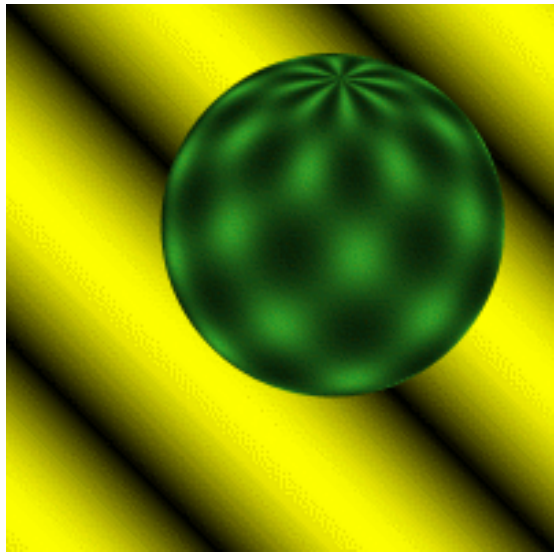
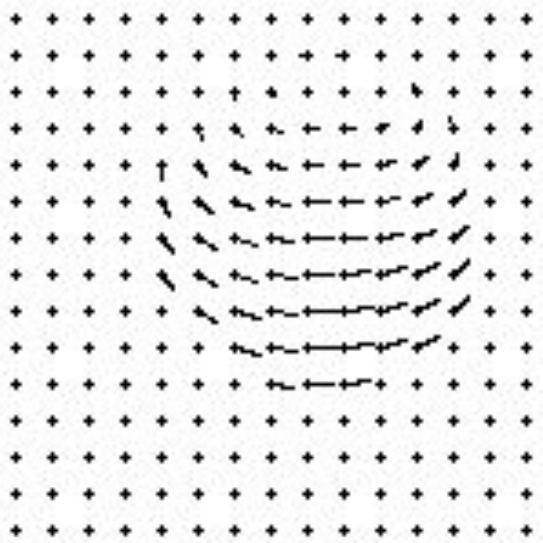
Computer Vision
- Motion & Optical flow

