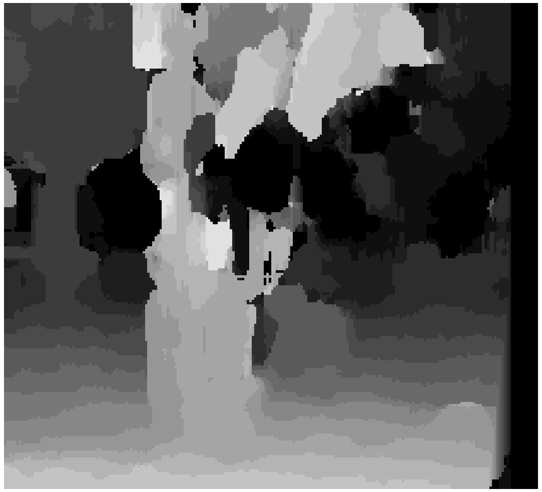
Can be implemented with dynamic programming

Ohta & Kanade ’85, Cox et al. ’96, Intille & Bobick, ‘01

Slide credit: Y. Boykov

Effect of window size

W = 3



W = 20

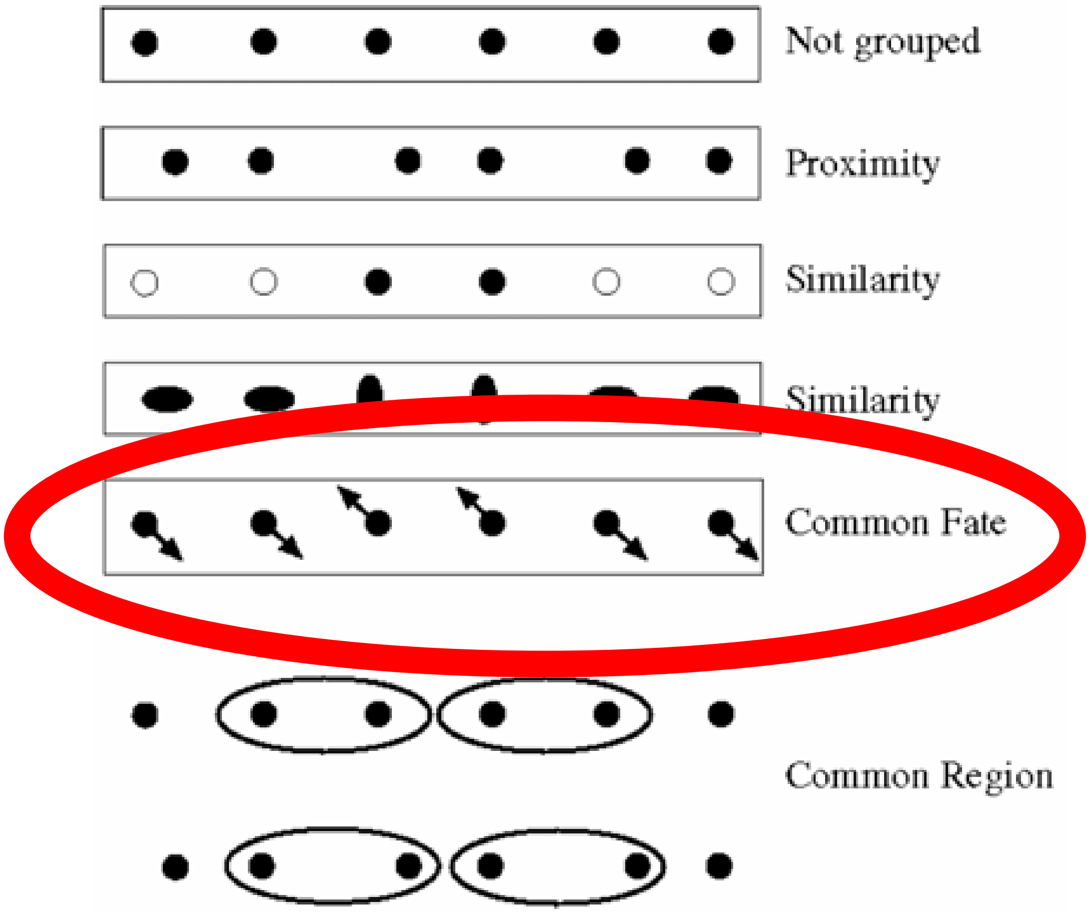
Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

Figures from Li Zhang

Motion and Gestalt laws of grouping

Gestalt psychology
(Max Wertheimer,
1880-1943)

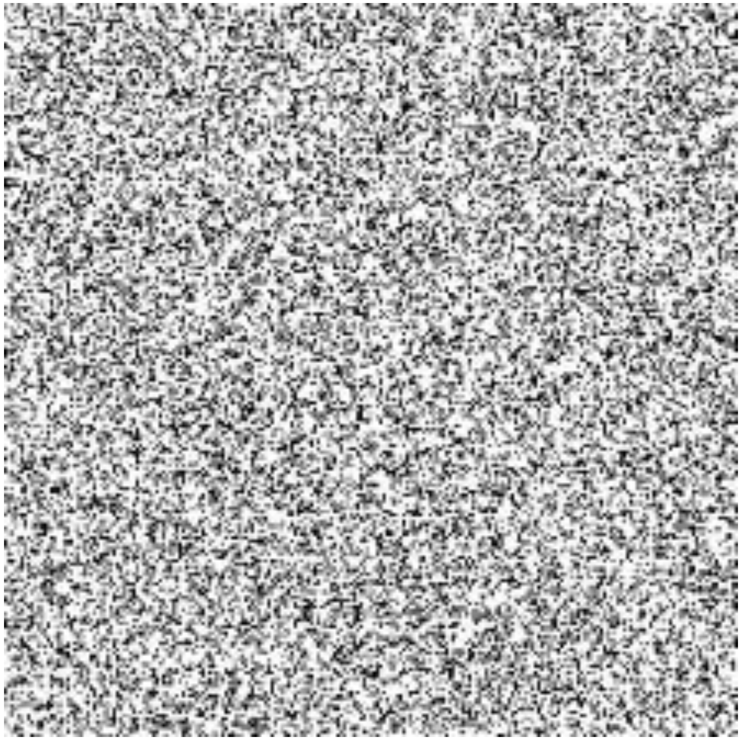
Motion and perceptual organization



Gestalt psychology (Max Wertheimer, 1880-1943)

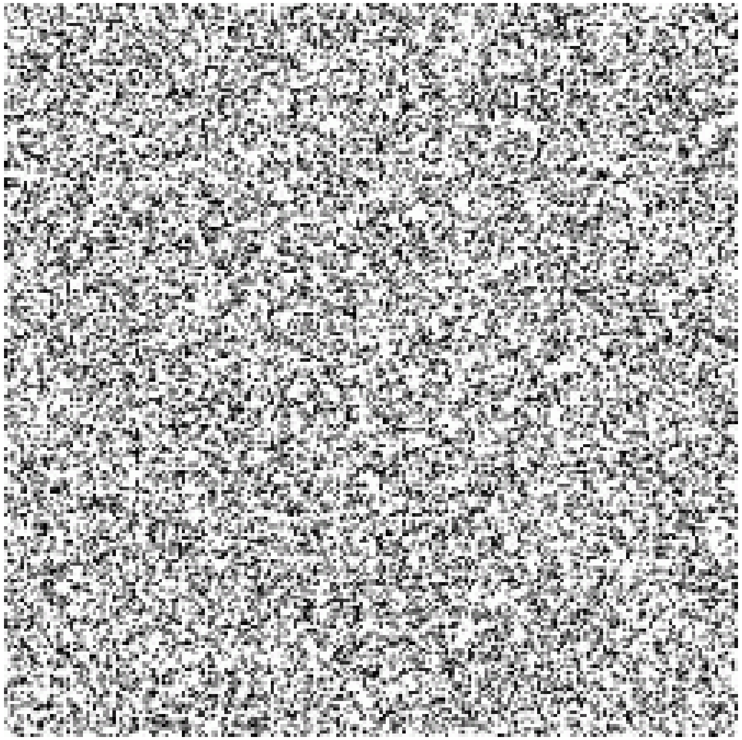
…plus closure, continuation, 'good form'

