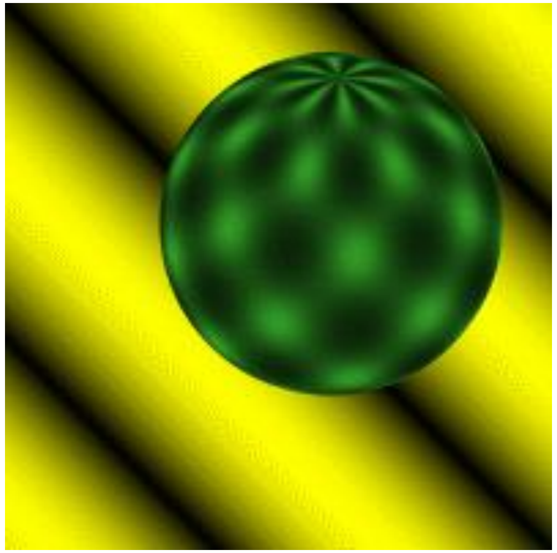
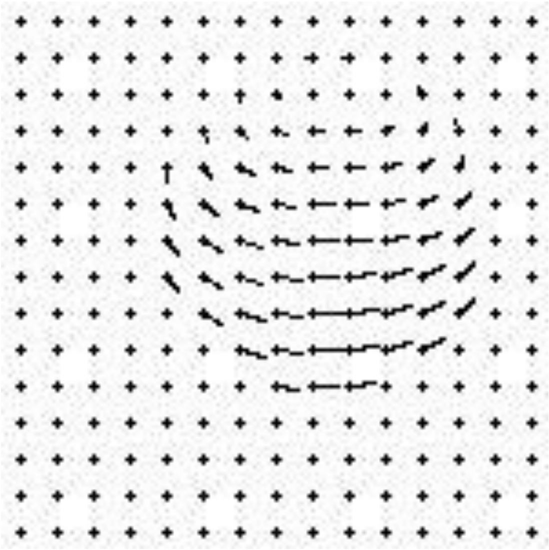
Computer Vision
- Motion & Optical flow

Junjie Cao @ DLUT

Spring 2018





We live in a moving world

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

-- from Linda Shapiro

Motion and perceptual organization

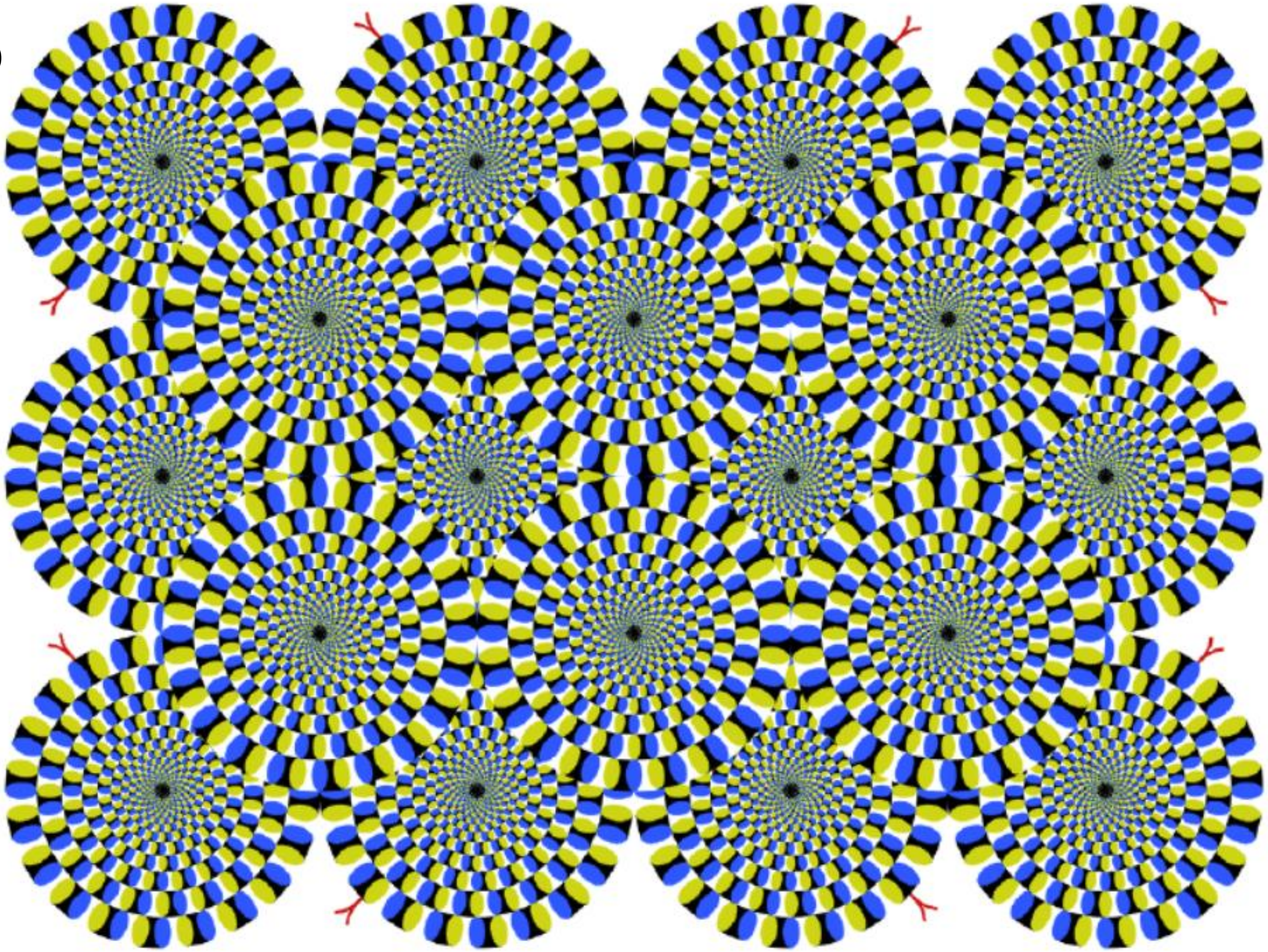
- Even “impoverished” motion data can evoke a strong percept

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

-- from Linda Shapiro

Seeing motion from a static picture?

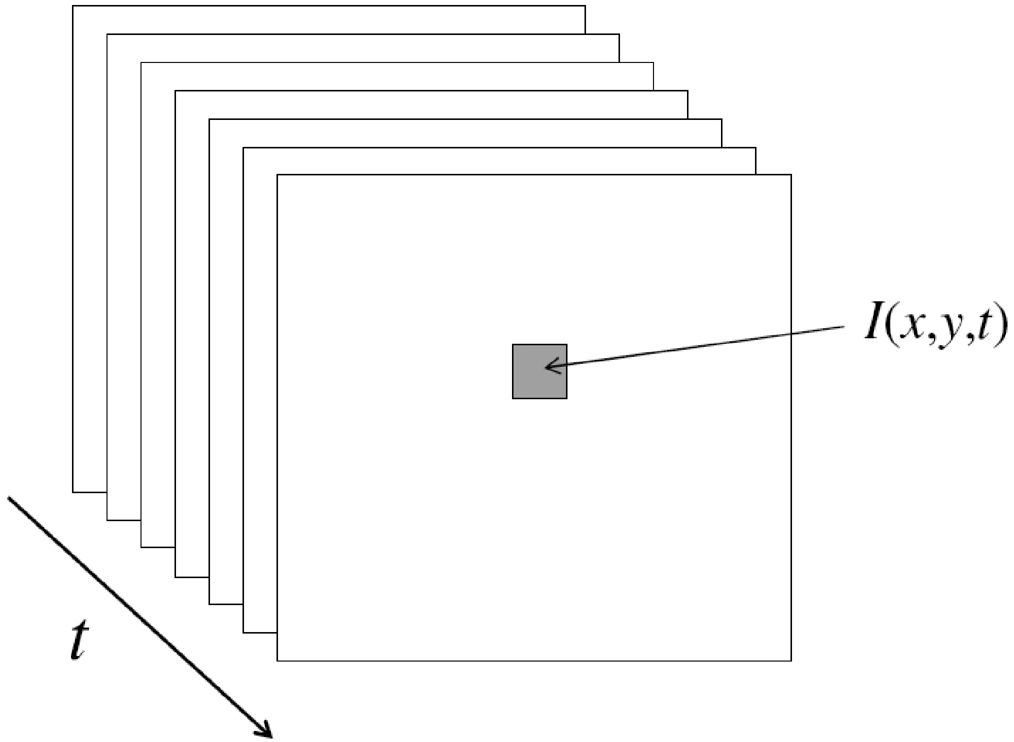
- The true mechanism is yet to be revealed
- FMRI data suggest that illusion is related to some component of eye movements
- We don’t expect computer vision to “see” motion from these stimuli, yet



-- from Linda Shapiro

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)



The cause of motion

- Three factors in imaging process
  - Light
  - Object
  - Camera

- Varying either of them causes motion
  - Static camera, moving objects (surveillance)
  - Moving camera, static scene (3D capture)
  - Moving camera, moving scene (sports, movie)
  - Static camera, moving objects, moving light (time lapse)

-- from Linda Shapiro

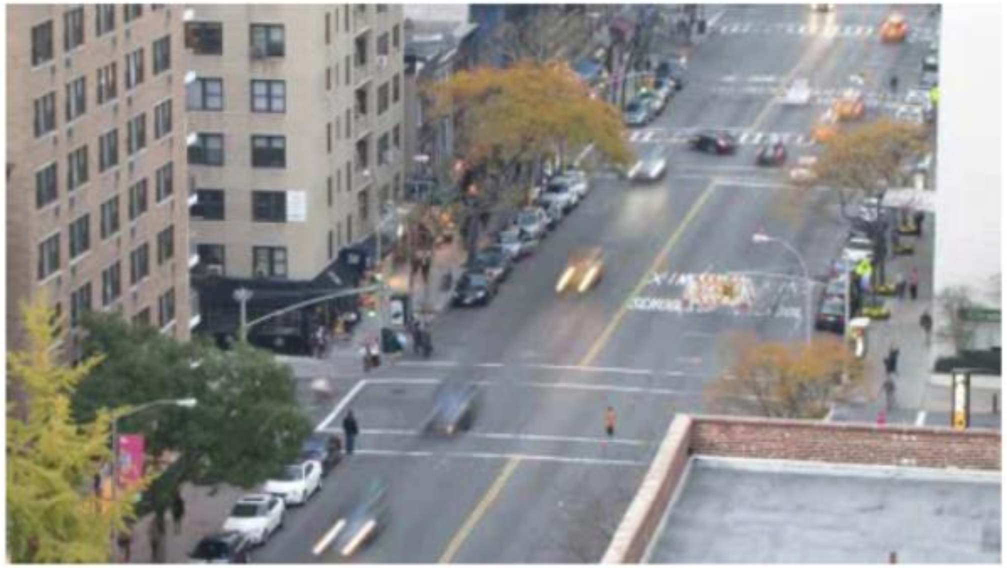
Motion scenarios (priors)

-- from Linda Shapiro

Static camera, moving objects (surveillance)

Moving camera, static scene (3D capture)

Moving camera, moving scene (sports, movie)



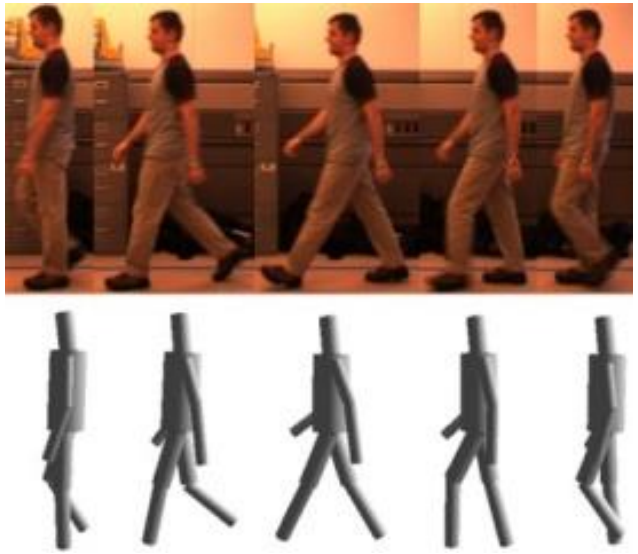
Static camera, moving objects, moving light (time lapse)

We still don't touch these areas

-- from Linda Shapiro

Uses of motion

- Estimating 3D structure
- Segmenting objects based on motion cues
- Learning dynamical models
- Recognizing events and activities
- Improving video quality (motion stabilization)



How can we recover motion?