Junjie Cao @ DLUT

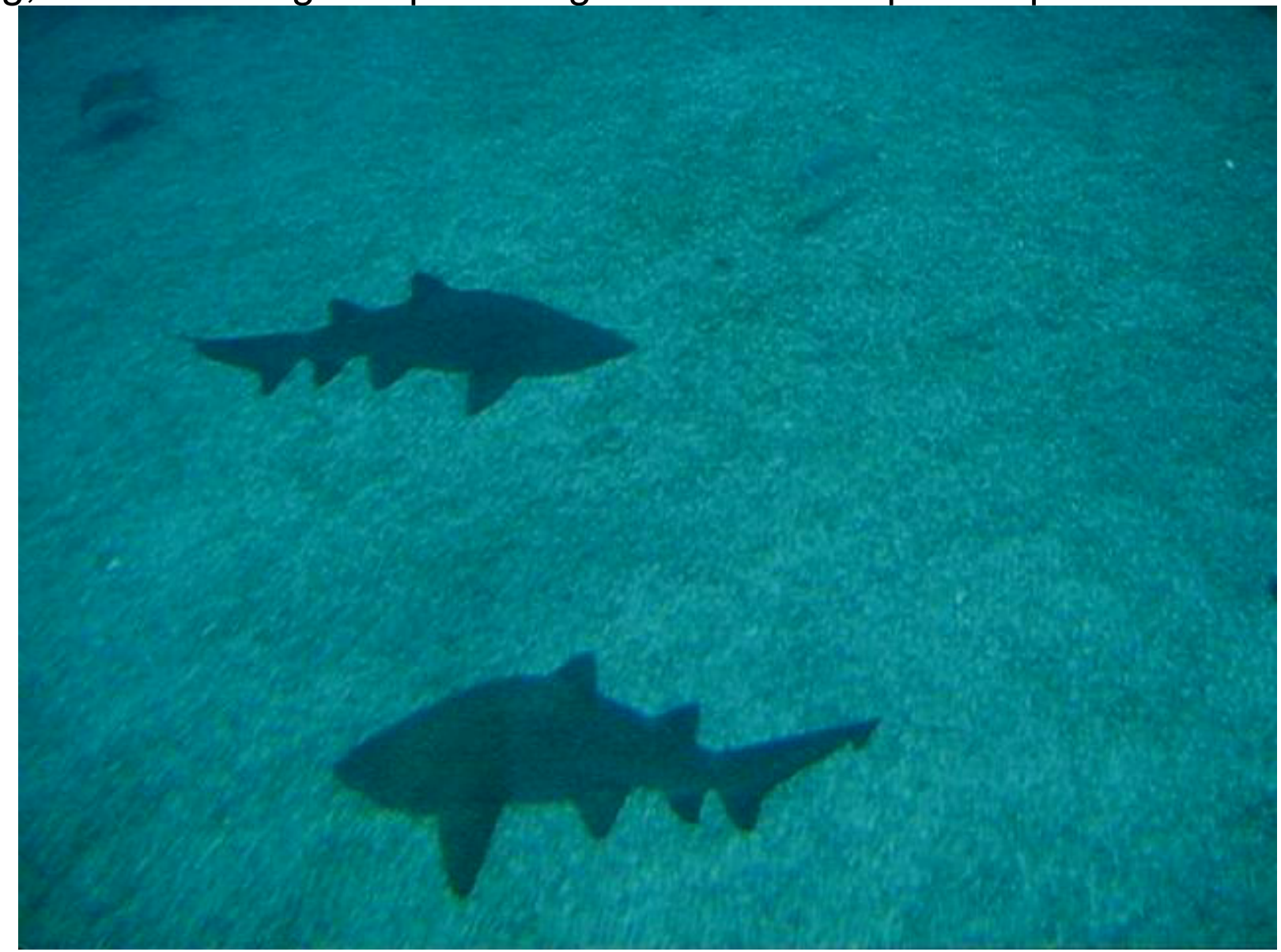
Spring 2019





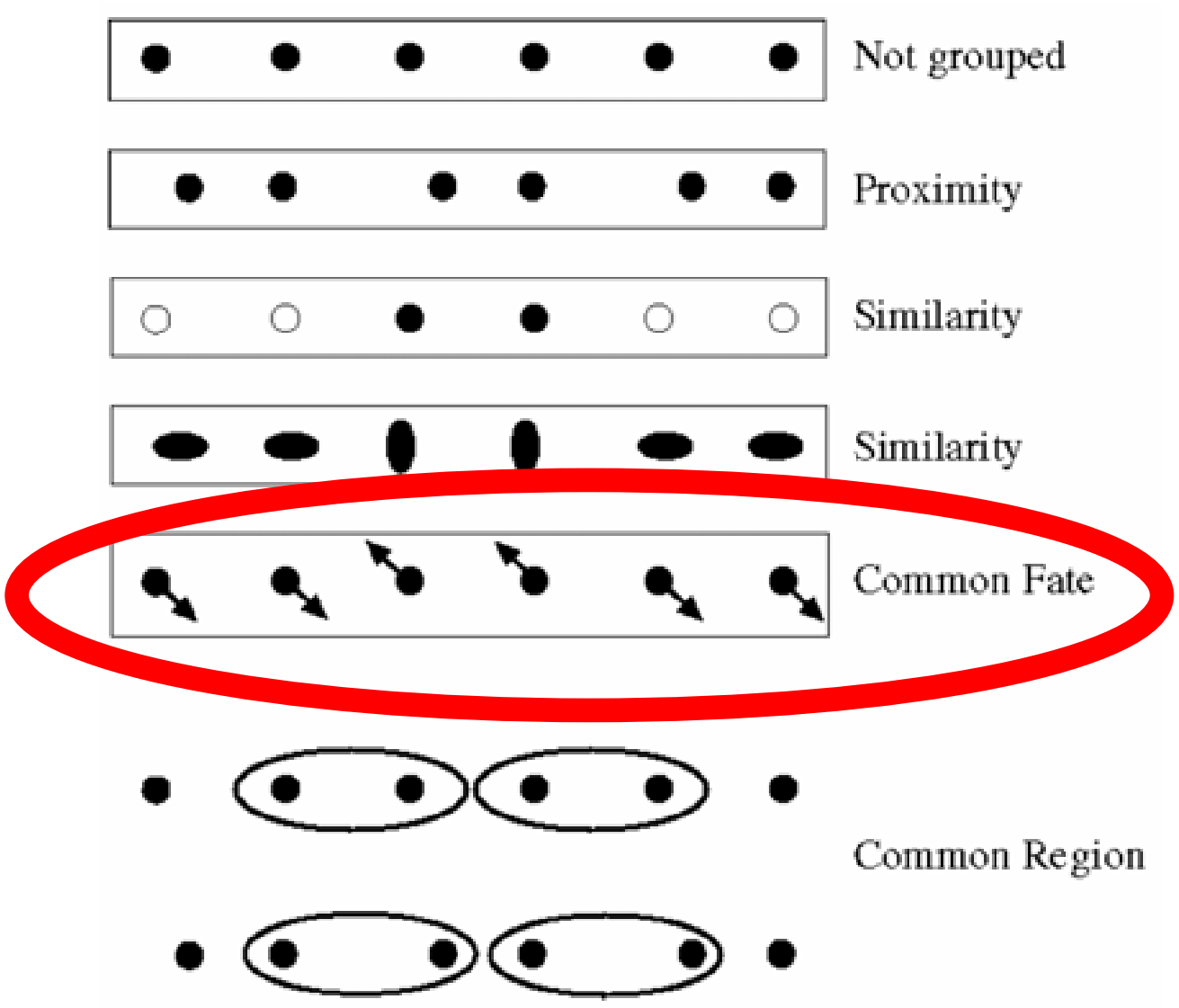
We live in a moving world

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



-- from Linda Shapiro

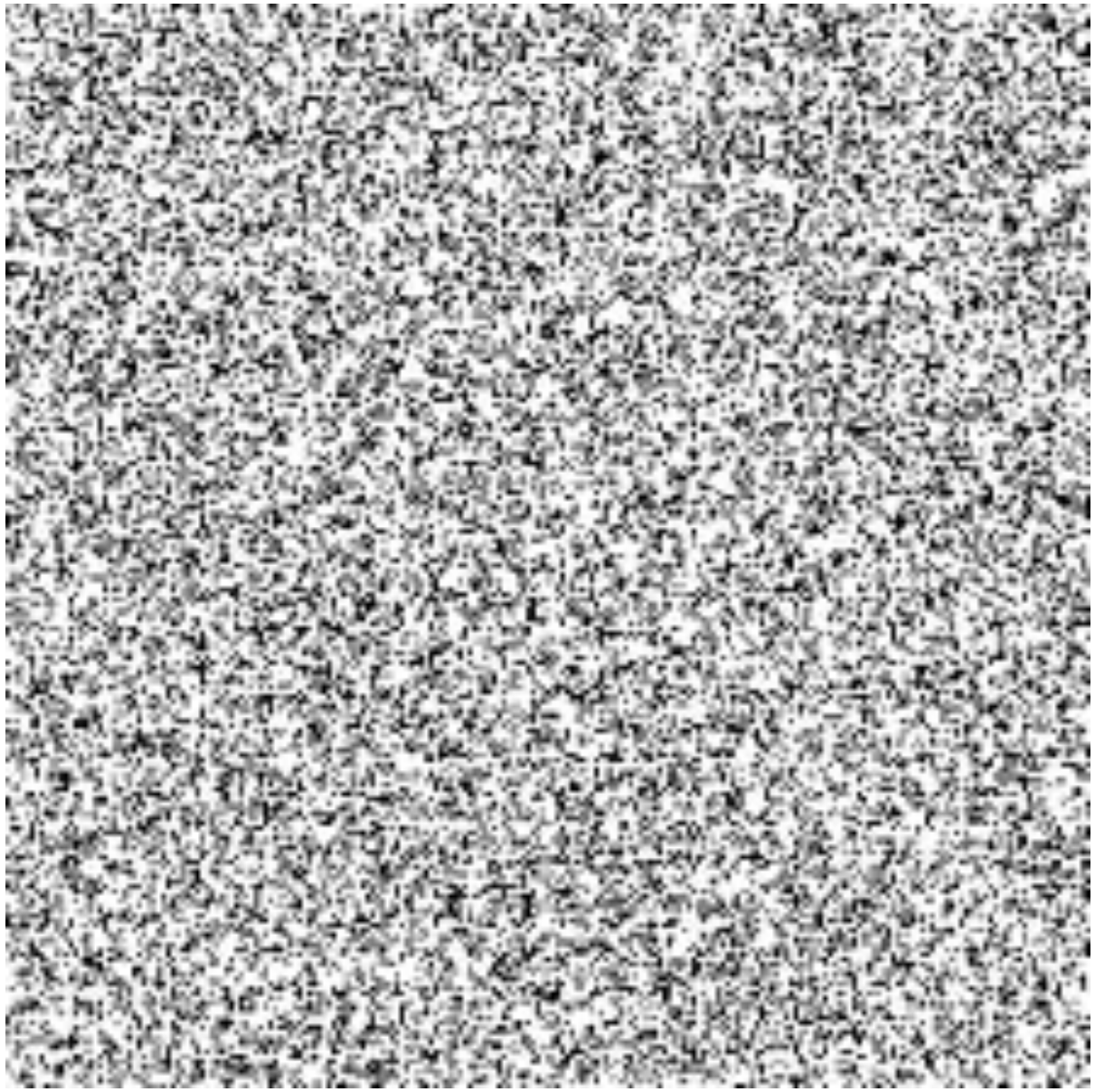
Motion and perceptual organization



Gestalt psychology (Max Wertheimer, 1880-1943)

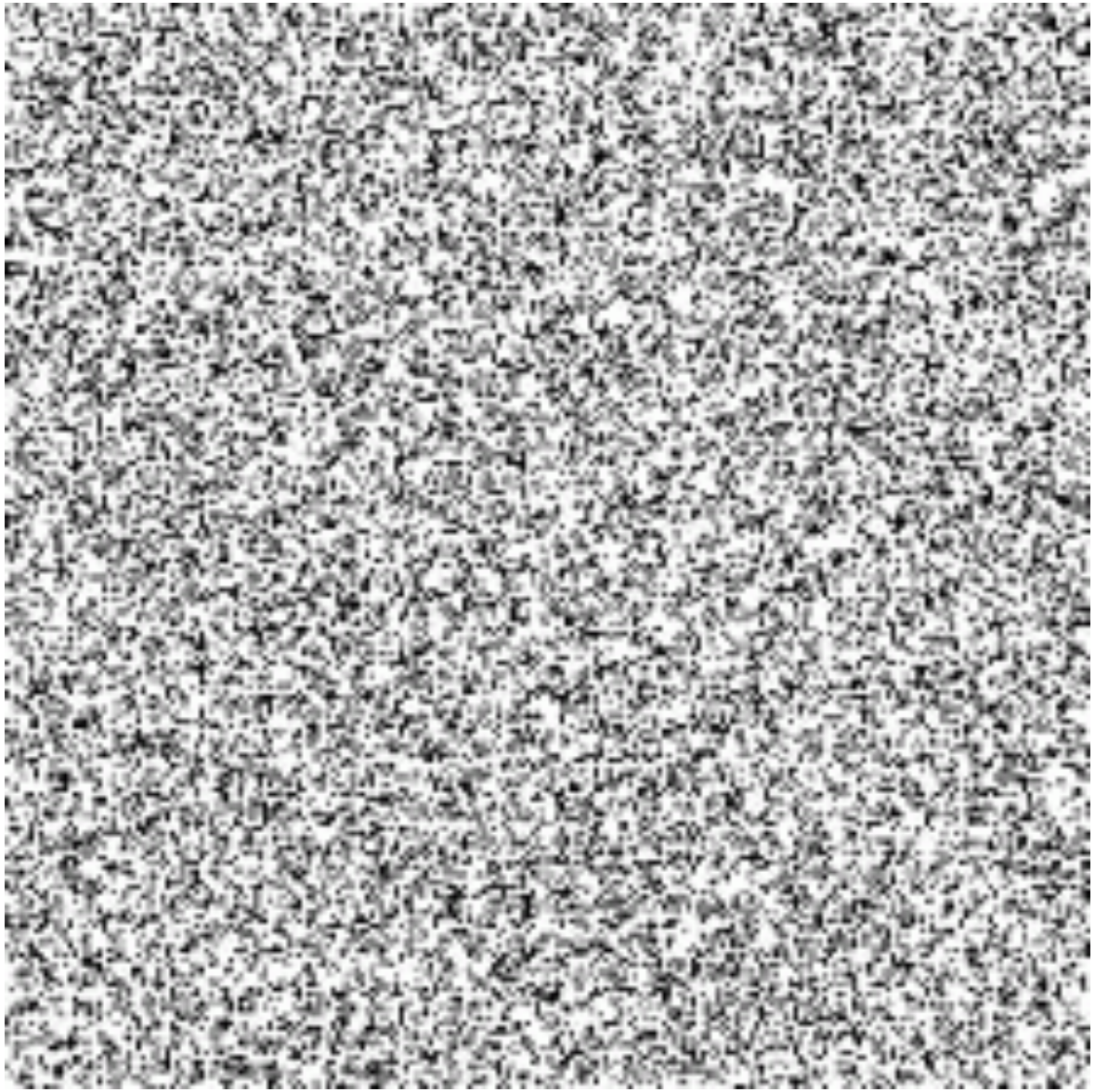
…plus closure, continuation, ‘good form’

Motion and perceptual organization



Motion and perceptual organization

- Sometimes motion is the only cue…



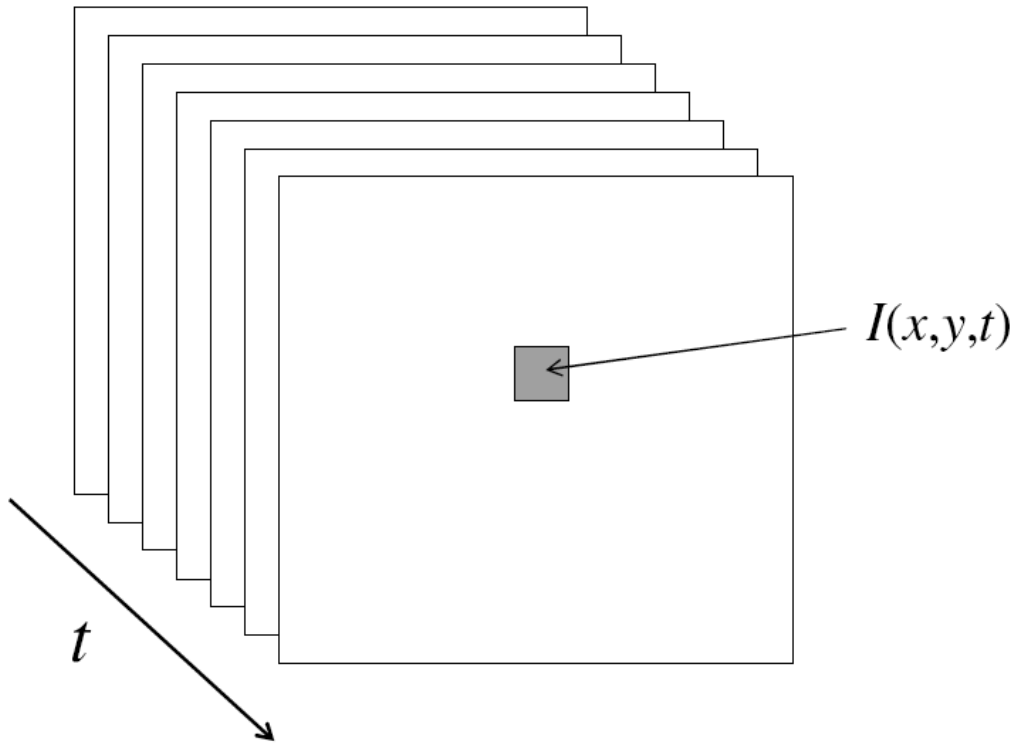
Motion and perceptual organization

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

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

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
  - **Object**
  - **Camera**
  - **Light**

- 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

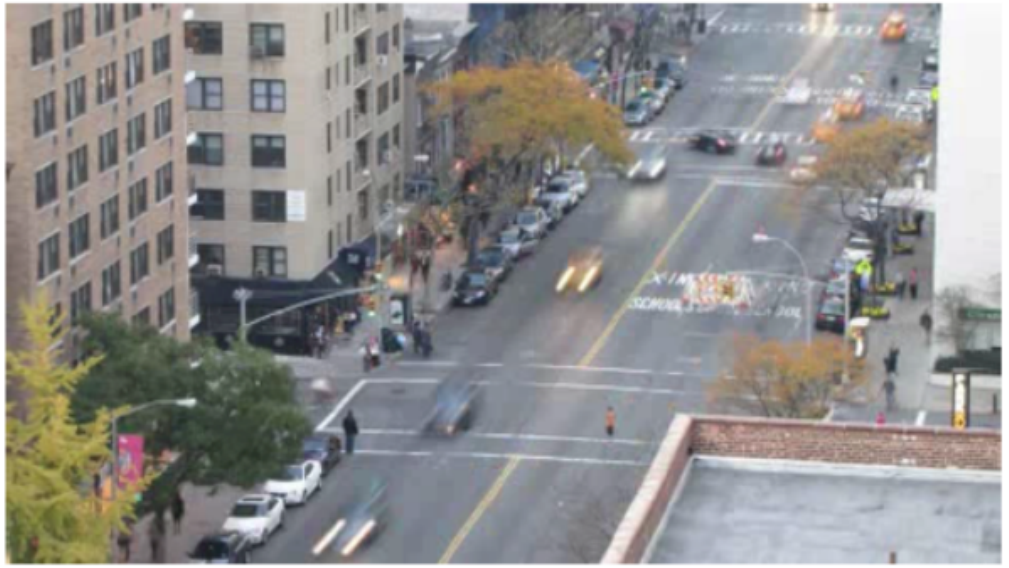
Motion scenarios (priors)

-- from Linda Shapiro

Static camera, moving objects (surveillance)

Moving camera, static scene (3D capture)

Moving camera, moving scene (sports, movie)



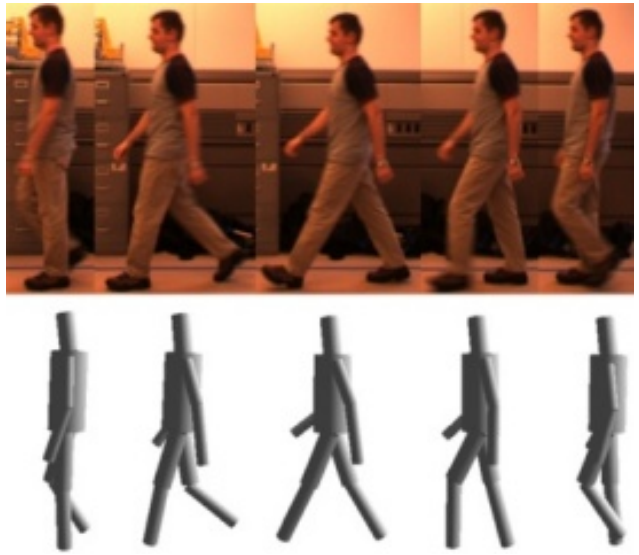
Static camera, moving objects, moving light (time lapse)

We still don’t touch these areas

-- from Linda Shapiro

Uses of motion

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



How can we recover motion?