Motion and perceptual organization



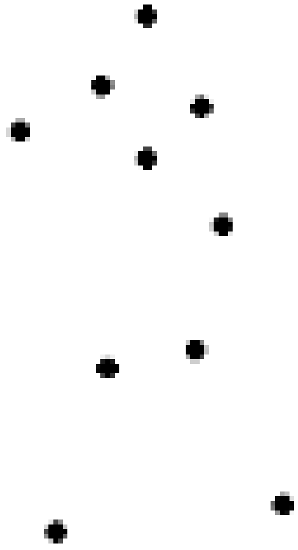
Motion and perceptual organization

- Sometimes motion is the only cue…



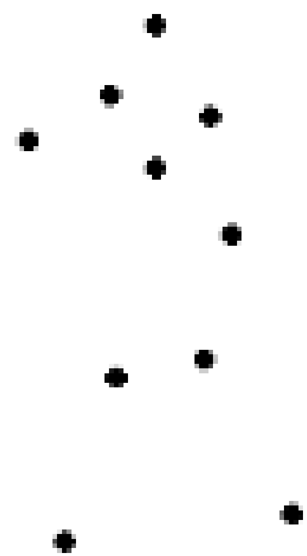
Motion and perceptual organization

Even “impoverished” motion data can evoke a strong percept



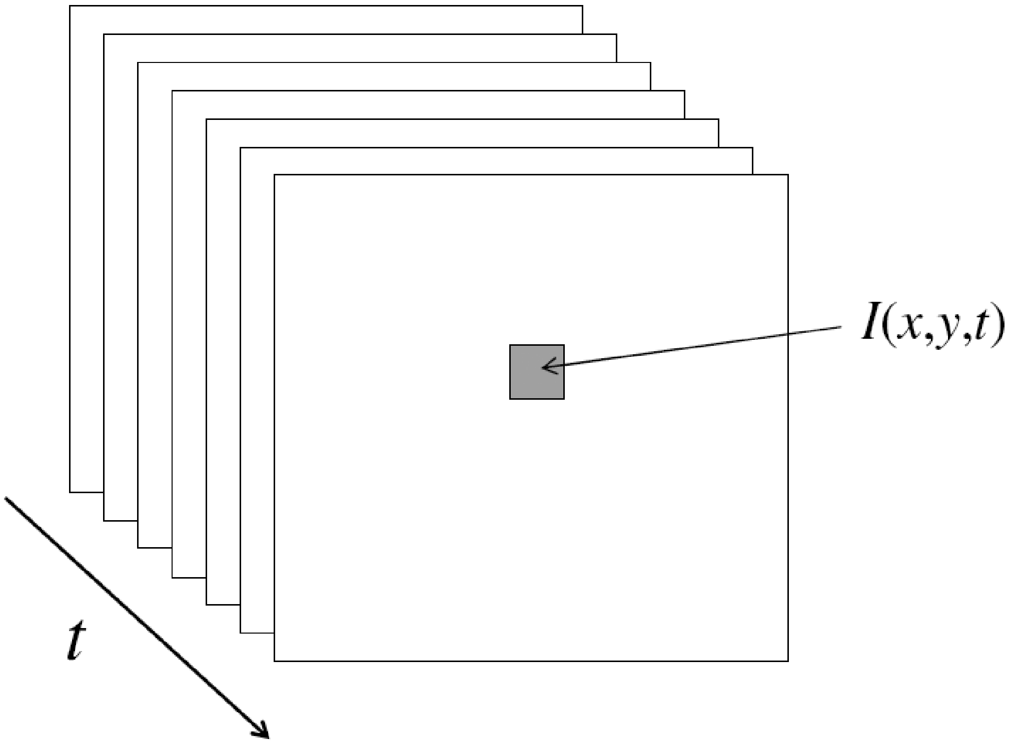
Motion and perceptual organization

Even “impoverished” motion data can evoke a strong percept



Video

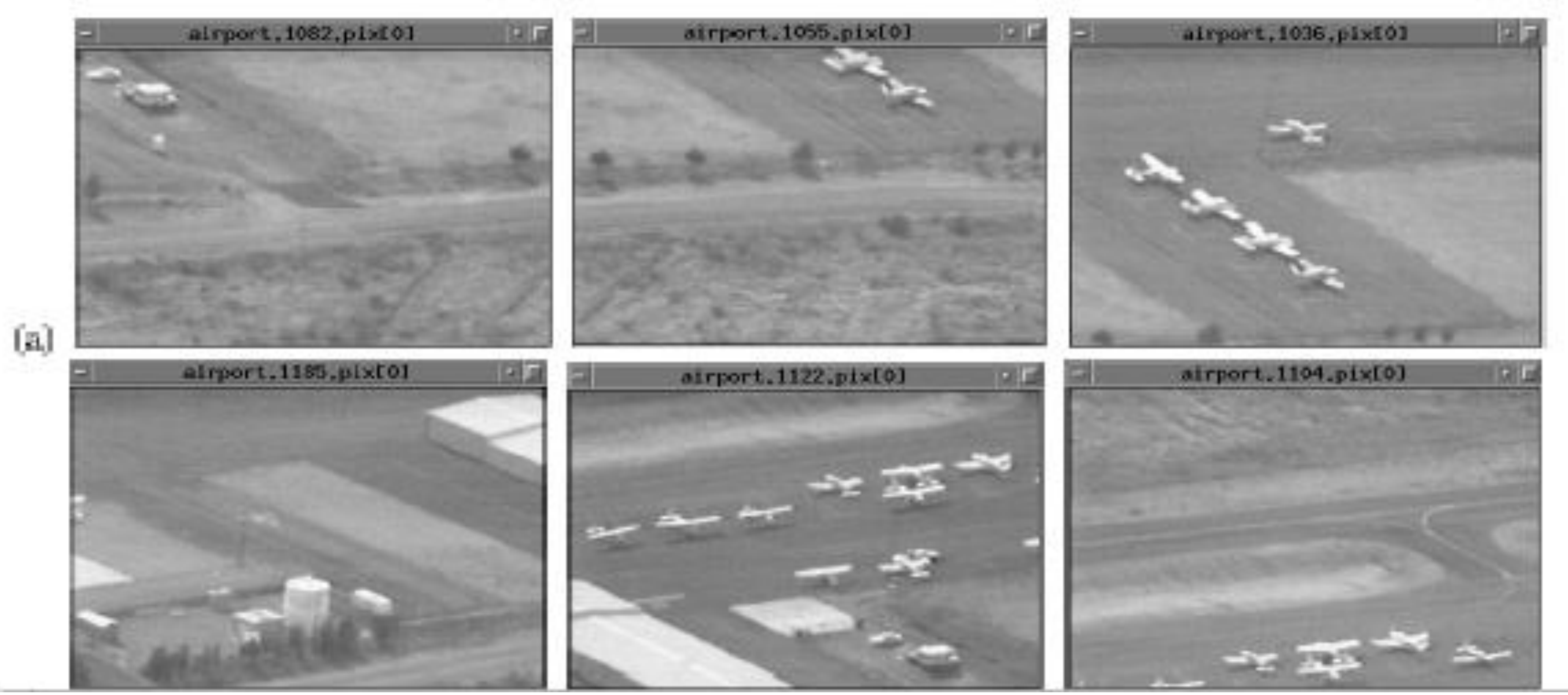
- A video is a sequence of frames captured over time
- A ‘function’ of space (x, y) and time (t)



Motion Applications

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

Mosaicing



(Michal Irani, Weizmann)

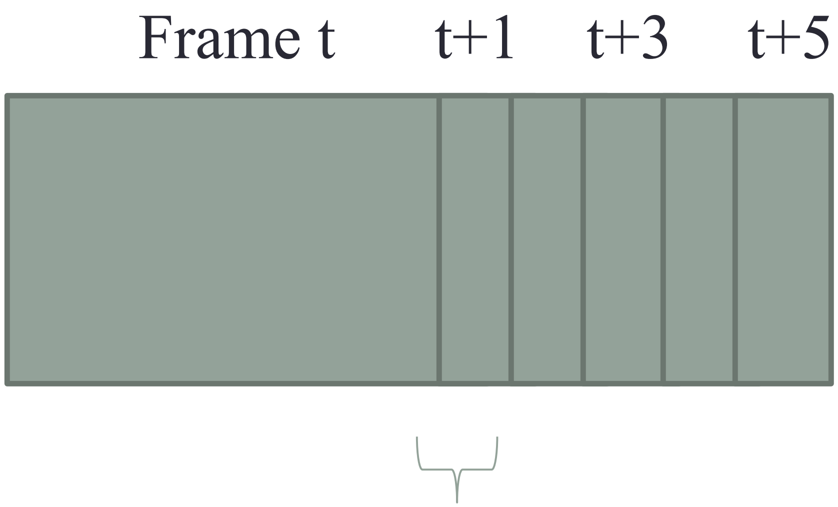
Mosaicing

. 1. Static background mosaic of an airport video clip.
(a) A few representative frames from the minute-long video clip. The video shows an airport being imaged from the air with a moving camera. The scene itself is static (i.e., no moving objects). (b) The static background mosaic image which provides an extended view of the entire scene imaged by the camera in the one-minute video clip.

(Michal Irani, Weizmann)

Mosaicing for Panoramas on Smartphones

Left to right sweep of video camera



Compare small
overlap for efficiency

Mosaicing for Panoramas on Smartphones

Mosaicing for Panoramas on Smartphones

Mosaicing for Panoramas on Smartphones

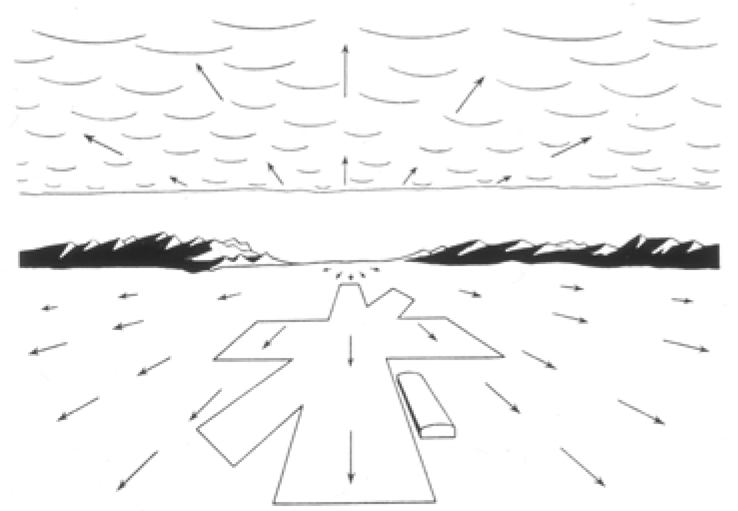
Mosaicing for Panoramas on Smartphones

Motion estimation techniques

- 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)

- Direct, dense 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
    - *Optical flow*!

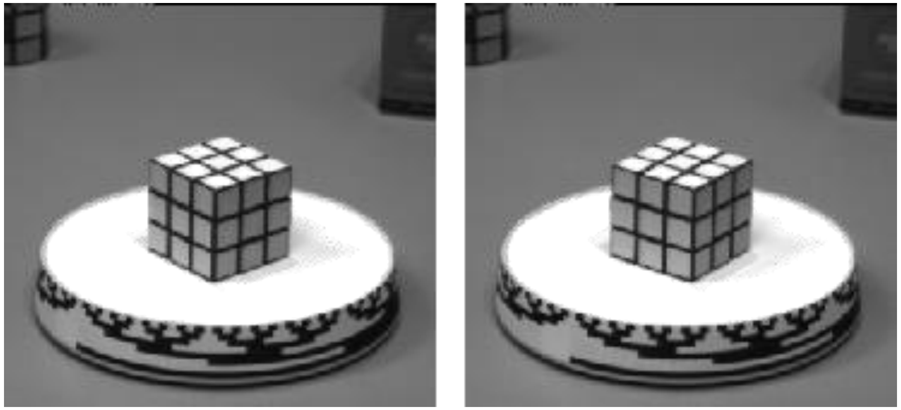
Computer Vision
Motion and Optical Flow

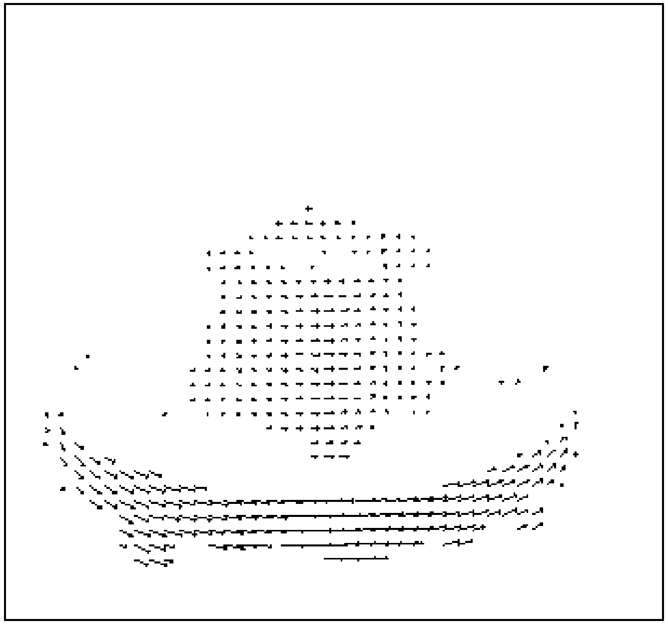


Many slides adapted from J. Hays, S. Seitz, R. Szeliski, M. Pollefeys, K. Grauman and others…

Motion estimation: Optical flow

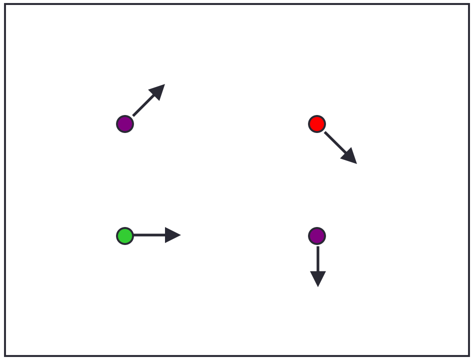
Optic flow is the **apparent** motion of objects or surfaces



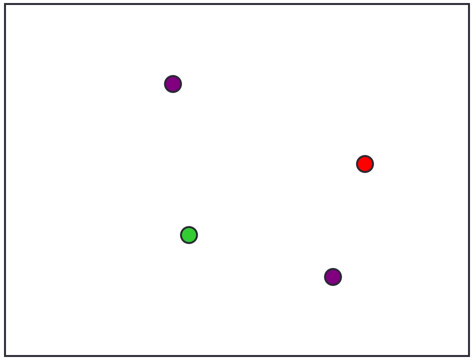


Will start by estimating motion of each pixel separately
Then will consider motion of entire image

Problem definition: optical flow



$I(x, y, t)$



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

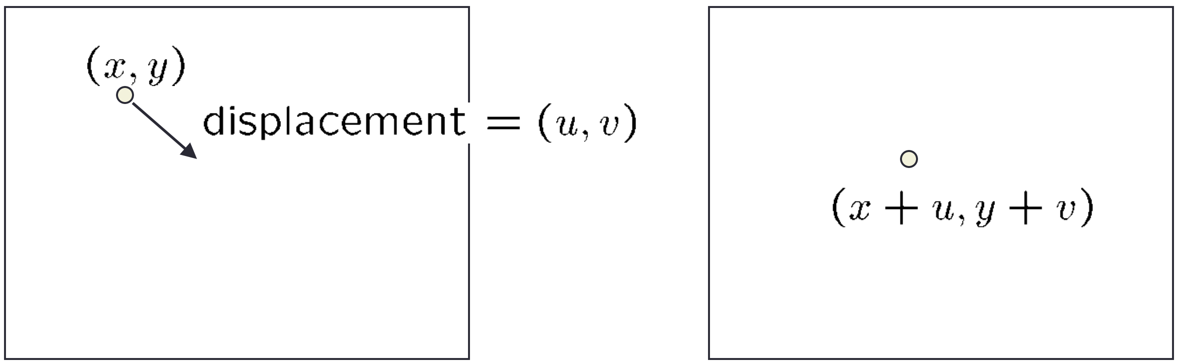
How to estimate pixel motion from image $I(x,y,t)$ to $I(x,y,t+1)$?