-- from Linda Shapiro

Recovering motion

- **Feature-tracking**
  - **Extract visual features (corners, textured areas) and “track” them over multiple frames**

- **Optical flow**
  - **Recover image motion at each pixel from spatio-temporal image brightness variations (optical flow)**

- **Two problems, one registration method**

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

-- from Linda Shapiro

Motion estimation techniques

- **Direct 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**

- **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)**

- **Two problems, one registration method**

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

Challenges

- **Figure out which features can be tracked**
- **Efficiently track across frames**

- **Some points may change appearance over time (e.g., due to rotation, moving into shadows, etc.)**

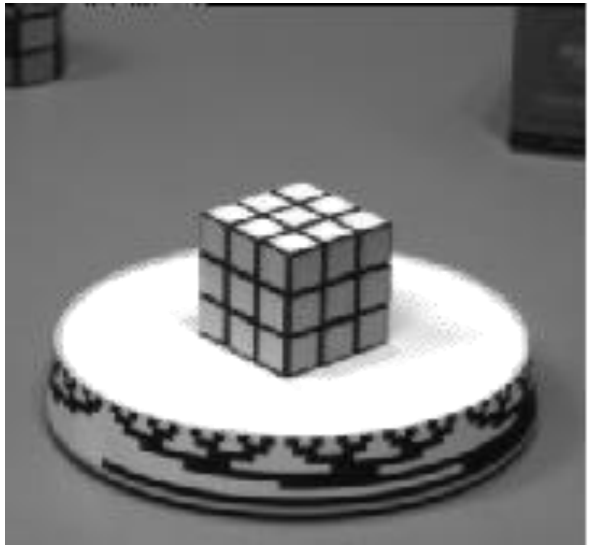
- **Drift: small errors can accumulate as appearance model is updated**

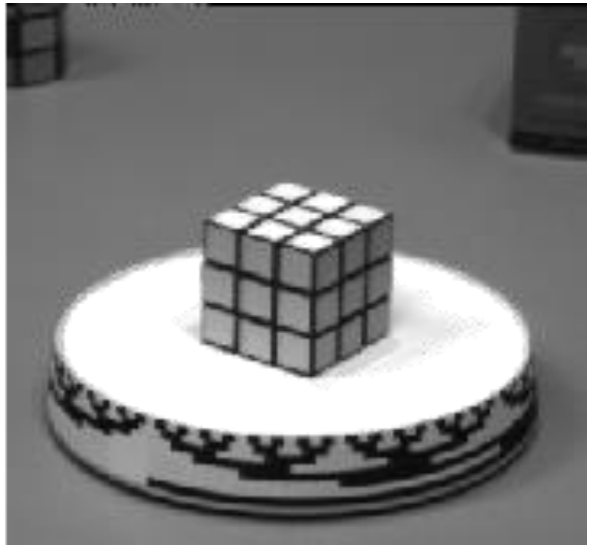
- **Points may appear or disappear: need to be able to add/delete tracked points**

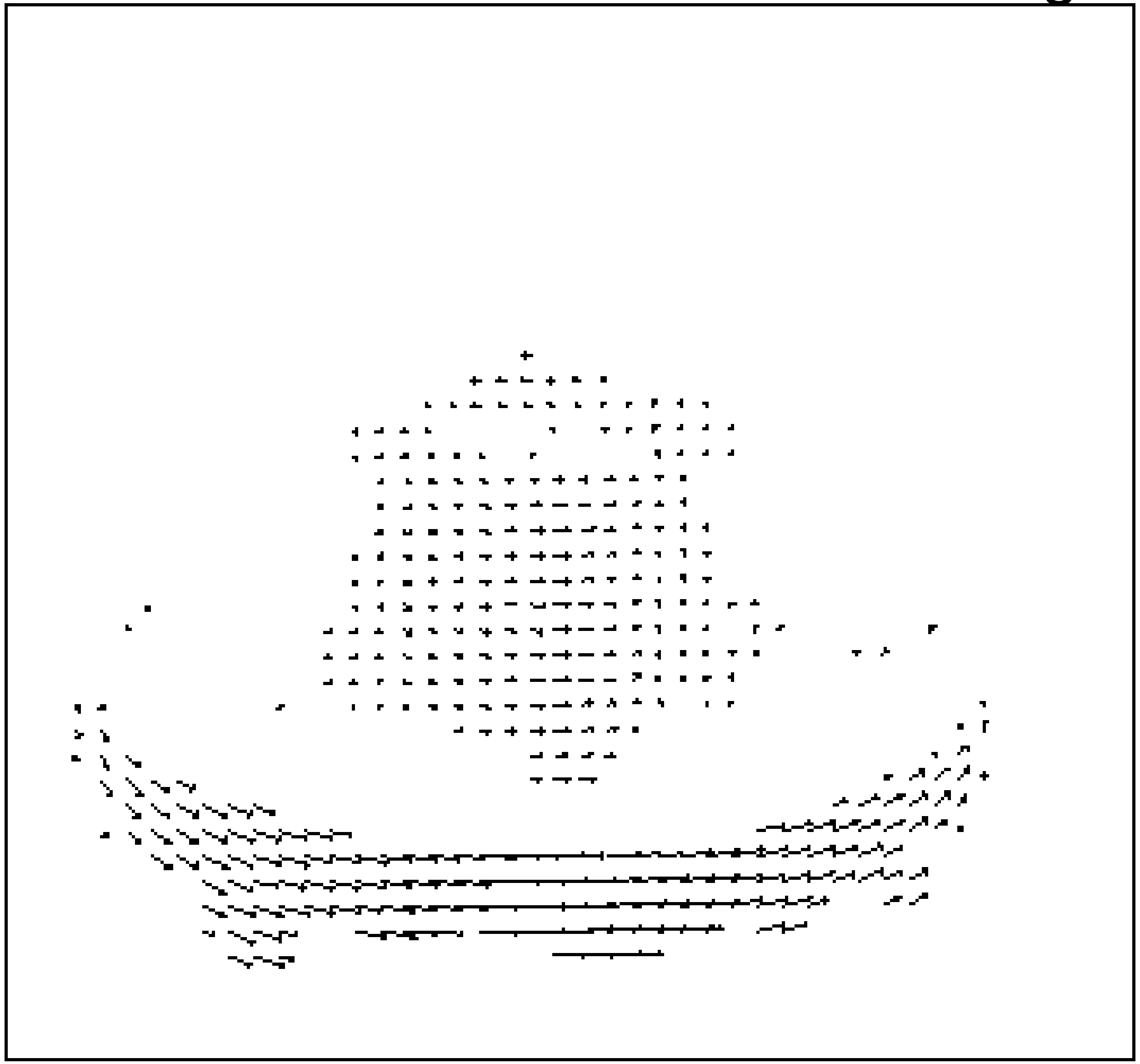
-- from Linda Shapiro

Motion field

- The motion field is the projection of the 3D scene motion into the image

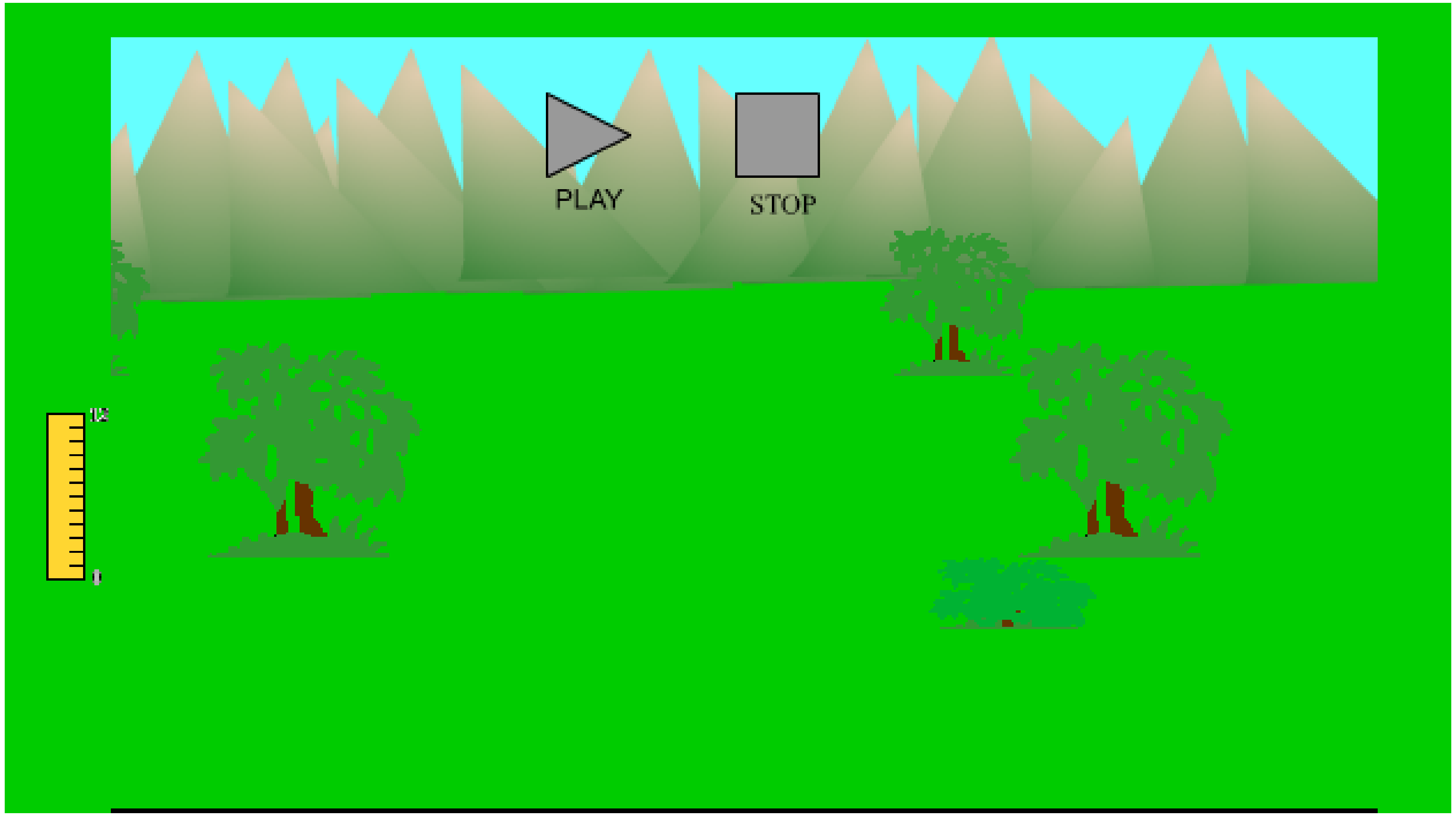


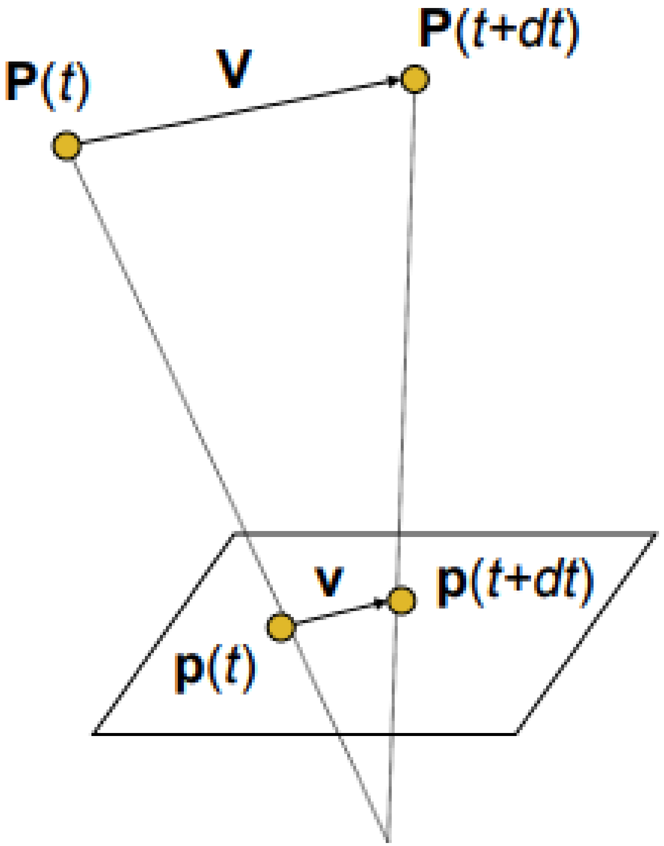




Motion parallax 运动视差

- Motion parallax is a **depth cue** that results from our motion.
  - As we move, objects that are closer to us move farther across our field of view than do objects that are in the distance.

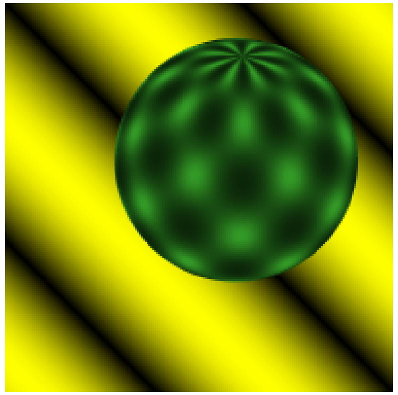
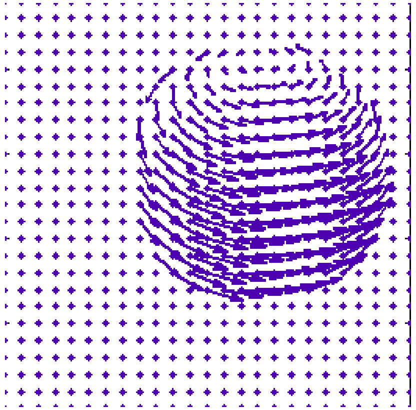
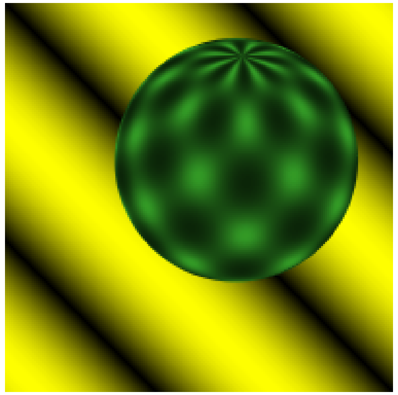




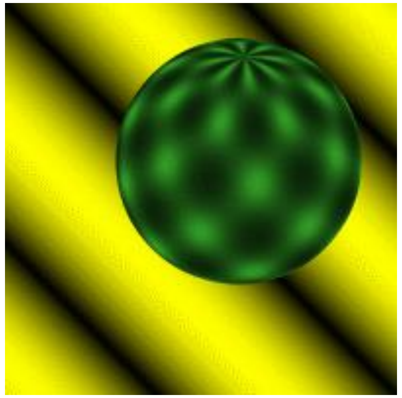
http://psych.hanover.edu/KRANTZ/MotionParallax/MotionParallax.html

What is Optical Flow?