-- from Linda Shapiro

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 methods – optical flow**
  - **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**

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.)**

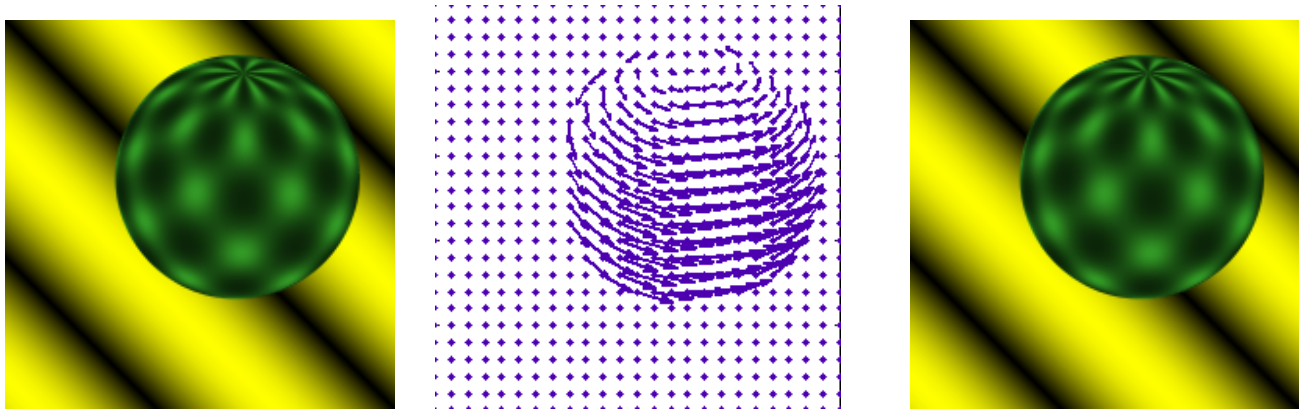
- **Drift: small errors can accumulate as appearance model is updated**

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

-- from Linda Shapiro

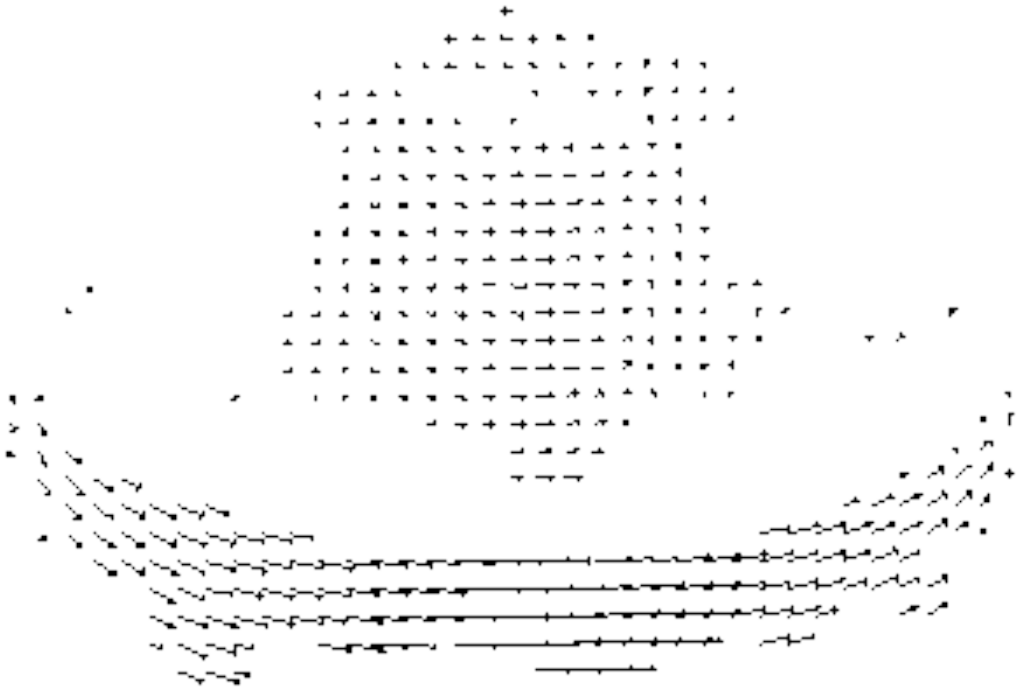
What is Optical Flow?

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



Motion field

- The **motion field** is the projection of the 3D scene motion into the image
- *Optic flow* is the **apparent** motion of objects or surfaces
- We try to discover motion via optic flow, but they are not the same. Why?



Optical flow (Apparent motion) != motion field

- **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**

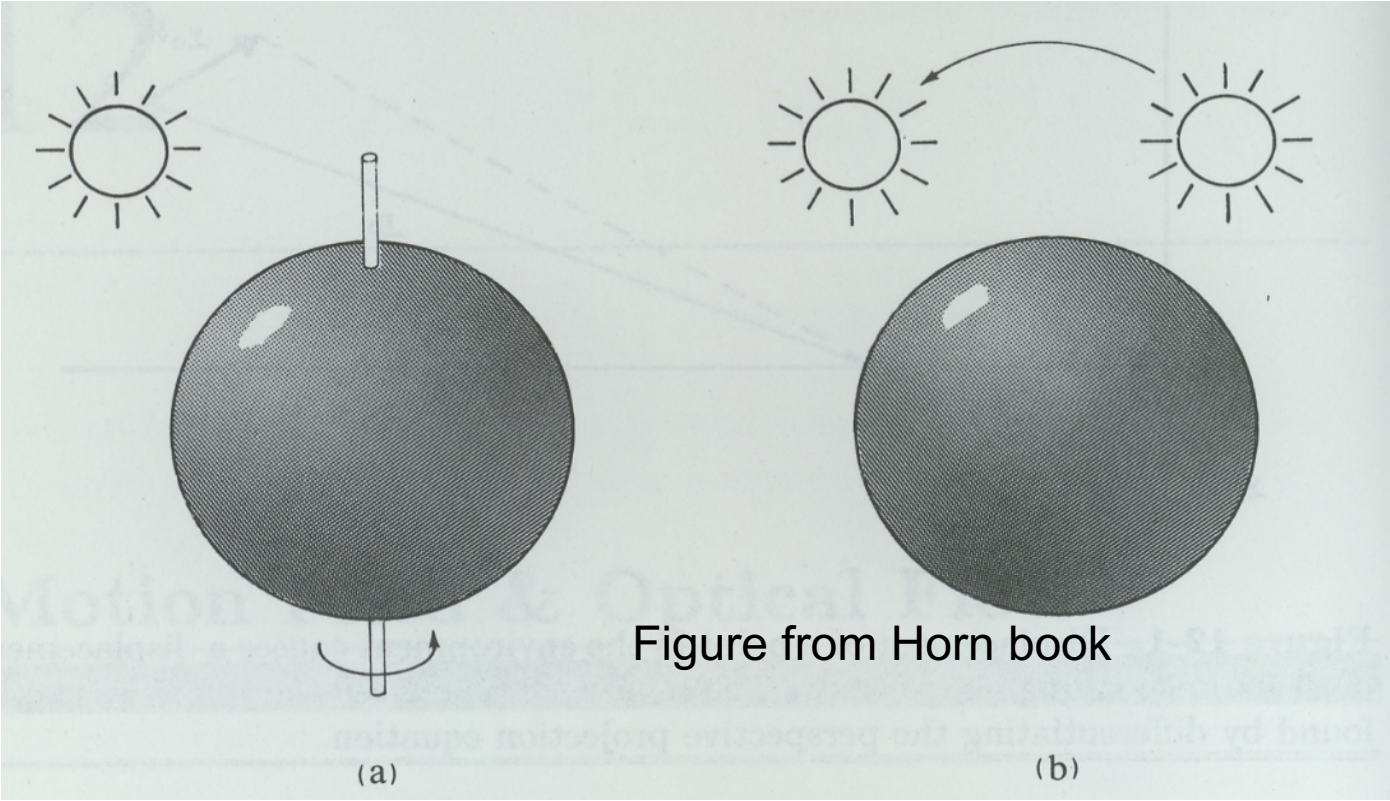
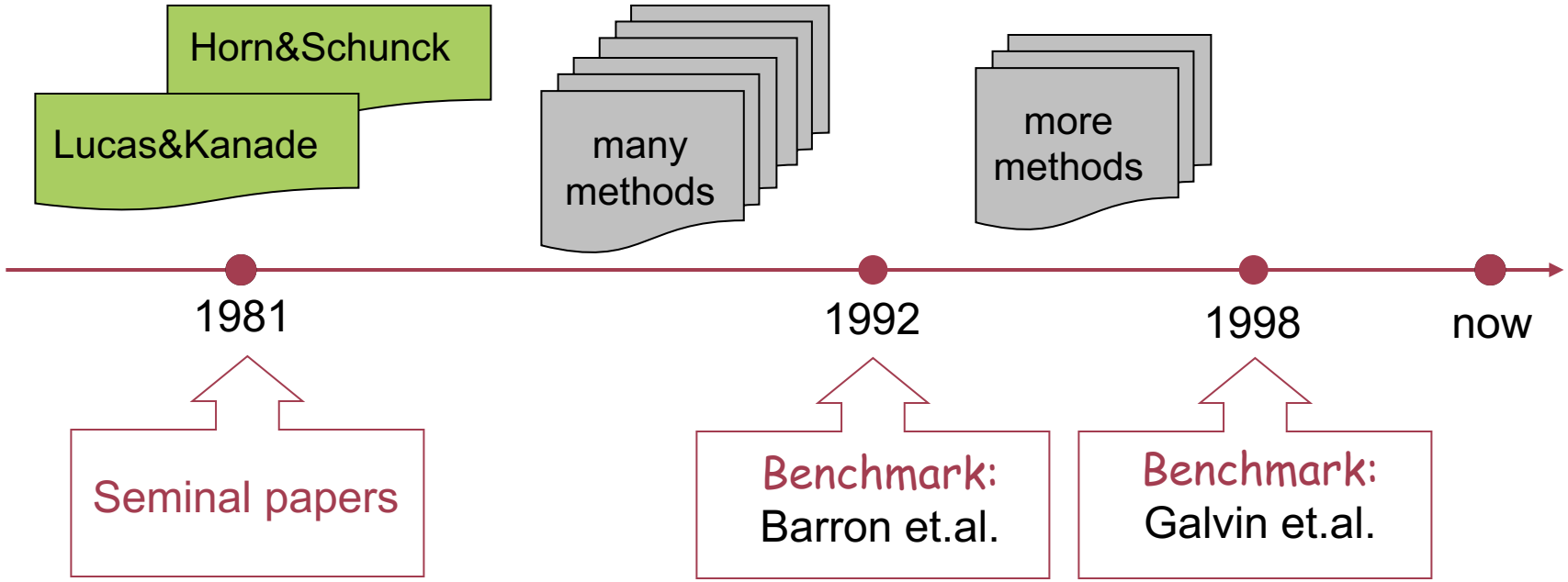


