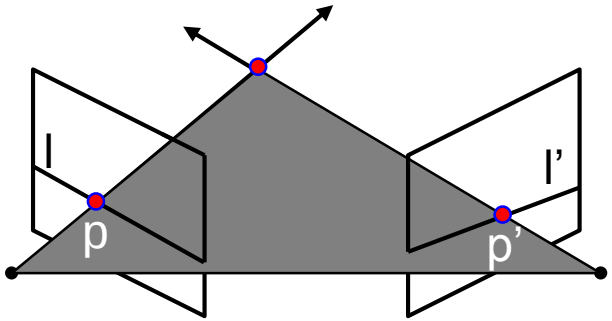
Fundamental matrix

Let *p* be a point in left image, *p'* in right image



Epipolar relation

- *p* maps to epipolar line *l'*
- *p'* maps to epipolar line *l*

Epipolar mapping described by a 3x3 matrix *F*

$$p'Fp = 0$$

Fundamental matrix

This matrix F is called

- the “Essential Matrix”
  - when image intrinsic parameters are known
- the “Fundamental Matrix”
  - more generally (uncalibrated case)

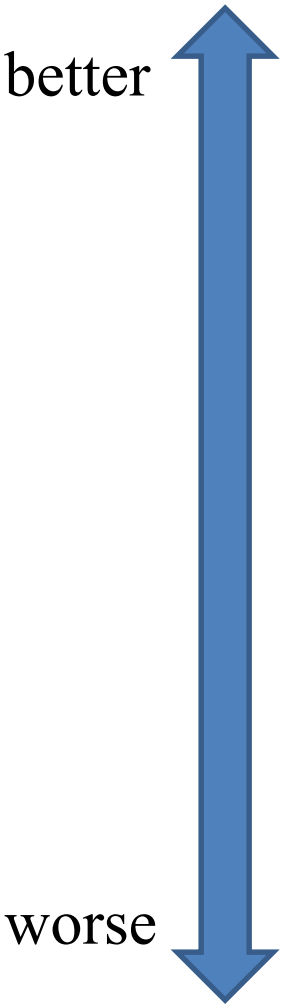
Can solve for F from point correspondences

- Each (p, p’) pair gives one linear equation in entries of F

$$p' F p = 0$$

- F has 9 entries, but really only 7 or 8 degrees of freedom.
- With 8 points it is simple to solve for F, but it is also possible with 7. See Marc Pollefey’s notes for a nice tutorial

The scale of algorithm name quality



RANSAC

SIFT

Deep Learning

Optical Flow

Hough Transform

Neural Networks

Essential and Fundamental Matrix

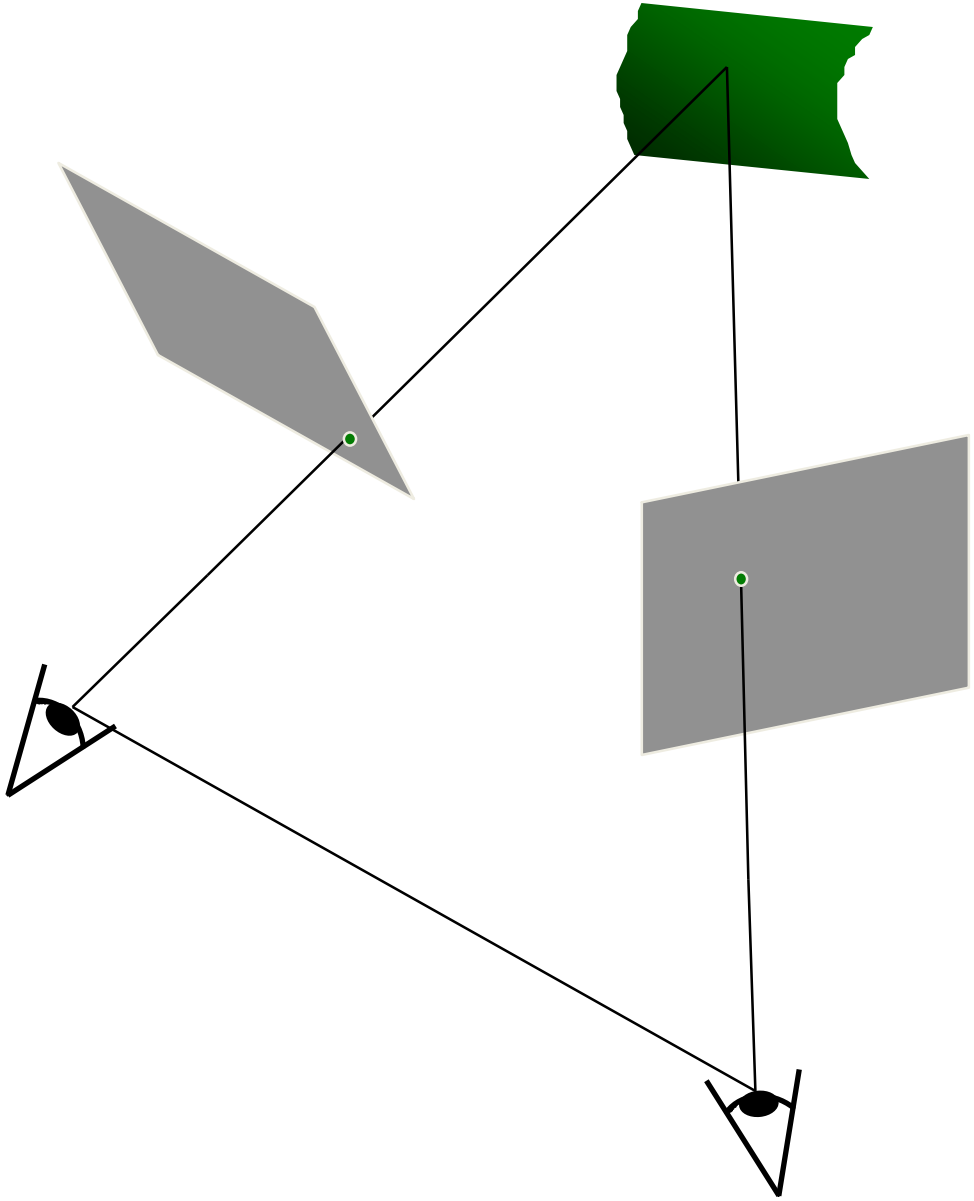
Dynamic Programming

Today’s lecture

- Stereo Matching (Spare correspondence to Dense Correspondence)
- Optical Flow (Dense motion estimation)

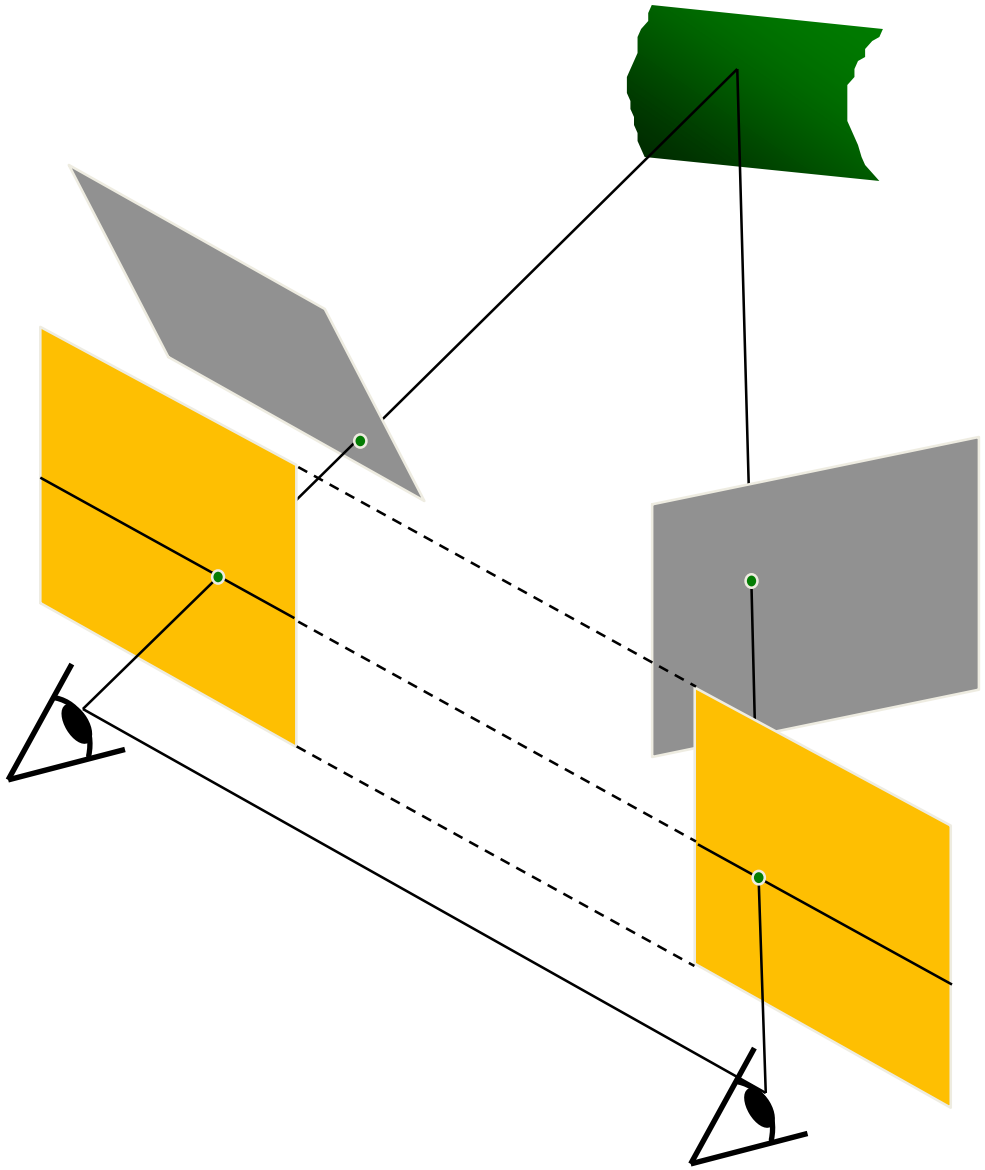
Stereo Matching

Stereo image rectification



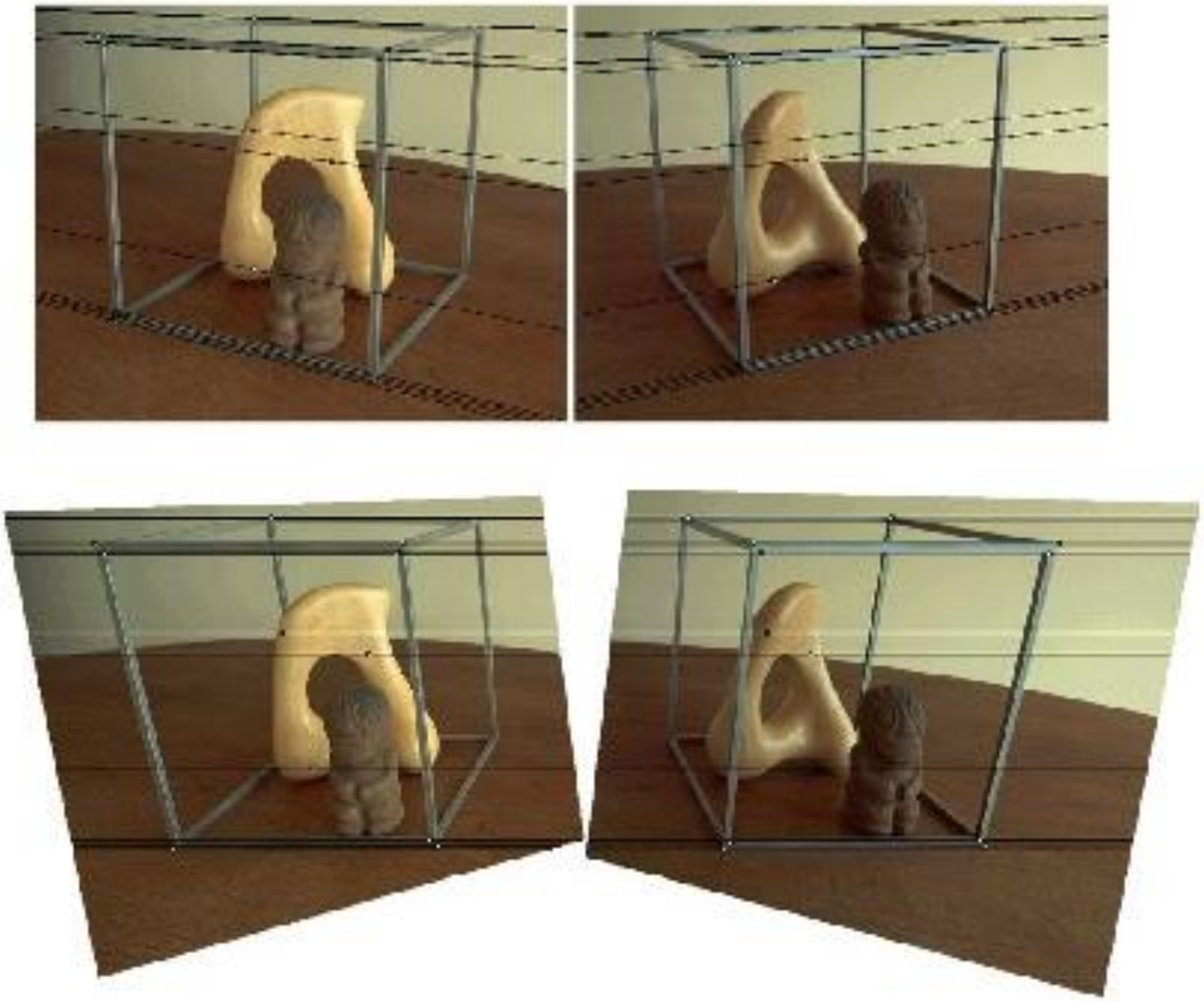
Stereo image rectification

- Reproject image planes onto a common plane parallel to the line between camera centers
- Pixel motion is horizontal after this transformation
- Two homographies (3x3 transform), one for each input image reprojection



➢ C. Loop and Z. Zhang. Computing Rectifying Homographies for Stereo Vision. IEEE Conf. Computer Vision and Pattern Recognition, 1999.

Rectification example

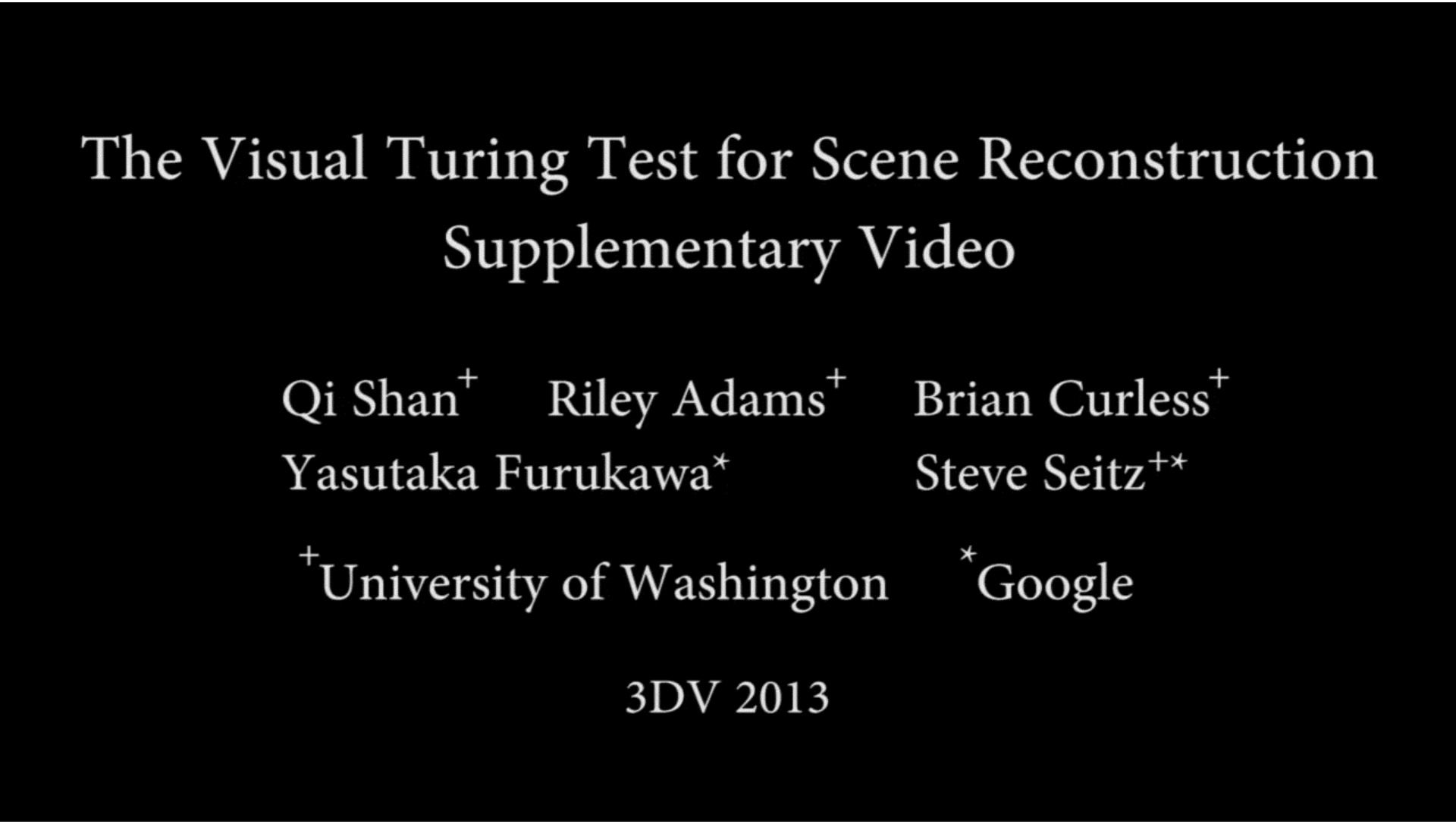


The correspondence problem

- Epipolar geometry constrains our search, but we still have a difficult correspondence problem.