The optical flow is a velocity field in the image which transforms one image into the next image in a sequence
[Horn&Schunck]

frame #1 + flow field = frame #2



Optical flow (Apparent motion) != motion field

- **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**

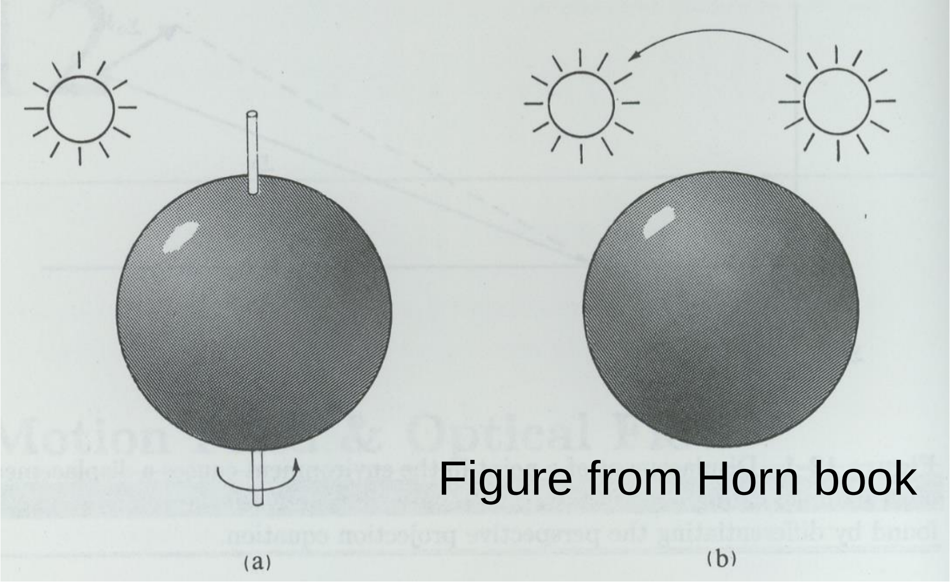
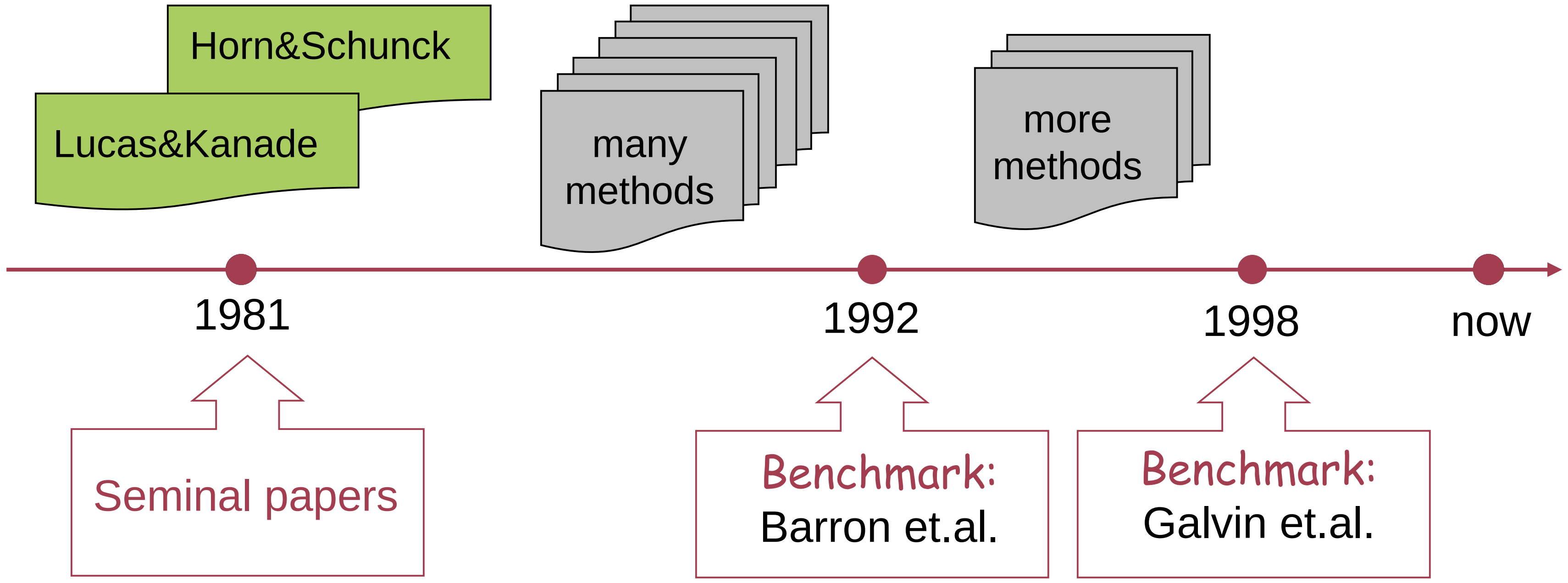


Figure from Horn book

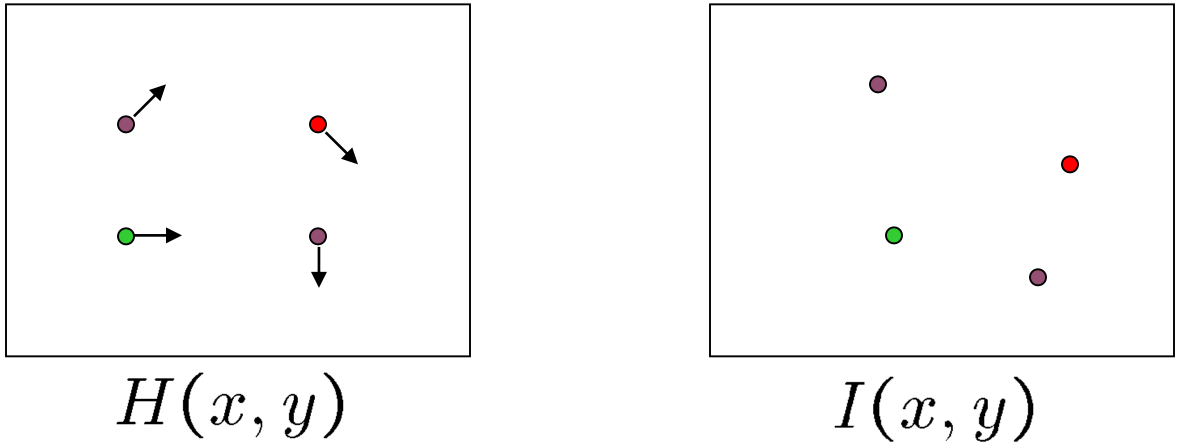
The motion field … is the projection into the image of three-dimensional motion vectors [Horn&Schunck]

Optical Flow Research: Timeline



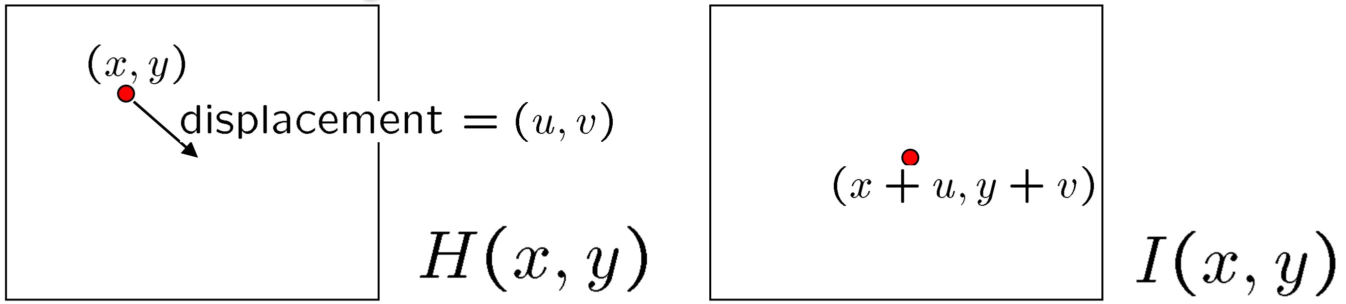
A slow and not very consistent improvement in results, but a lot of useful ingredients were developed

Problem Definition: Optical Flow



- **How to estimate pixel motion from image H to image I?**
  - **Find pixel correspondences**
    - **Given a pixel in H, look for nearby pixels of the same color in I**
- **Key assumptions**
  - **color constancy: a point in H looks "the same" in image I**
    - **For grayscale images, this is brightness constancy**
  - **small motion: points do not move very far**

Brightness constancy constraint



- **Brightness Constancy Equation:**

$$\boxed{I(x,y,t) = I(x+u, y+v, t+1)}$$

Take Taylor expansion of $I(x+u, y+v, t+1)$ at (x,y,t) to linearize the right side:

Image derivative along x   Difference over frames

$$\boxed{I(x+u, y+v, t+1) \approx I(x,y,t) + \boxed{I_x} \cdot u + I_y \cdot v + \boxed{I_t}}$$

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

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

Brightness constancy constraint (for gray image)

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

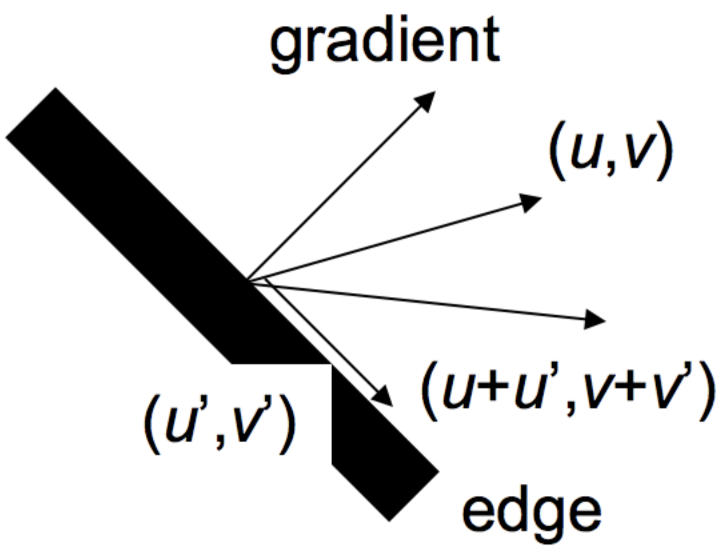
$$\nabla I \cdot [u \;\; v]^T + I_t = 0$$

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

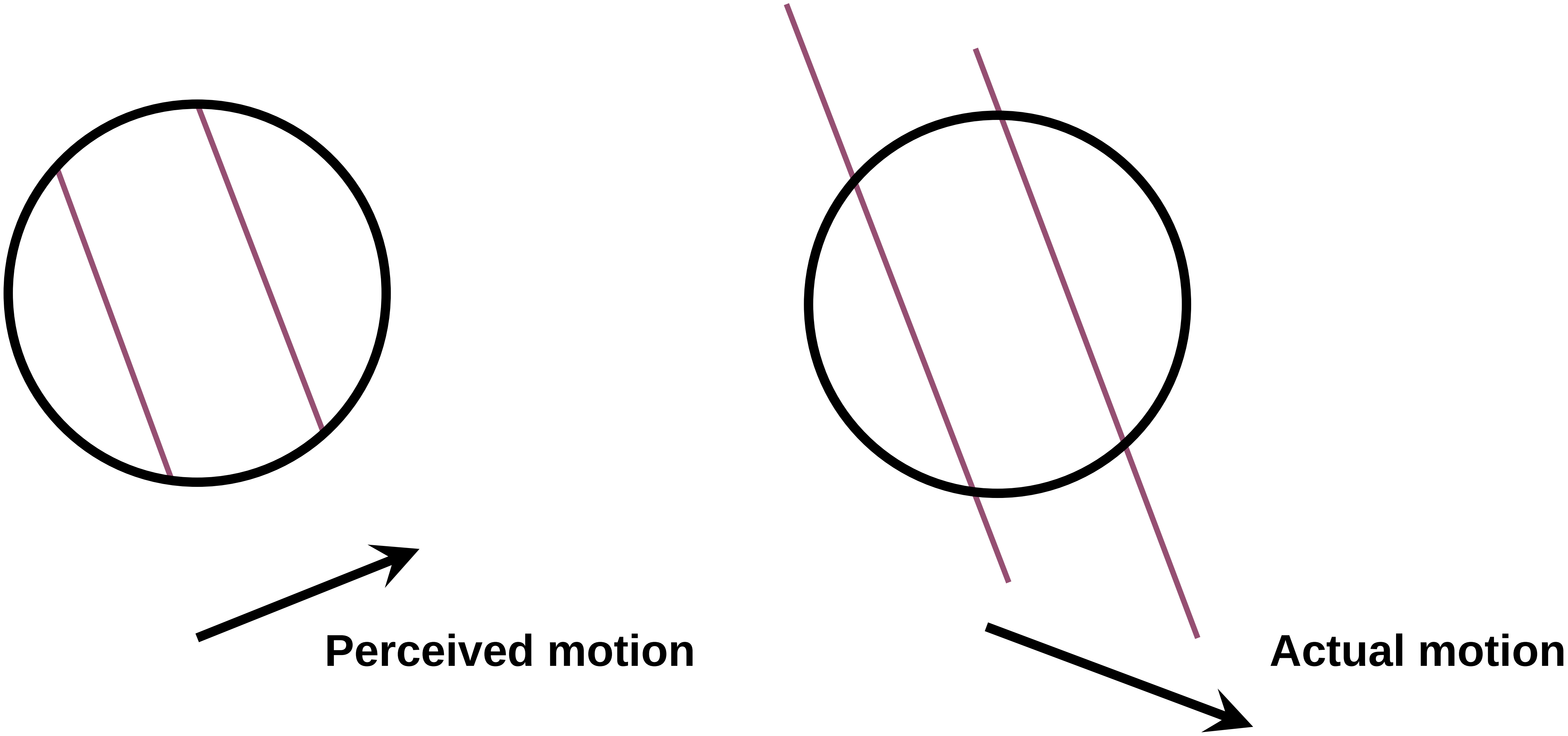
- The component of the motion perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