- 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

- Small motion: Points do not move very far
- 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

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

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

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

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

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

Brightness constancy constraint equation

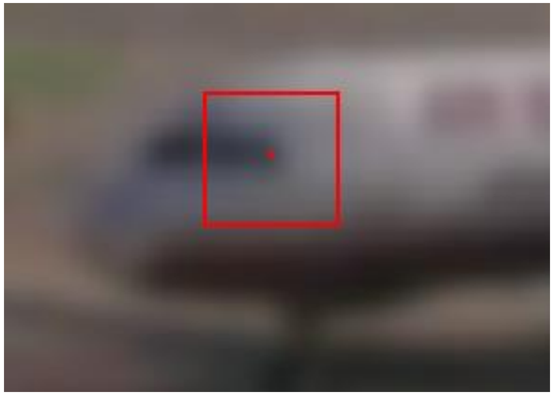
$$I_x u + I_y v + I_t = 0$$

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?



The brightness constancy constraint

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

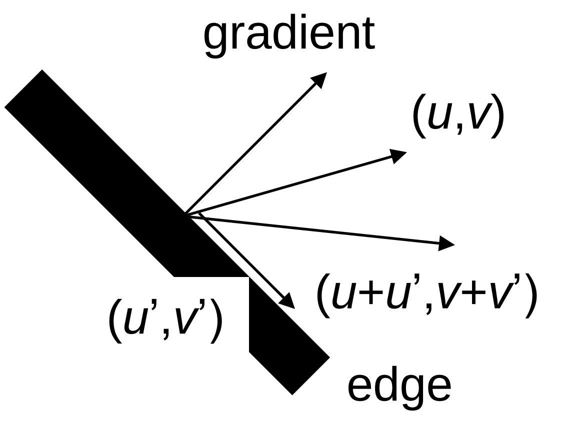
$$0 = I_t + \nabla I \cdot < u, v > \quad \text{or} \quad I_x u + I_y v + I_t = 0$$

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

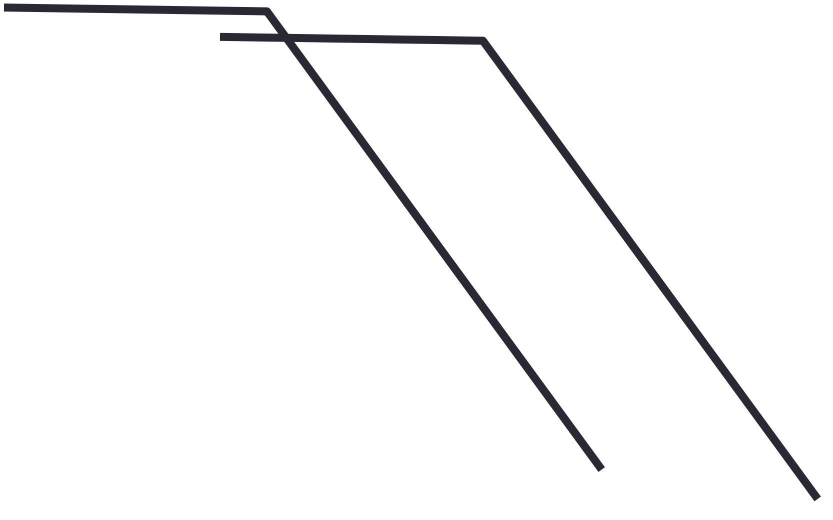
The component of the motion perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

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

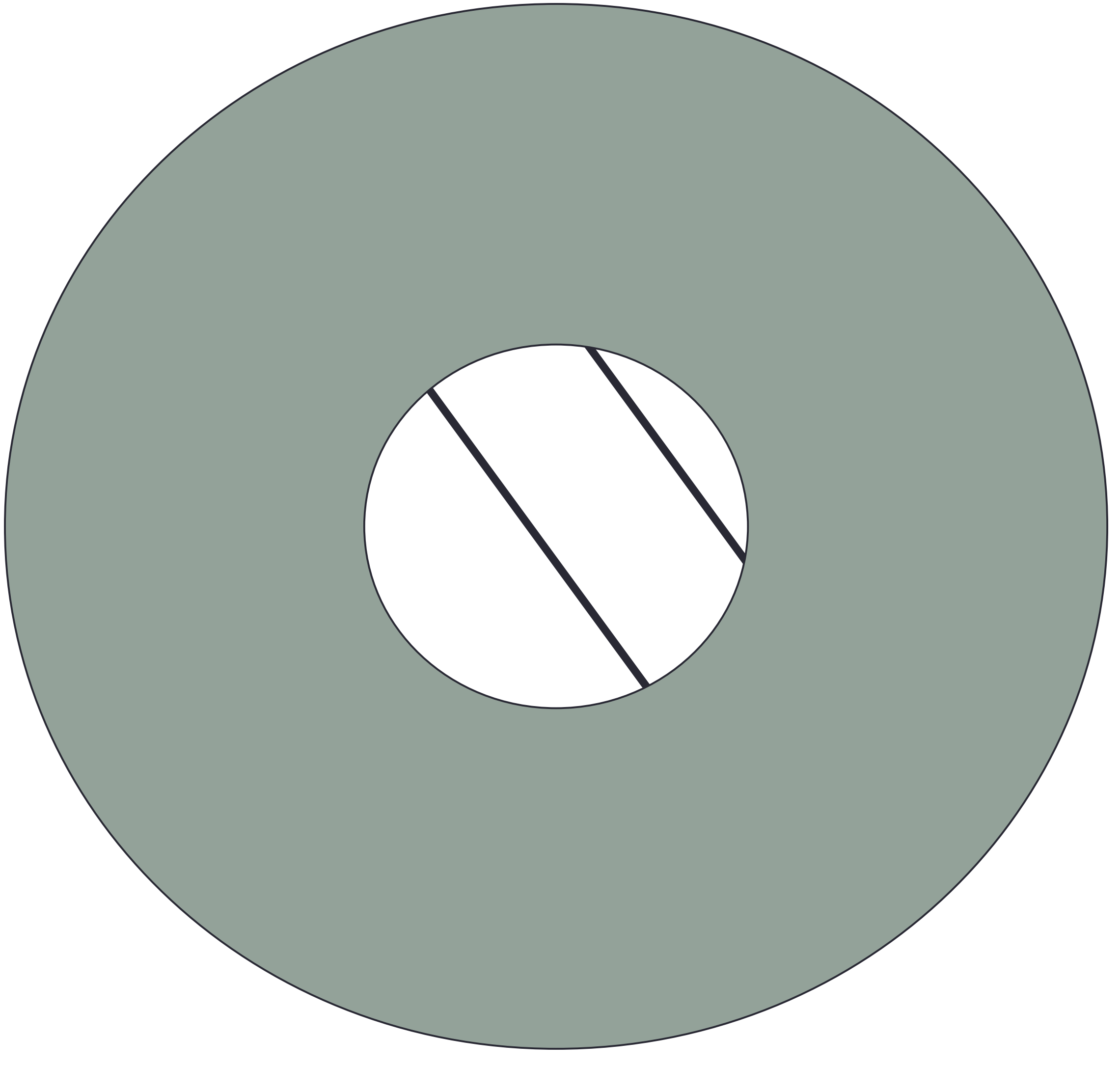
$$\nabla \mathrm{I} \cdot \left[\mathrm{u'} \ \ \mathrm{v'} \right]^{\mathrm{T}} = 0$$