Fundamental Matrix + Sparse correspondence

Fundamental Matrix + Dense correspondence



SIFT + Fundamental Matrix + RANSAC

Despite their scale invariance and robustness to appearance changes, SIFT features are *local* and do not contain any global information about the image or about the location of other features in the image. Thus feature matching based on SIFT features is still prone to errors. However, since we assume that we are dealing with rigid scenes, there are strong geometric constraints on the locations of the matching features and these constraints can be used to clean up the matches. In particular, when a rigid scene is imaged by two pinhole cameras, there exists a 3×3 matrix F, the *Fundamental matrix*, such that corresponding points x_{ij} and x_{ik} (represented in homogeneous coordinates) in two images j and k satisfy[10]:

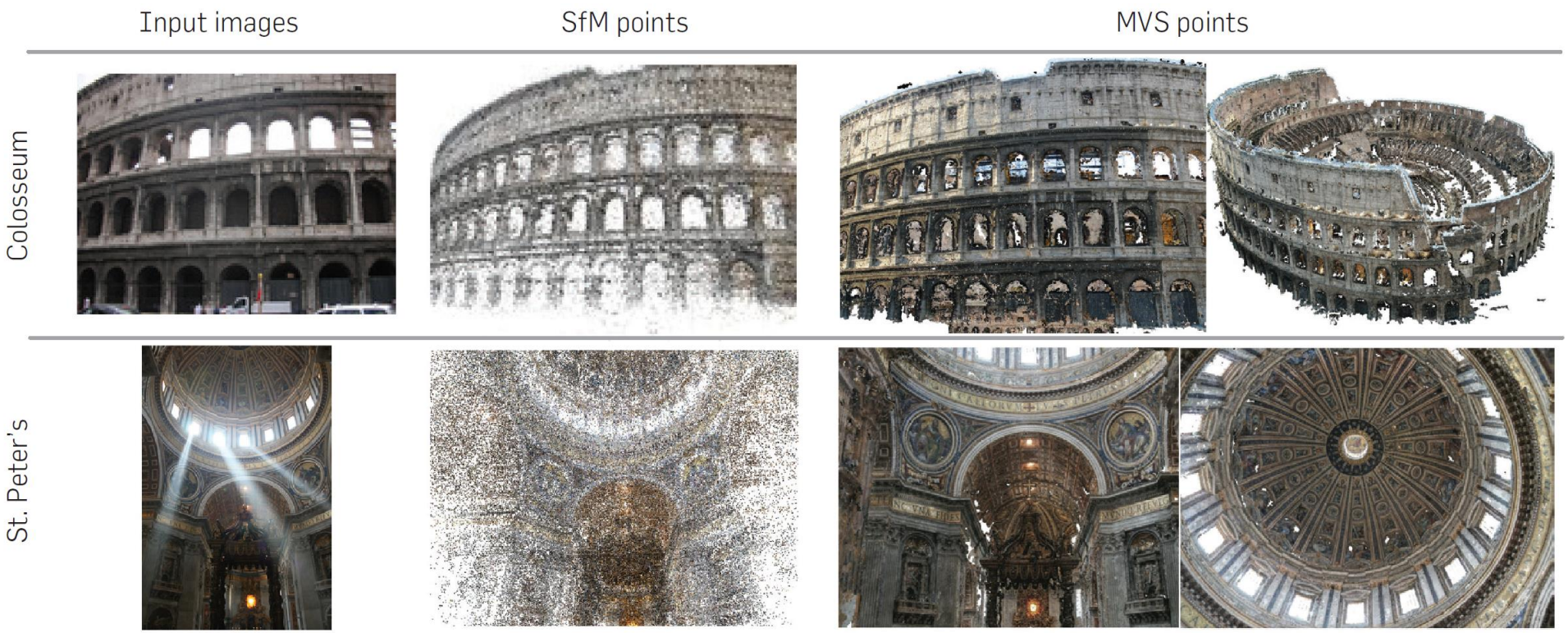
$$x_{ij}^{\top} F x_{ij} = 0. \tag{3}$$

A common way to impose this constraint is to use a greedy randomized algorithm to generate suitably chosen random estimates of F and choose the one that has the largest support among the matches, i.e., the one for which the most matches satisfy (3). This algorithm is called Random Sample Consensus (RANSAC)[6] and is used in many computer vision problems.

Building Rome in a Day

By Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, Richard Szeliski
Communications of the ACM, Vol. 54 No. 10, Pages 105-112

Sparse to Dense Correspodence

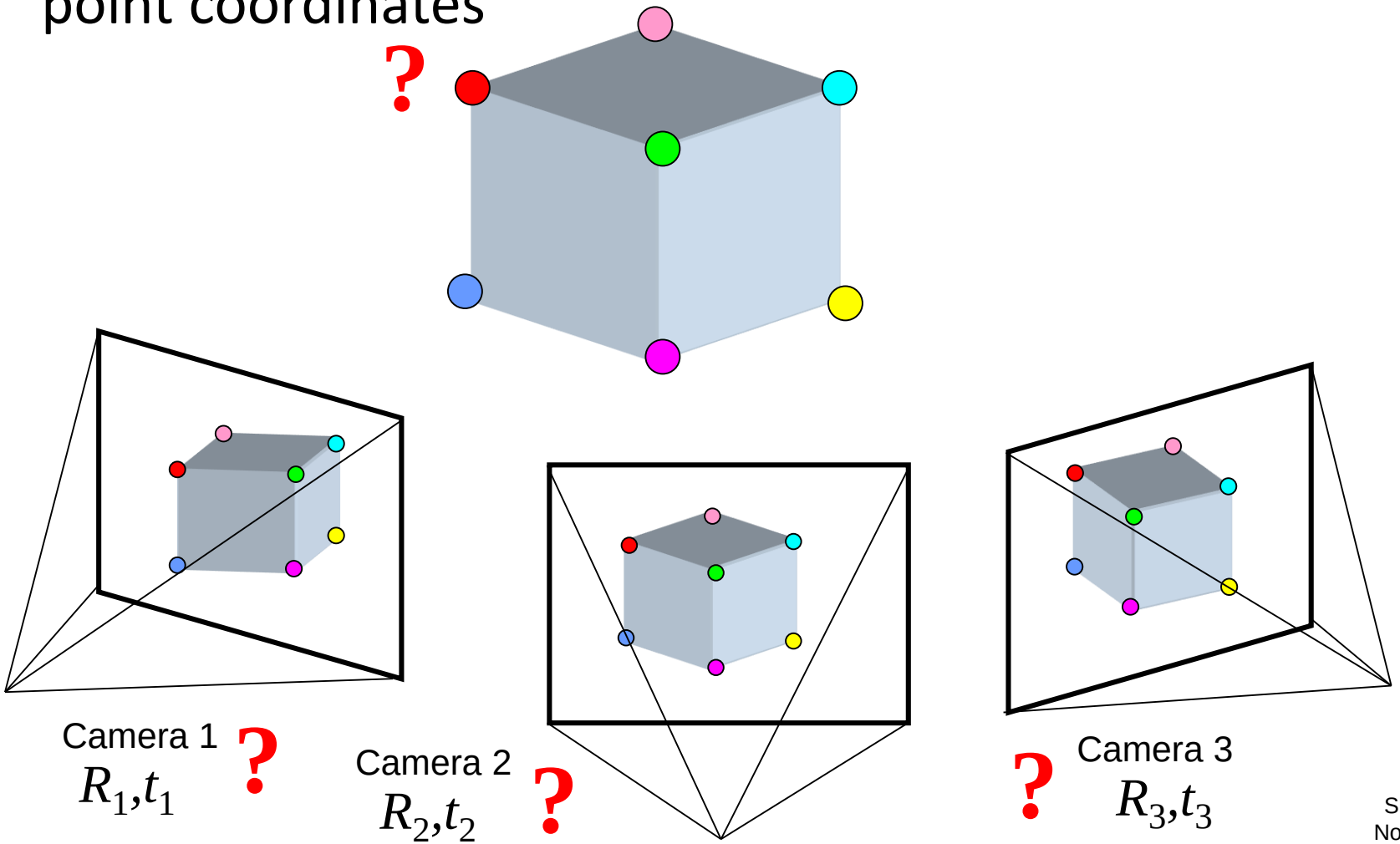


Building Rome in a Day

By Sameer Agarwal, Yasutaka Furukawa, Noah Snavely, Ian Simon, Brian Curless, Steven M. Seitz, Richard Szeliski
Communications of the ACM, Vol. 54 No. 10, Pages 105-112

Structure from motion (or SLAM)

- Given a set of corresponding points in two or more images, compute the camera parameters and the 3D point coordinates



Slide credit: Noah Snavely

Structure from motion ambiguity

- If we scale the entire scene by some factor *k* and, at the same time, scale the camera matrices by the factor of 1/*k*, the projections of the scene points in the image remain exactly the same:

$$\mathbf{x} = \mathbf{P}\mathbf{X} = \left(\frac{1}{k}\mathbf{P}\right)(k\mathbf{X})$$

It is impossible to recover the absolute scale of the scene!

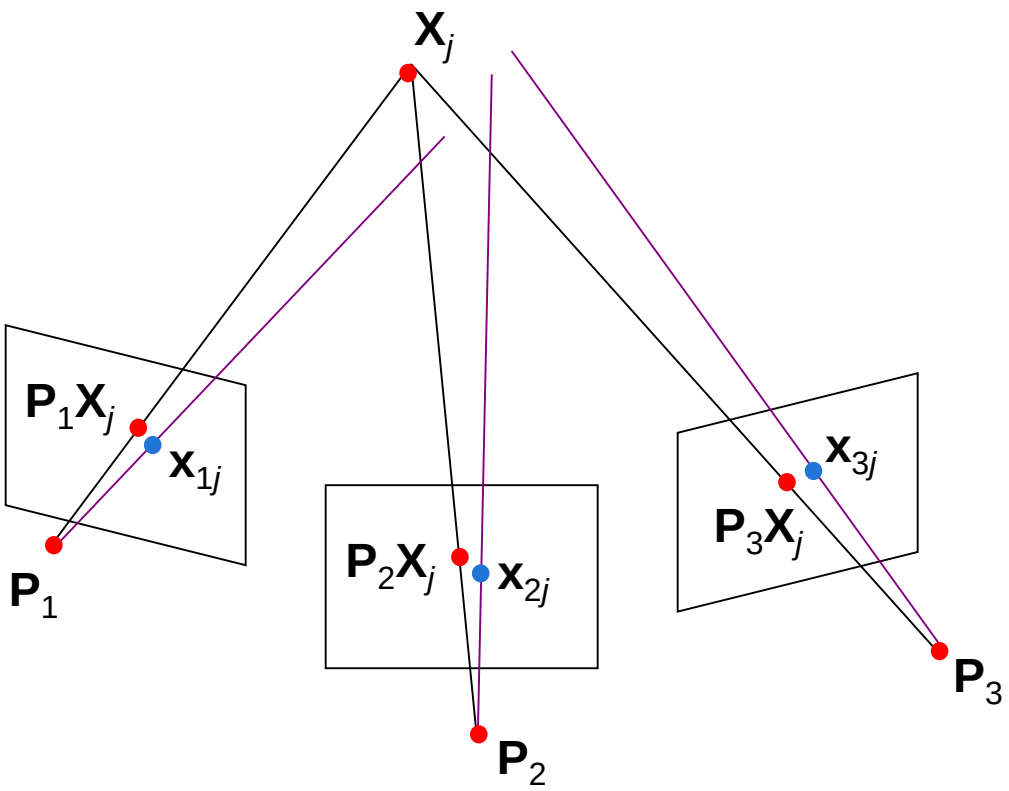
How do we know the scale of image content?


Photography
Your Eyes Deserve It
TOTAL
RECLAIMS
TOTAL
RECLAIMS
TOTAL
RECLAIMS

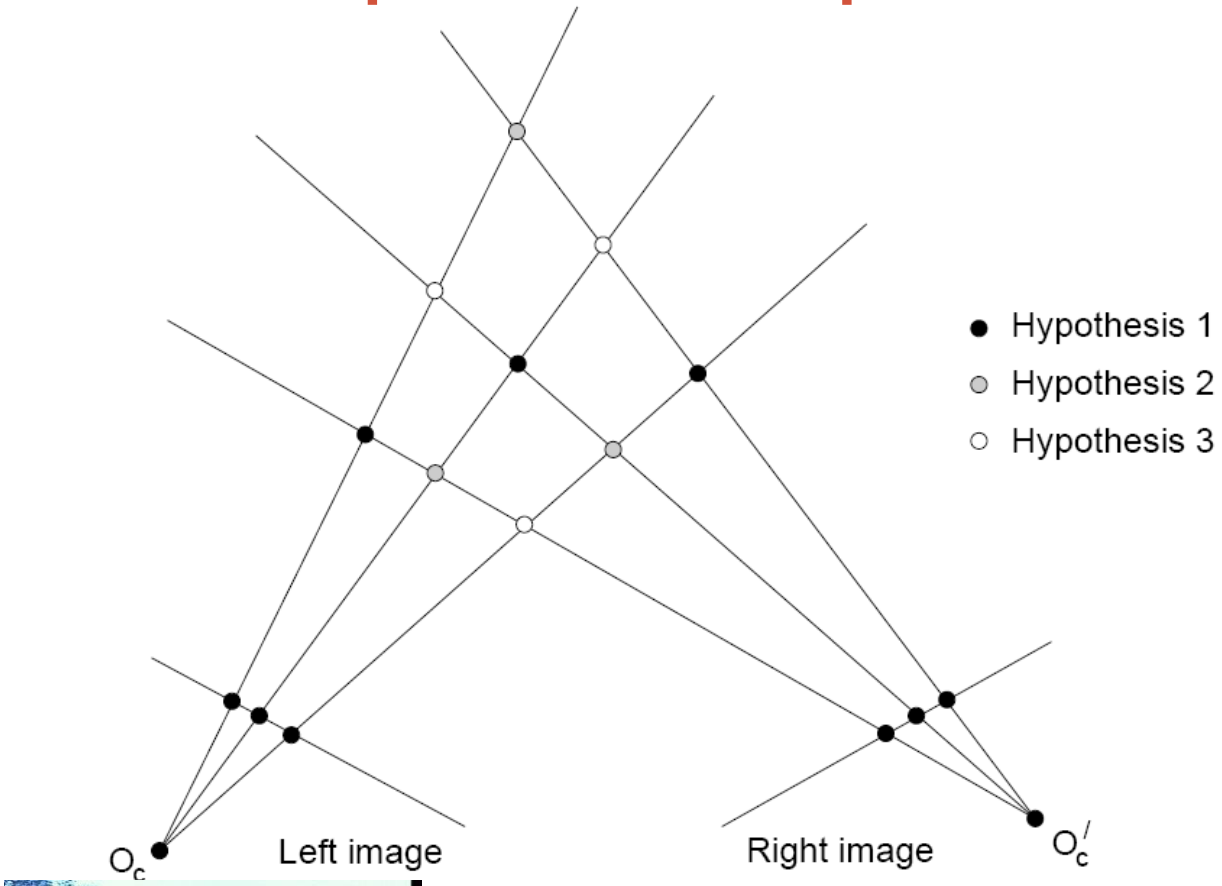
Bundle adjustment

- Non-linear method for refining structure and motion
- Minimizing reprojection error

$$E(\mathbf{P}, \mathbf{X}) = \sum_{i=1}^{m} \sum_{j=1}^{n} D\left(\mathbf{x}_{ij}, \mathbf{P}_i \mathbf{X}_j\right)^2$$



Correspondence problem



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

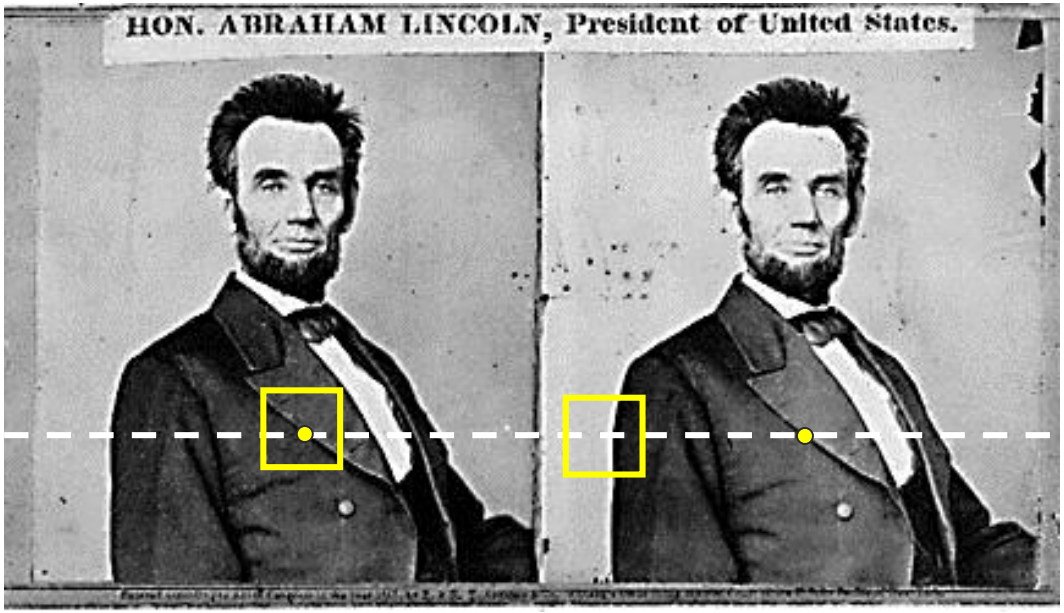
Figure from Gee & Cipolla 1999

Correspondence problem

- Beyond the hard constraint of epipolar geometry, there are “soft” constraints to help identify corresponding points
  - Similarity
  - Uniqueness
  - Ordering
  - Disparity gradient

- To find matches in the image pair, we will assume
  - Most scene points visible from both views
  - Image regions for the matches are similar in appearance