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

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



The aperture problem 孔径问题

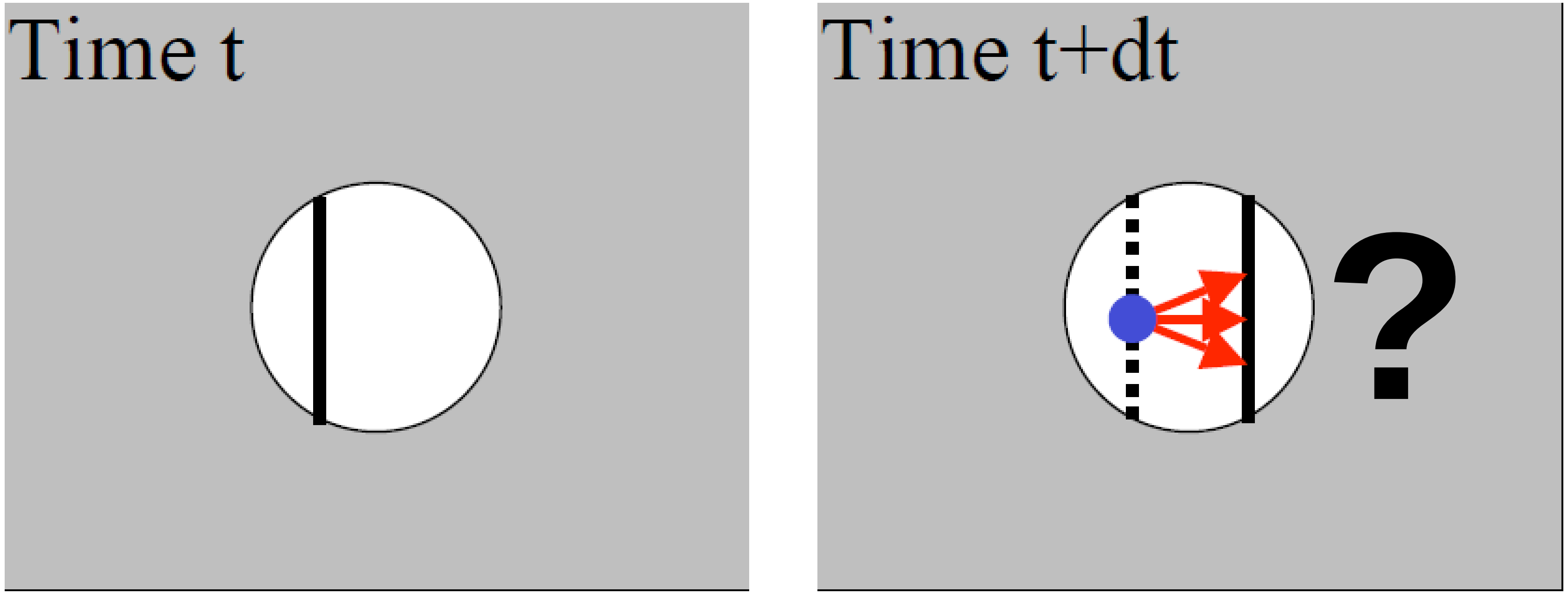


The component of the flow perpendicular to the gradient (i.e., parallel to the edge) is unknown

Aperture Problem

The aperture problem

- For points on a line of fixed intensity we can only recover the normal flow



Where did the blue point move to?

We need additional constraints.

Solving the aperture problem

- Basic idea: assume motion field is smooth
- Horn & Schunk: add smoothness term

$$\int\int (I_t + \nabla I \cdot [u\ v])^2 + \lambda^2 (\|\nabla u\|^2 + \|\nabla v\|^2)\ dx\ dy$$

- Lucas & Kanade: assume locally constant motion
  - pretend the pixel's neighbors have the same (u,v)

- Many other methods exist. Here's an overview: • S. Baker, M. Black, J. P. Lewis, S. Roth, D. Scharstein, and R. Szeliski. A database and evaluation methodology for optical flow. In Proc. ICCV, 2007 • http://vision.middlebury.edu/flow/

Solving the aperture problem - Lucas-Kanade flow

- **How to get more equations for a pixel?**
- **Spatial coherence constraint: pretend the pixel's neighbors have the same (u,v)**

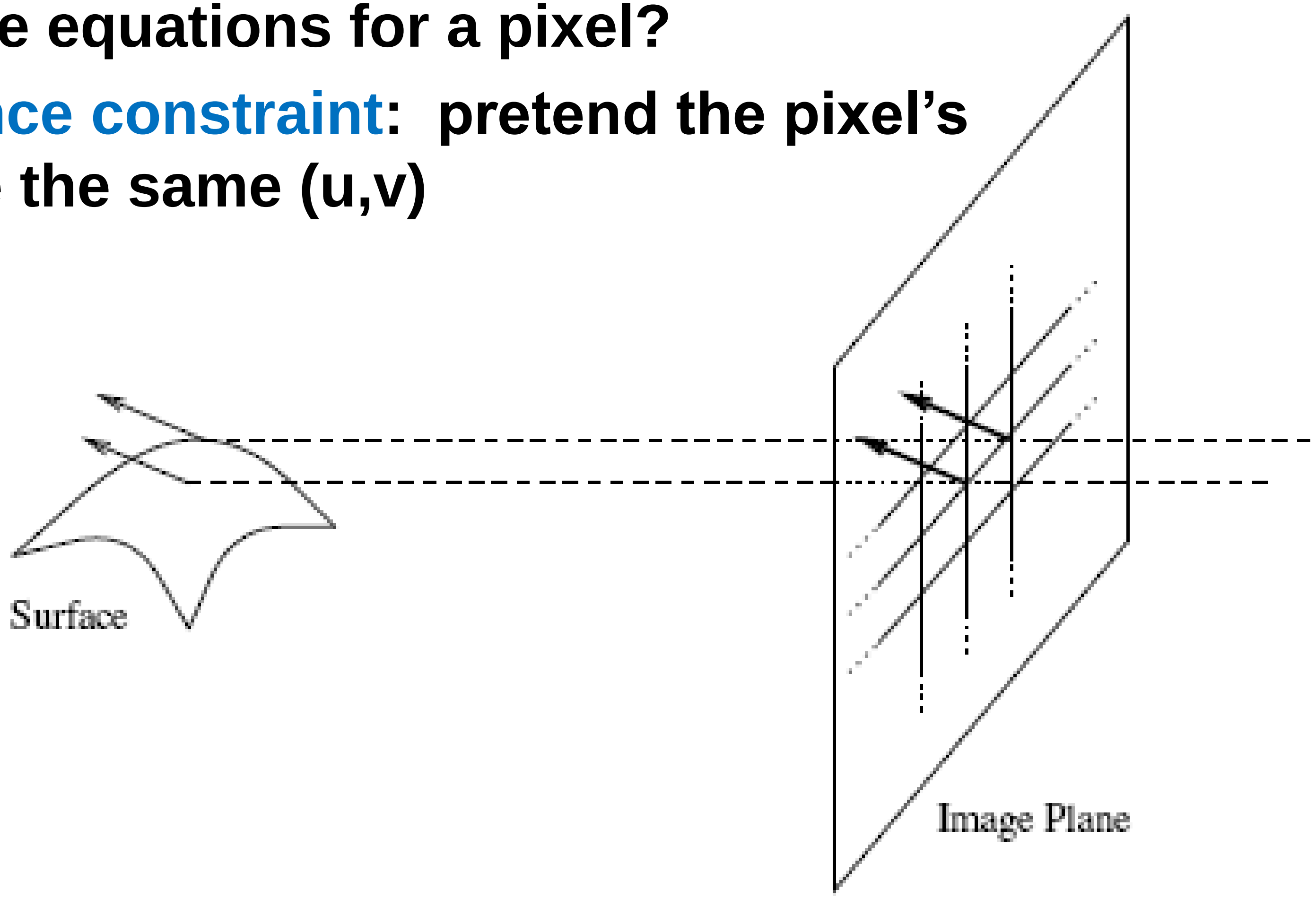


Figure 1.7: Spatial coherence assumption. Neighboring points in the image are assumed to belong to the same surface in the scene.

Figure by Michael Black

Solving the aperture problem

- **Spatial coherence constraint: pretend the pixel's neighbors have the same (u,v)**
  - **If we use a 5x5 window, that gives us 25 equations per pixel**

$$0 = I_t(\mathbf{p_i}) + \nabla I(\mathbf{p_i}) \cdot [u \;\; v]$$

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

$$\underset{25\times 2}{A} \;\; \underset{2\times 1}{d} = \underset{25\times 1}{b}$$

Steve Seitz

RGB version

- **How to get more equations for a pixel?**
  - **Basic idea: impose additional constraints**
    - **most common is to assume that the flow field is smooth locally**
    - **one method: pretend the pixel's neighbors have the same (u,v)**
      - **If we use a 5x5 window, that gives us 25*3 equations per pixel!**

$$0 = I_t(\mathbf{p_i})[0,1,2] + \nabla I(\mathbf{p_i})[0,1,2] \cdot [u \;\; v]$$

$$\underset{\substack{A \\ 75\times 2}}{\begin{bmatrix} I_x(\mathbf{p}_1)[0] & I_y(\mathbf{p}_1)[0] \\ I_x(\mathbf{p}_1)[1] & I_y(\mathbf{p}_1)[1] \\ I_x(\mathbf{p}_1)[2] & I_y(\mathbf{p}_1)[2] \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25})[0] & I_y(\mathbf{p}_{25})[0] \\ I_x(\mathbf{p}_{25})[1] & I_y(\mathbf{p}_{25})[1] \\ I_x(\mathbf{p}_{25})[2] & I_y(\mathbf{p}_{25})[2] \end{bmatrix}} \underset{\substack{d \\ 2\times 1}}{\begin{bmatrix} u \\ v \end{bmatrix}} = - \underset{\substack{b \\ 75\times 1}}{\begin{bmatrix} I_t(\mathbf{p}_1)[0] \\ I_t(\mathbf{p}_1)[1] \\ I_t(\mathbf{p}_1)[2] \\ \vdots \\ I_t(\mathbf{p}_{25})[0] \\ I_t(\mathbf{p}_{25})[1] \\ I_t(\mathbf{p}_{25})[2] \end{bmatrix}}$$

Solving the aperture problem

Prob: we have more equations than unknowns

$$\underset{25\text{x}2}{A}\ \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b} \longrightarrow \text{minimize } \|Ad - b\|^2$$

Solution: solve least squares problem

- **minimum least squares solution given by solution (in d) of:**

$$\underset{2\text{x}2}{(A^T A)}\ \underset{2\text{x}1}{d} = \underset{2\text{x}1}{A^T b}$$

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

- **The summations are over all pixels in the K x K window**
- **This technique was first proposed by Lucas & Kanade (1981)**

Conditions for solvability

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

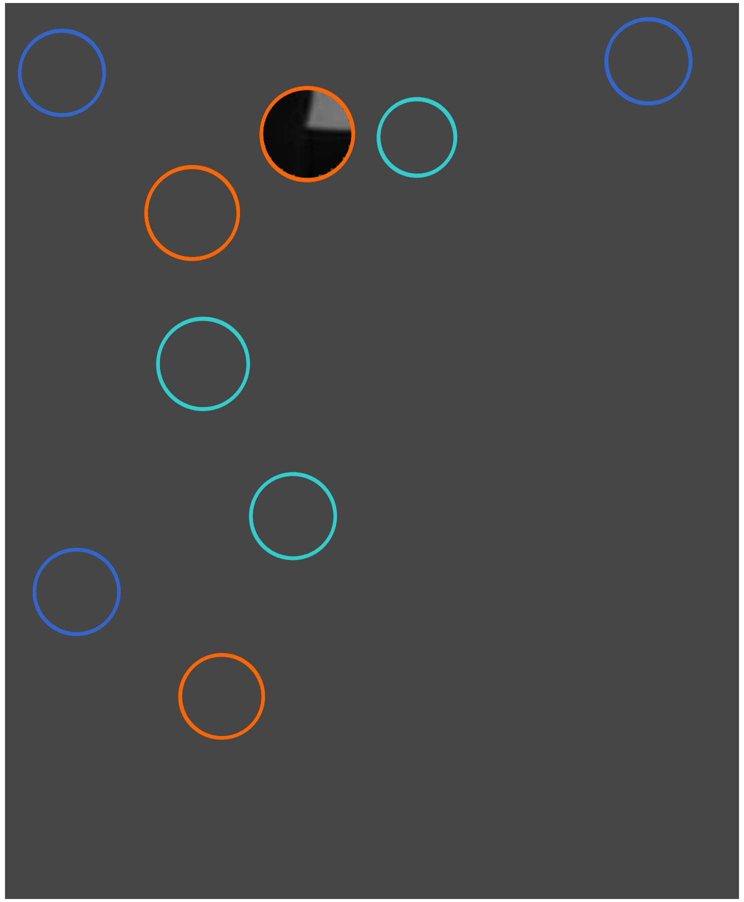
$$A^T A \qquad\qquad\qquad\qquad A^T b$$

When is this solvable?