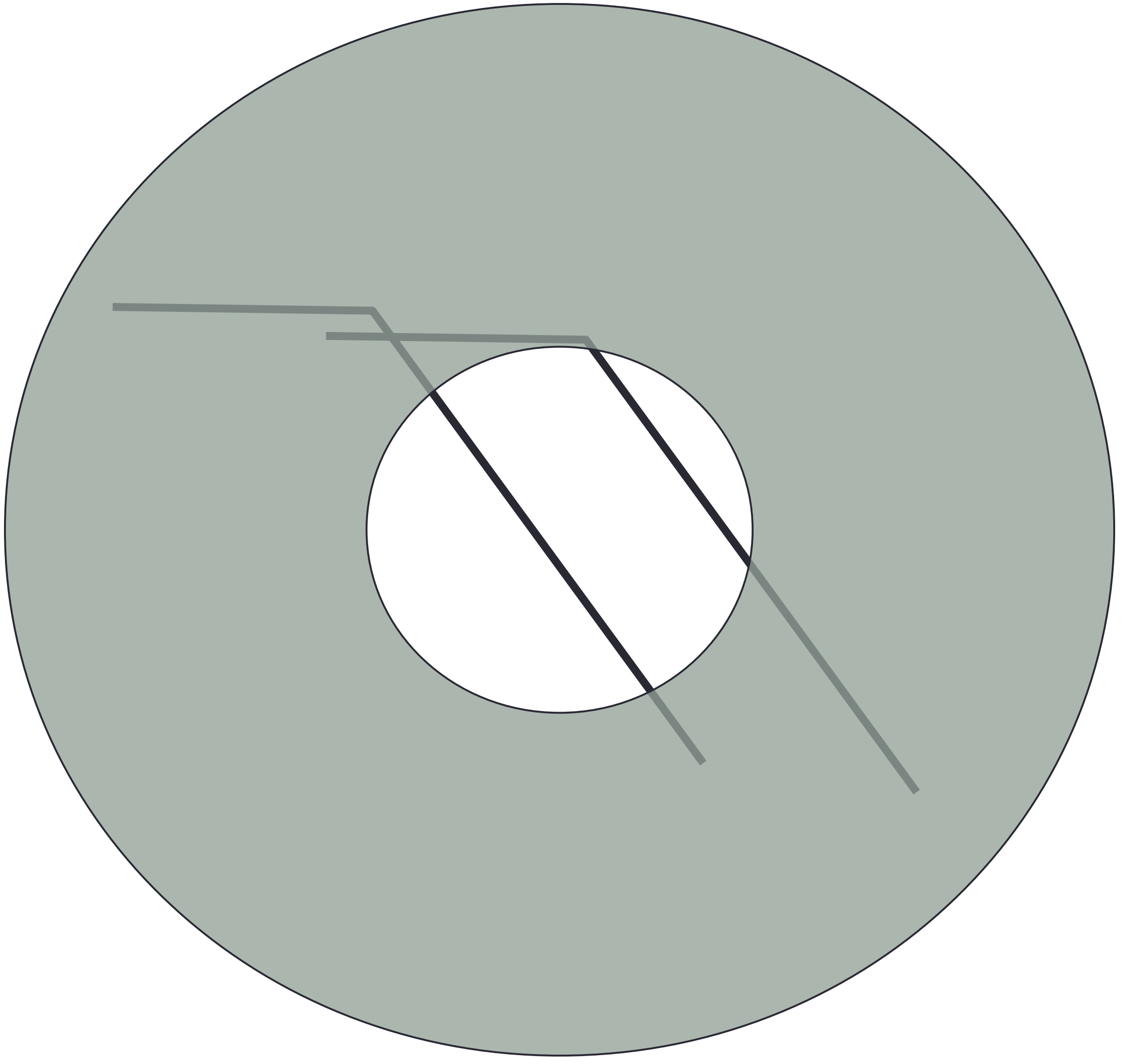
Aperture problem



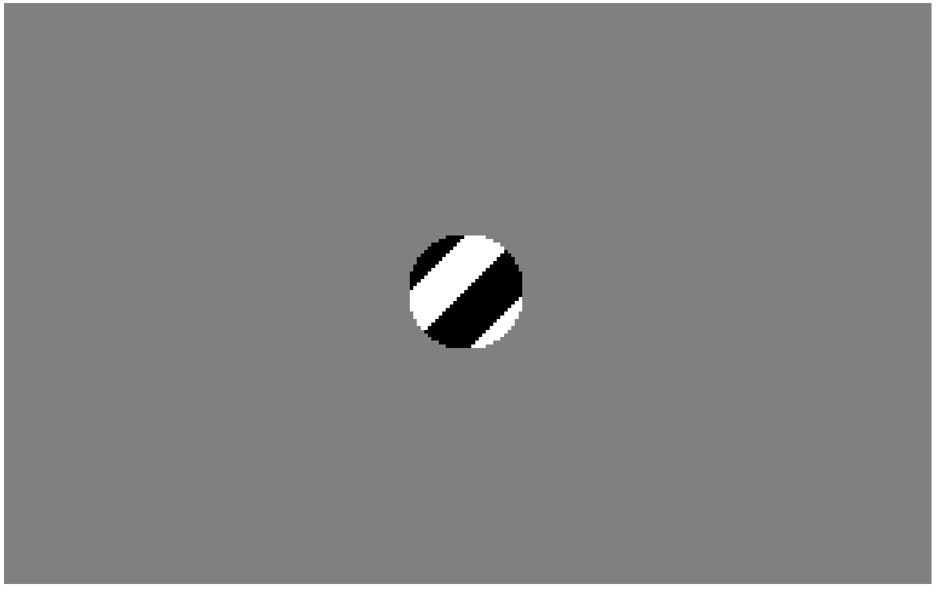
Aperture problem



Aperture problem

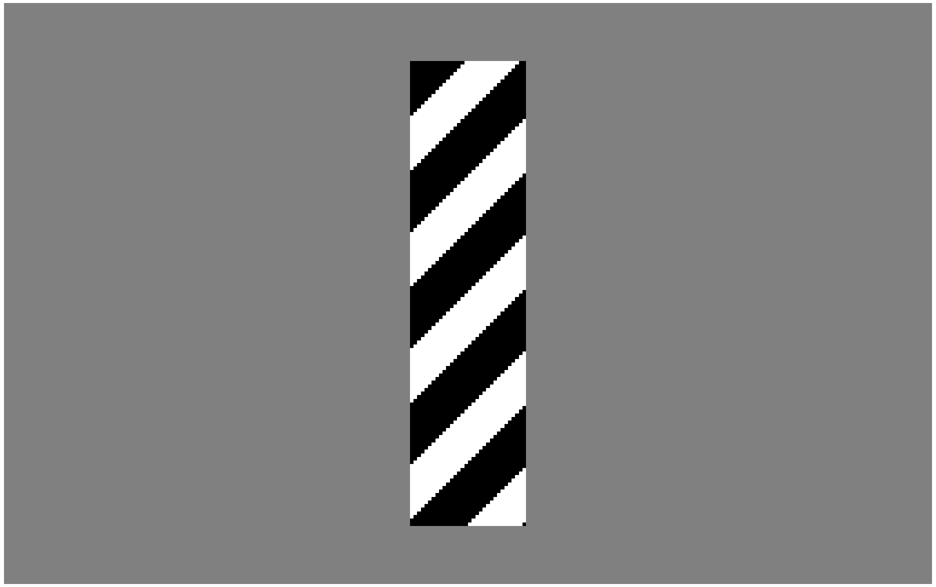


The barber pole illusion



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:

$$\begin{bmatrix} I_x(\mathbf{p_1}) & I_y(\mathbf{p_1}) \\ I_x(\mathbf{p_2}) & I_y(\mathbf{p_2}) \\ \vdots & \vdots \\ I_x(\mathbf{p_{25}}) & I_y(\mathbf{p_{25}}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p_1}) \\ I_t(\mathbf{p_2}) \\ \vdots \\ I_t(\mathbf{p_{25}}) \end{bmatrix} \qquad \underset{25\text{x}2}{A} \; \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b}$$

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

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

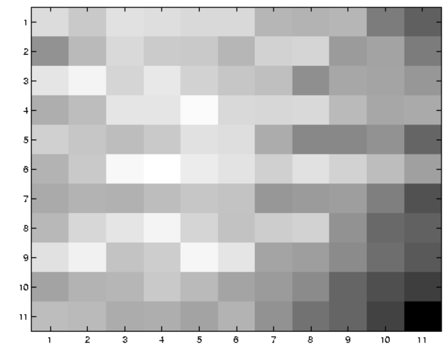
When is this solvable? What are good points to track?

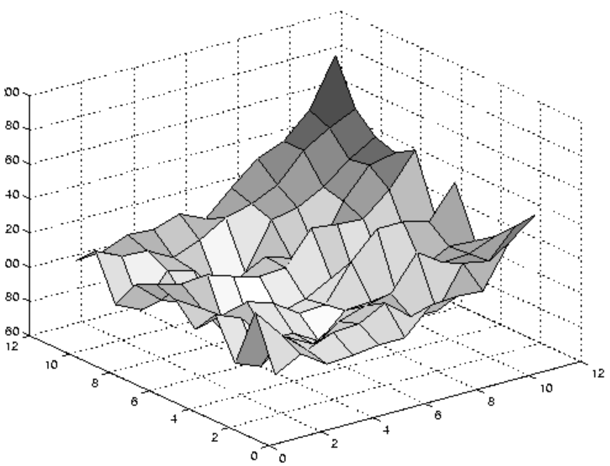
- **A^TA** should be invertible
- **A^TA** should not be too small due to noise
  - eigenvalues λ_1 and λ_2 of **A^TA** should not be too small
- **A^TA** should be well-conditioned
  - λ_1 / λ_2 should not be too large (λ_1 = larger eigenvalue)

Does this remind you of anything?