Dense correspondence search



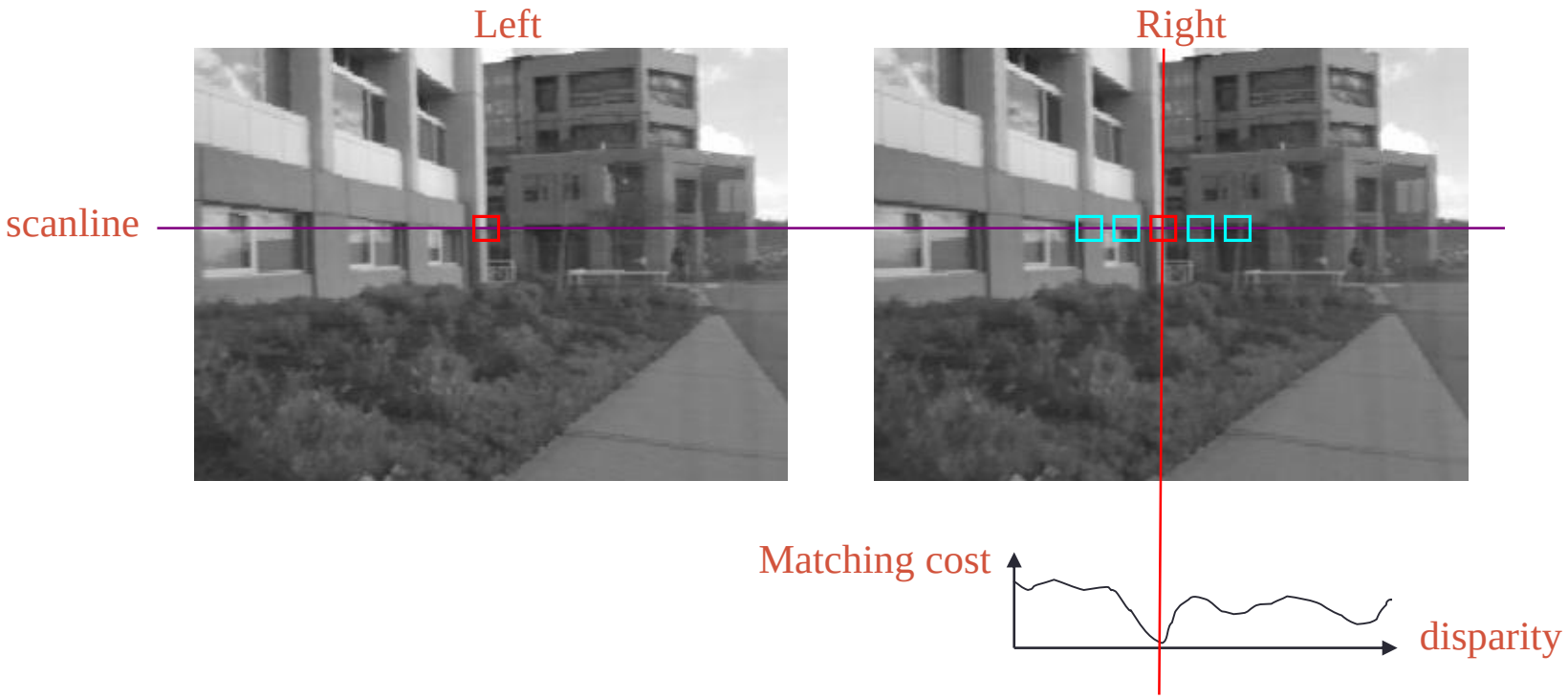
For each epipolar line

For each pixel / window in the left image

- compare with every pixel / window on same epipolar line in right image
- pick position with minimum match cost (e.g., SSD, normalized correlation)

Adapted from Li Zhang

Correspondence search with similarity constraint



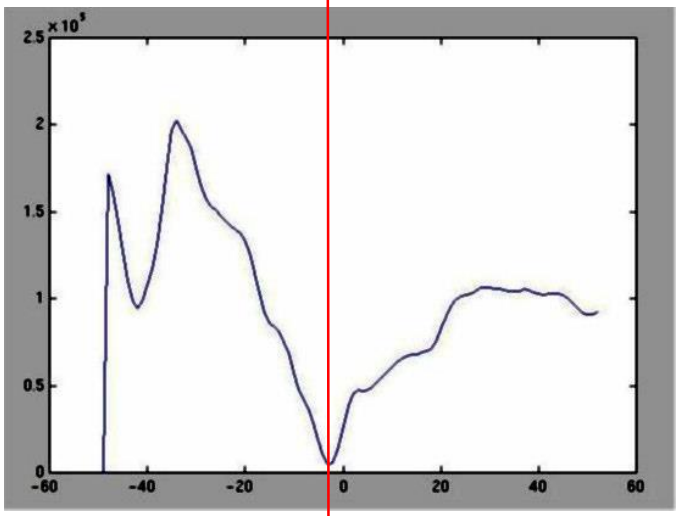
- Slide a window along the right scanline and compare contents of that window with the reference window in the left image
- Matching cost: SSD or normalized correlation

Correspondence search with similarity constraint

Left

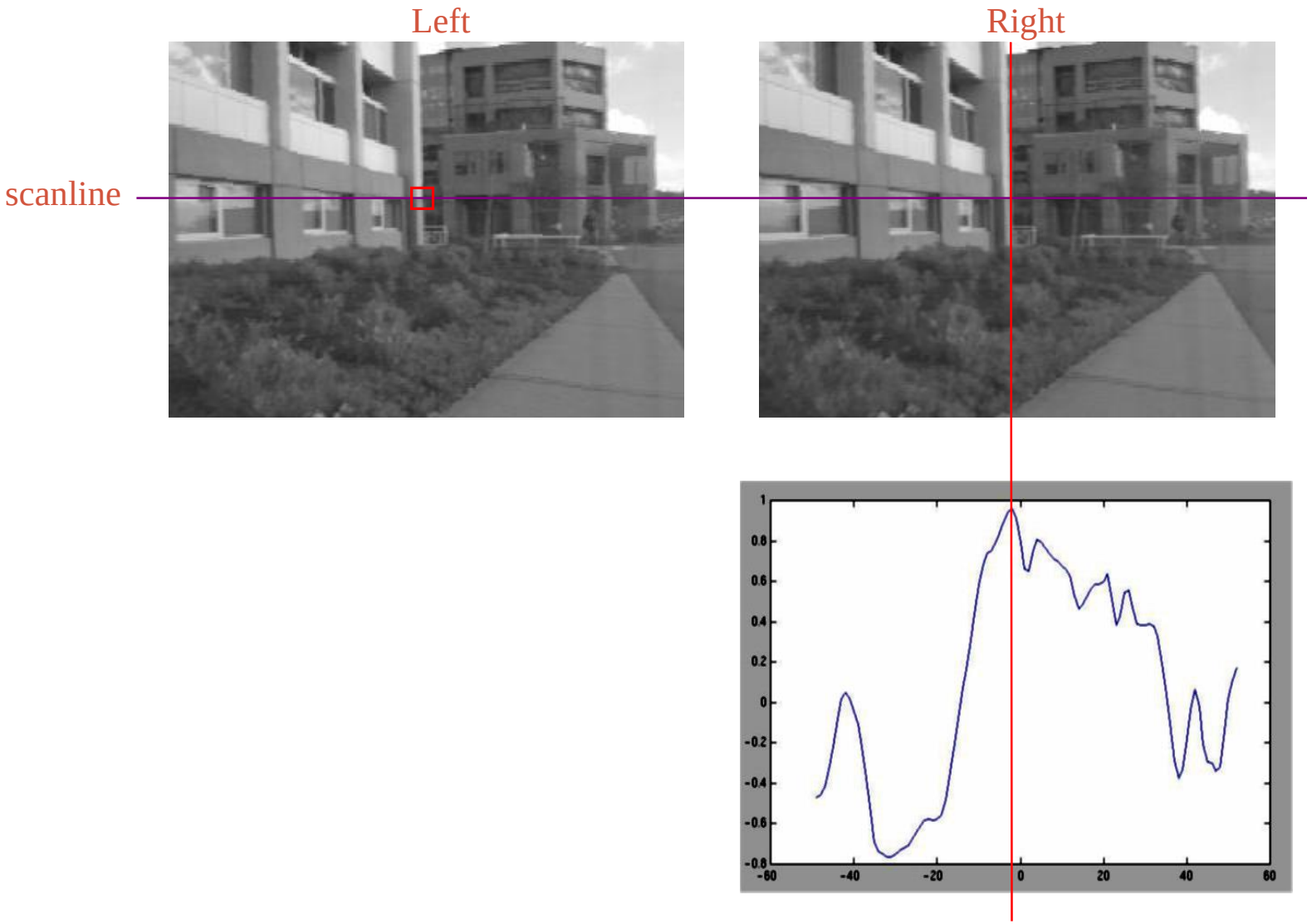
Right

scanline



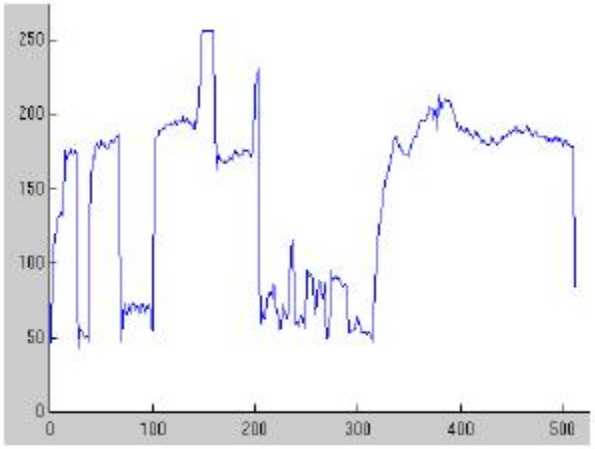
SSD

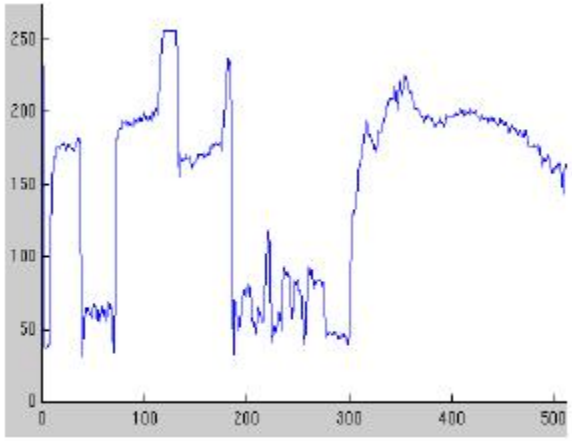
Correspondence search with similarity constraint



Correspondence problem

Intensity profiles

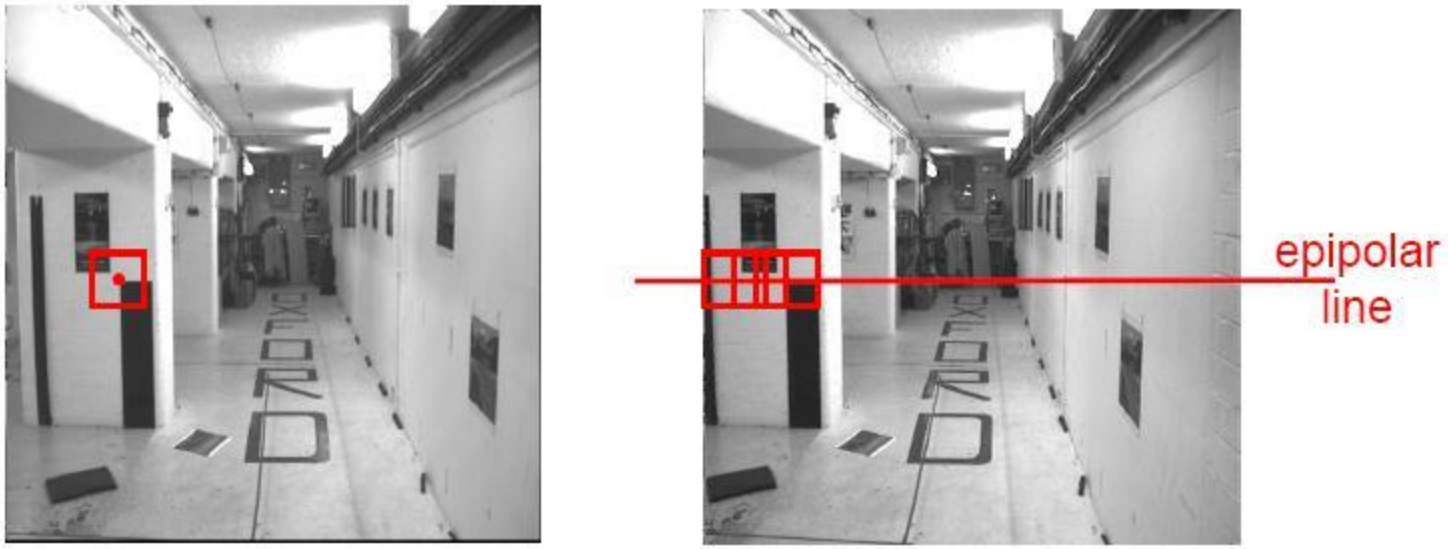




- Clear correspondence between intensities, but also noise and ambiguity

Source: Andrew Zisserman

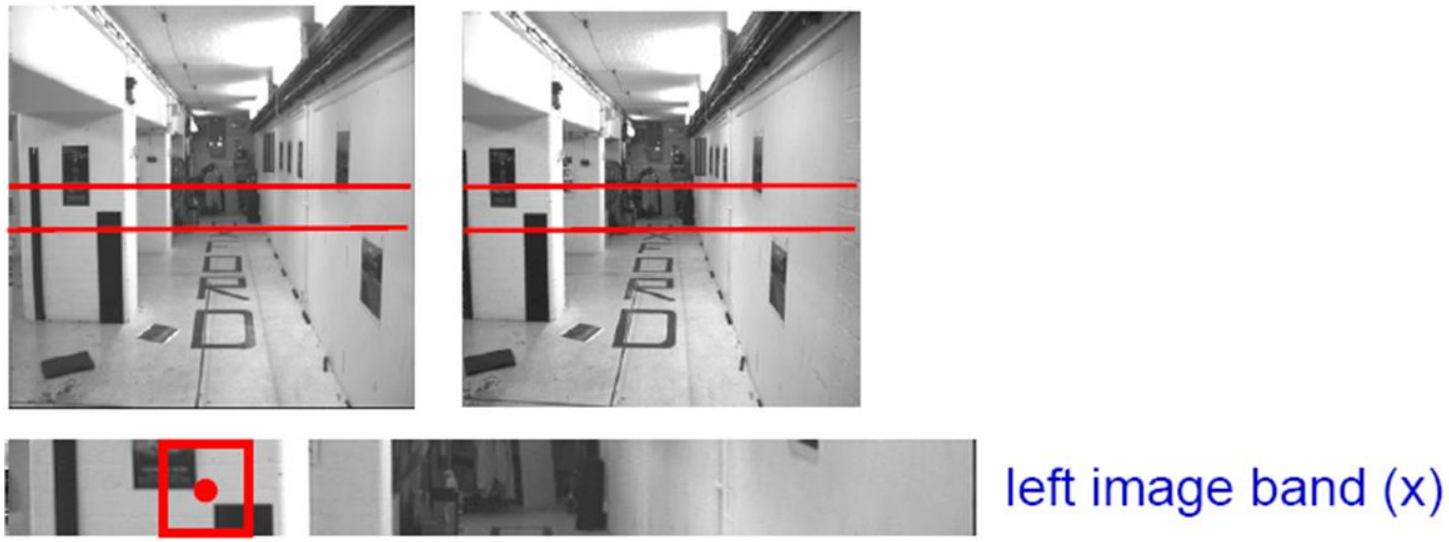
Correspondence problem



Neighborhoods of corresponding points are similar in intensity patterns.