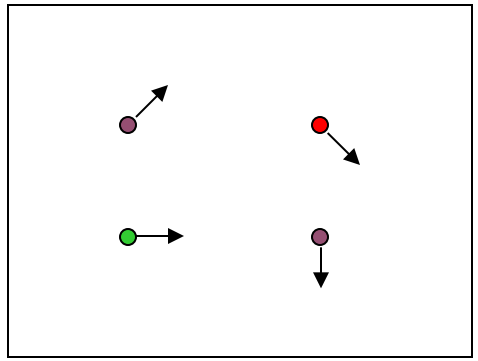
Figure from Horn book

Optical Flow Research: Timeline

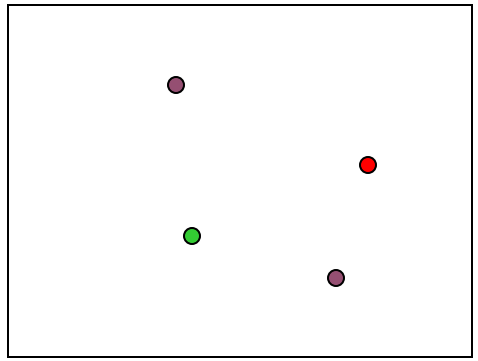


A slow and not very consistent improvement in results, but a lot of useful ingredients were developed

Problem definition: optical flow



$I(x, y, t)$



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

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

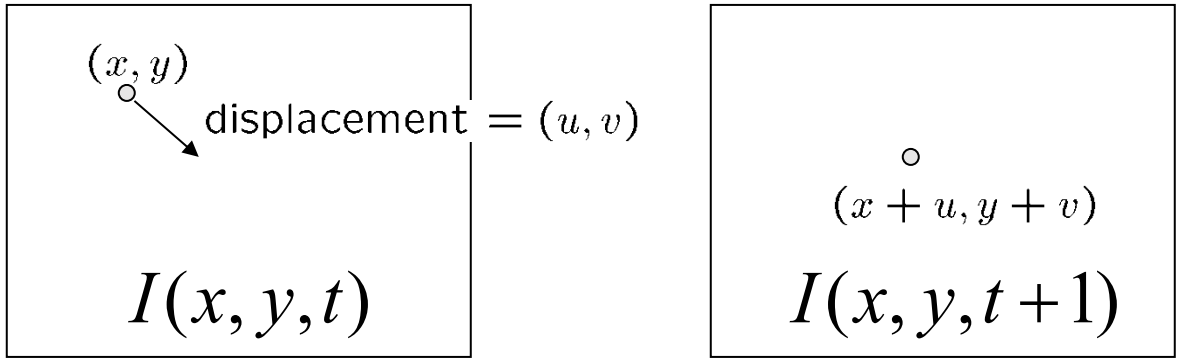
Solve pixel correspondence problem

– 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

Brightness constancy constraint (grayscale images)



- Brightness Constancy Equation:

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

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

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

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

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

Brightness constancy constraint (for gray image)

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

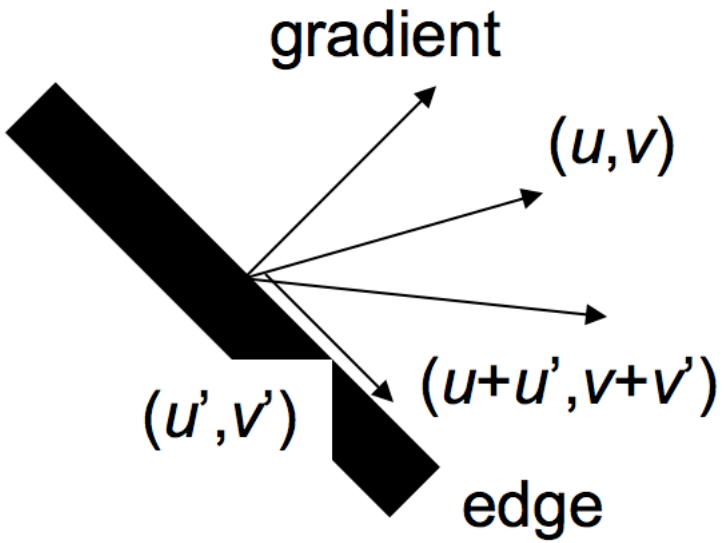
$$\nabla I \cdot \begin{bmatrix} u & v \end{bmatrix}^T + I_t = 0$$

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

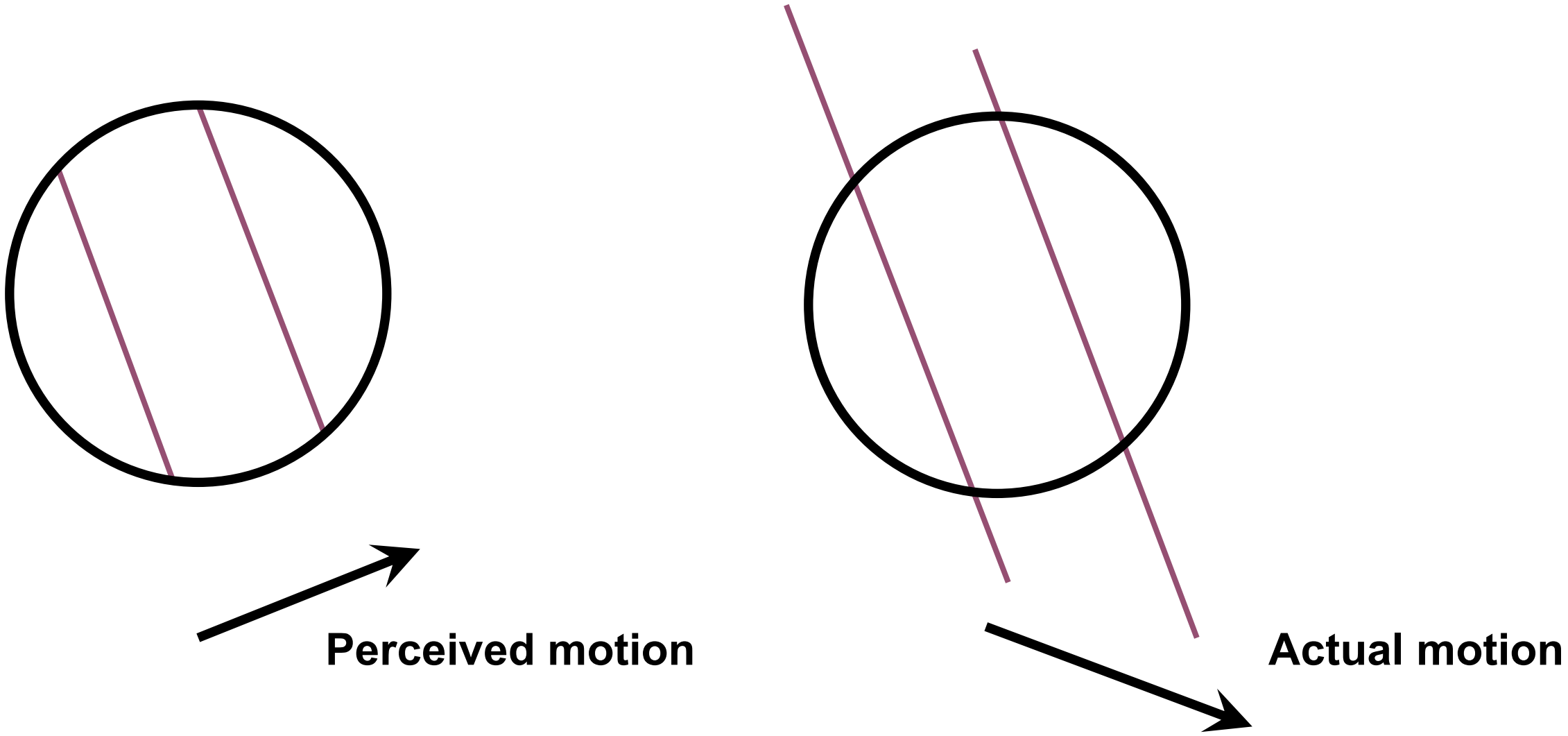
- The component of the motion perpendicular to the gradient (i.e., parallel to the edge) cannot be measured

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

$$\nabla I \cdot \begin{bmatrix} u' & v' \end{bmatrix}^T = 0$$



The aperture problem 孔径问题

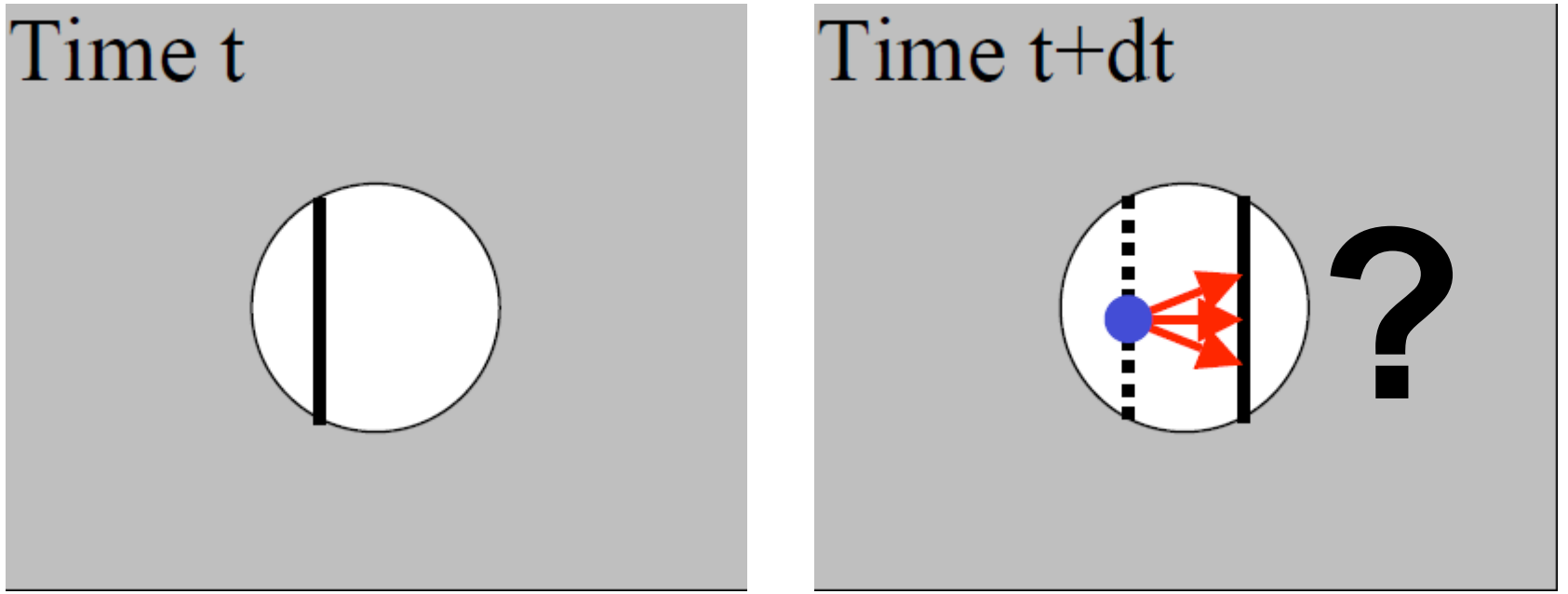


The component of the flow perpendicular to the gradient (i.e., parallel to the edge) is unknown