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

Slide by Steve Seitz, UW

Aperture problem

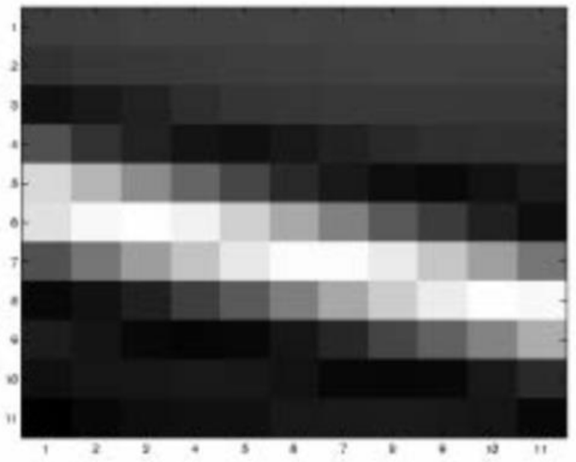


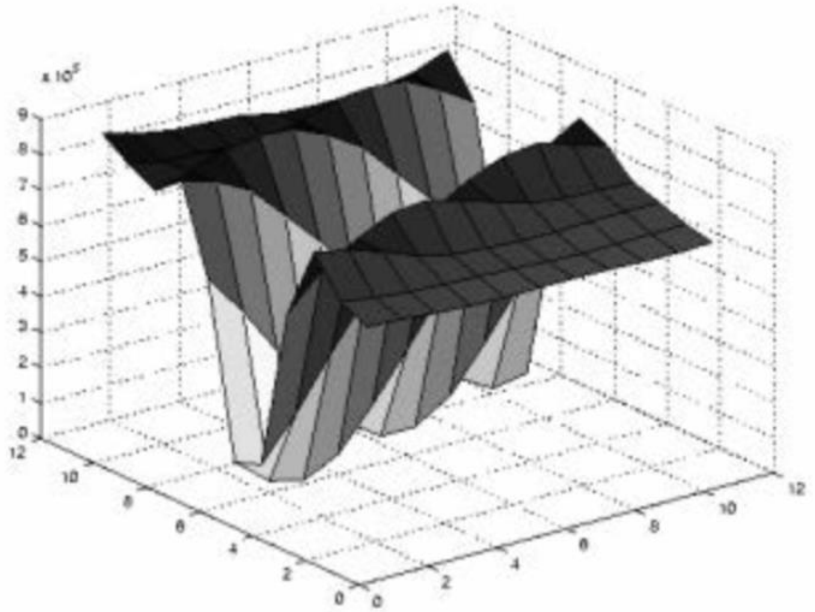
Corners

Lines

Flat regions

Edge

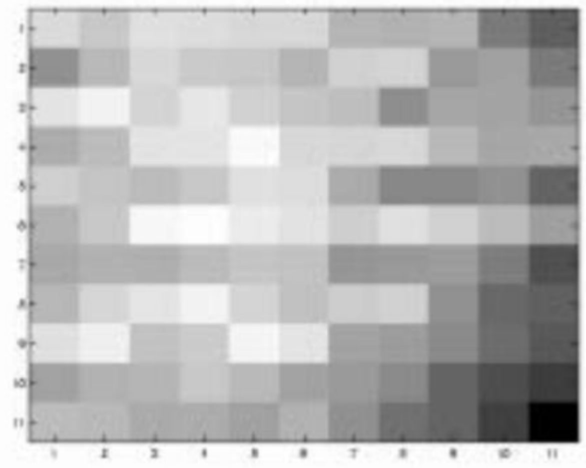


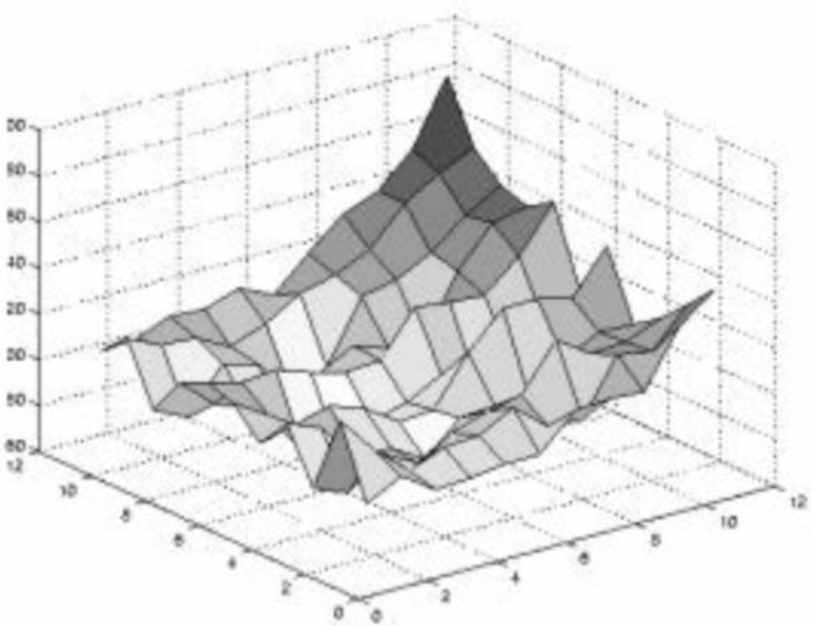


– **gradients very large or very small**
– **large λ_1, small λ_2**

A^TA always becomes singular

Low-texture region



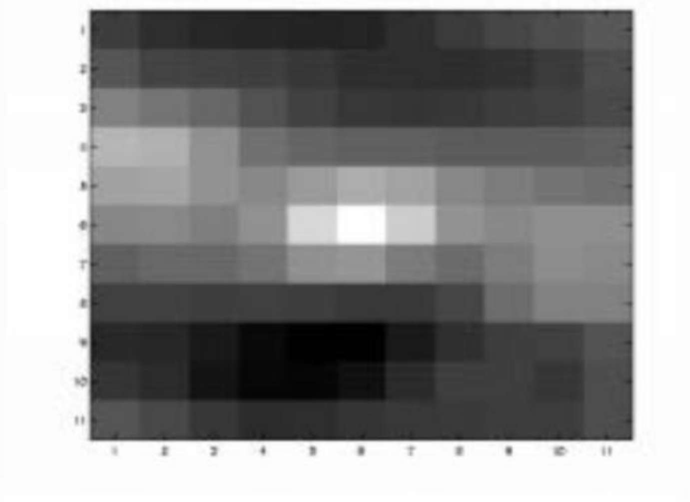


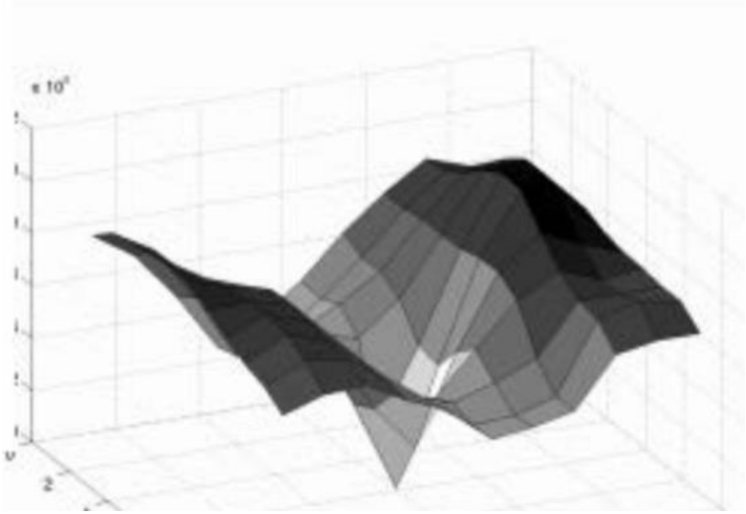
– gradients have small magnitude

– small λ_1, small λ_2

$A^TA \sim 0$

High-texture region





– **gradients are different, large magnitudes**

– **large λ_1, large λ_2**

Computing Optical Flow-from Energy: Horn & Schunk

- **Formulate Error in Optical Flow Constraint:**

$$e_c = \iint_{image} (I_x u + I_y v + I_t)^2 \, dx \, dy$$

- **We need additional constraints!**

- **Smoothness Constraint (as in shape from shading and stereo):**

Usually motion field varies smoothly in the image.
So, penalize departure from smoothness:

$$e_s = \iint_{image} (u_x^2 + u_y^2) + (v_x^2 + v_y^2) \;\; dx \, dy$$

- **Find (u,v) at each image point that MINIMIZES:**

$$e = e_s + \lambda e_c$$

weighting
factor

Errors in Lukas-Kanade

What are the potential causes of errors in this procedure?

- **Suppose ATA is easily invertible**
  Suppose there is not much noise in the image

When our assumptions are violated

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

Improving accuracy

Recall our small motion assumption

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

This is not exact

- To do better, we need to add higher order terms back in:

$$= I(x,y) + I_x u + I_y v + \text{higher order terms} - H(x,y)$$

This is a polynomial root finding problem

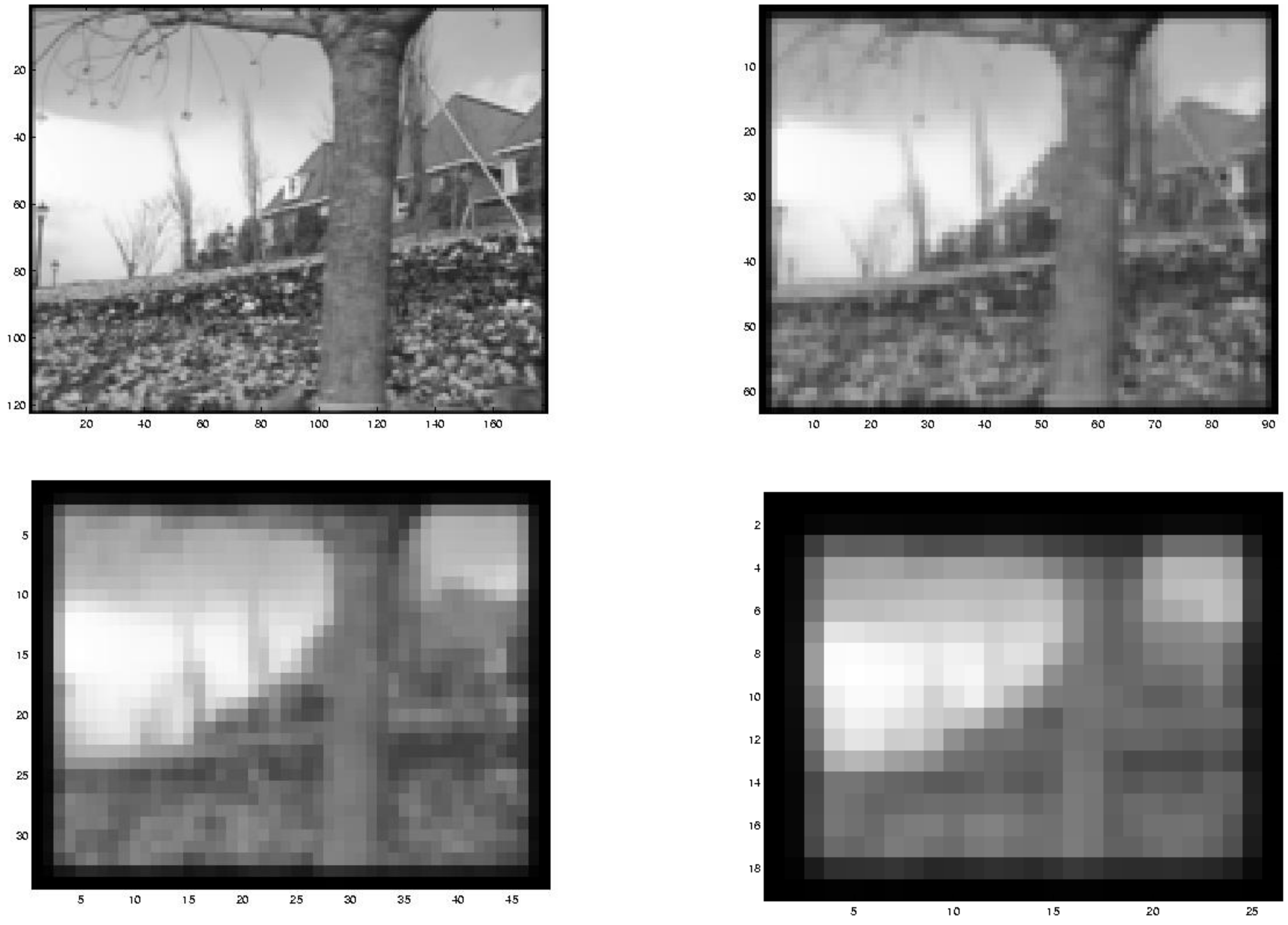
- Can solve using **Newton's method**
  - Also known as **Newton-Raphson** method
  - For more on Newton-Raphson, see (first four pages)
    - » http://www.ulib.org/webRoot/Books/Numerical_Recipes/bookcpdf/c9-4.pdf
- Lucas-Kanade method does one iteration of Newton's method
  - Better results are obtained via more iterations