Criteria for Harris corner detector

Low texture region



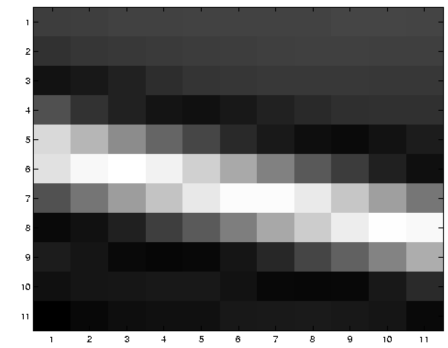


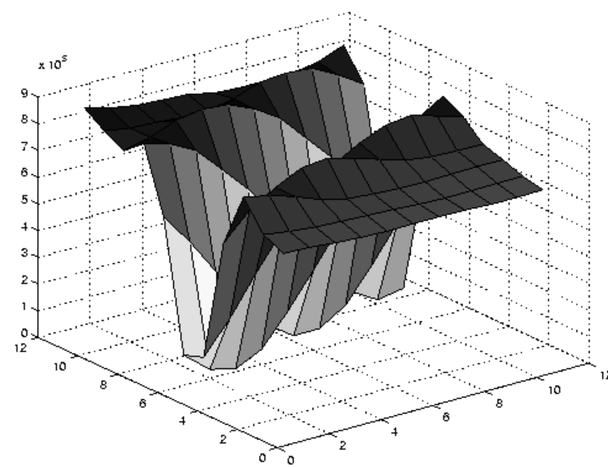
$$\sum \nabla I (\nabla I)^T$$

– gradients have small magnitude

– small λ_1, small λ_2

Edge



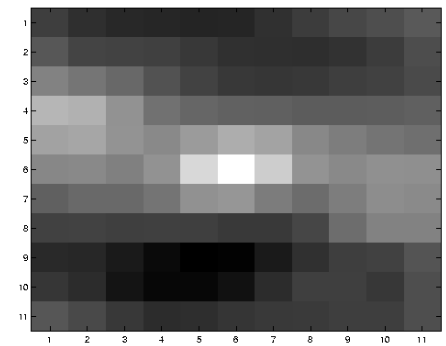


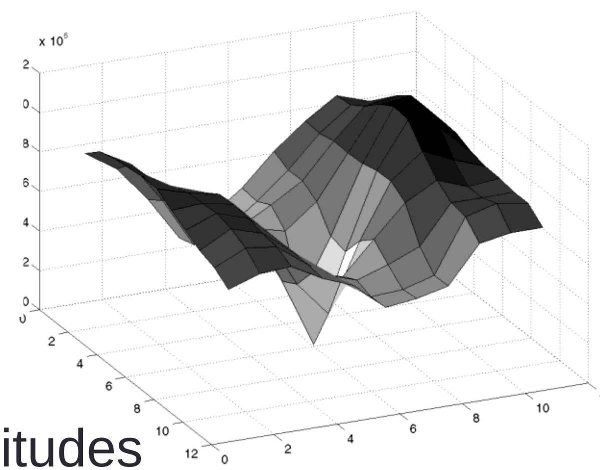
$$\sum \nabla I (\nabla I)^T$$

– large gradients, all the same

– large λ_1, small λ_2

High textured region



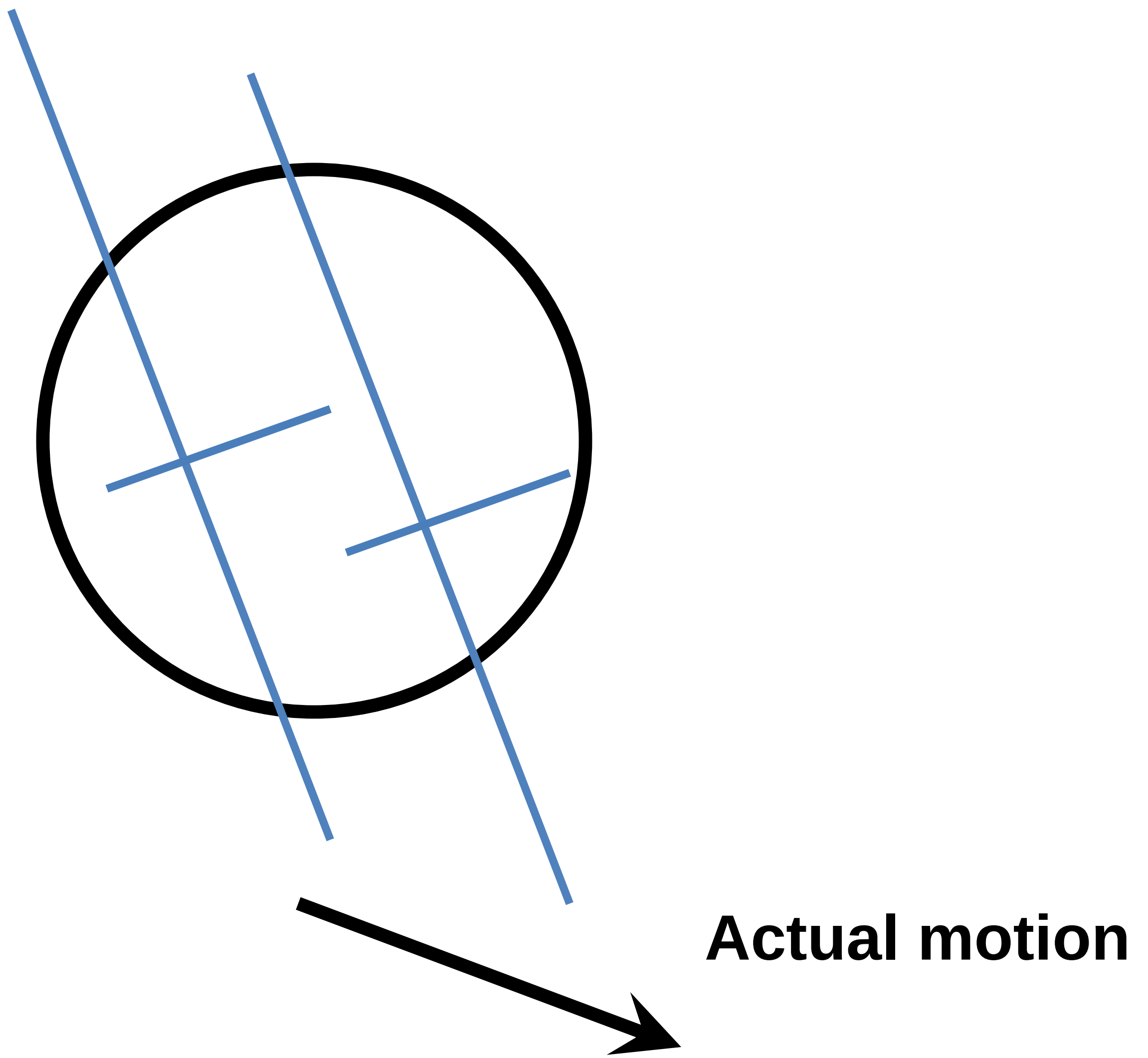


$$\sum \nabla I (\nabla I)^T$$

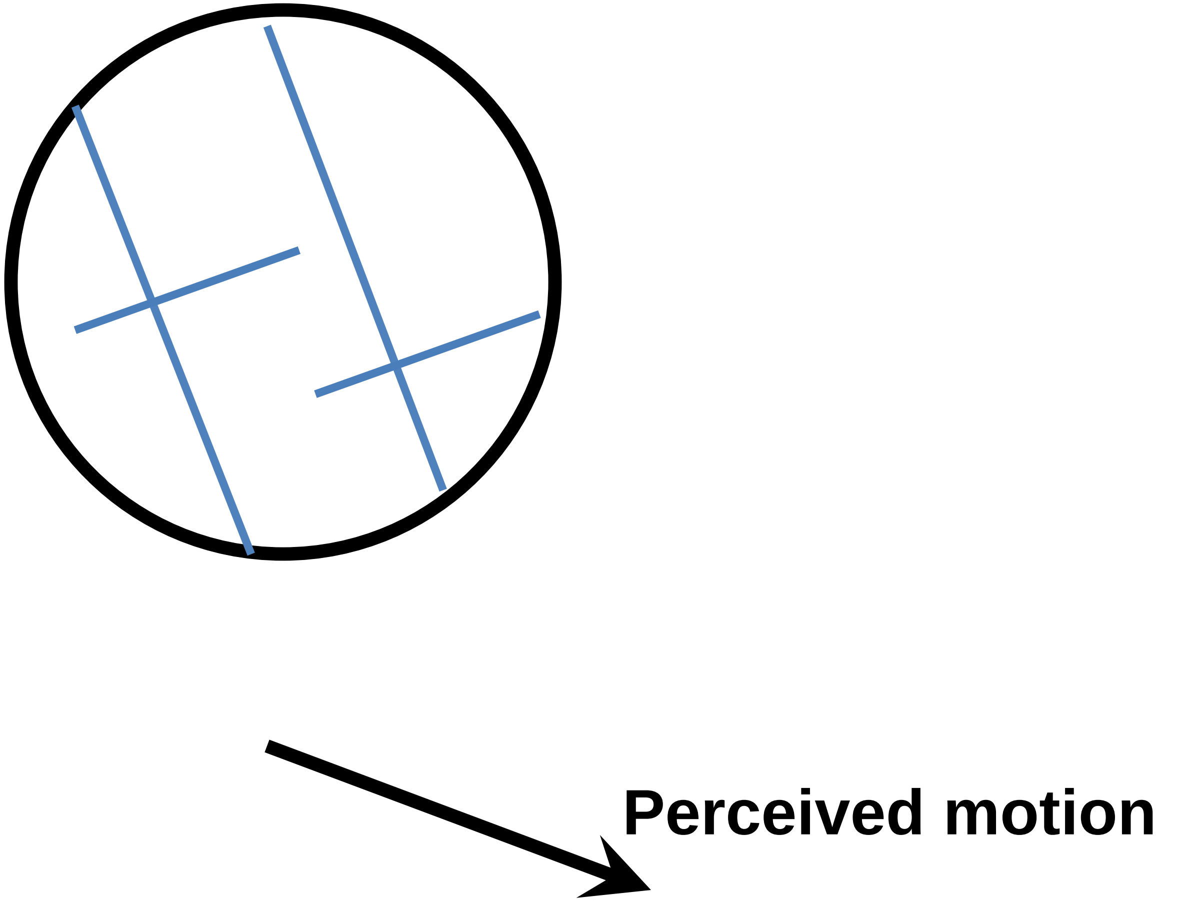
– gradients are different, large magnitudes

– large λ_1, large λ_2

The aperture problem resolved



The aperture problem resolved



Errors in assumptions

- A point does not move like its neighbors
  - Motion segmentation
- Brightness constancy does not hold
  - Do exhaustive neighborhood search with normalized correlation - tracking features – maybe SIFT – more later….
- The motion is large (larger than a pixel)
  1. Not-linear: Iterative refinement
  2. Local minima: coarse-to-fine estimation

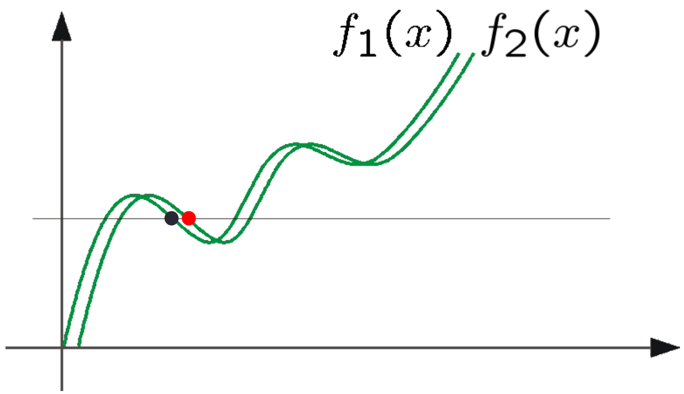
Revisiting the small motion assumption

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

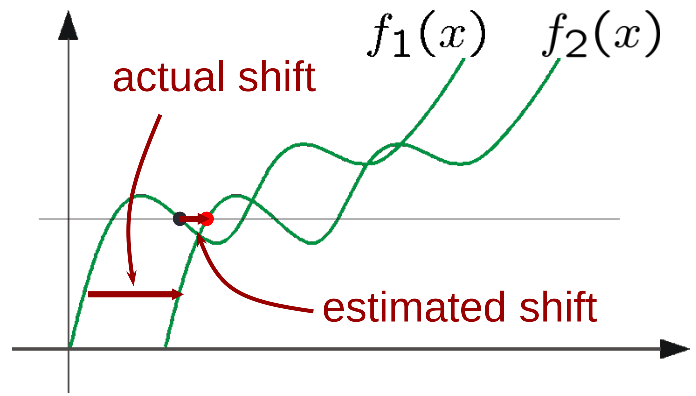
Optical Flow: Aliasing

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?



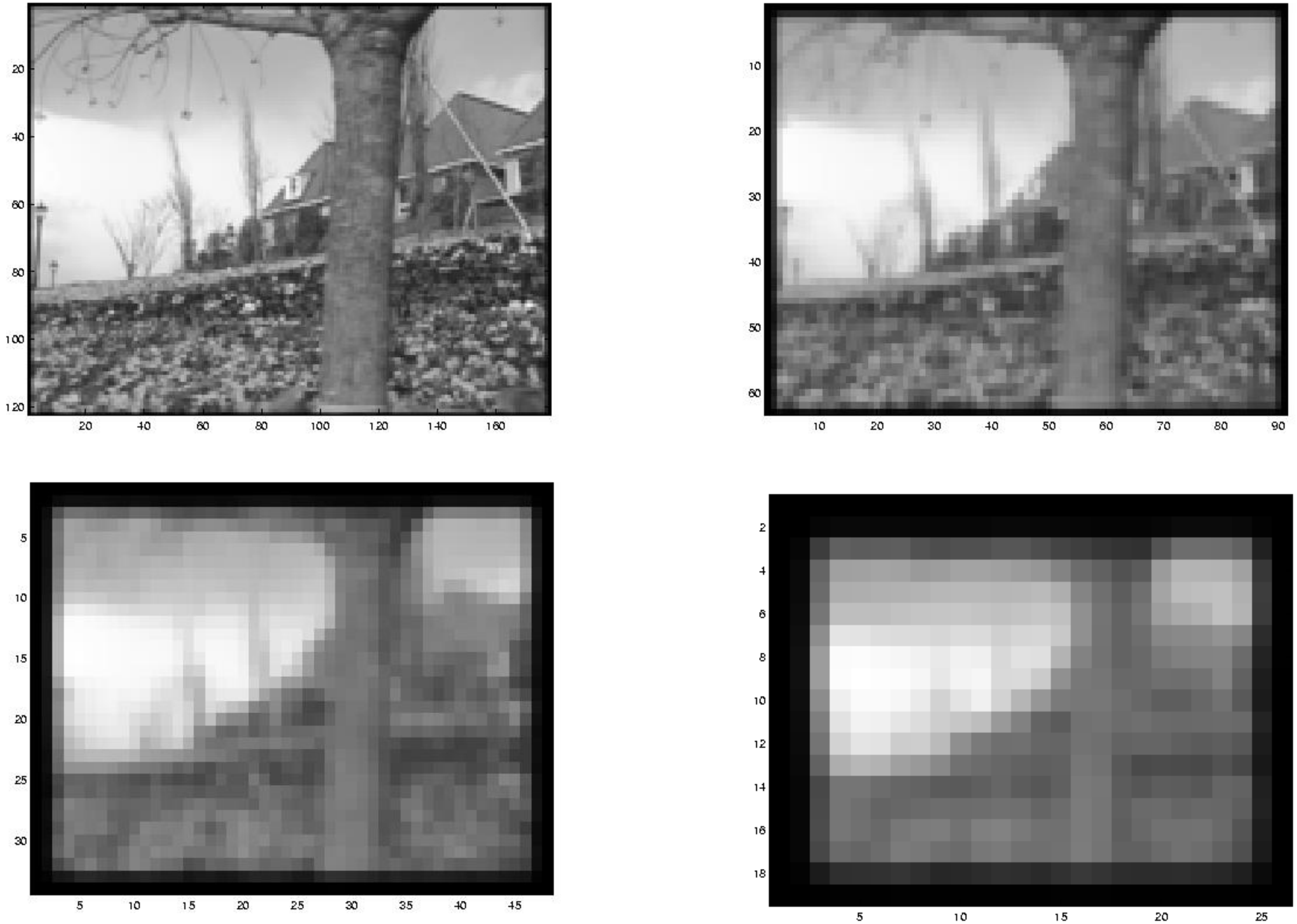
nearest match is correct (no aliasing)