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 ambiguity… - Lucas-Kanade flow

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

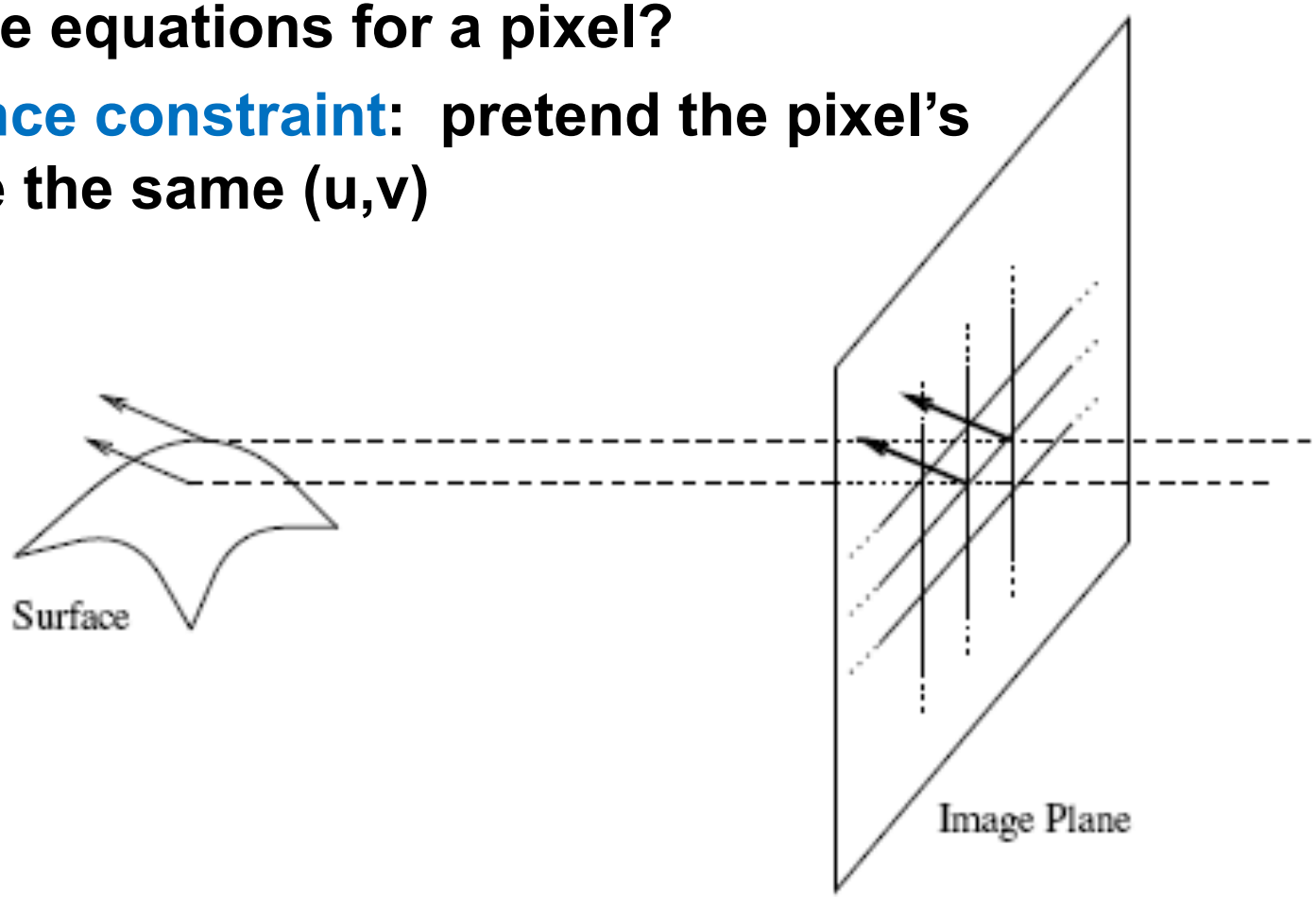


Figure 1.7: Spatial coherence assumption. Neighboring points in the image are assumed to belong to the same surface in the scene.

Figure by Michael Black

Solving the ambiguity…

- **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**

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

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

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

RGB version

- 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!

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

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

A (75x2) $\quad d$ (2x1) $\quad b$ (75x1)

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:**

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

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

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

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.

Conditions for solvability

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

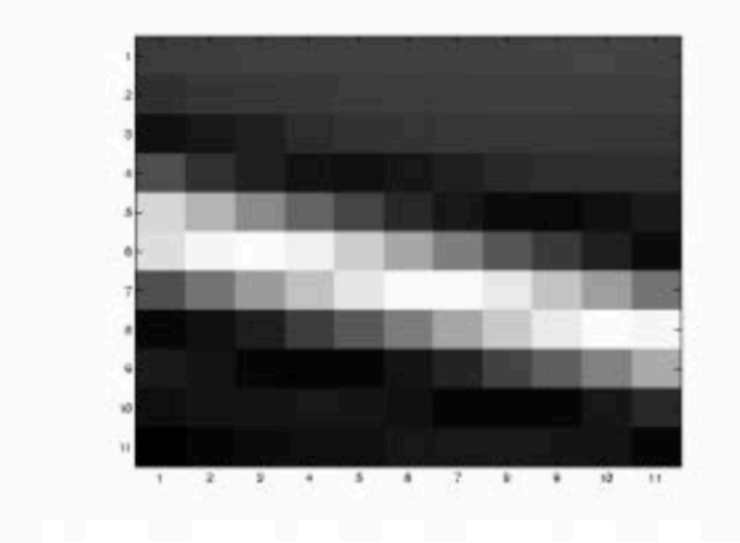
$$A^T A \qquad\qquad A^T b$$

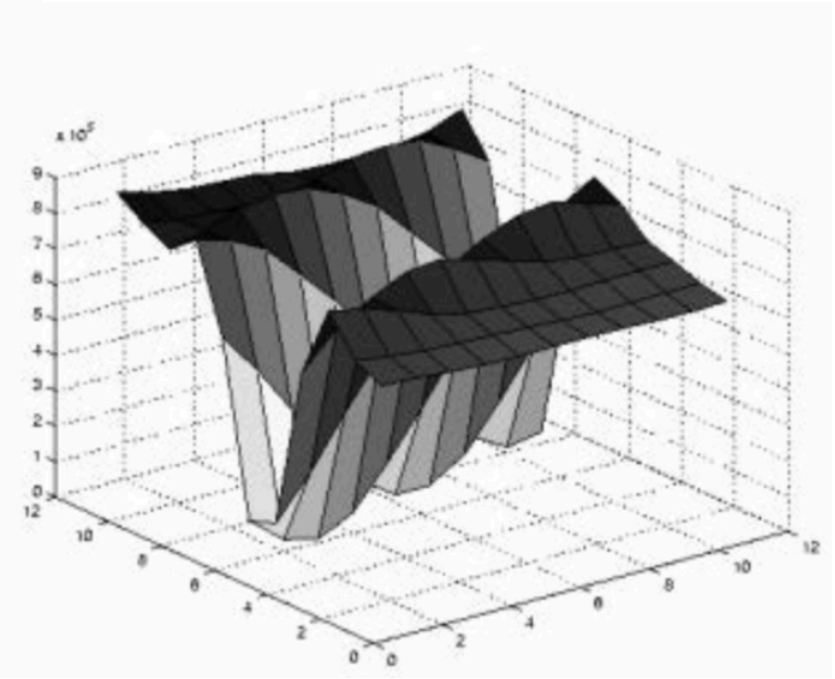
When is this solvable?

- 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)