1D case
on board

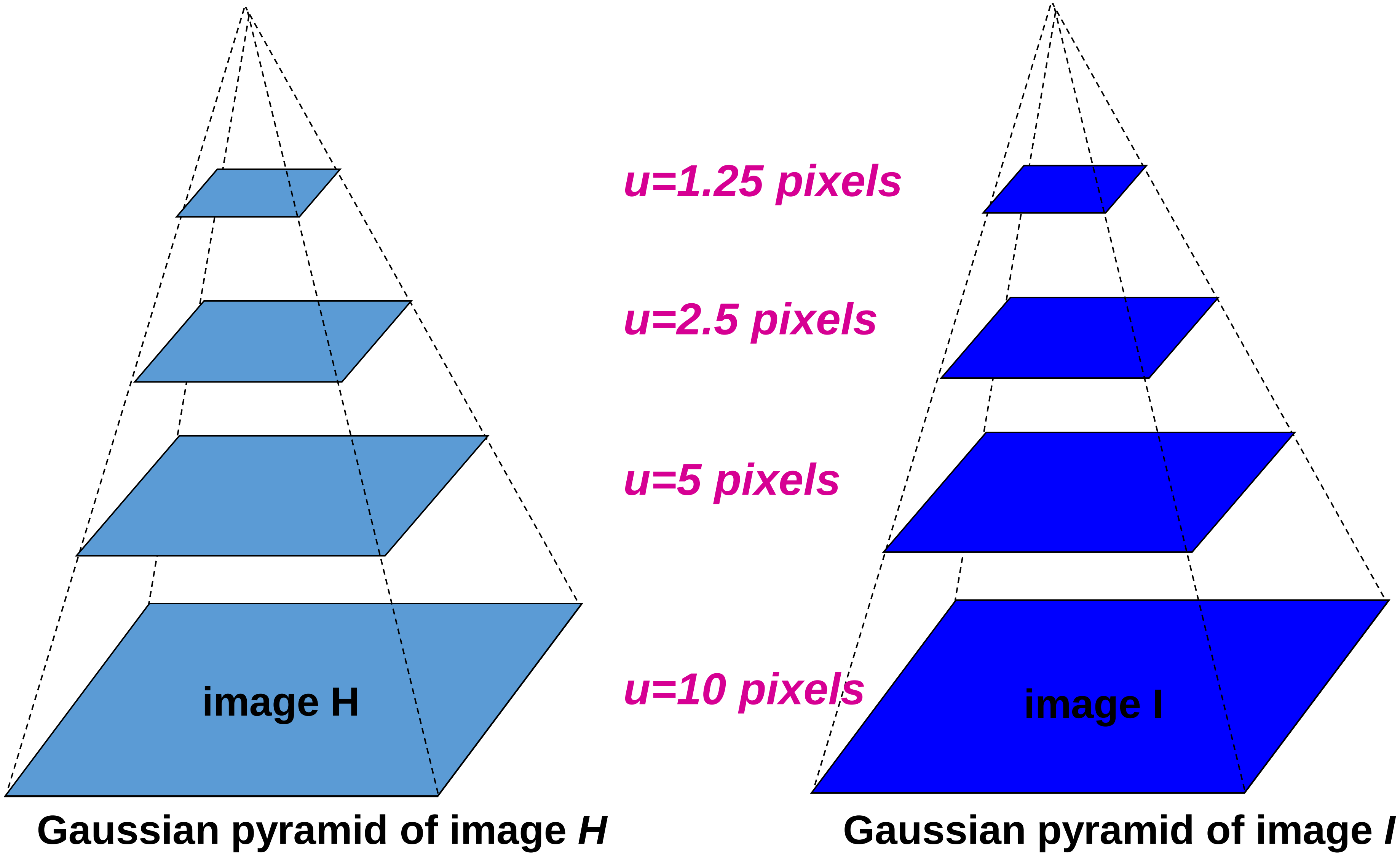
Revisiting the Small Motion Assumption

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

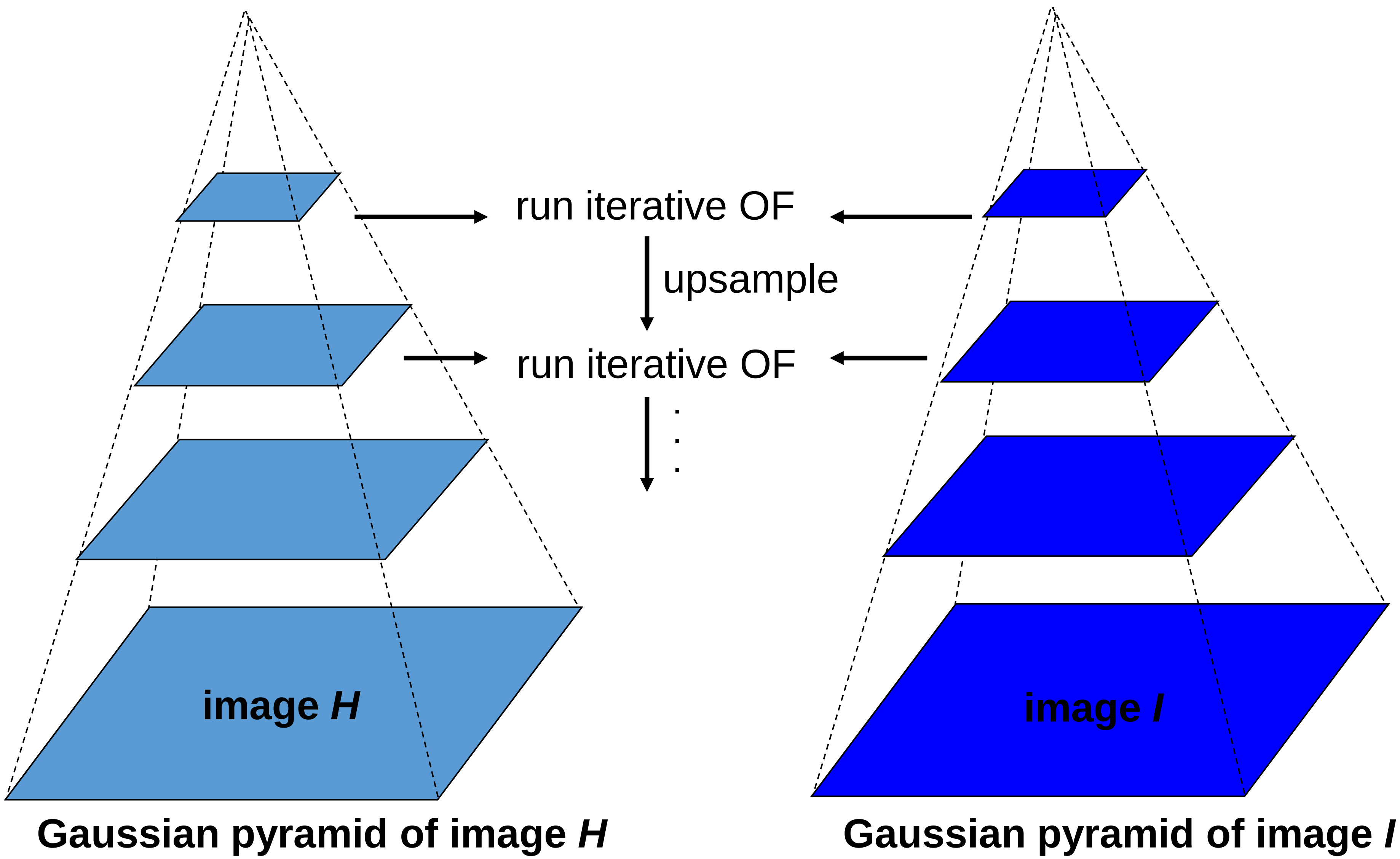
Reduce the Resolution!



Coarse-to-fine Optical Flow Estimation



Coarse-to-fine Optical Flow Estimation

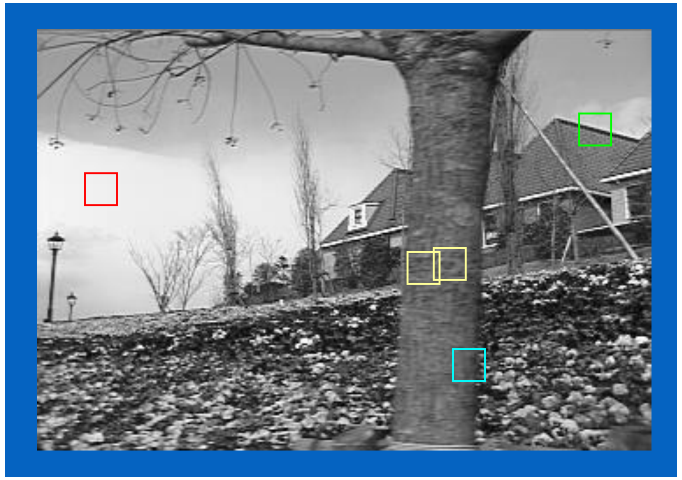


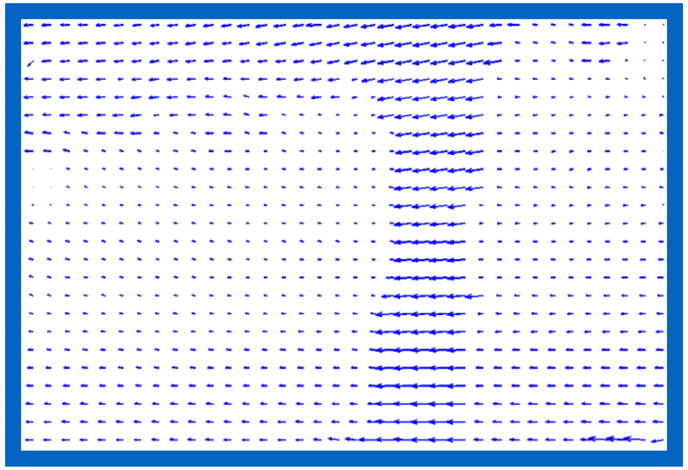
Iterative Lucas-Kanade Algorithm

- Top level
  - Apply L-K to get a flow field representing the flow from the first frame to the second frame.
  - Apply this flow field to warp the first frame toward the second frame.
  - Rerun L-K on the new warped image to get a flow field from it to the second frame. Repeat till convergence
- NextLevel
  - Upsample the flow field to the next level as the first guess of the flow at that level. Apply this flow field to warp the first frame toward the second frame. Rerun L-K and warping till convergence as above
- Etc

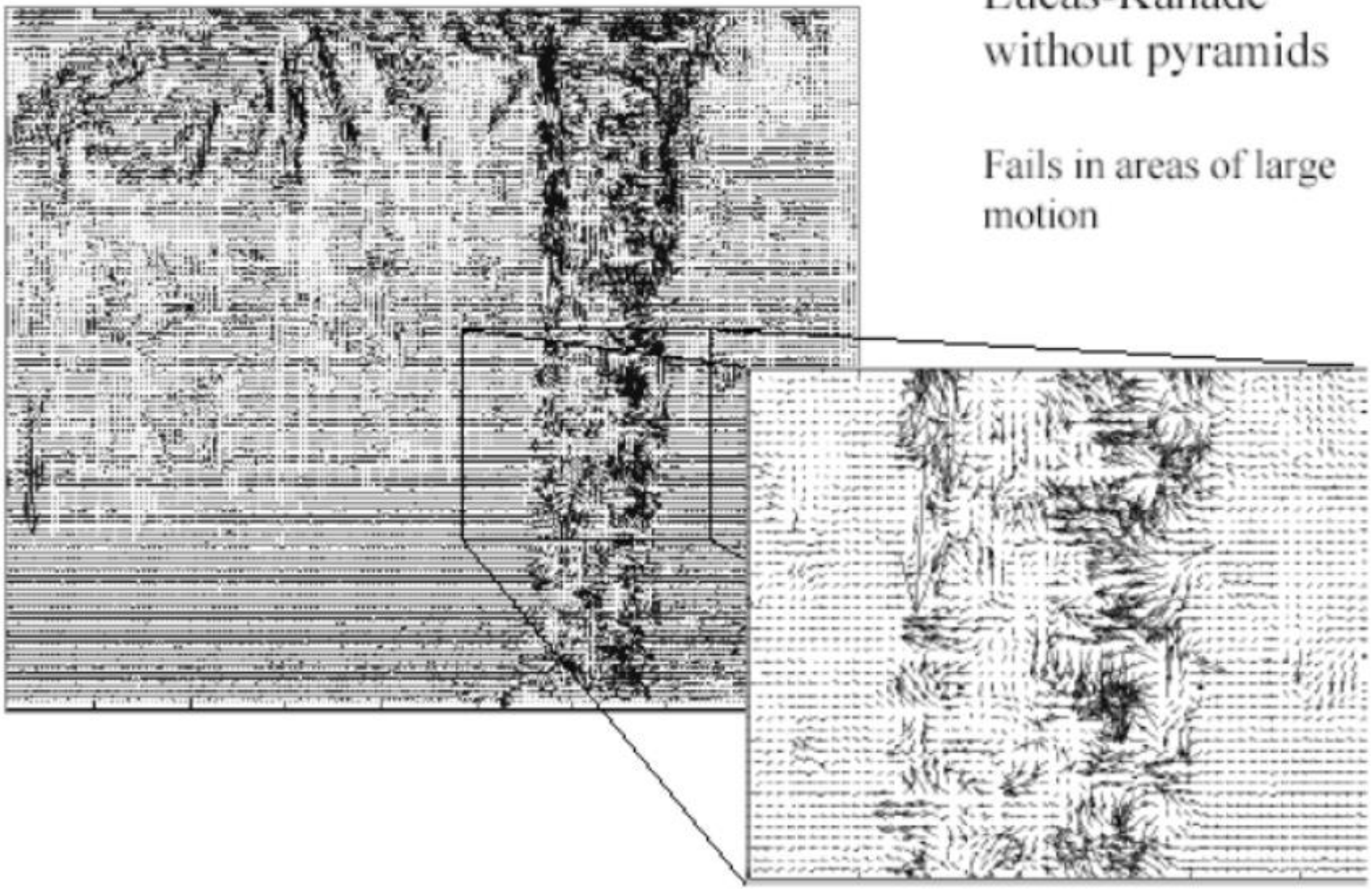
The Flower Garden Video

What should the optical flow be?