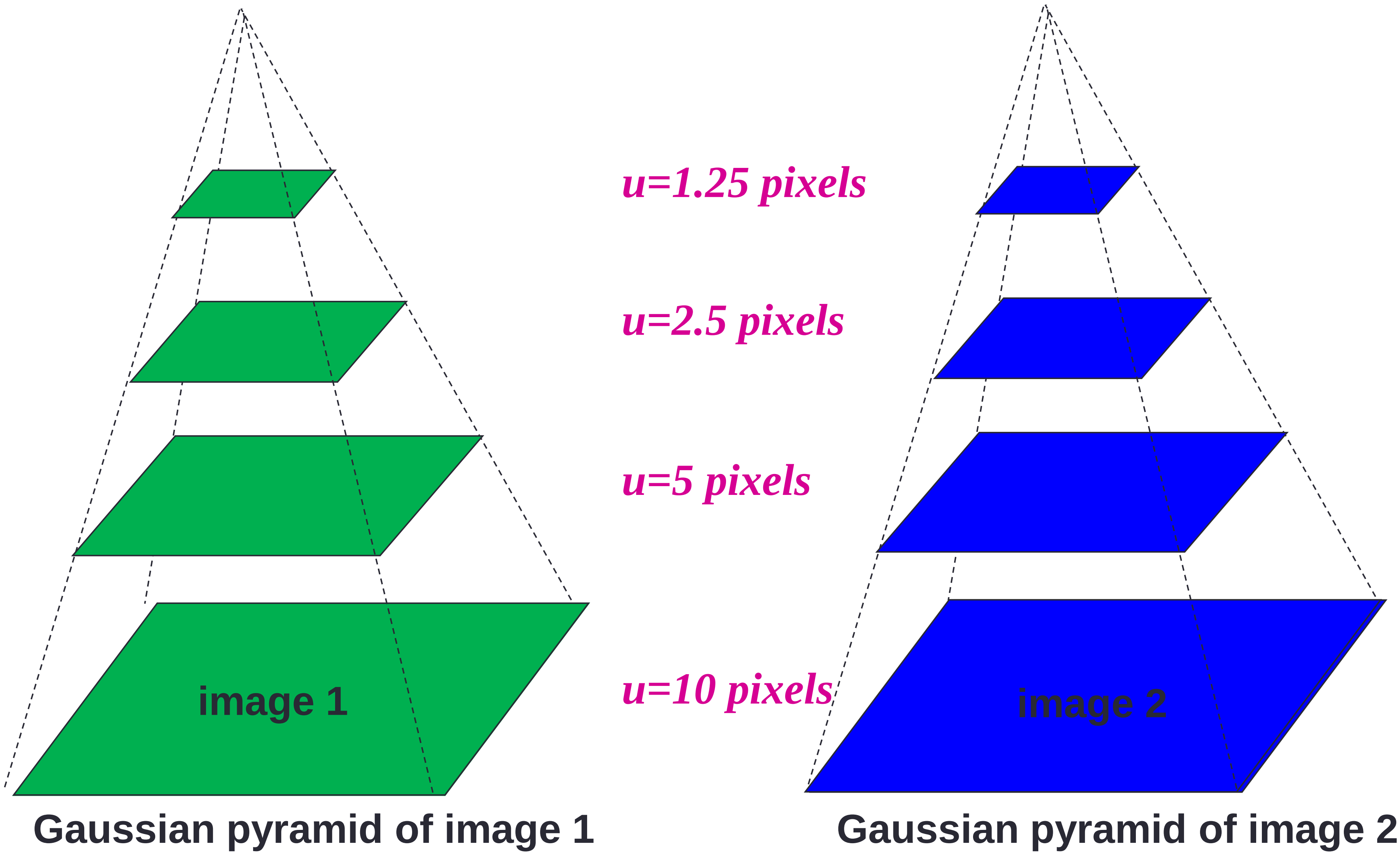
nearest match is incorrect (aliasing)

To overcome aliasing: coarse-to-fine estimation.

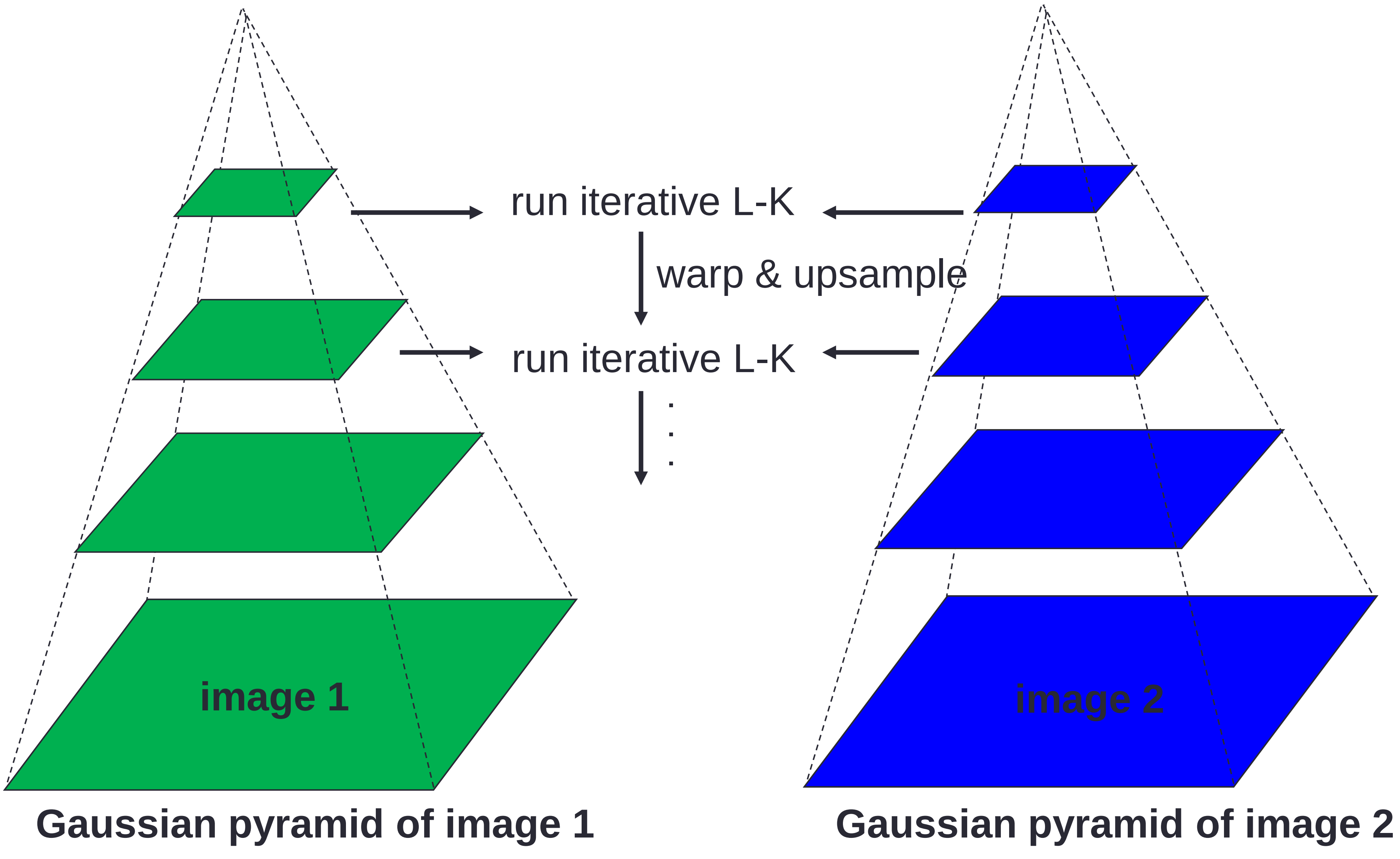
Reduce the resolution!