Criteria for Harris corner detector

Edge

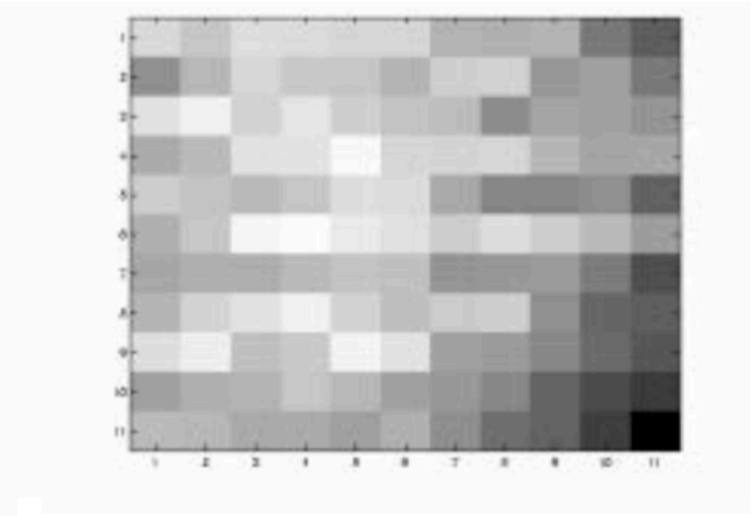


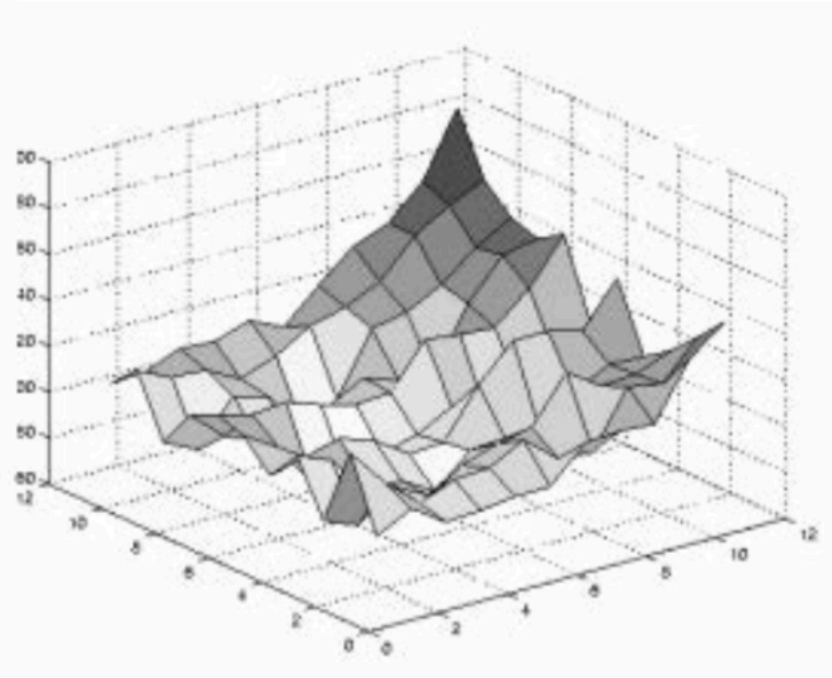


– **gradients very large or very small**
– **large λ_1, small λ_2**

A^TA always becomes singular

Low-texture region



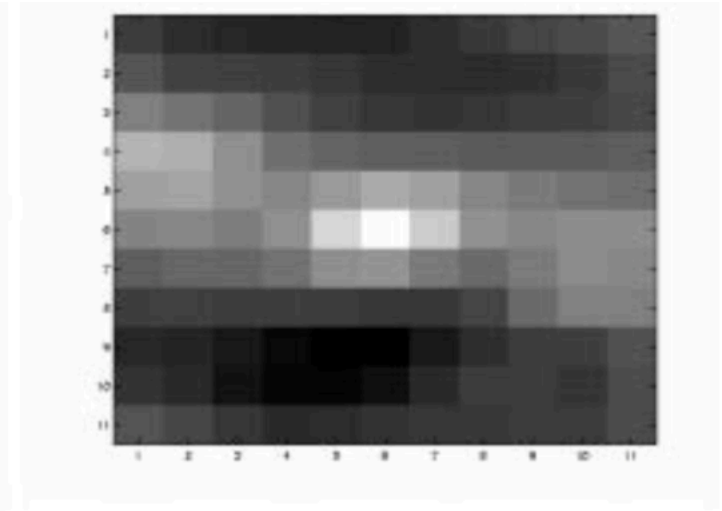


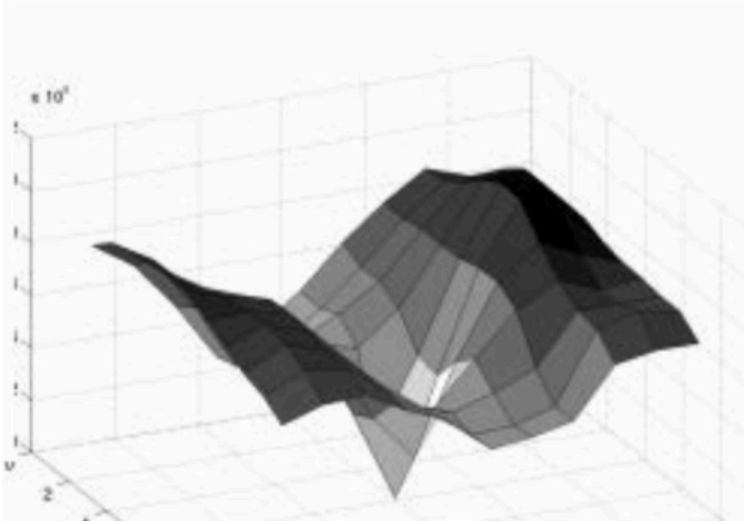
– **gradients have small magnitude**

– **small λ_1, small λ_2**

$A^TA \sim 0$

High-texture region

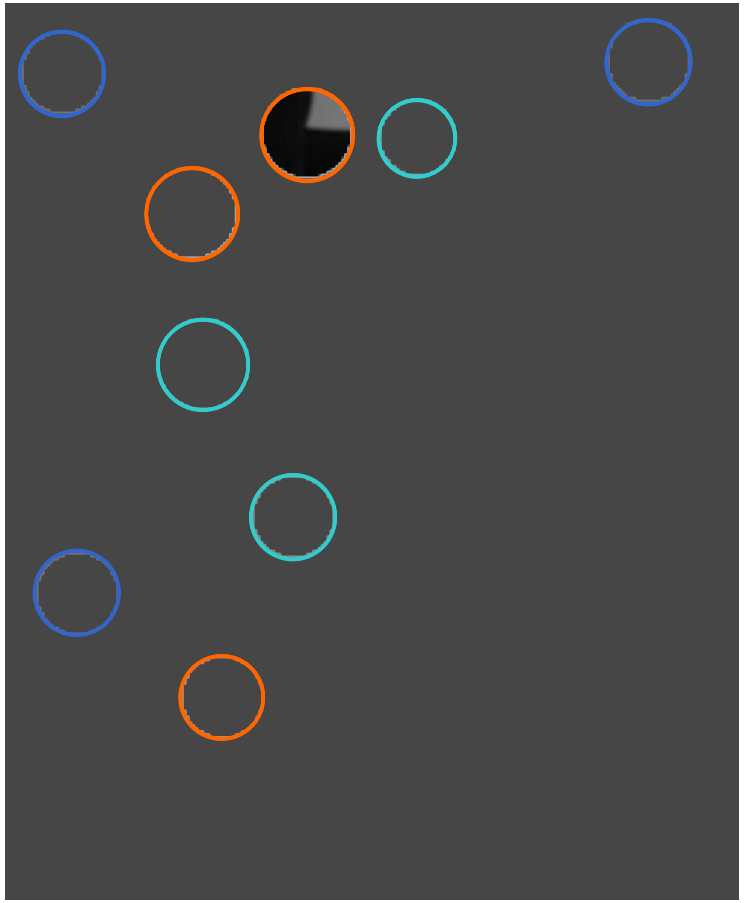




– **gradients are different, large magnitudes**

– **large λ_1, large λ_2**

Aperture problem

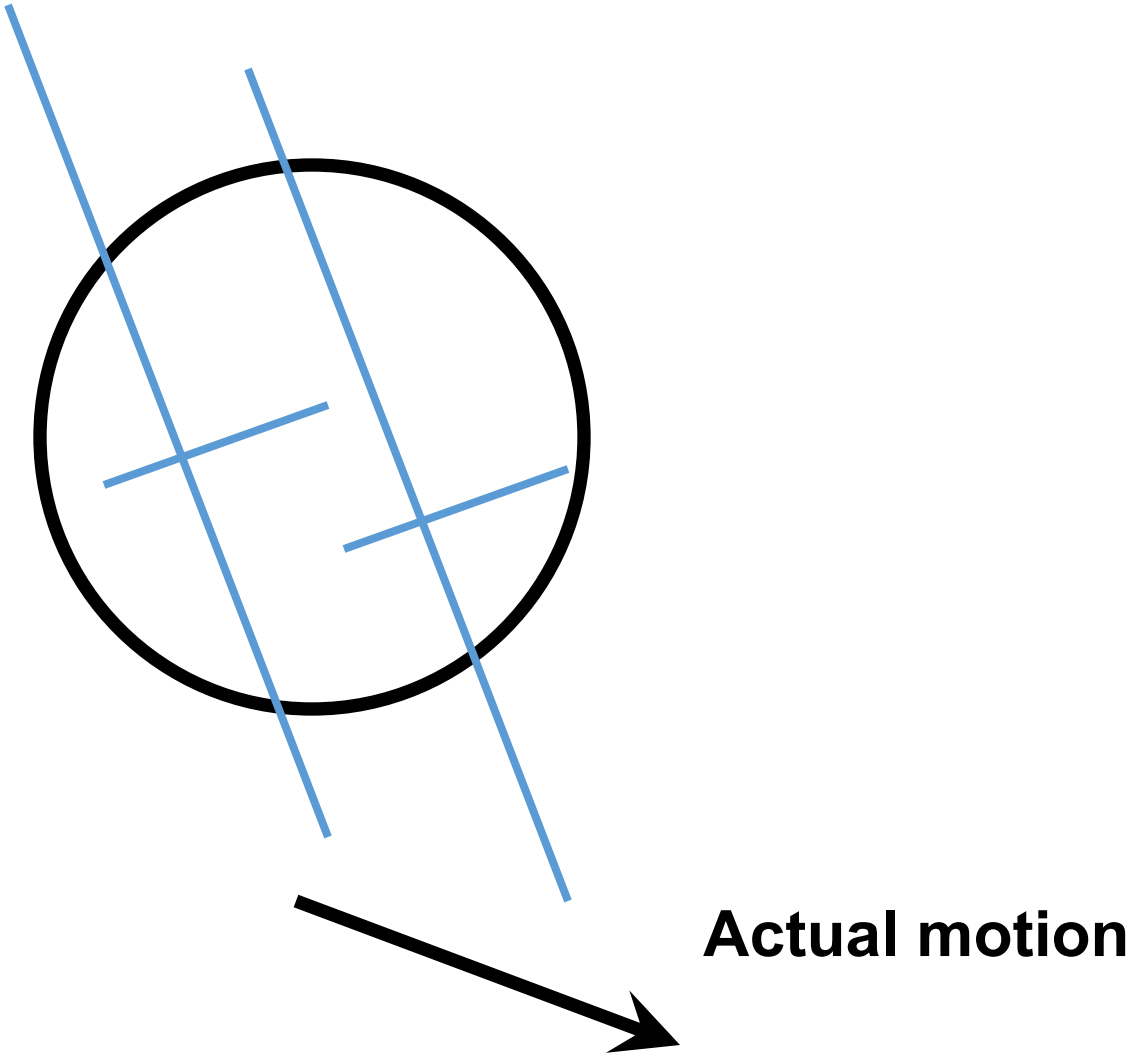


Corners

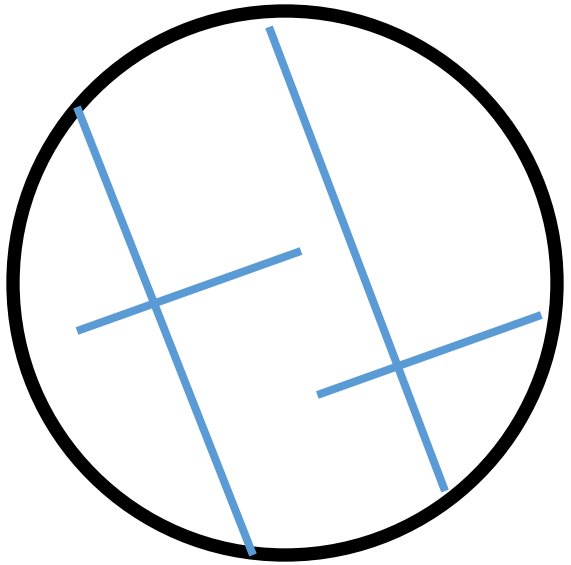
Lines

Flat regions

The aperture problem resolved



The aperture problem resolved



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

Solving the ambiguity…

- 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/

Errors in in assumptions

What are the potential causes of errors in this procedure?

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

Errors in in assumptions

When our assumptions are violated

- **The motion is large** (larger than a pixel)
- **Brightness constancy is not satisfied**

$$I(x,y,t) = I(x+u, y+v, t+1) \rightarrow \nabla I \cdot [u \;\; v]^T + I_t = 0$$