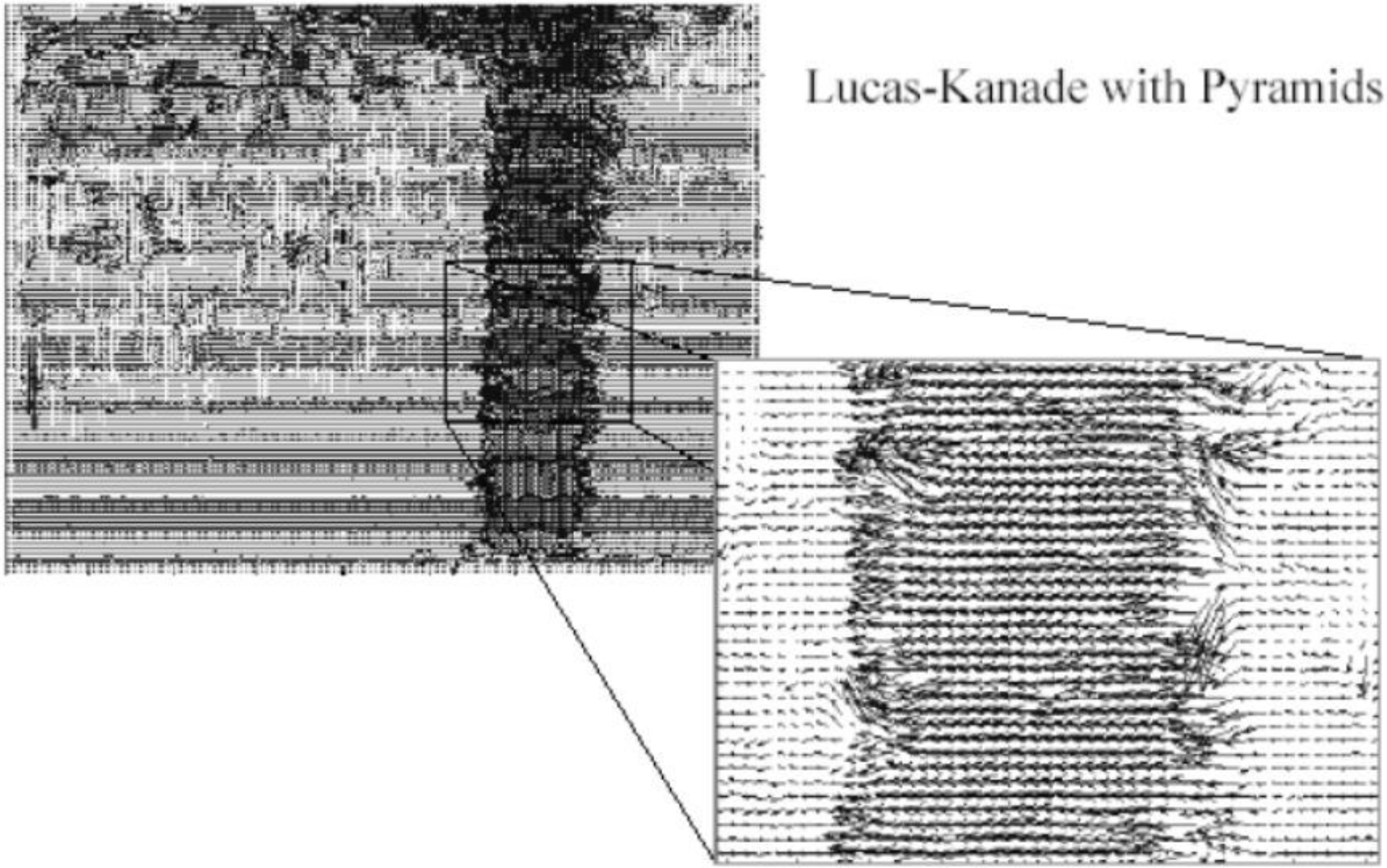
Optical Flow Results



Lucas-Kanade without pyramids

Fails in areas of large motion

Optical Flow Results



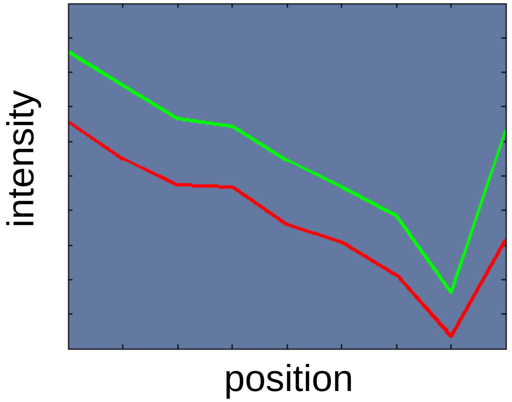
Lucas-Kanade with Pyramids

Brightness is not always constant

Rotating cylinder

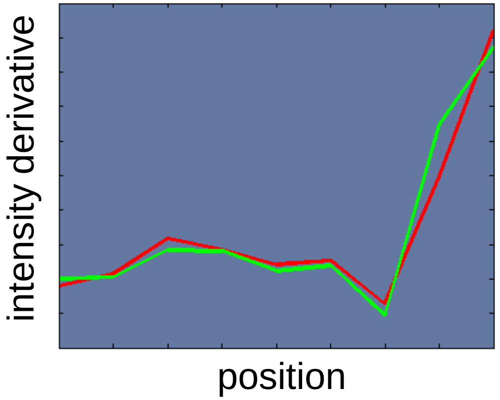
Brightness constancy does not always hold

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



Gradient constancy holds

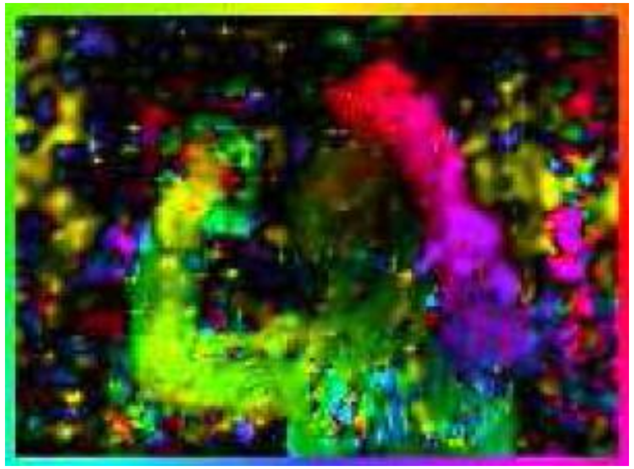
$$\nabla I(x+u, y+v, t+1) = \nabla I(x, y, t)$$



Local constraints work poorly

input video

Optical flow direction using only local constraints



color encodes direction as marked on the boundary

Where local constraints fail

Occlusions **We have not seen where some points moved**

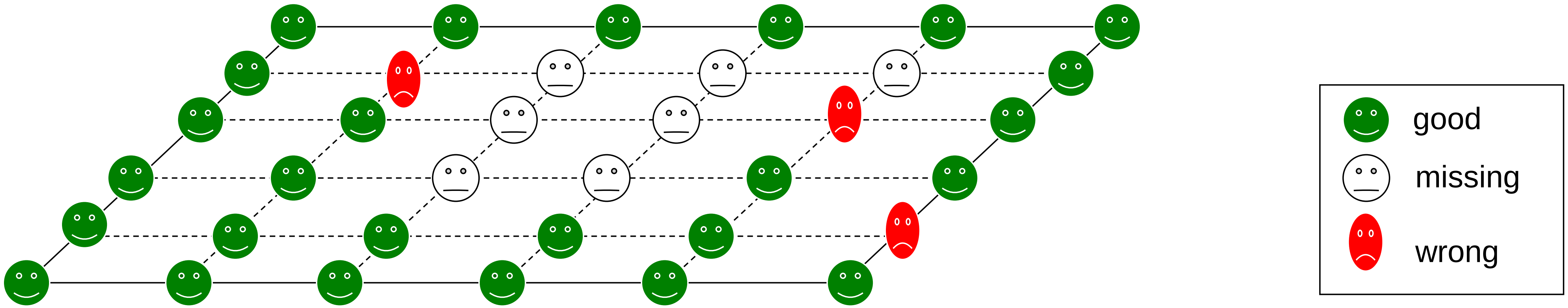
Occluded regions are marked in red

Obtaining support from neighbors

Two main problems with local constraints:

- information about motion is missing in some points
  => need spatial coherency

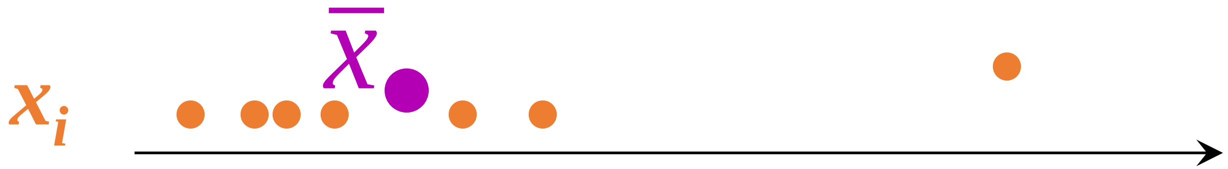
- constraints do not hold everywhere
  => need methods to combine them robustly