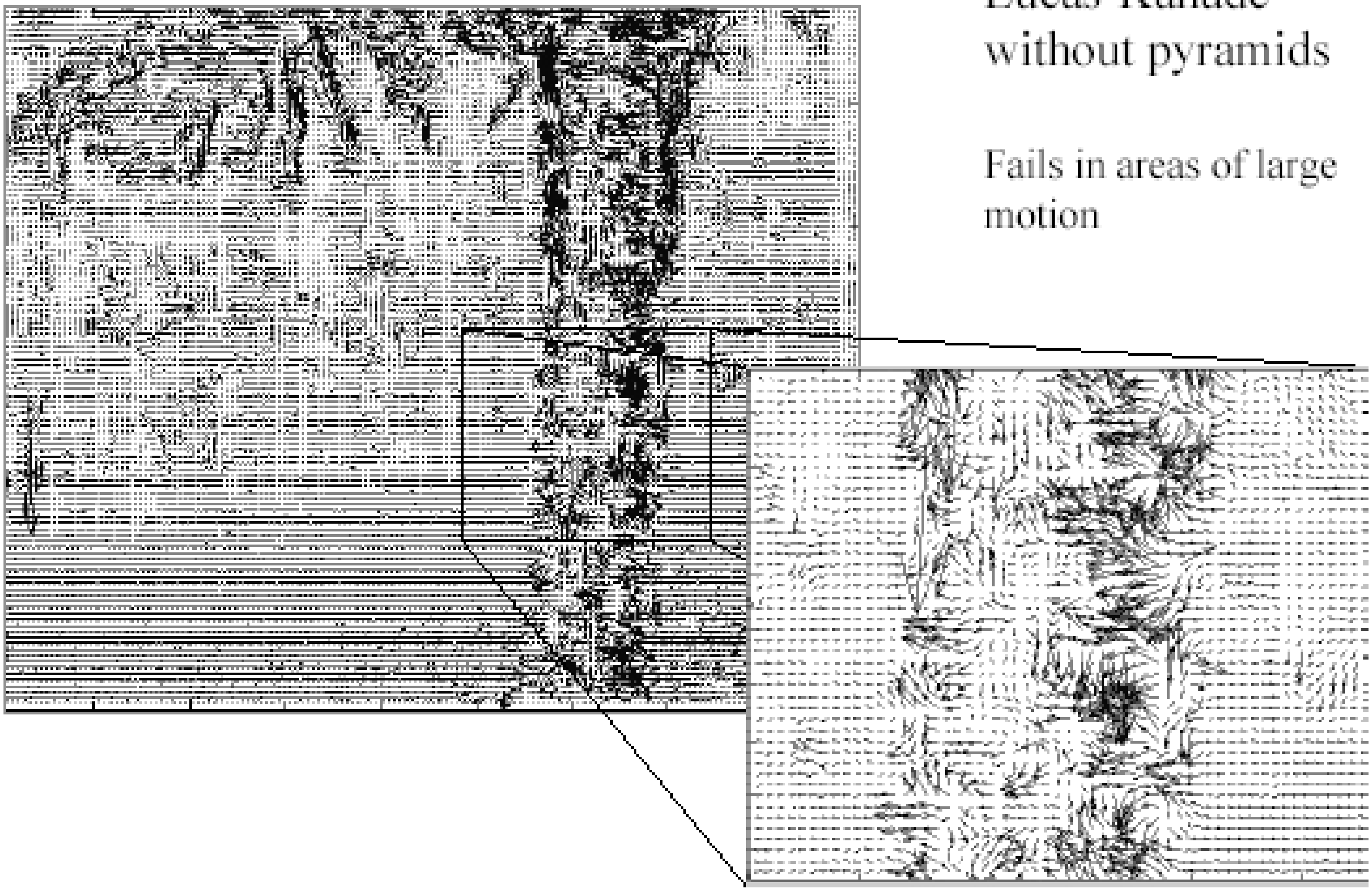
Coarse-to-fine optical flow estimation



Coarse-to-fine optical flow estimation



Optical Flow Results

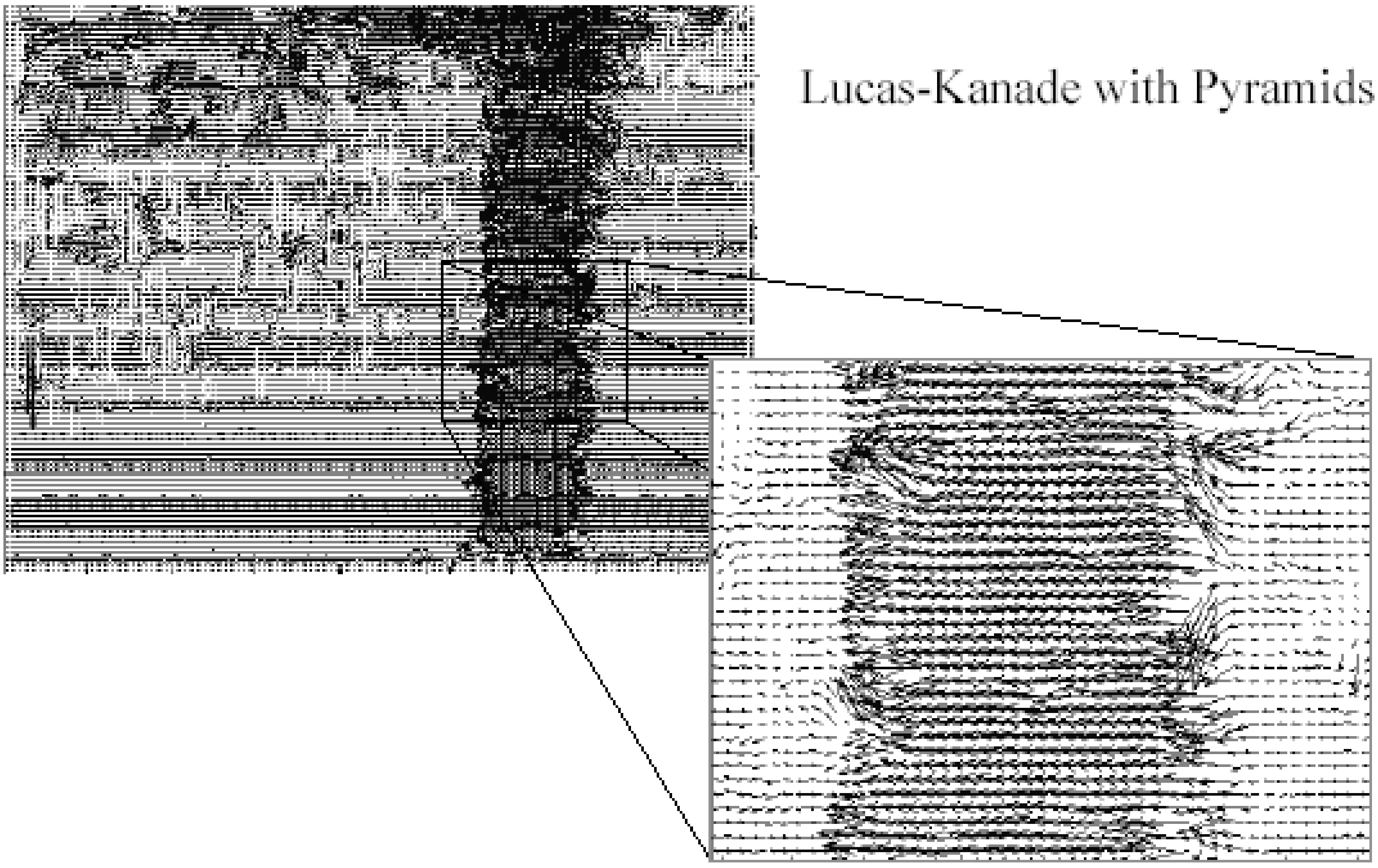


Lucas-Kanade without pyramids

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, 2009

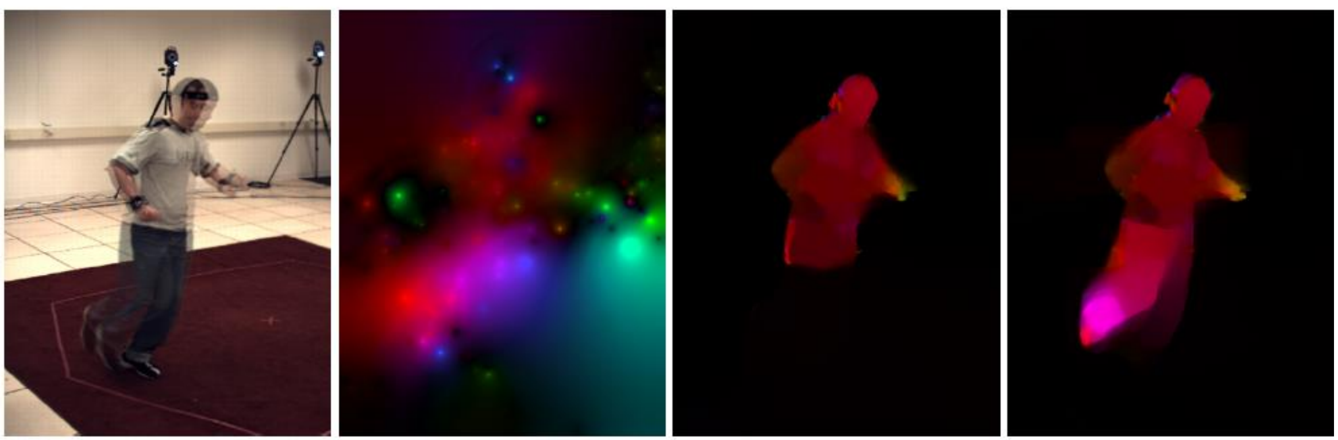
Start with something similar to Lucas-Kanade

+ 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

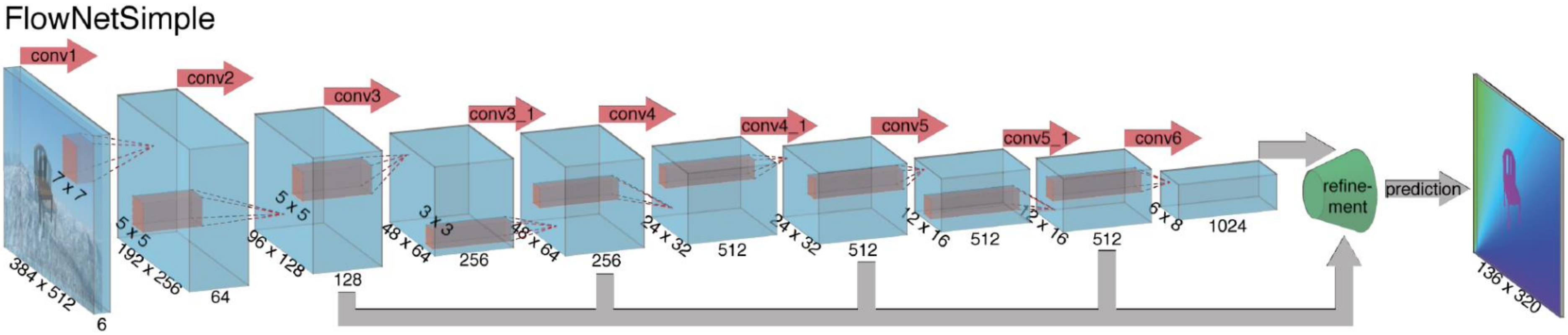
State-of-the-art optical flow, 2015

CNN

Pair of input frames

Upsample estimated flow back to input resolution

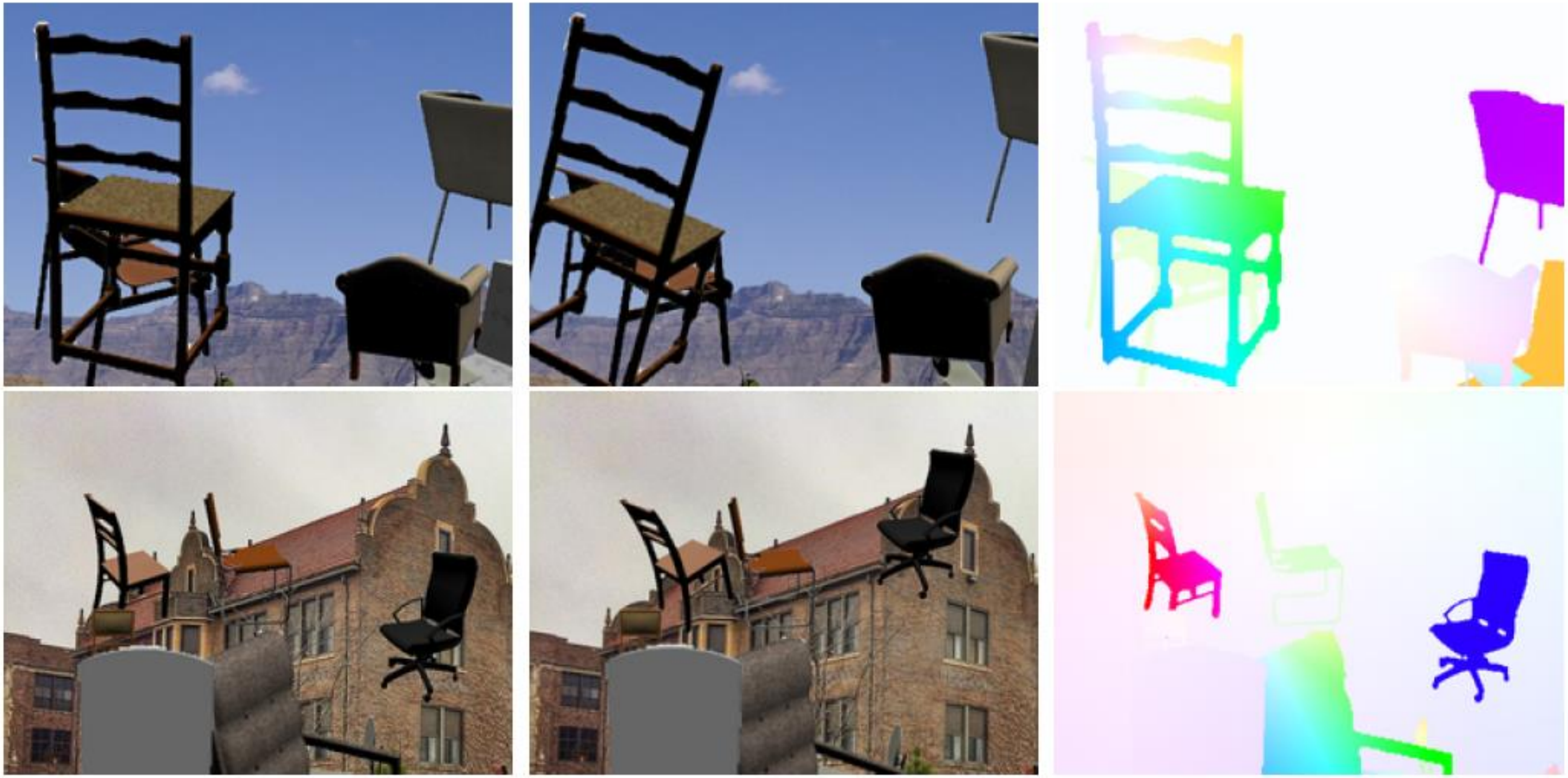
Near state-of-the-art in terms of end-point-error