- **A point does not move like its neighbors**

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

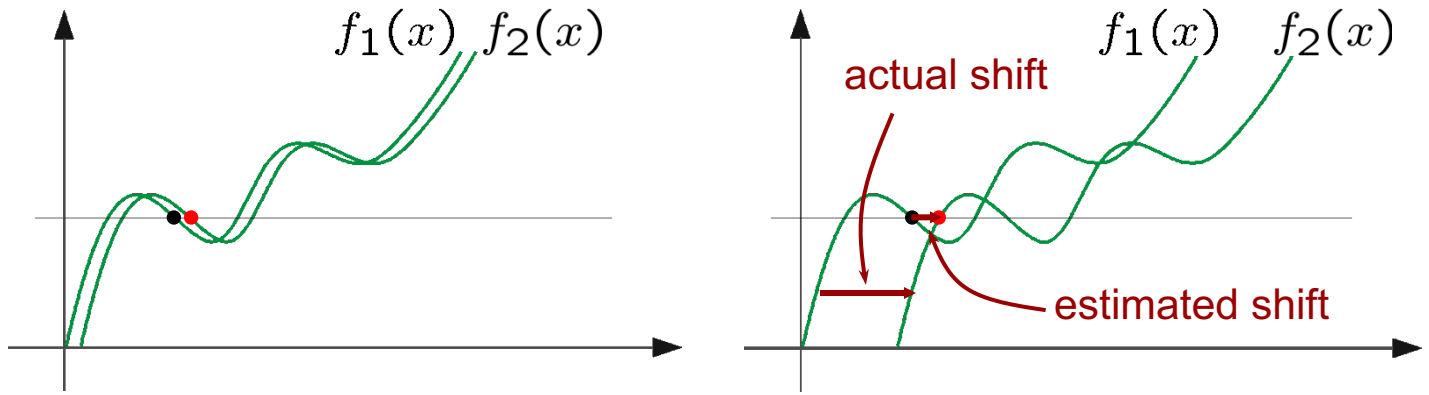
Revisiting the Small Motion Assumption

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

Optical Flow: Aliasing

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

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

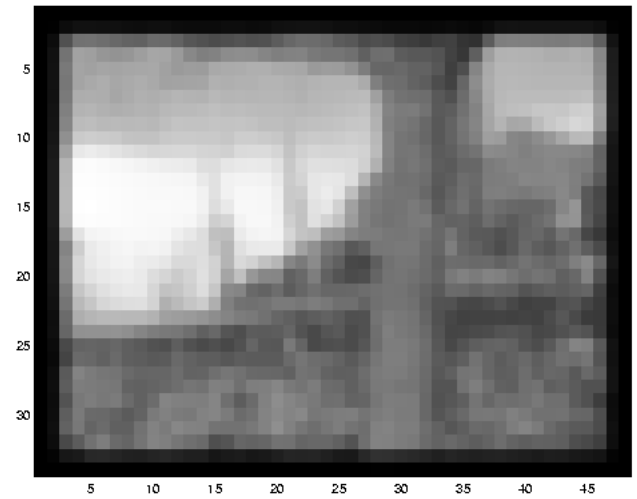


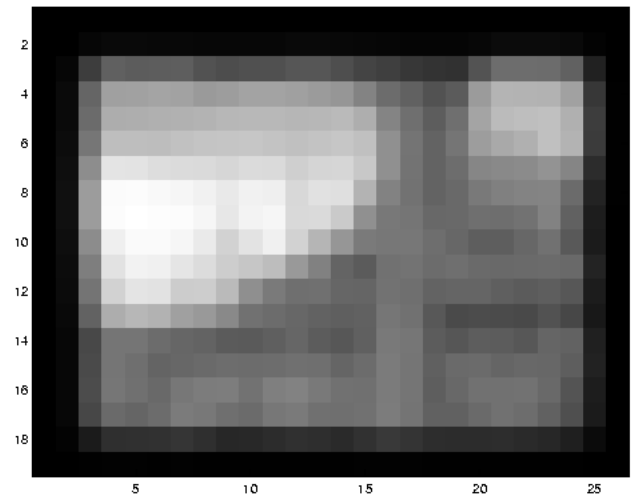
nearest match is correct (no aliasing)

nearest match is incorrect (aliasing)

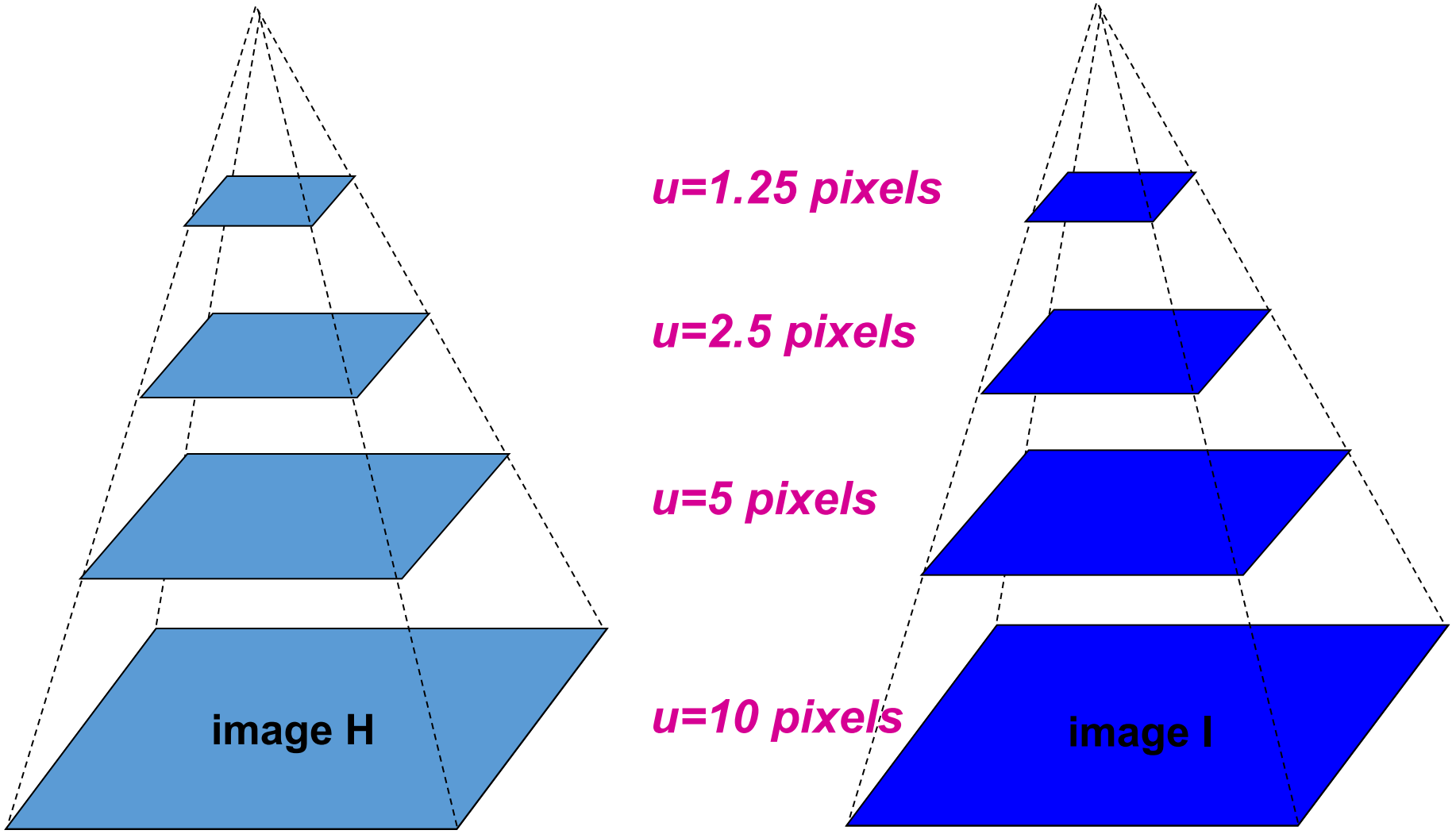
To overcome aliasing: coarse-to-fine estimation.

Reduce the Resolution!





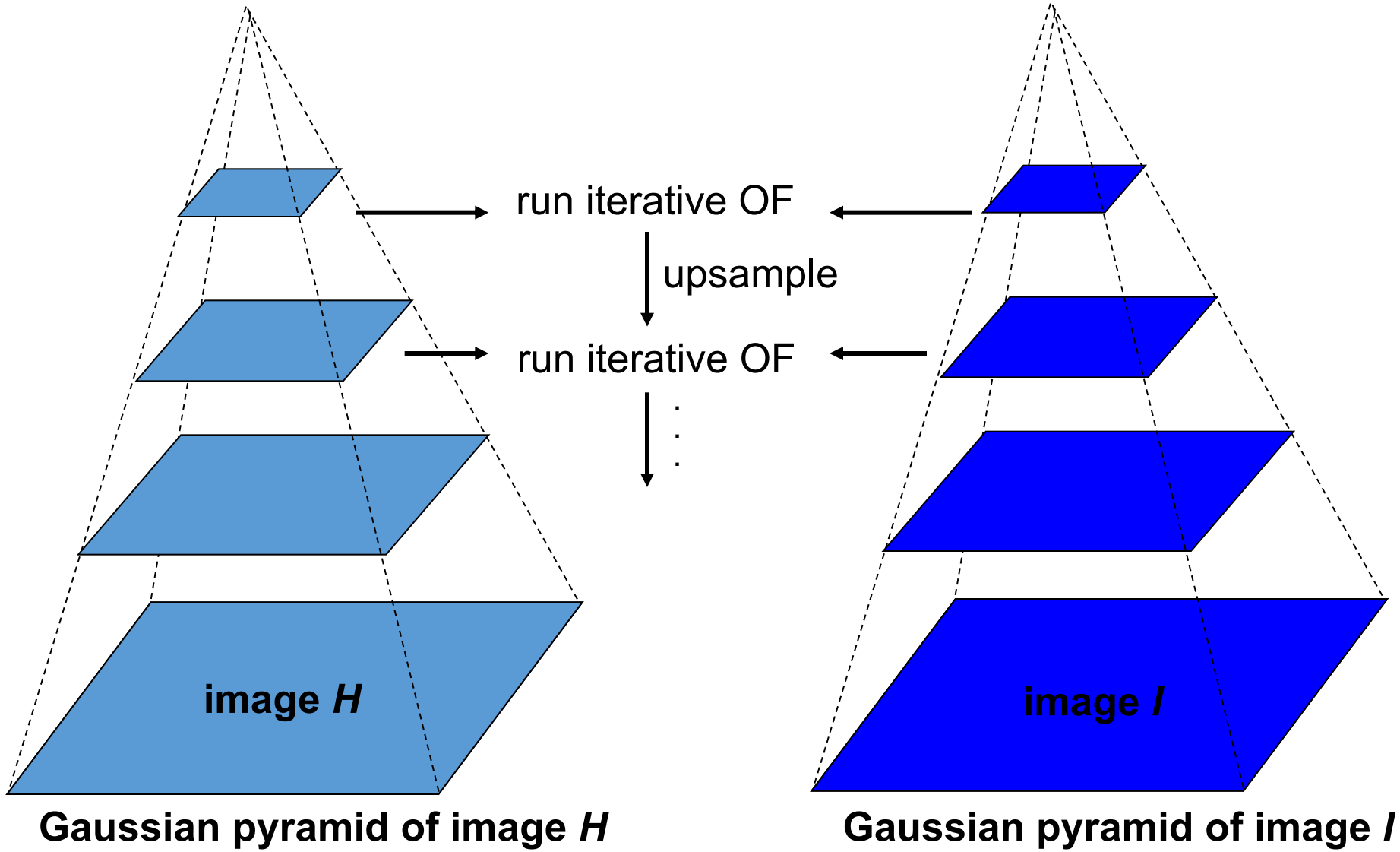
Coarse-to-fine Optical Flow Estimation



Gaussian pyramid of image *H*

Gaussian pyramid of image *I*

Coarse-to-fine Optical Flow Estimation

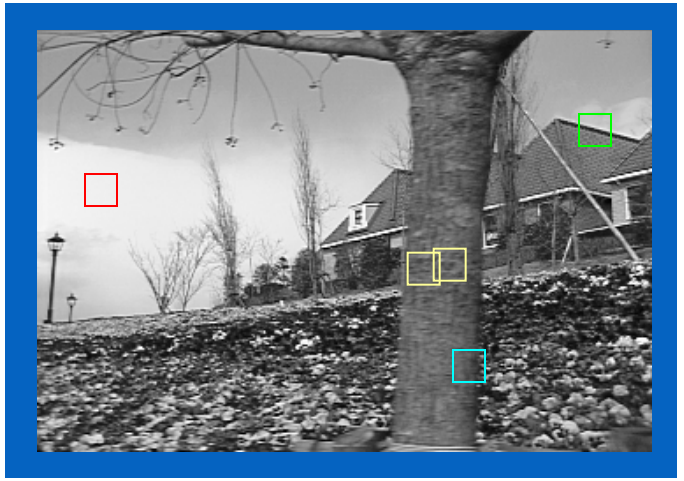


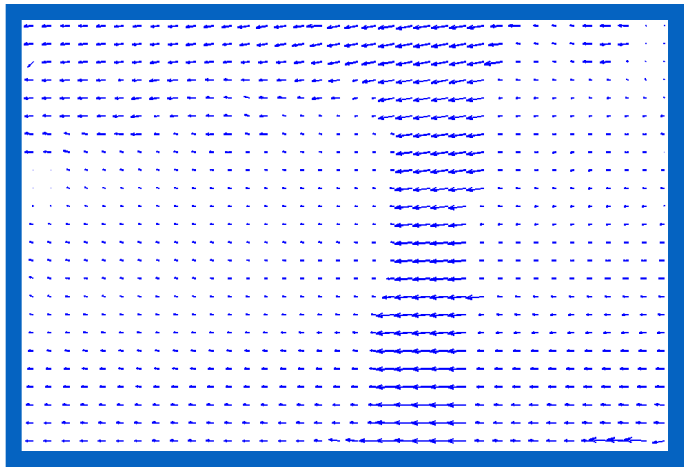
Iterative Lucas-Kanade Algorithm

- 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.
  - Re-run 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