Source: Andrew Zisserman

Correlation-based window matching

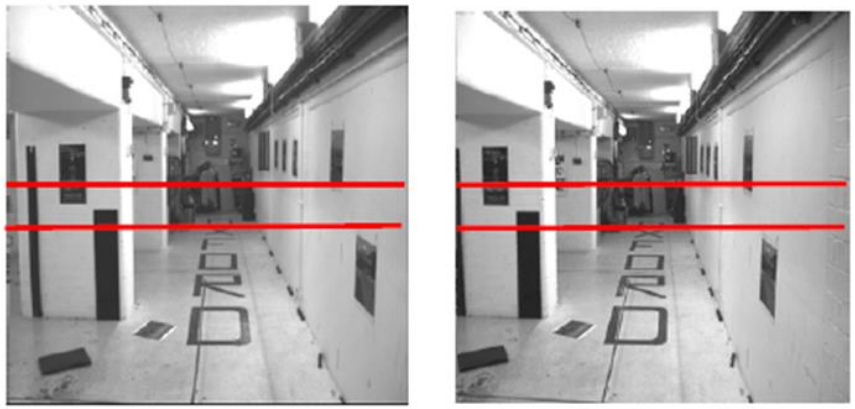


Correlation-based window matching

left image band (x)

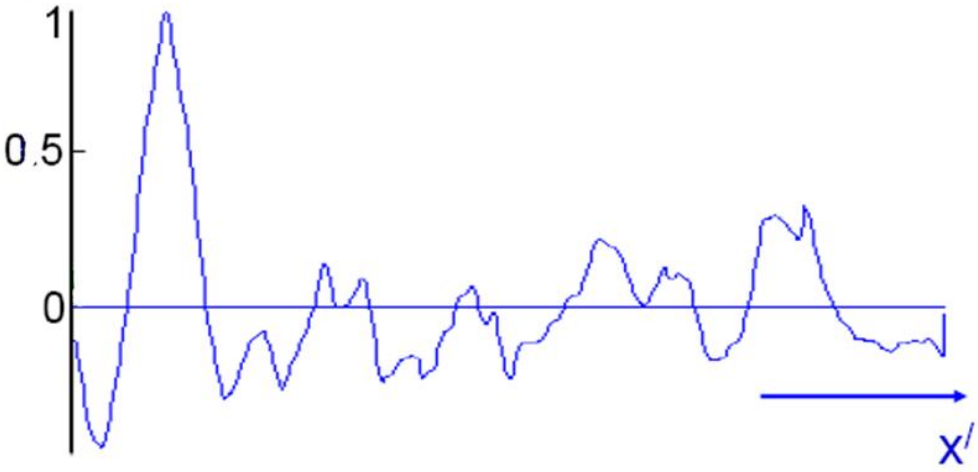
right image band (x')

Correlation-based window matching



left image band (x)

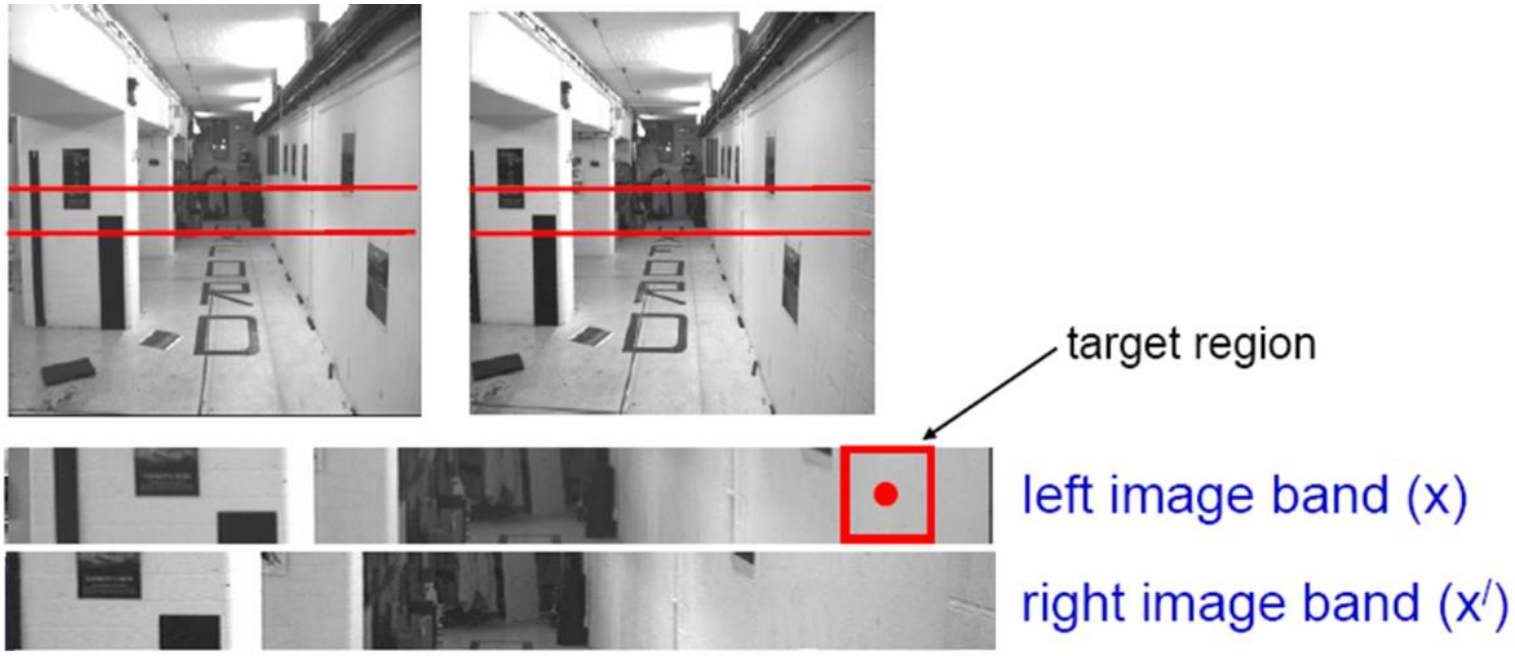
right image band (x')



cross correlation

disparity = $x' - x$

Correlation-based window matching



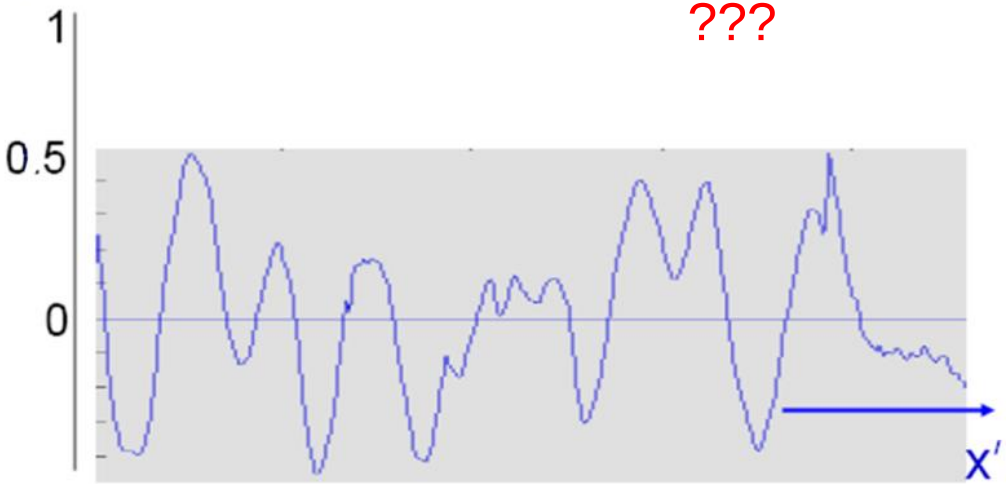
Correlation-based window matching

target region

left image band (x)

right image band (x')

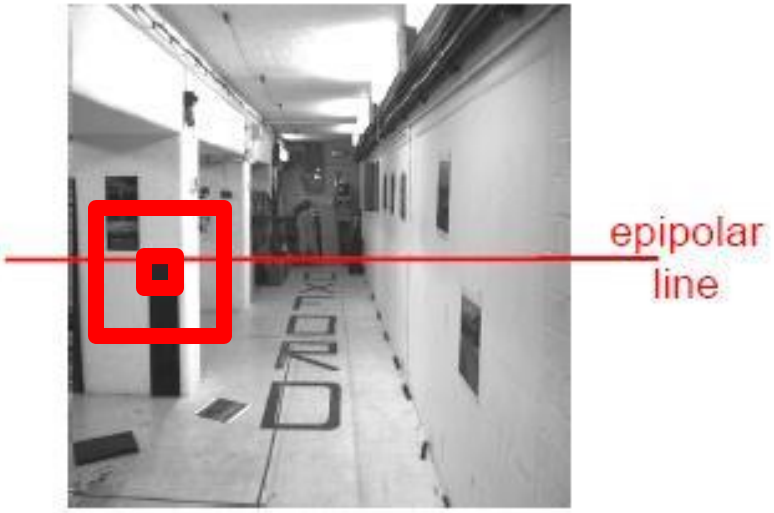
???



cross correlation

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

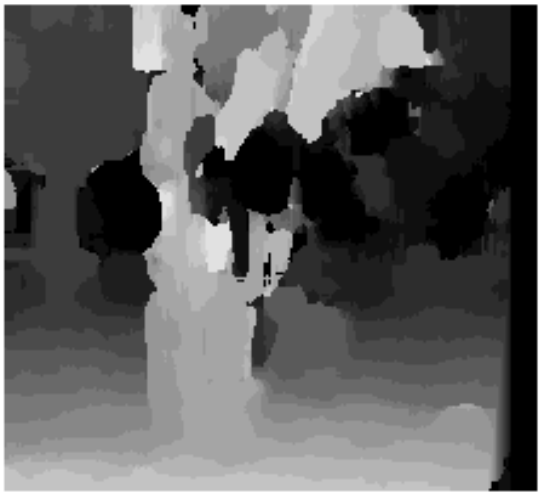
Effect of window size



Source: Andrew Zisserman

Effect of window size

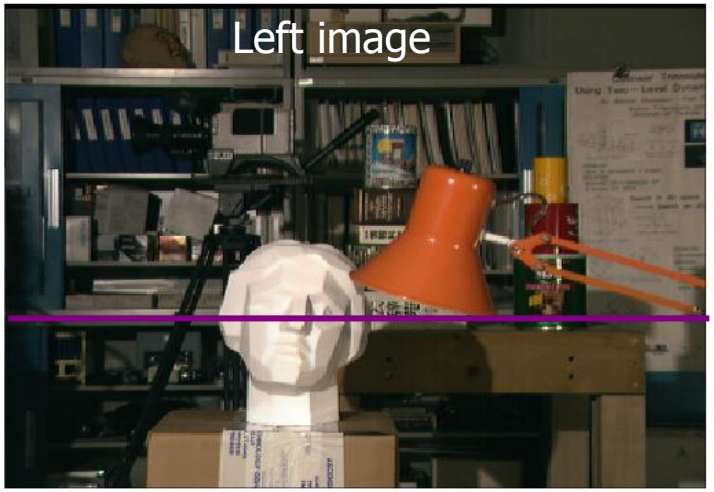
W = 3

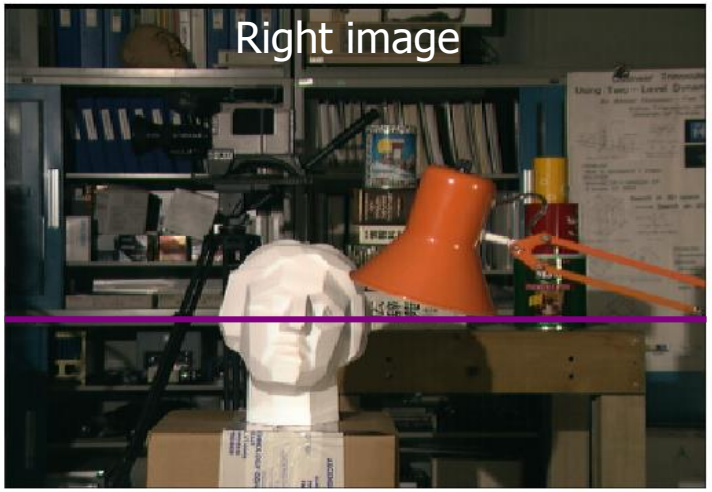


W = 20

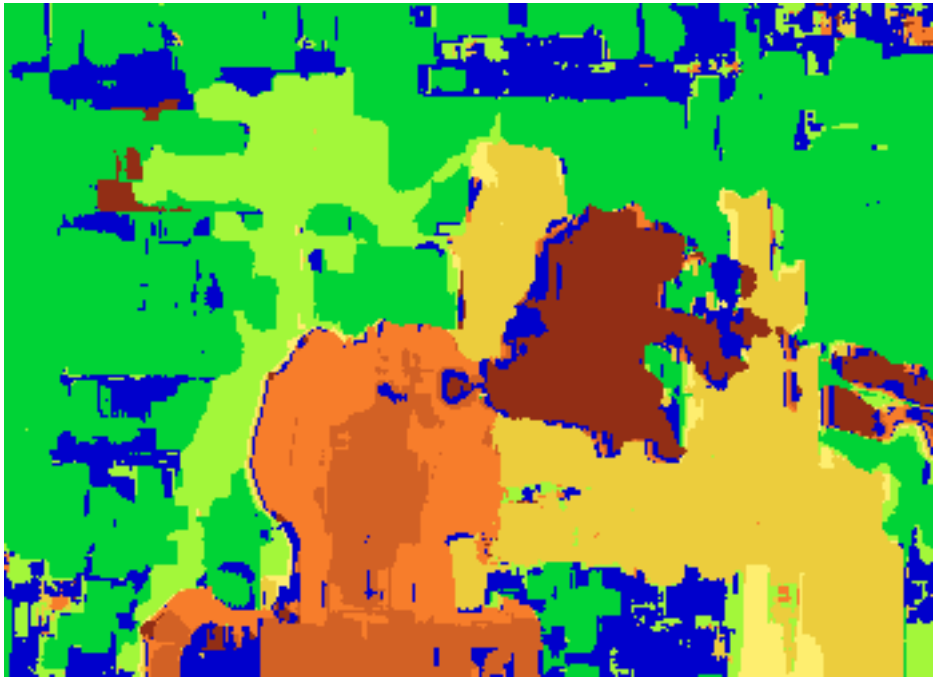
Want window large enough to have sufficient intensity variation, yet small enough to contain only pixels with about the same disparity.

Figures from Li Zhang


Left image


Right image

Results with window search



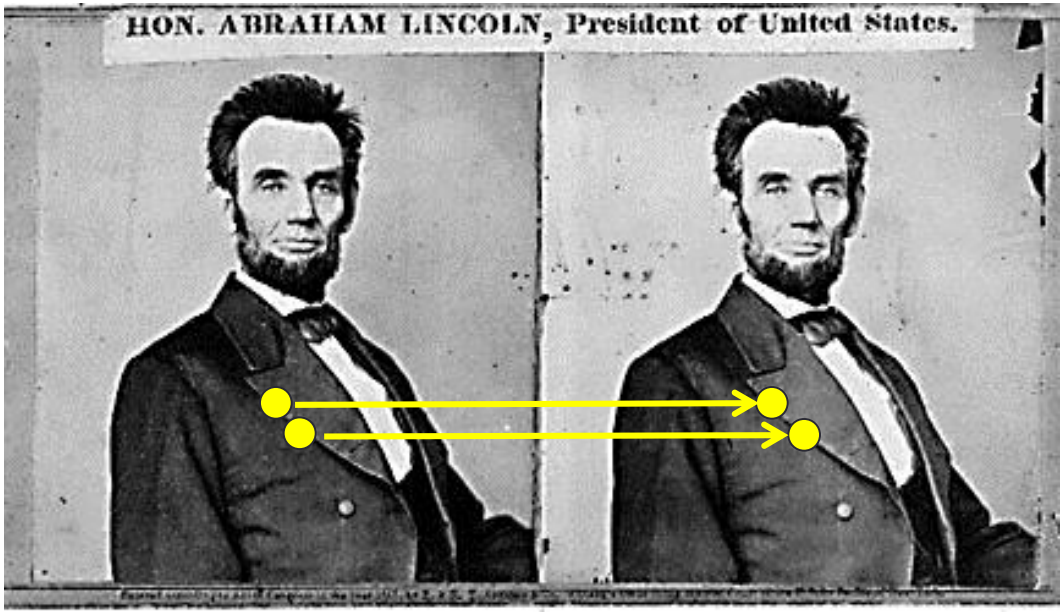
Window-based matching
(best window size)

Ground truth

Better solutions

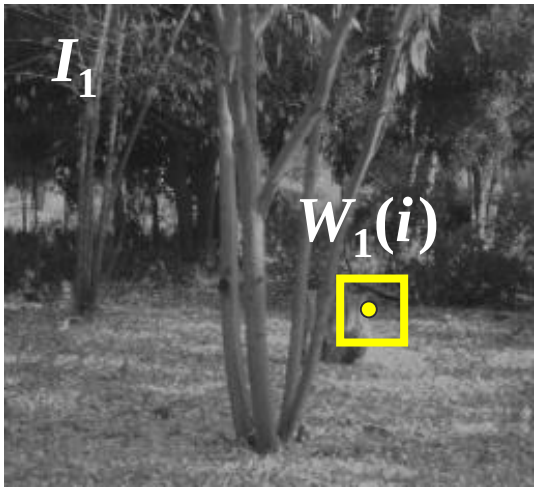
- Beyond individual correspondences to estimate disparities:
- Optimize correspondence assignments jointly
  - Scanline at a time (DP)
  - Full 2D grid (graph cuts)

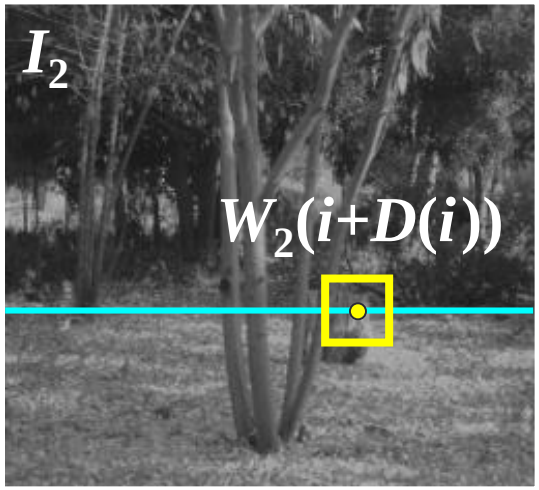
Stereo as energy minimization

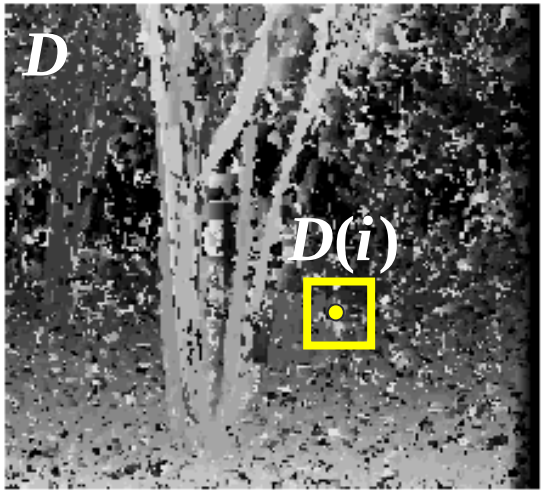


- What defines a good stereo correspondence?
  1. Match quality
     - Want each pixel to find a good match in the other image
  2. Smoothness
     - If two pixels are adjacent, they should (usually) move about the same amount

Stereo matching as energy minimization







$$E = \alpha E_{\text{data}}(I_1, I_2, D) + \beta E_{\text{smooth}}(D)$$

$$E_{\text{data}} = \sum_i \left(W_1(i) - W_2(i + D(i))\right)^2$$

$$E_{\text{smooth}} = \sum_{\text{neighbors}\, i,j} \rho\left(D(i) - D(j)\right)$$

- Energy functions of this form can be minimized using *graph cuts*

Y. Boykov, O. Veksler, and R. Zabih, Fast Approximate Energy Minimization via Graph Cuts, PAMI 2001

Source: Steve Seitz

Better results…

Graph cut method

Ground truth

Boykov et al., Fast Approximate Energy Minimization via Graph Cuts, International Conference on Computer Vision, September 1999.