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

State-of-the-art optical flow, 2015

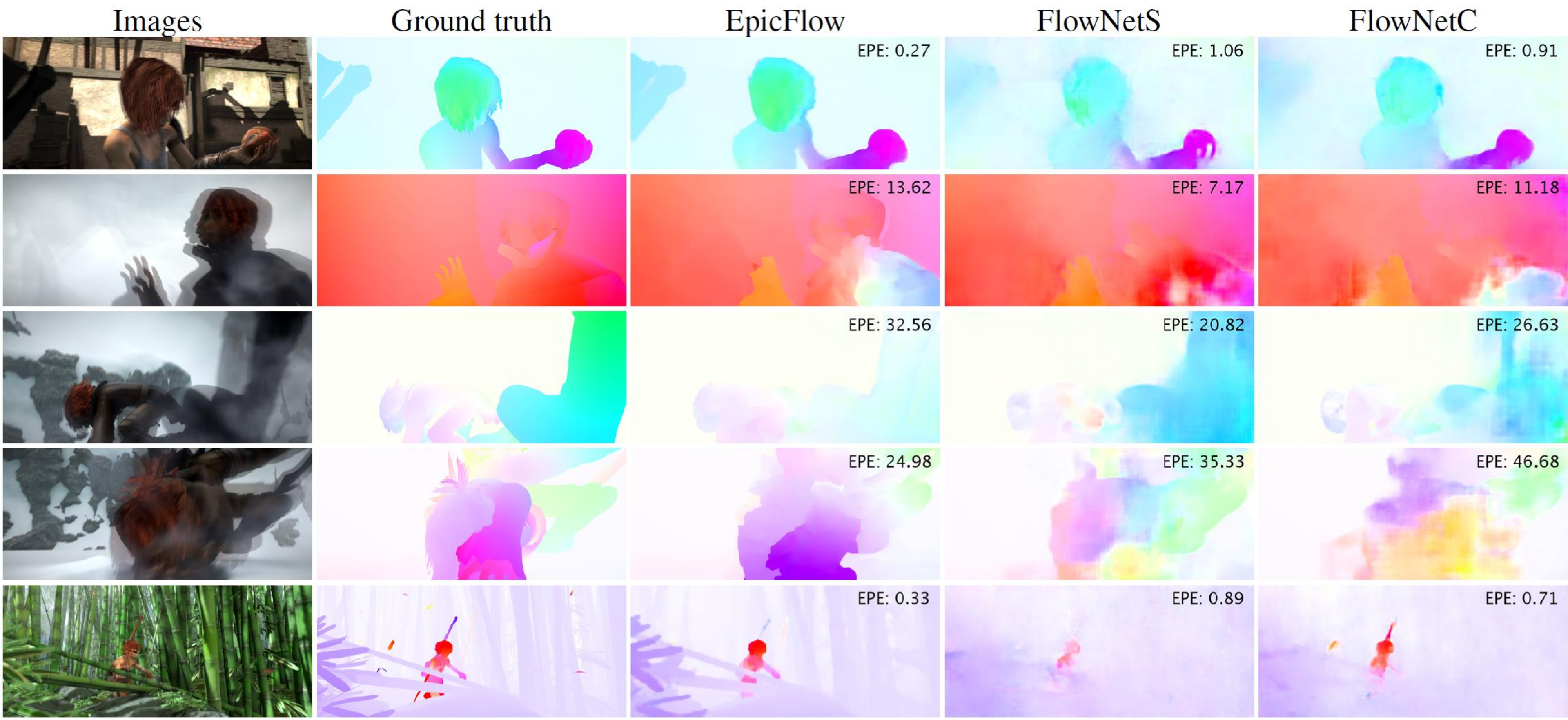
Synthetic Training data



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

State-of-the-art optical flow, 2015

Results on Sintel



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

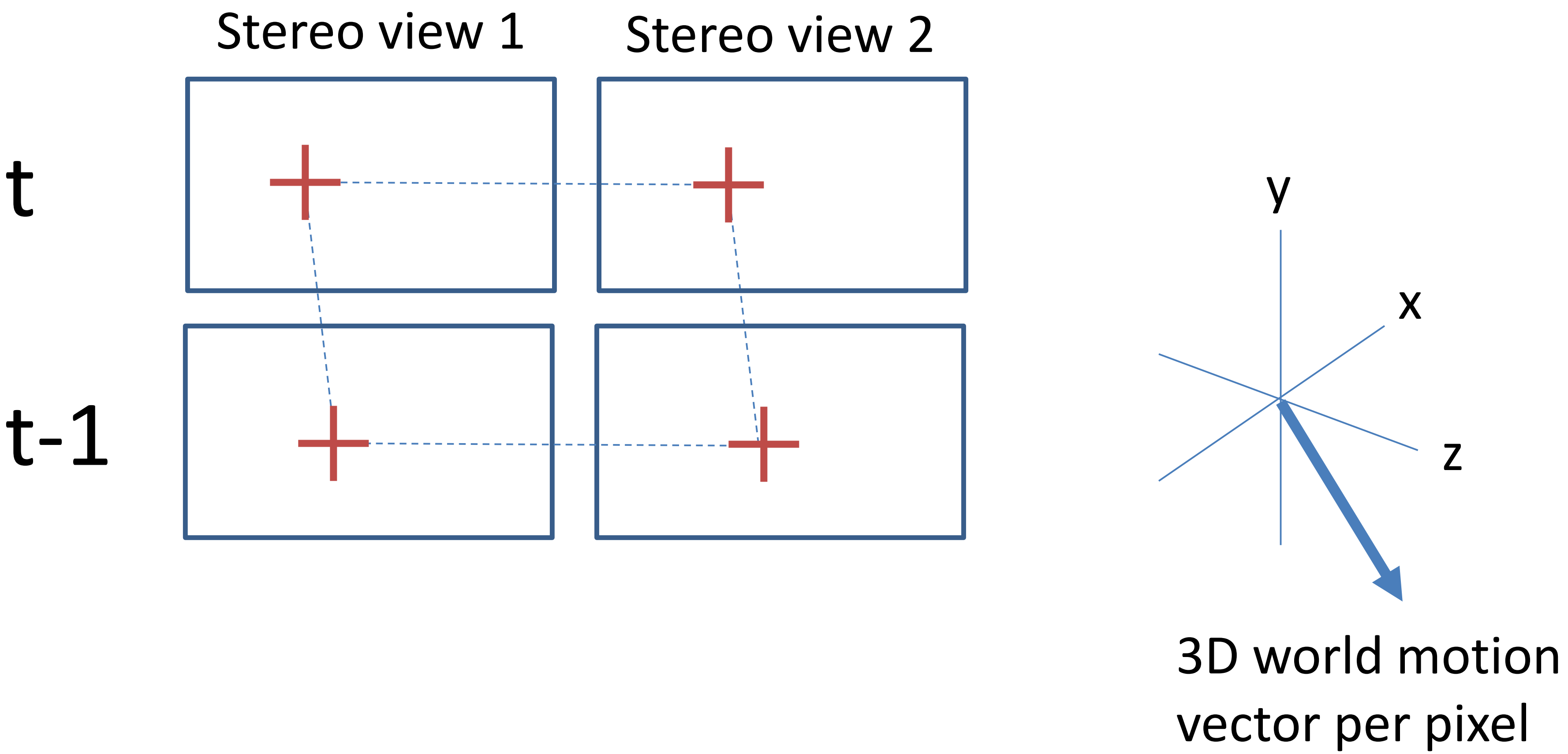
Optical flow

- Definition: the *apparent* motion of brightness patterns in the image
- Ideally, the same as the projected motion field
- Take care: apparent motion can be caused by lighting changes without any actual motion
  - Imagine a uniform rotating sphere under fixed lighting vs. a stationary sphere under moving illumination.

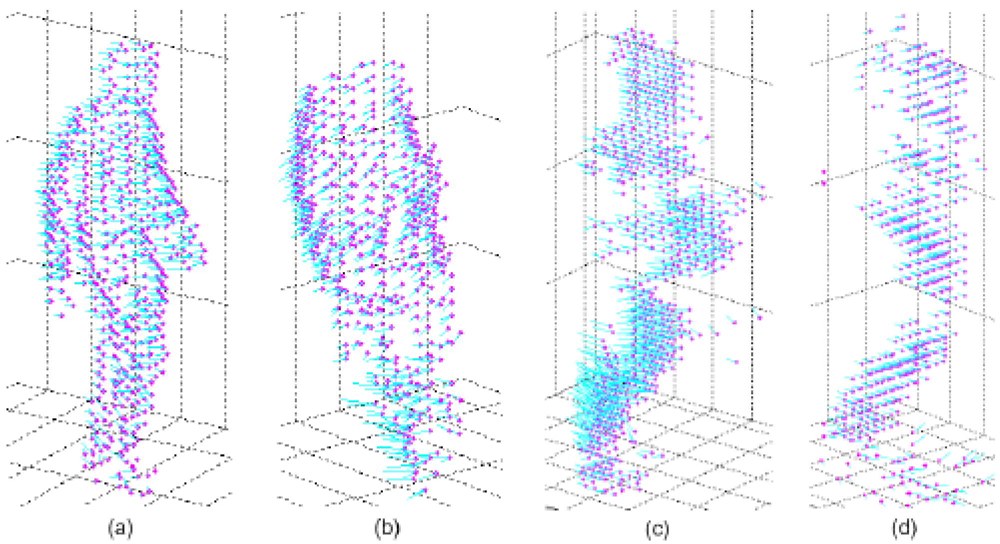
Can we do more? *Scene flow*

Combine spatial stereo & temporal constraints

Recover 3D vectors of world motion



Scene flow example for human motion



Estimating 3D Scene Flow from Multiple 2D Optical Flows, Ruttle et al., 2009

Scene Flow

https://www.youtube.com/watch?v=RL_TK_Be6_4

https://vision.in.tum.de/research/sceneflow

[Estimation of Dense Depth Maps and 3D Scene Flow from Stereo Sequences, M. Jaimez et al., TU Munchen]