Robust combination of partially reliable data

Toy example

Find "best" representative for the set of numbers



L2: $E = \sum_i \left|\bar{x} - x_i\right|^2 \rightarrow \min$

L1: $E = \sum_i \left|\bar{x} - x_i\right| \rightarrow \min$

Influence of x_i on E: $x_i \rightarrow x_i + \Delta$

$$E_{new} \cong E_{old} + 2\left(x_i - \bar{x}\right) \cdot \Delta$$

proportional to $\left|\bar{x} - x_i\right|$

$$E_{new} \cong E_{old} + \Delta$$

equal for all x_i

Outliers influence the most

$\bar{x} = \mathrm{mean}(x_i)$

Majority decides

$\bar{x} = \mathrm{median}(x_i)$

Elections and robust statistics

many ordinary people

a very rich man

wealth

Oligarchy

Votes proportional to the wealth

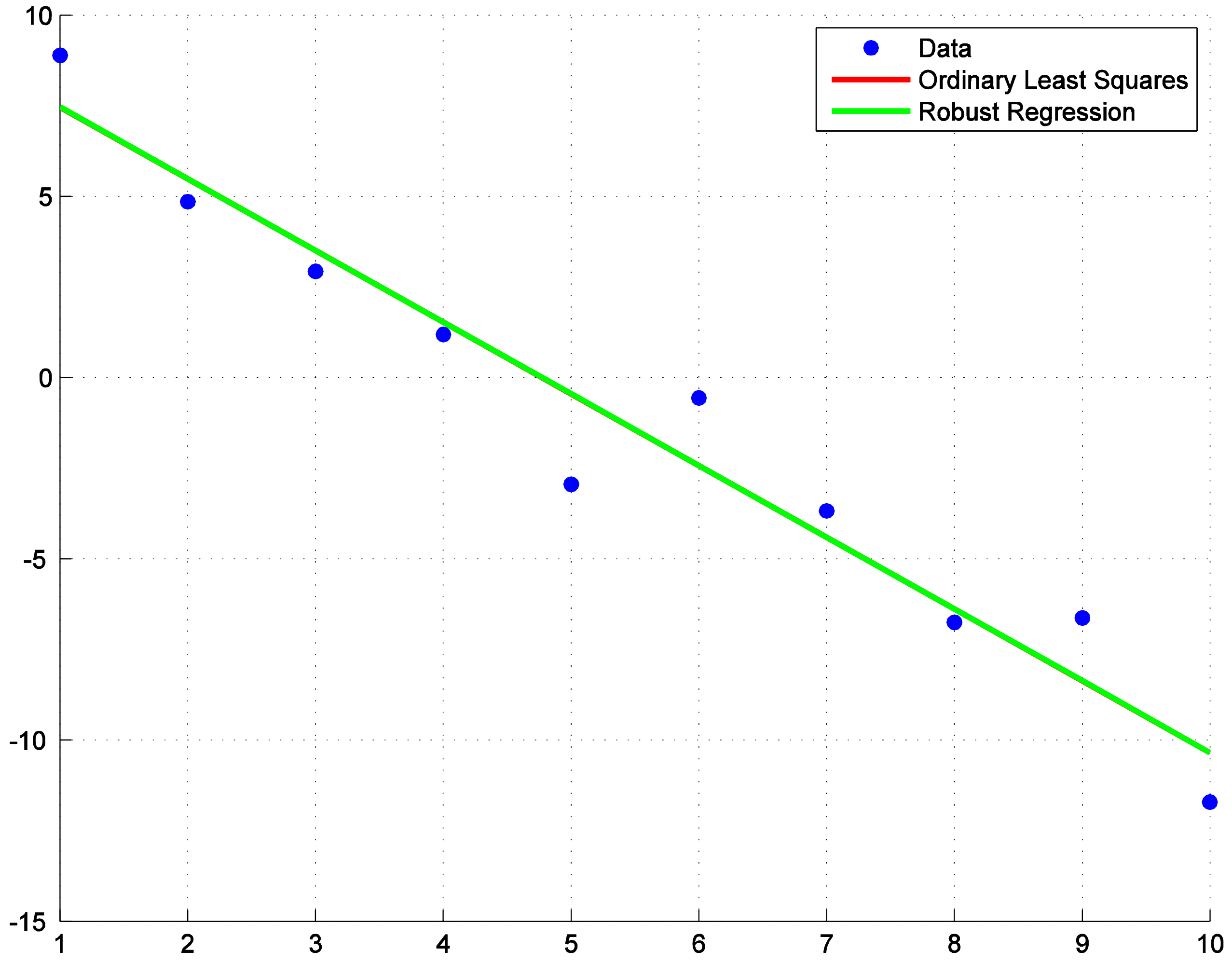
like in L2 norm minimization

Democracy

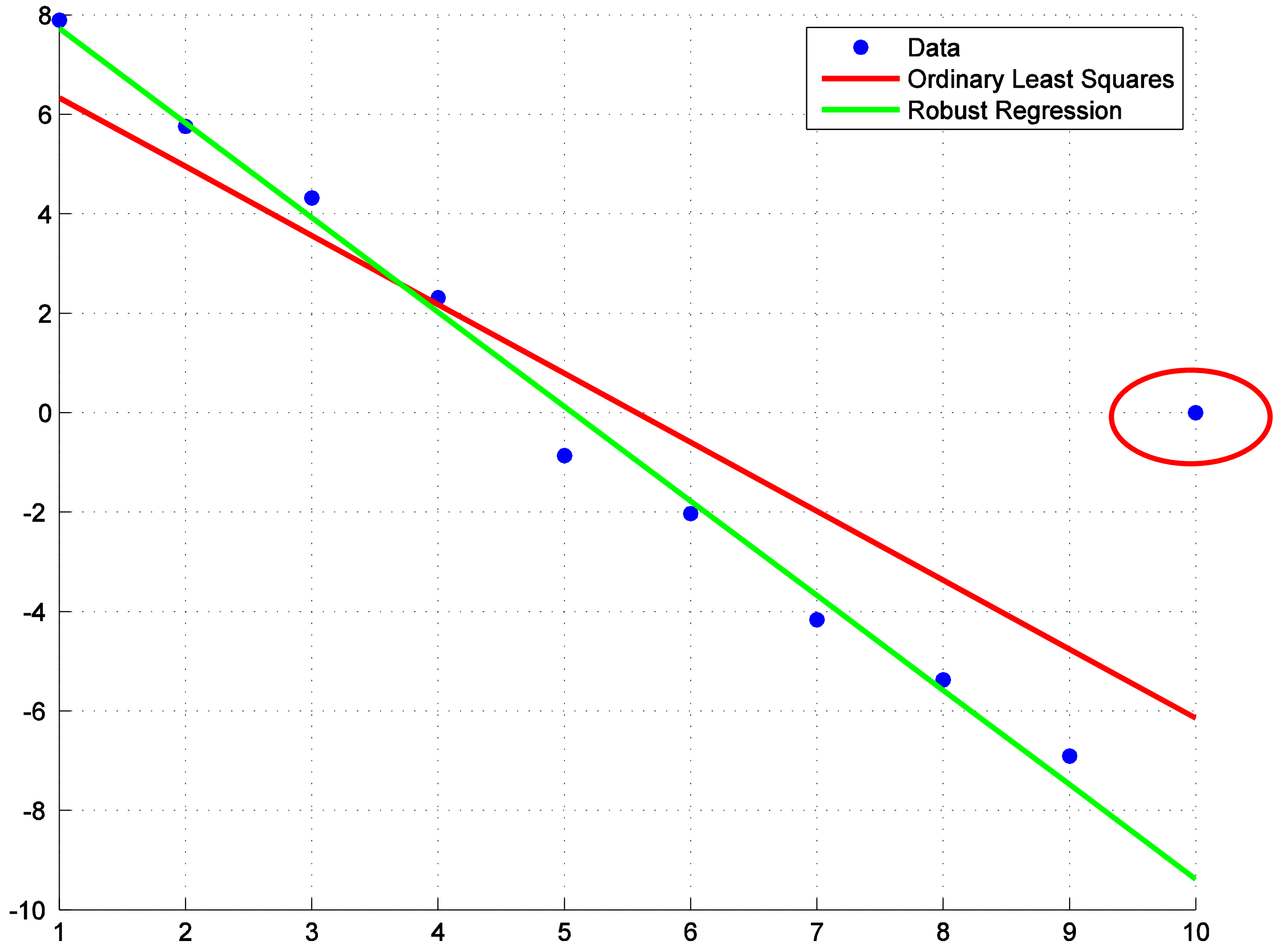
One vote per person

like in L1 norm minimization

A Simple Example



A Simple Example

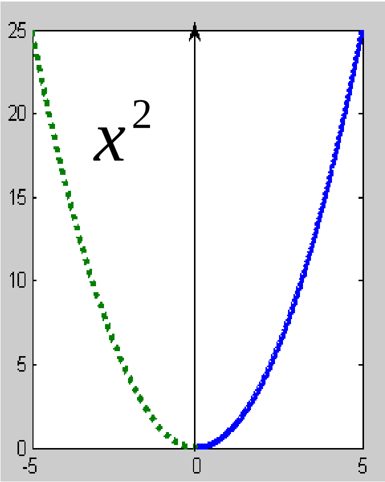


Combination of two flow constraints

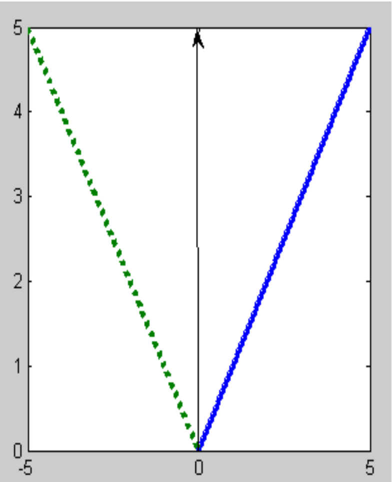
$$\min \int_{video} \phi\left(\left|I_{warped} - I\right|\right) + \alpha\phi\left(\left|\nabla I_{warped} - \nabla I\right|\right)$$

$$I_{warped} = I(x+u, y+v, t+1); \quad I = I(x,y,t)$$

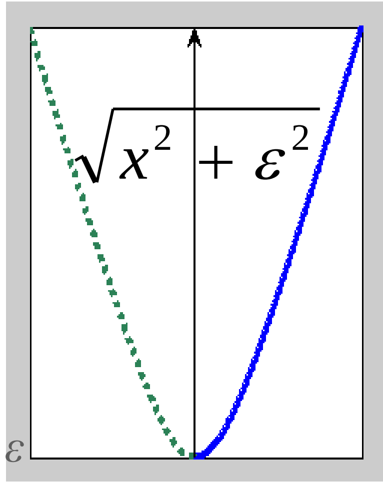
usual: L2



robust: L1



robust regularized



usual: L2	robust: L1	robust regularized
✓ easy to analyze and minimize – sensitive to outliers	✓ robust in presence of outliers – non-smooth: hard to analyze	✓ smooth: easy to analyze ✓ robust in presence of outliers

[A. Bruhn, J. Weickert, 2005]
Towards ultimate motion estimation: Combining highest accuracy with real-time performance

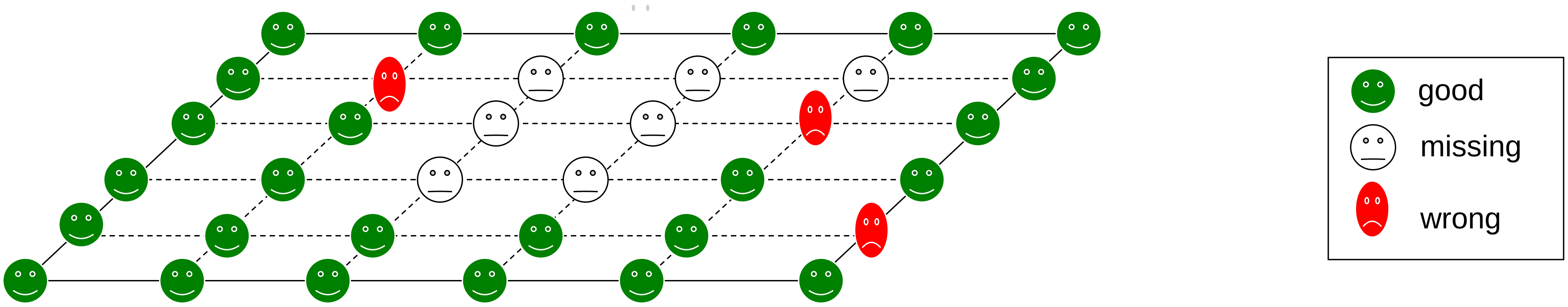
Spatial Propagation

Obtaining support from neighbors

Two main problems with local constraints:

- **information about motion is missing in some points**
  => need spatial coherency

- **constraints do not hold everywhere**
  => need methods to combine them robustly



Homogeneous propagation

$$\min \int_{video} |\nabla u|^2 + |\nabla v|^2$$

$u(x, y, t)$ - flow in the x direction
$v(x, y, t)$ - flow in the y direction
∇ - gradient

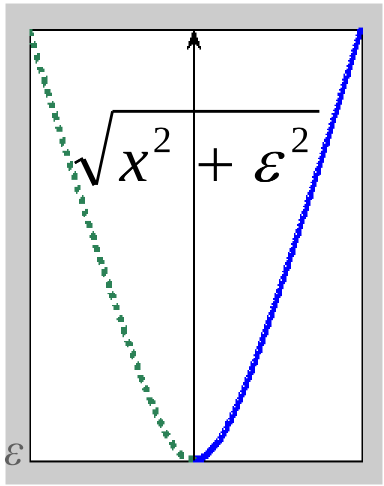
This constraint is not correct on motion boundaries
=> over-smoothing of the resulting flow

[Horn&Schunck 1981]

Robustness to flow discontinuities

$$\min \int_{video} \phi\left(\sqrt{|\nabla u|^2 + |\nabla v|^2}\right)$$

ϕ :



(also known as isotropic flow-driven regularization)

[T. Brox, A. Bruhn, N. Papenberg, J. Weickert, 2004]
High accuracy optical flow estimation based on a theory for warping

Combining ingredients

Local constraints

–Brightness constancy

–Image gradient constancy

Spatial coherency

–Homogeneous

–Flow-driven

$$\text{Energy} = \int \phi(\text{Data}) + \int \phi(\text{“Smoothness”})$$

Combined using robust statistics ϕ

Computed coarse-to-fine

Use several frames

The more ingredients - the better

brightness constancy

gradient constancy

spatial coherence

$$\int_{video} \phi\left(\left|I_{warped} - I\right|\right) + \alpha\phi\left(\left|\nabla I_{warped} - \nabla I\right|\right) + \beta \int_{video} \phi\left(\left\|\nabla \text{flow}\right\|\right)$$

[Bruhn, Weickert, 2005]
Towards ultimate motion estimation: Combining highest accuracy with real-time performance

How to minimize energy

$$\text{minimize} \quad E(u) = \int F(x, u, u')\, dx$$

Analogy:

$$\text{minimize} \quad f(x)$$

Necessary condition

$$f'(x) = 0$$

Necessary condition:

$$\frac{\partial F}{\partial u} - \frac{d}{dx}\frac{\partial F}{\partial u'} = 0$$

Euler-Lagrange equation

Recent & interesting papers

- Large displacement optical flow with nearest neighbor field, cvpr 2013
- Optical Flow Fields: Dense Correspondence Fields for Highly Accurate Large Displacement Optical Flow Estimation, cvpr 2017

Summary

- Major contributions from Lucas, Tomasi, Kanade
  - Tracking feature points
  - Optical flow
  - Stereo
  - Structure from motion

- Key ideas
  - By assuming brightness constancy, truncated Taylor expansion leads to simple and fast patch matching across frames
  - Coarse-to-fine registration

Thanks

- More information can be found in
- http://vision.middlebury.edu/flow/eval

Advanced topics

- Particles: combining features and flow
  - Peter Sand et al.
  - http://rvsn.csail.mit.edu/pv/

- State-of-the-art feature tracking/SLAM
  - Georg Klein et al.
  - http://www.robots.ox.ac.uk/~gk/