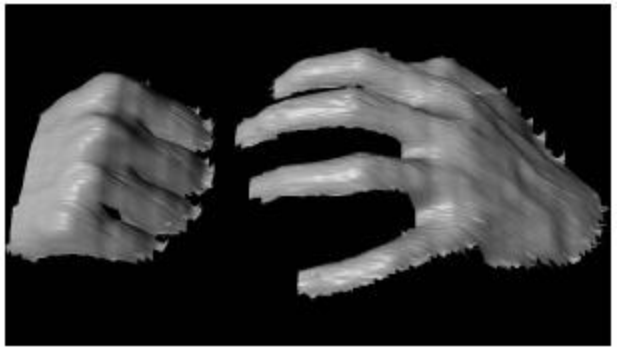
For the latest and greatest: http://www.middlebury.edu/stereo/

Challenges

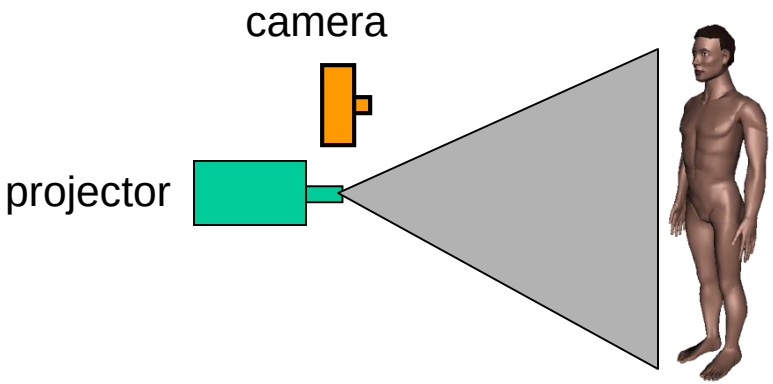
- Low-contrast ; textureless image regions
- Occlusions
- Violations of brightness constancy (e.g., specular reflections)
- Really large baselines (foreshortening and appearance change)
- Camera calibration errors

Active stereo with structured light





- Project “structured” light patterns onto the object
  - Simplifies the correspondence problem
  - Allows us to use only one camera

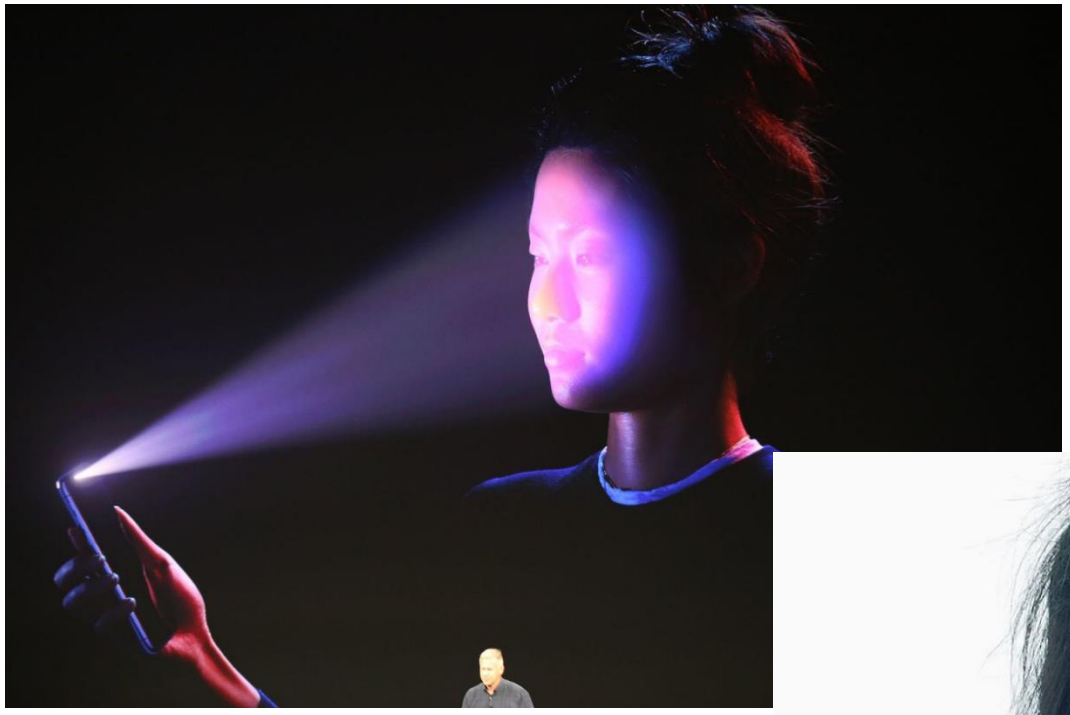


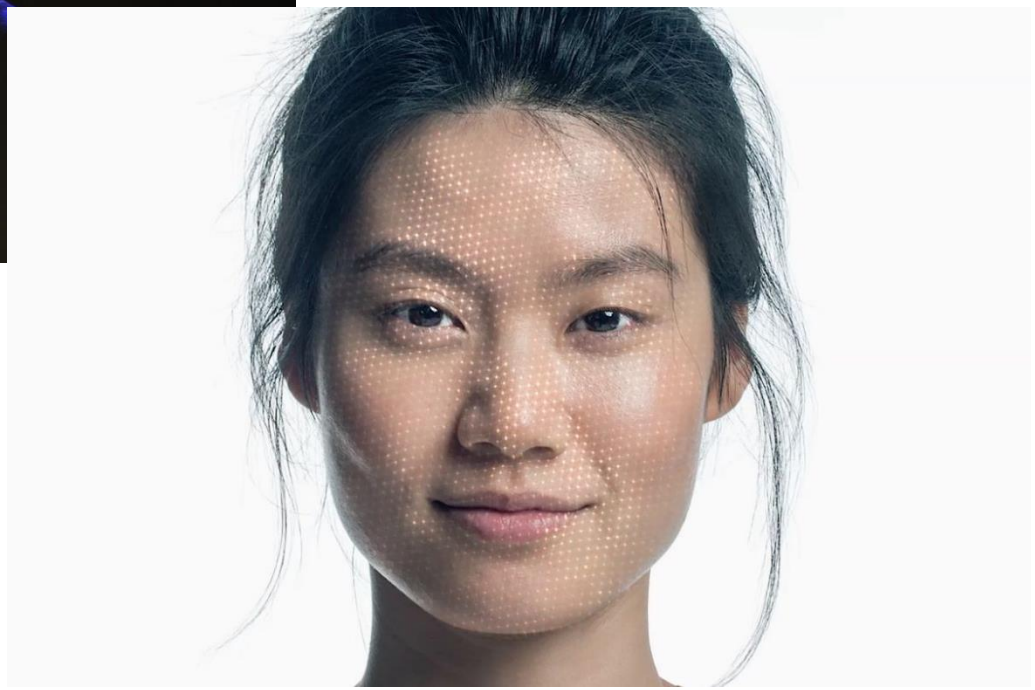
L. Zhang, B. Curless, and S. M. Seitz. Rapid Shape Acquisition Using Color Structured Light and Multi-pass Dynamic Programming. *3DPVT* 2002

Kinect: Structured infrared light

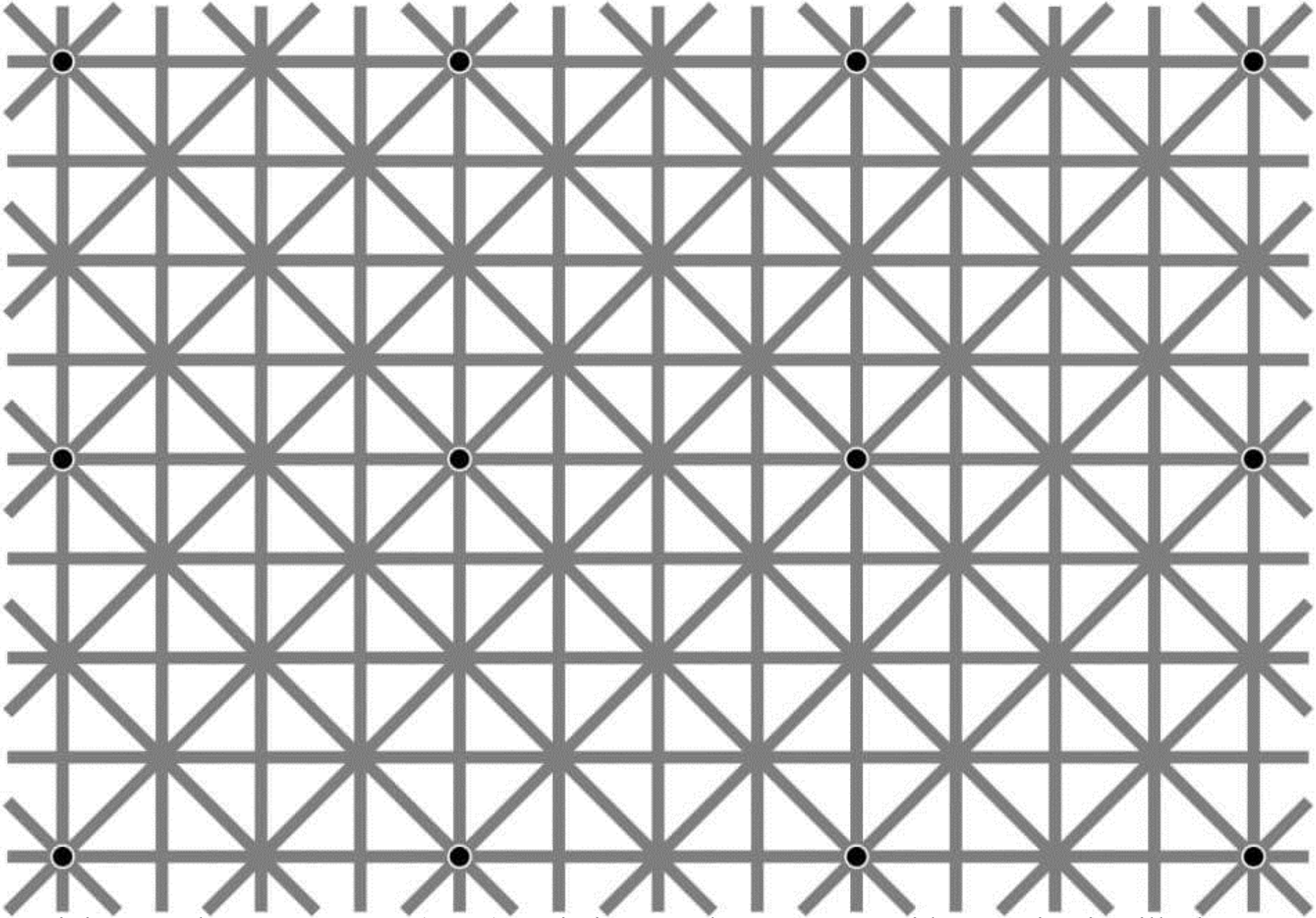
http://bbzippo.wordpress.com/2010/11/28/kinect-in-infrared/

iPhone X





3 minute break



Ninio, J. and Stevens, K. A. (2000) Variations on the Hermann grid: an extinction illusion. Perception, 29, 1209-1217.

Perception, 2000, volume 29, pages 1209 – 1217

DOI:10.1068/p2985

Variations on the Hermann grid: an extinction illusion

Jacques Ninio
Laboratoire de Physique Statistique[1], École Normale Supérieure, 24 rue Lhomond, 75231 Paris cedex 05, France; e-mail: jacques.ninio@lps.ens.fr

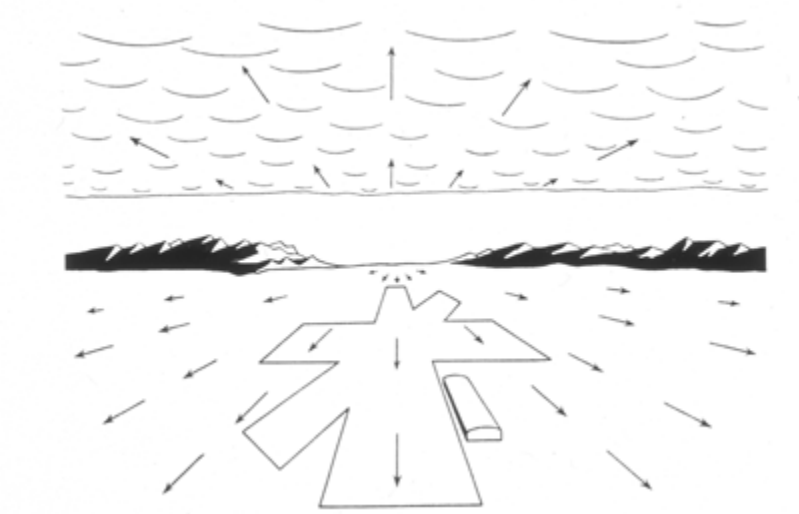
Kent A Stevens
Department of Computer Science, Deschutes Hall, University of Oregon, Eugene, OR 97403, USA; e-mail: kent@cs.uoregon.edu
Received 21 September 1999, in revised form 21 June 2000

Abstract. When the white disks in a scintillating grid are reduced in size, and outlined in black, they tend to disappear. One sees only a few of them at a time, in clusters which move erratically on the page. Where they are not seen, the grey alleys seem to be continuous, generating grey crossings that are not actually present. Some black sparkling can be seen at those crossings where no disk is seen. The illusion also works in reverse contrast.

The Hermann grid (Brewster 1844; Hermann 1870) is a robust illusion. It is classically presented as a two-dimensional array of black squares, separated by rectilinear alleys. It is thought to be caused by processes of local brightness computation in arrays of

Computer Vision

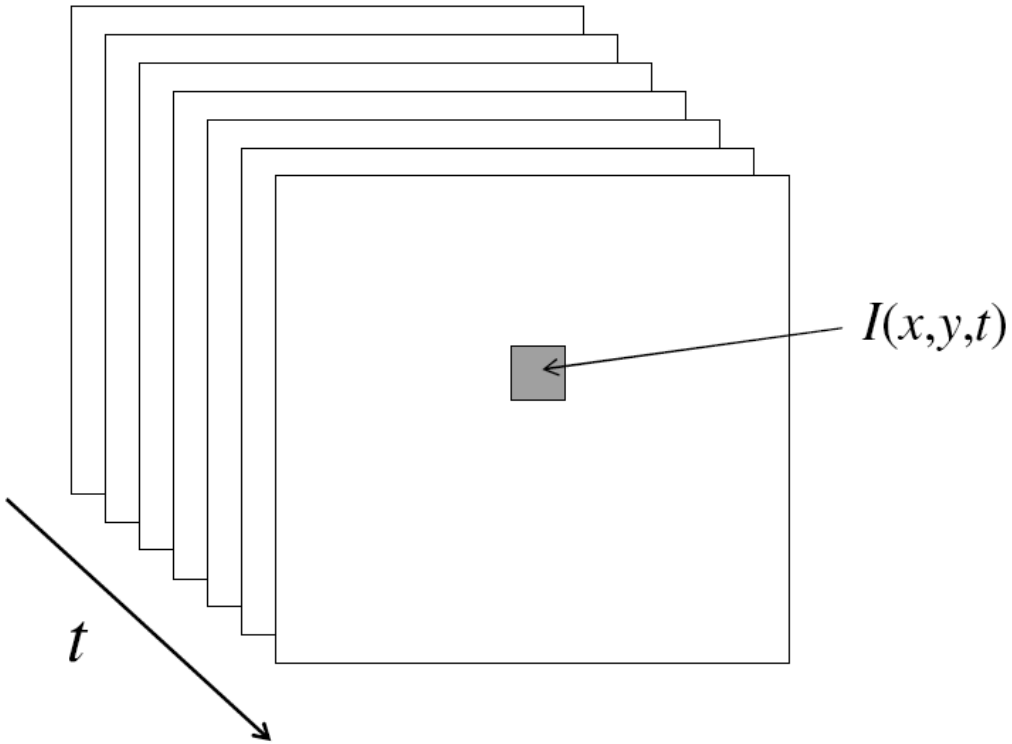
Motion and Optical Flow



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)

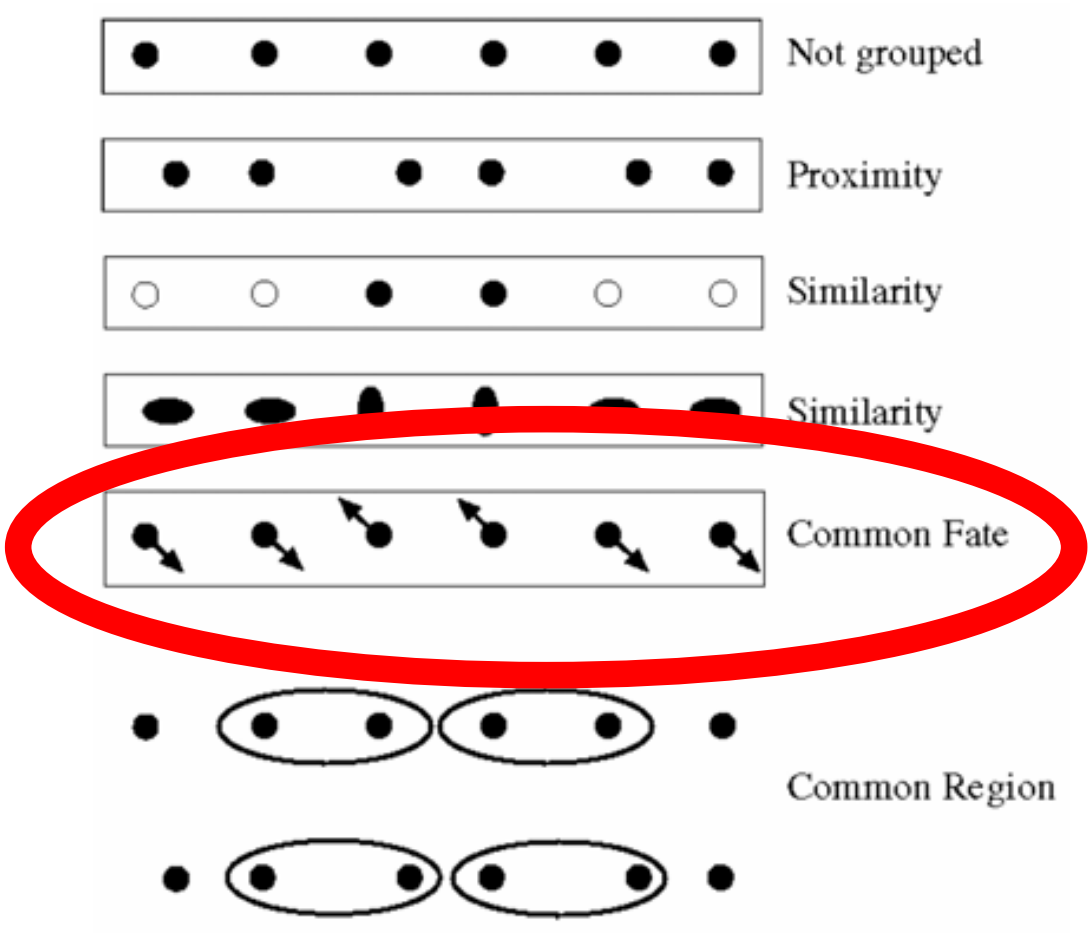


Motion and perceptual organization

Gestalt psychology (Max Wertheimer, 1880-1943)

Motion and perceptual organization

- Sometimes, motion is the only cue



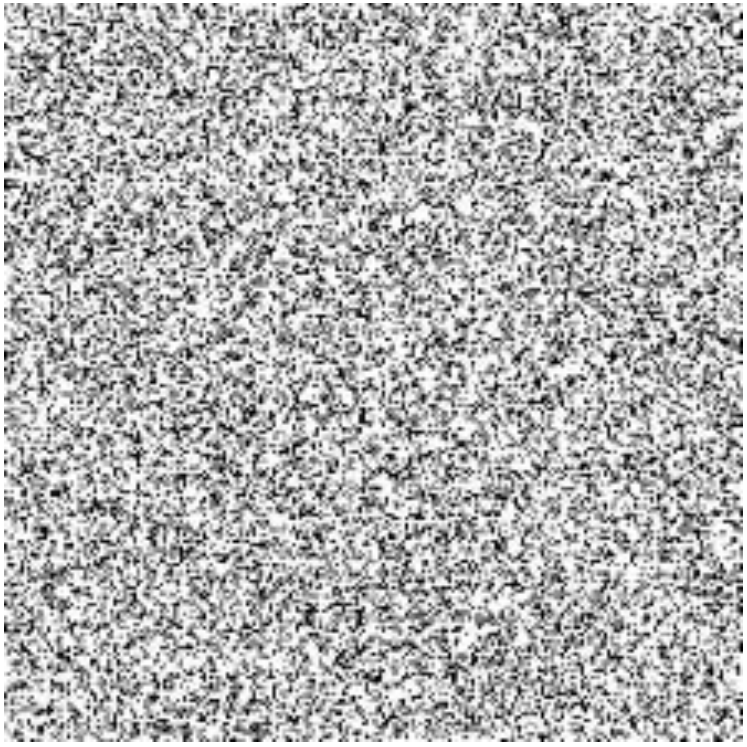
Gestalt psychology (Max Wertheimer, 1880-1943)

Motion and perceptual organization

- Sometimes, motion is the only cue

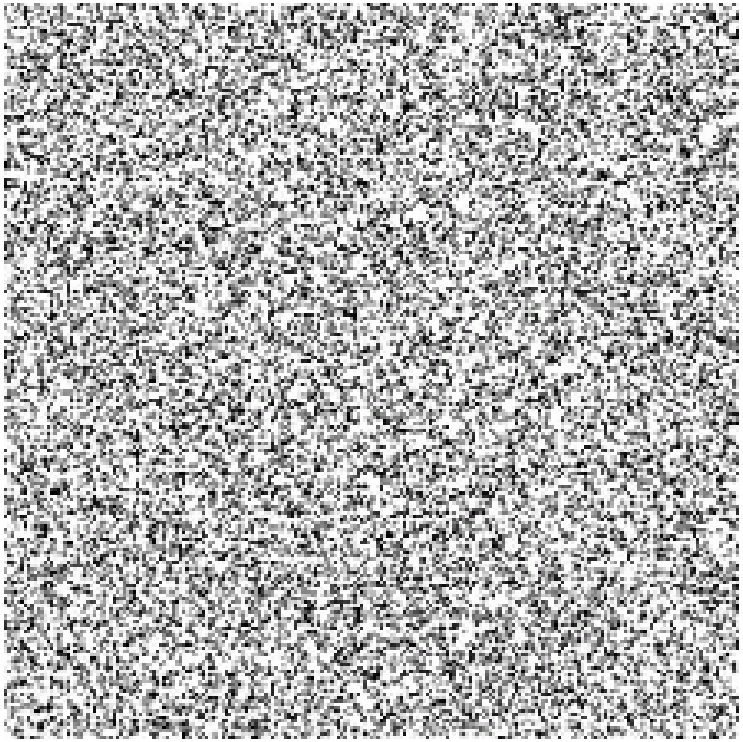
Motion and perceptual organization

- Sometimes, motion is the only cue



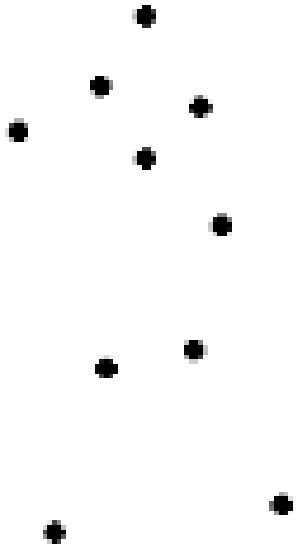
Motion and perceptual organization

- Sometimes, motion is the only cue



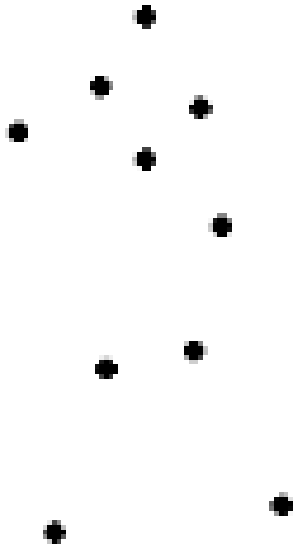
Motion and perceptual organization

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

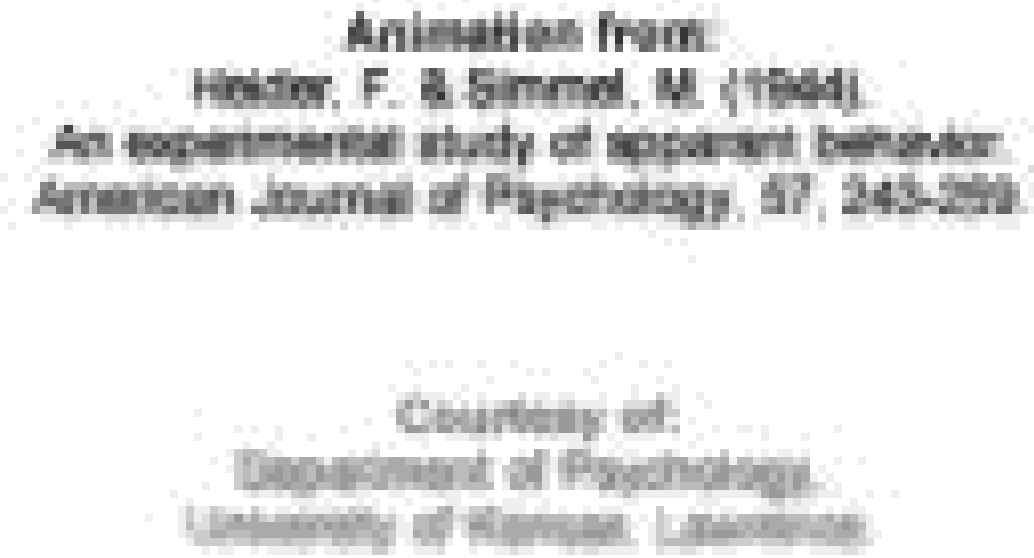


Motion and perceptual organization

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



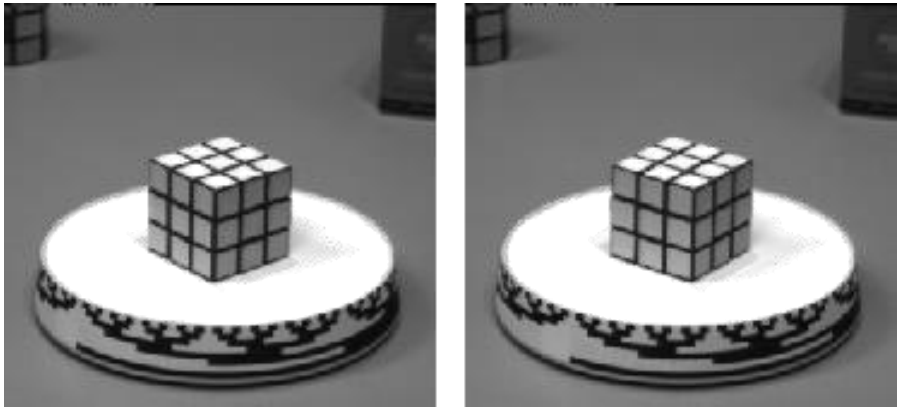
Motion and perceptual organization

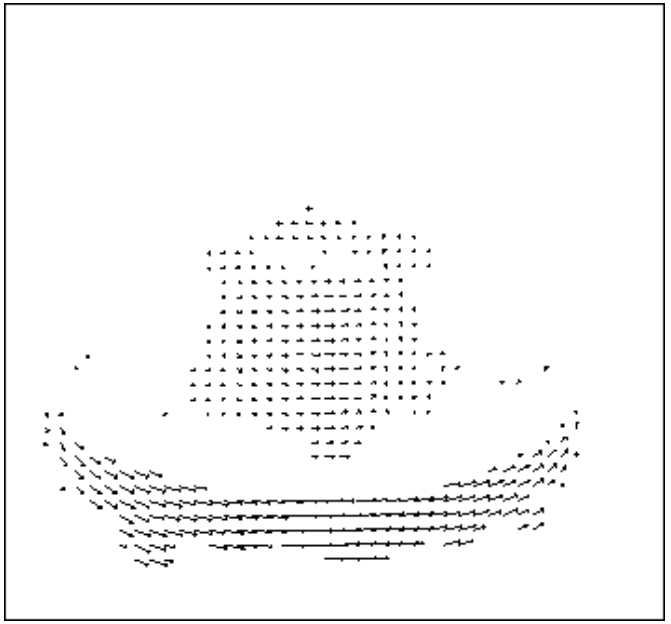


Experimental study of apparent behavior.
Fritz Heider & Marianne Simmel. 1944

Motion estimation: Optical flow

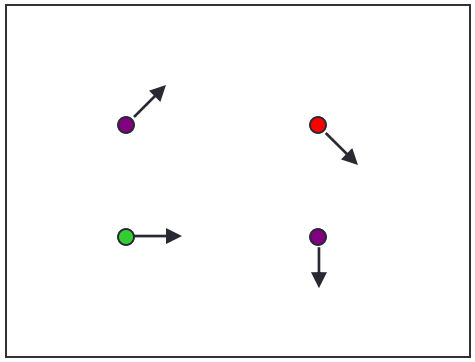
Optic flow is the **apparent** motion of objects or surfaces



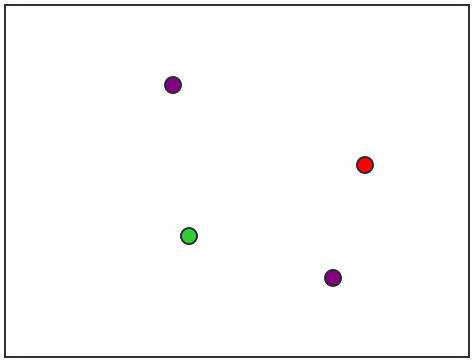


Will start by estimating motion of each pixel separately
Then will consider motion of entire image

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

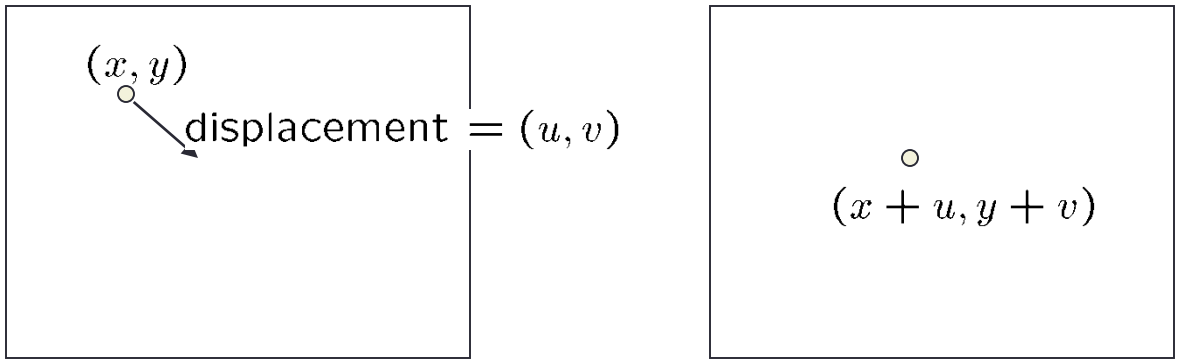
- Solve pixel correspondence problem
  - given a pixel in I(x,y,t), look for nearby pixels of the same color in *I(x,y,t+1)*

Key assumptions

- 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)



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

- 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 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

$$0 = I(x+u, y+v, t+1) - I(x, y, t)$$
$$\approx I(x, y, t+1) + I_x u + I_y v - I(x, y, t)$$

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

Optical flow equation

- Combining these two equations

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

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

Optical flow equation

- Combining these two equations

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

(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

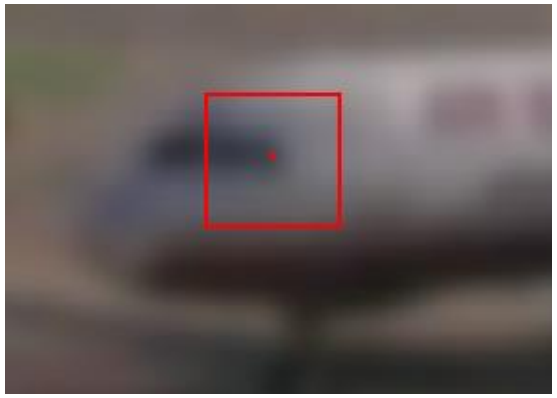
$$I_x\, u + I_y\, v + I_t = 0$$

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?

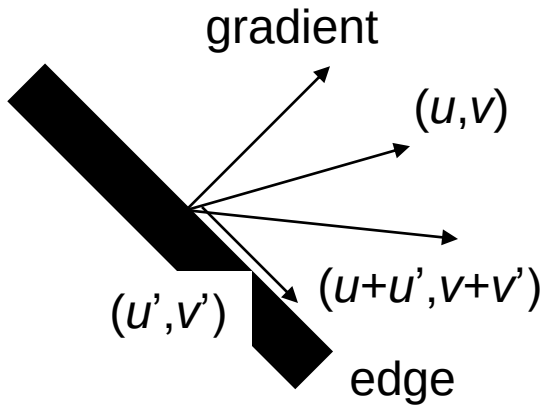
$$0 = I_t + \nabla I \cdot < u, v > \quad \text{or} \quad I_x u + I_y v + I_t = 0$$

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

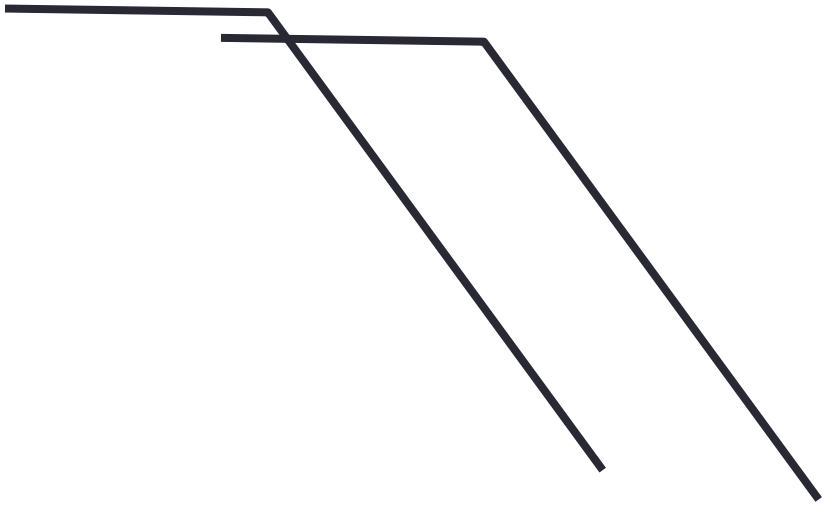
The component of the motion perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

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

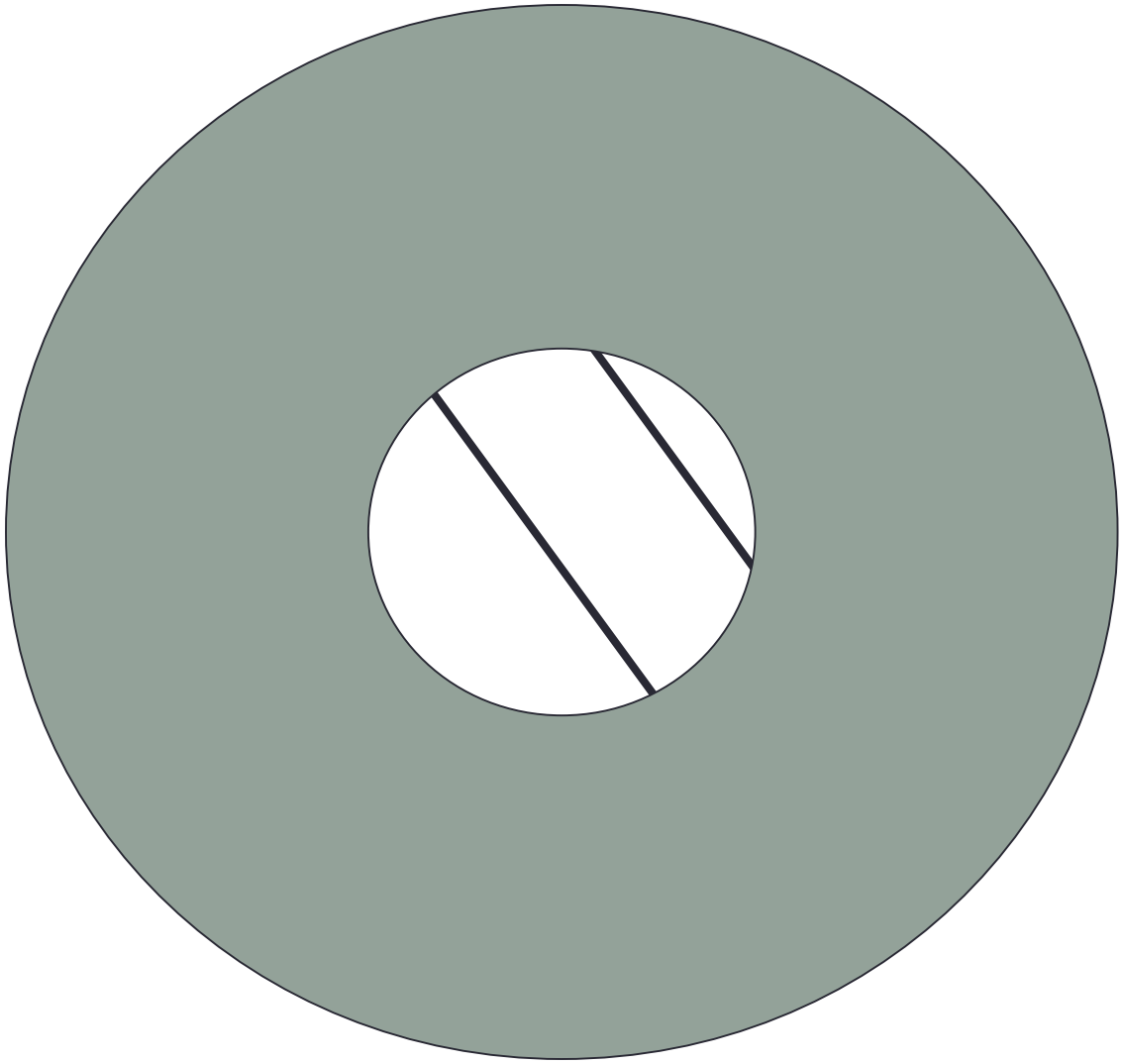
$$\nabla \mathrm{I} \cdot \left[\mathrm{u'}\ \ \mathrm{v'}\right]^{\mathrm{T}} = 0$$



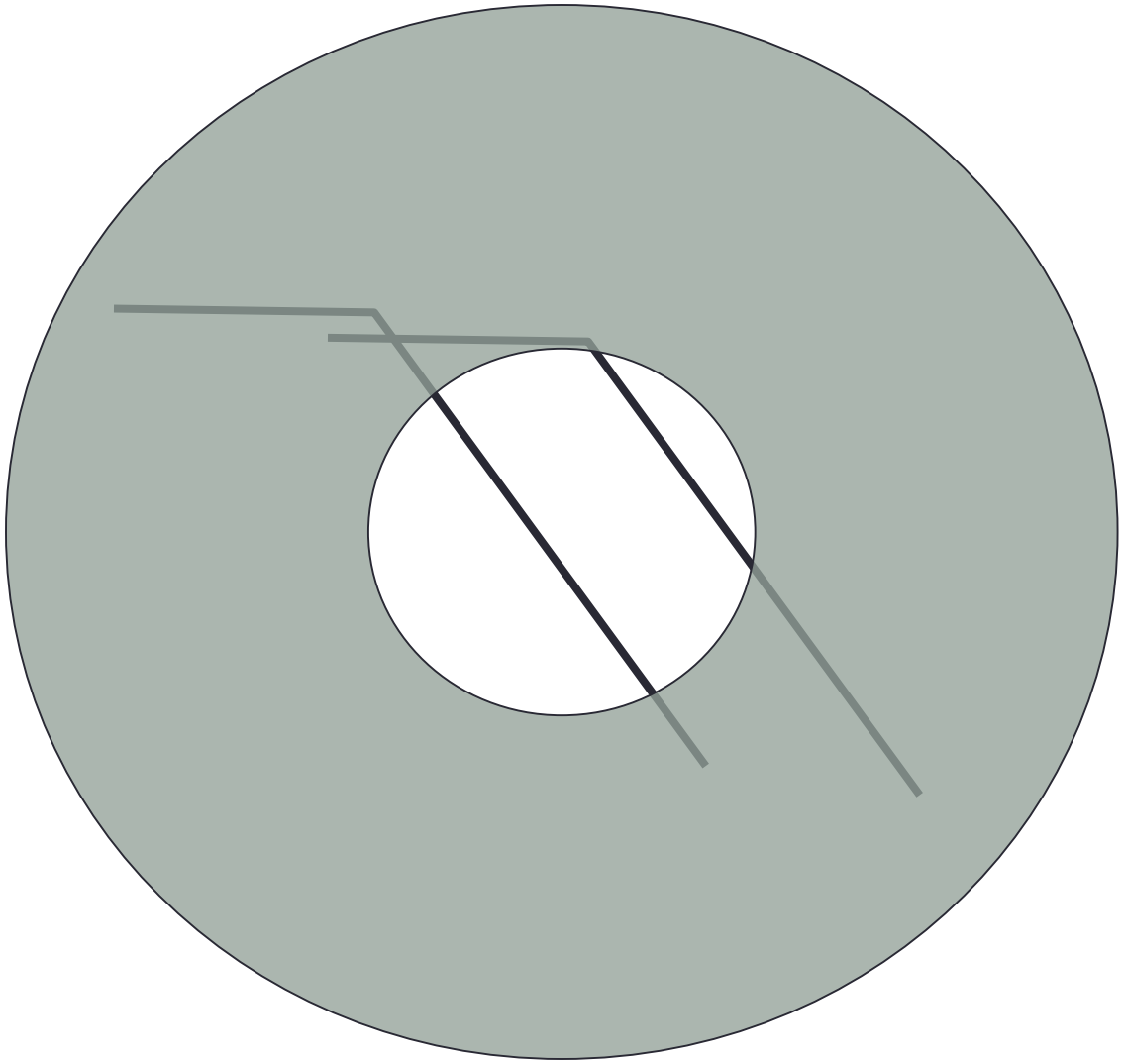
Aperture problem



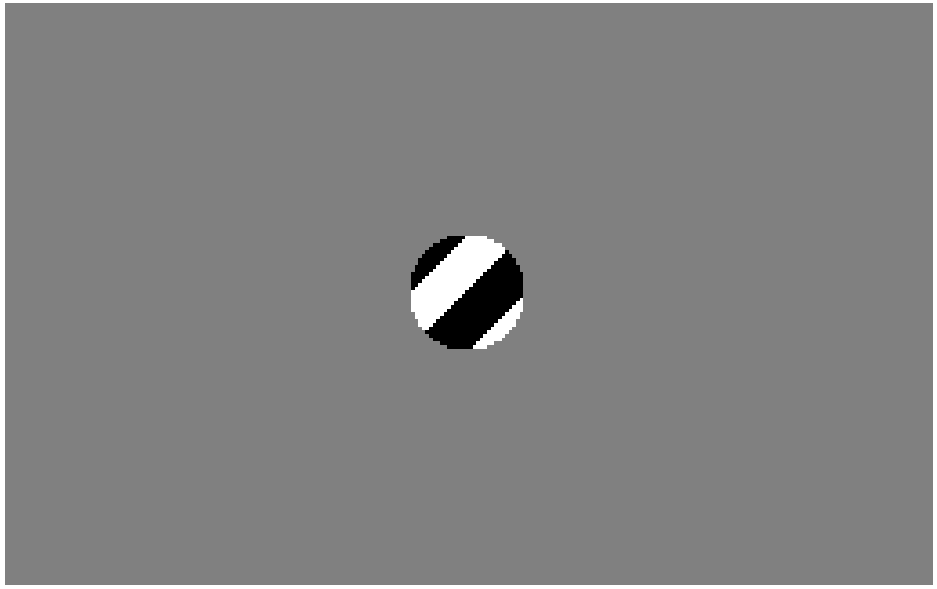
Aperture problem



Aperture problem

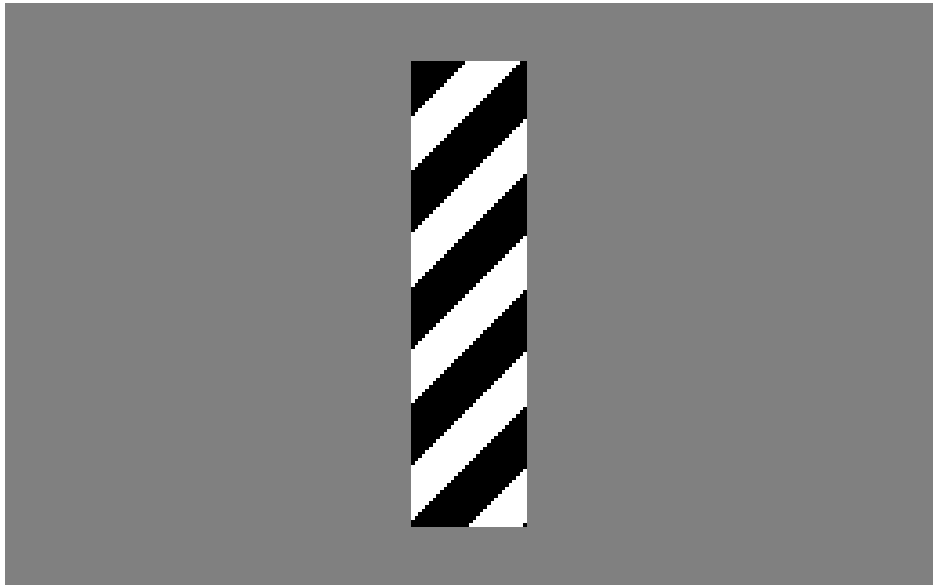


The barber pole illusion



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:

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

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

Matching patches across images

- Overconstrained linear system

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

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

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

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

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

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

Conditions for solvability

Optimal (u, v) satisfies Lucas-Kanade equation

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

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

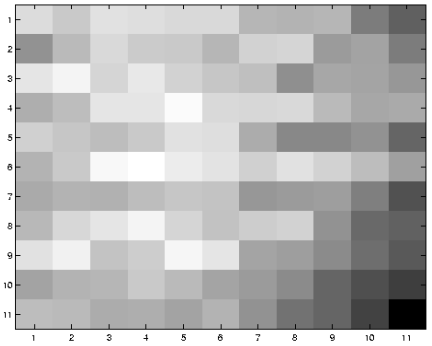
When is this solvable? I.e., what are good points to track?

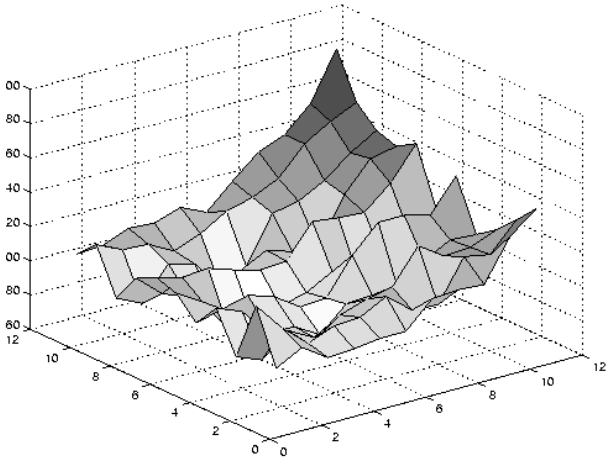
- $\mathbf{A^T A}$ should be invertible
- $\mathbf{A^T A}$ should not be too small due to noise
  - eigenvalues λ_1 and λ_2 of $\mathbf{A^T A}$ should not be too small
- $\mathbf{A^T A}$ should be well-conditioned
  - λ_1 / λ_2 should not be too large (λ_1 = larger eigenvalue)

Does this remind you of anything?

Criteria for Harris corner detector

Low texture region



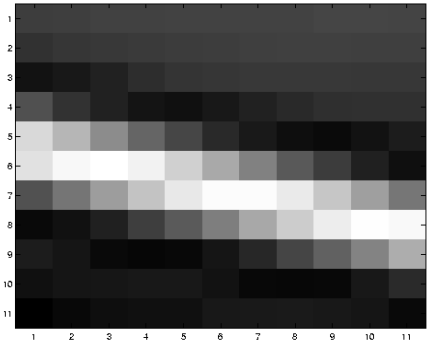


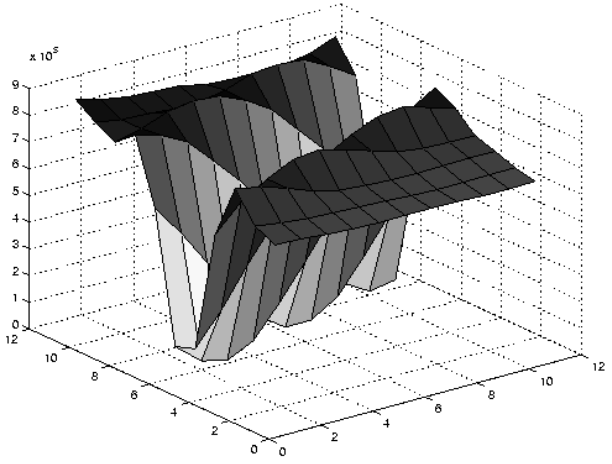
$$\sum \nabla I (\nabla I)^T$$

– gradients have small magnitude

– small λ_1, small λ_2

Edge

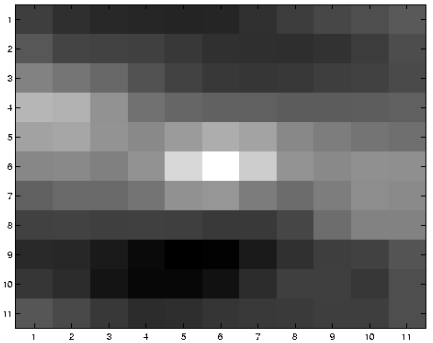


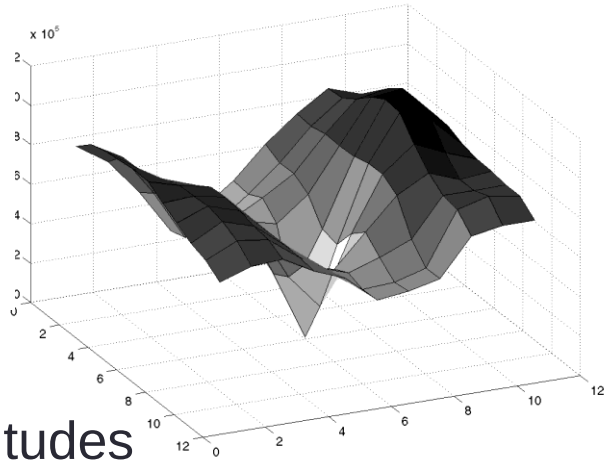


$$\sum \nabla I (\nabla I)^T$$

- large gradients, all the same
- large λ_1, small λ_2

High textured region



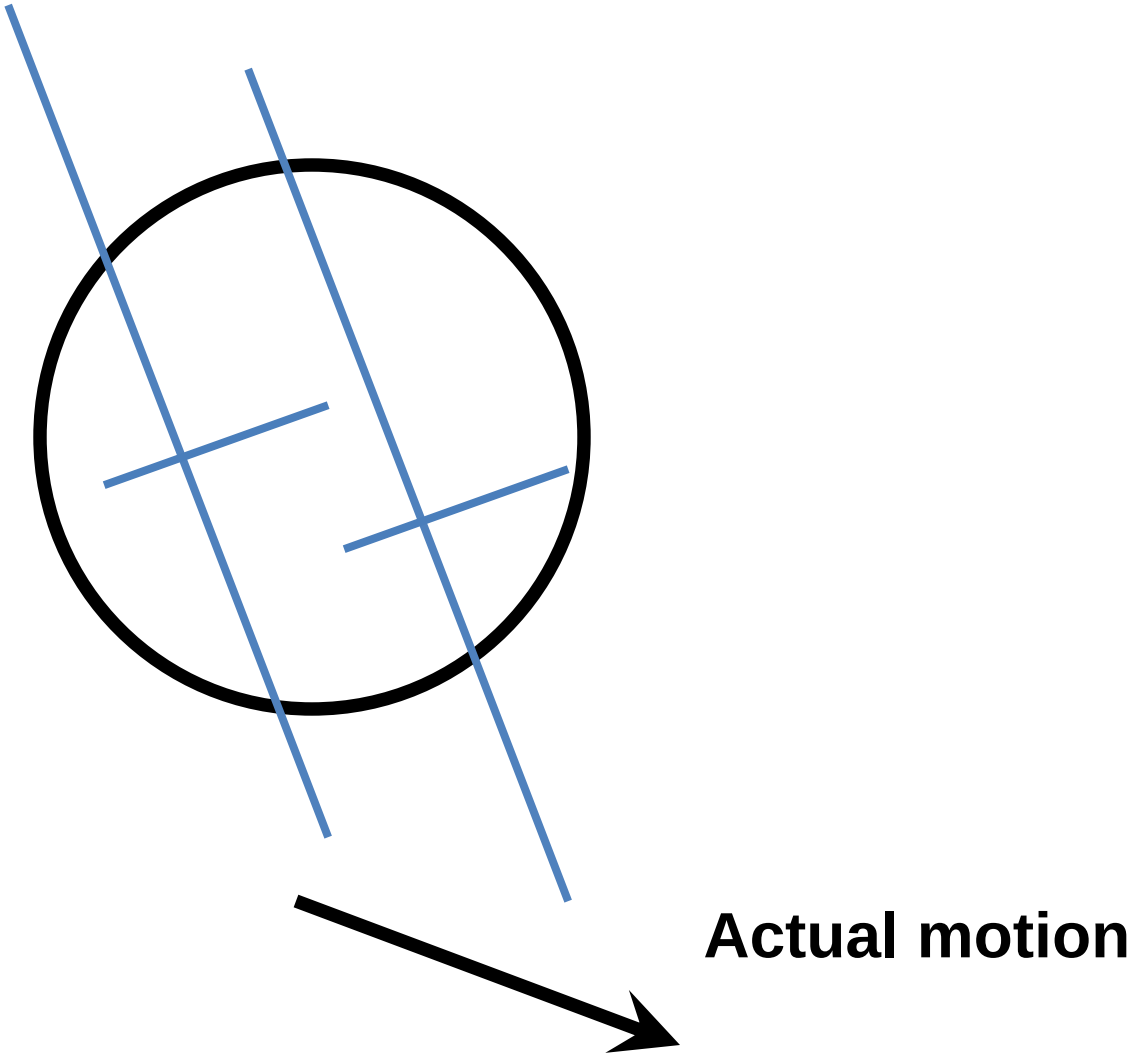


$$\sum \nabla I (\nabla I)^T$$

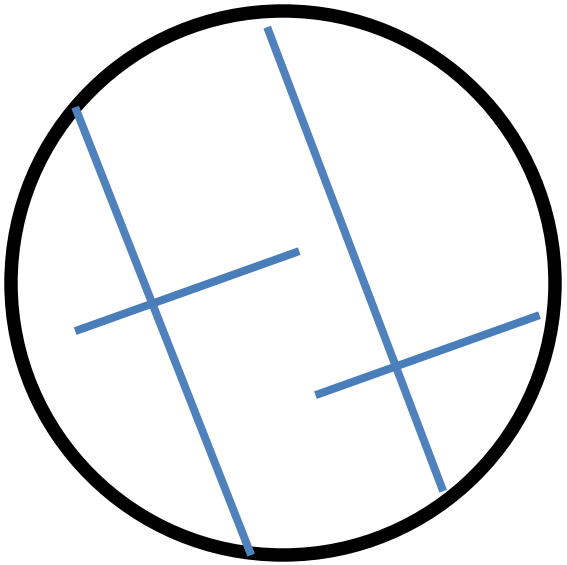
– gradients are different, large magnitudes

– large λ_1, large λ_2

The aperture problem resolved



The aperture problem resolved



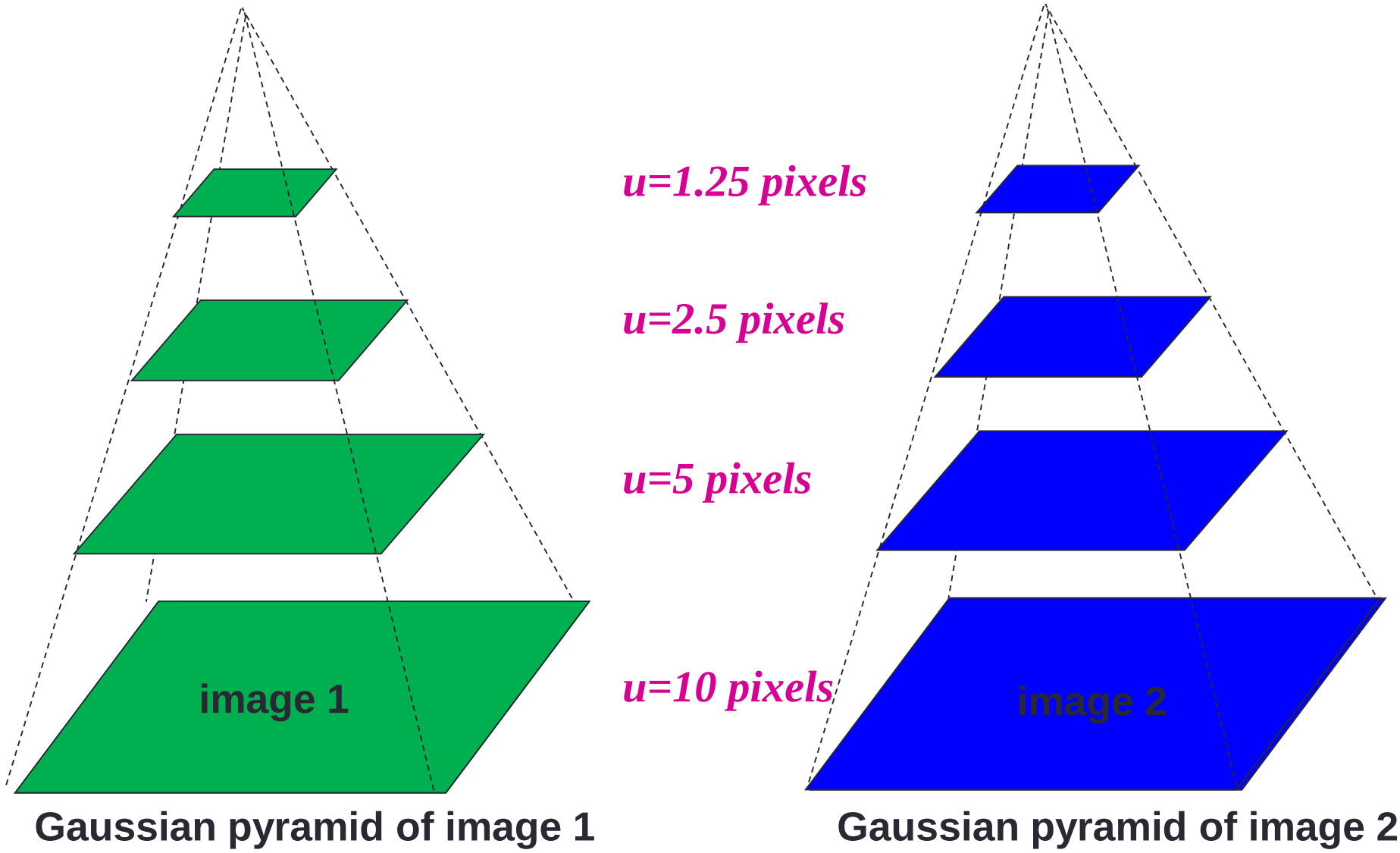
Errors in Lucas-Kanade

- 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

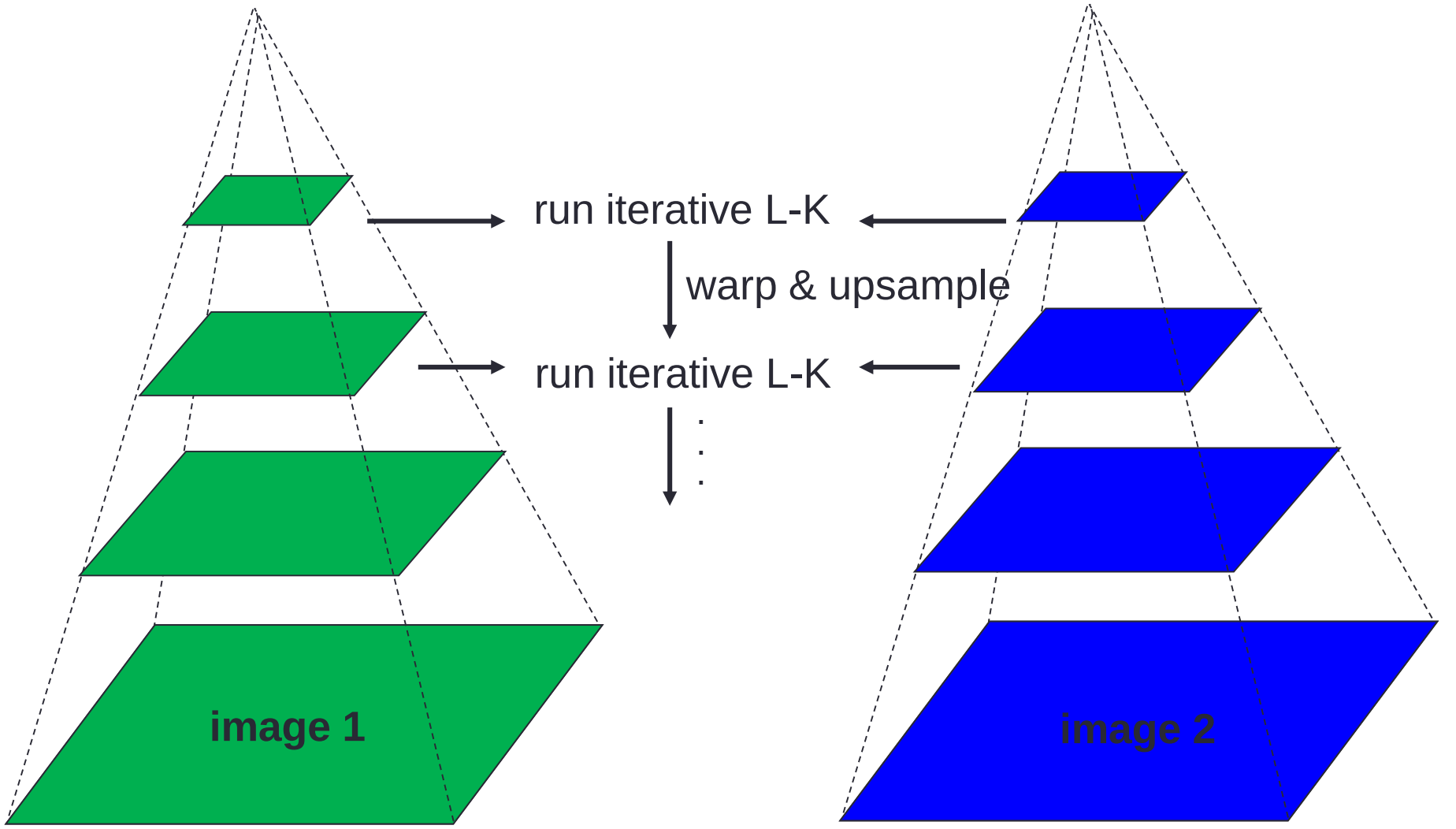
Revisiting the small motion assumption

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

Coarse-to-fine optical flow estimation



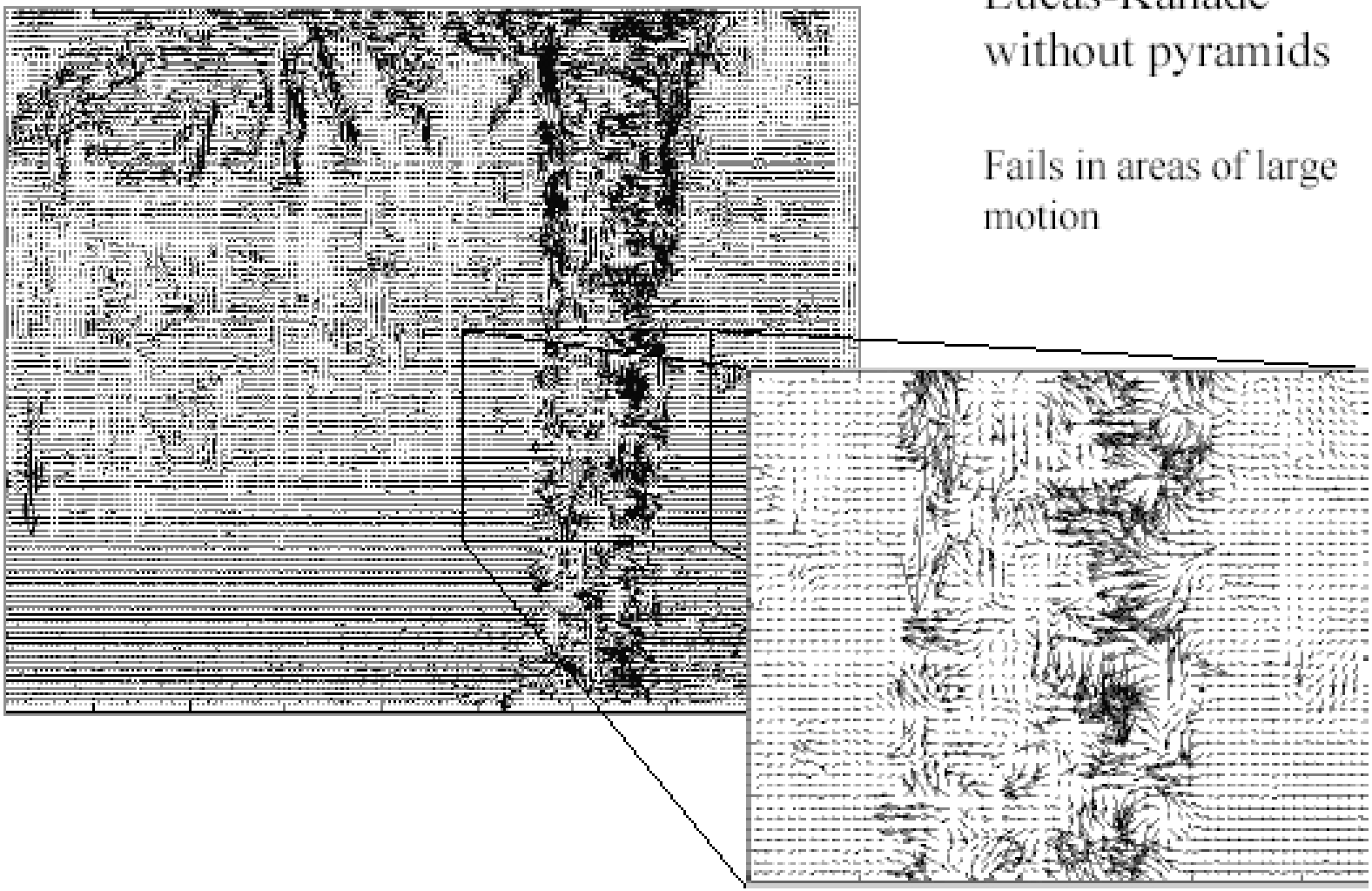
Coarse-to-fine optical flow estimation



Gaussian pyramid of image 1

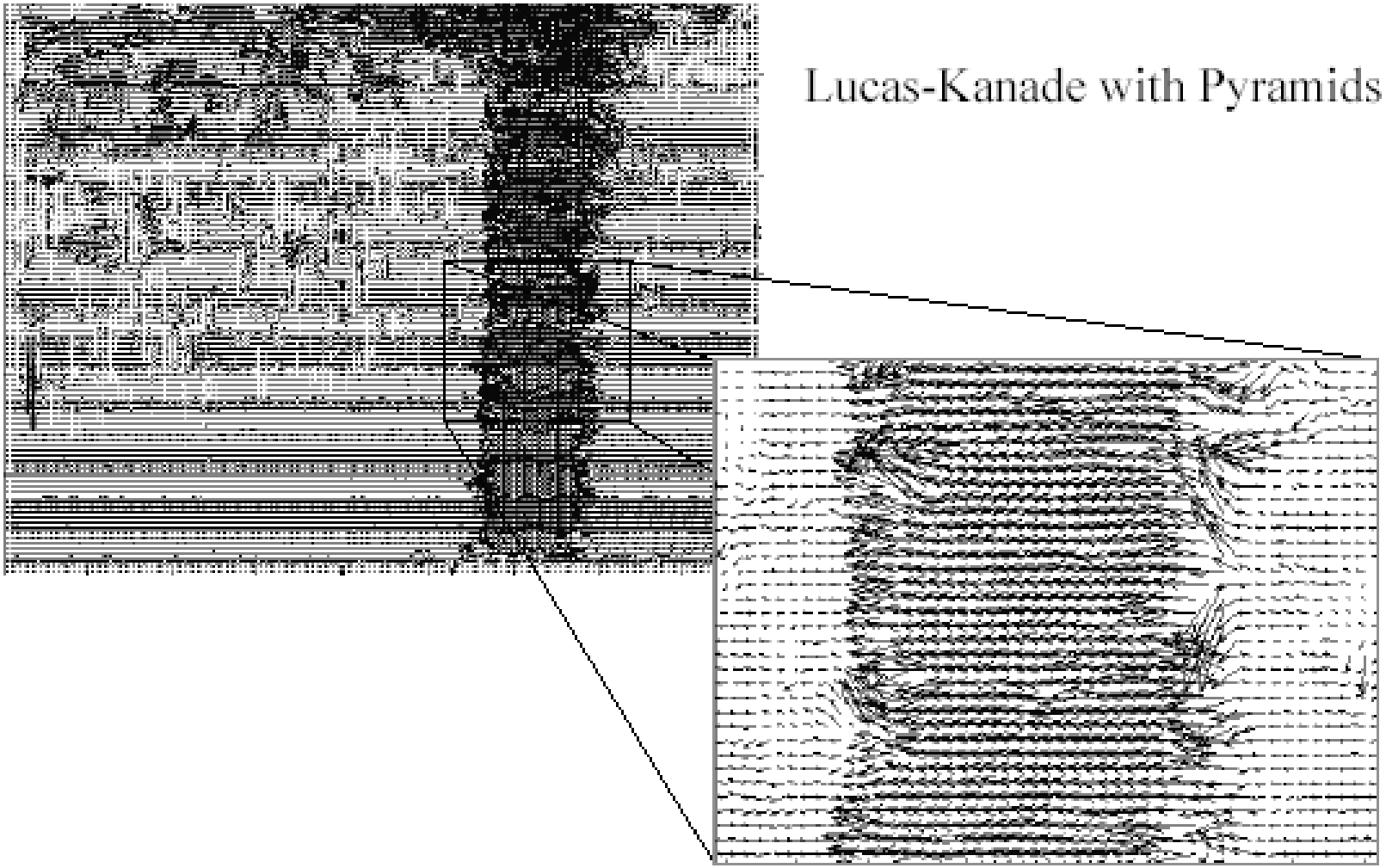
Gaussian pyramid of image 2

Optical Flow Results



* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003

Optical Flow Results



* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003