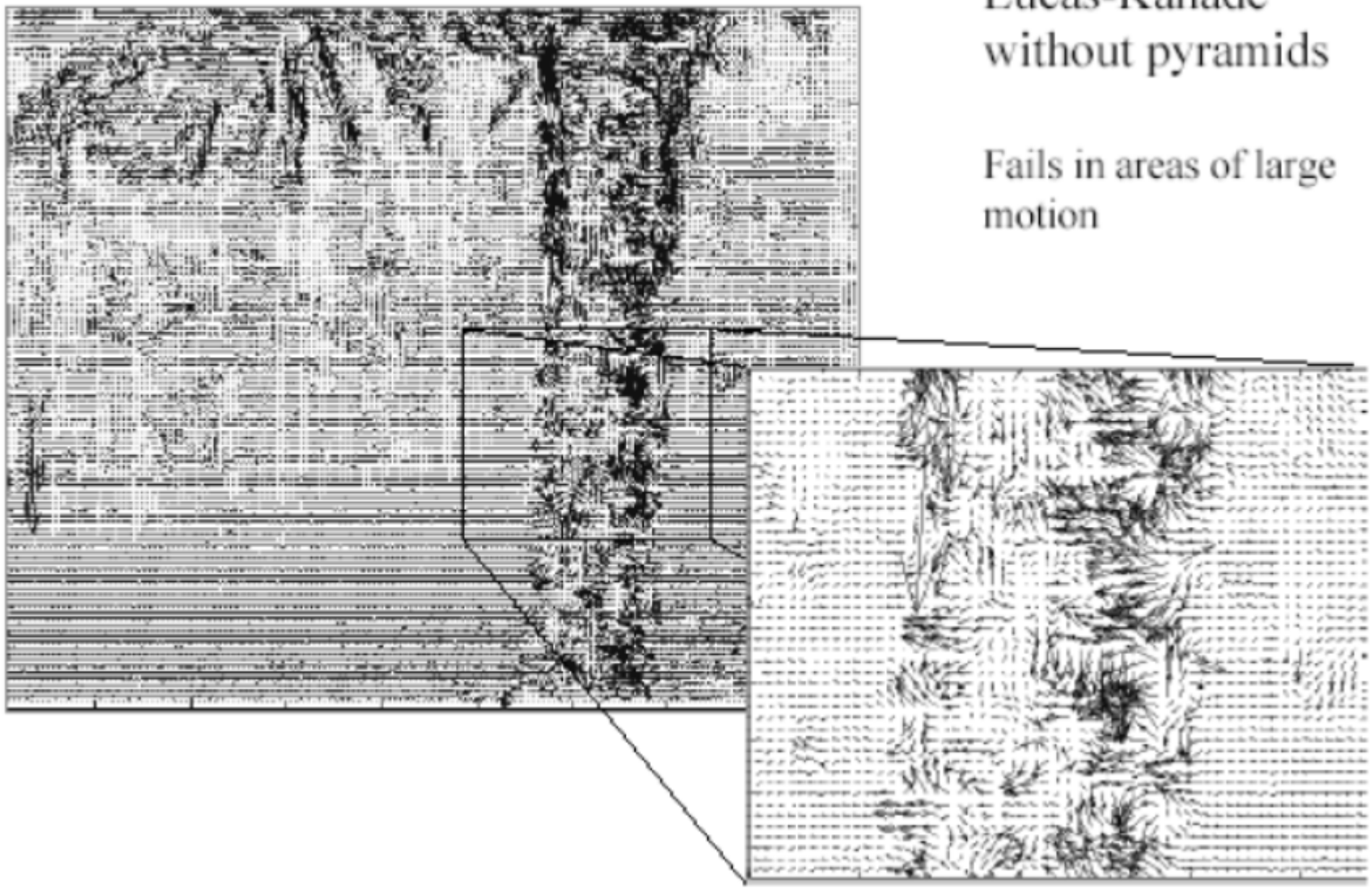
The Flower Garden Video

What should the optical flow be?

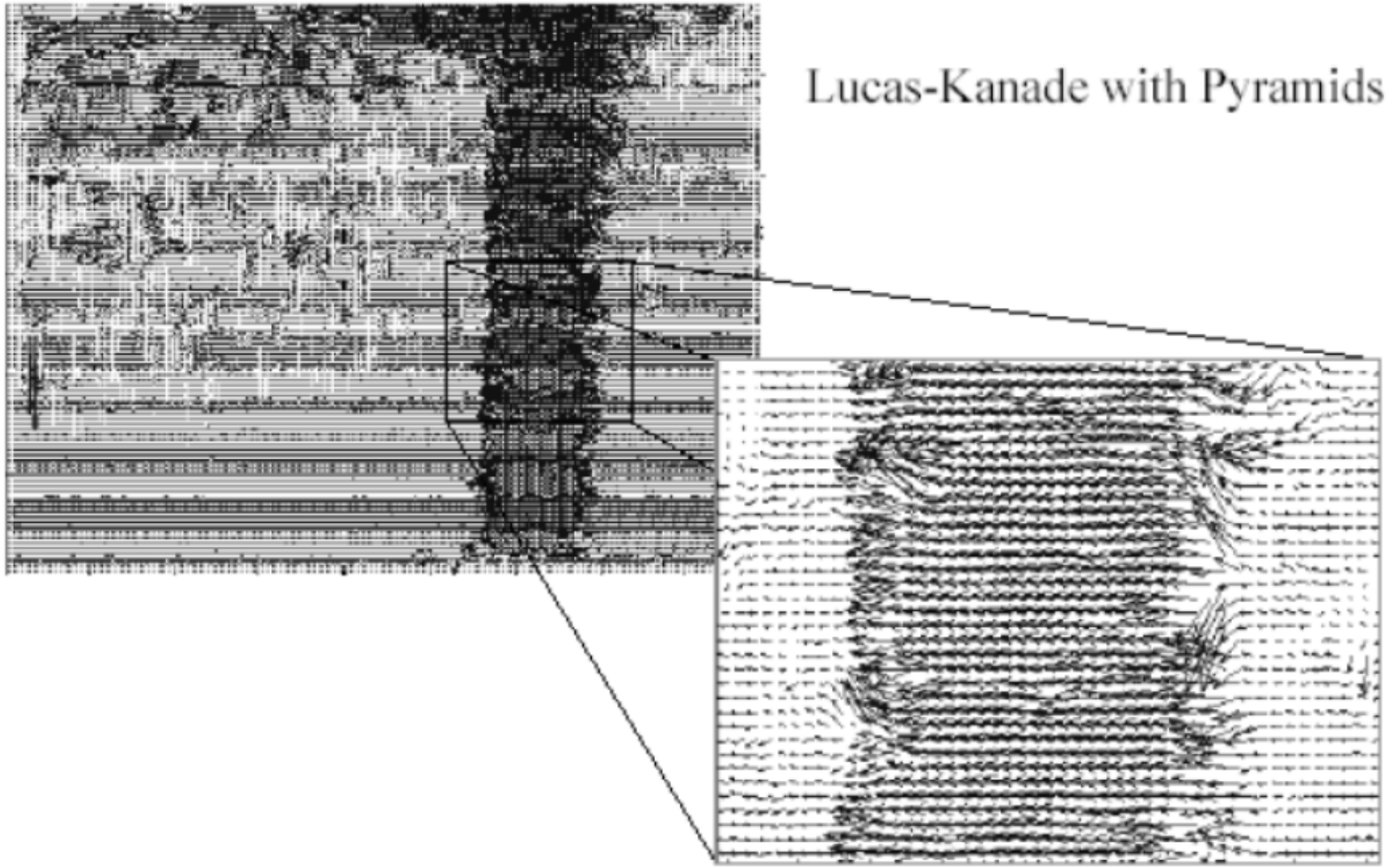




Optical Flow Results

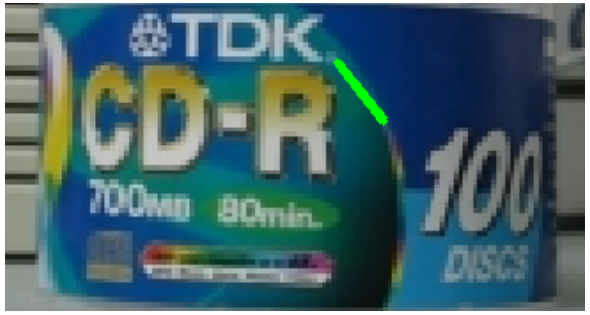


Optical Flow Results



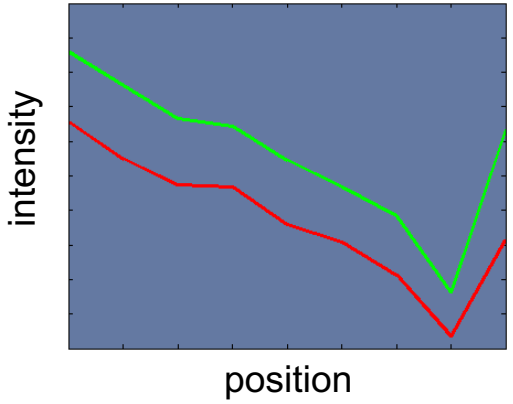
Brightness is not always constant

Rotating cylinder



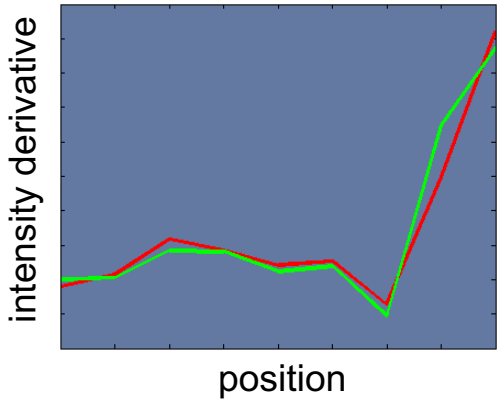
Brightness constancy does not always hold

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



Gradient constancy holds

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



Local constraints work poorly

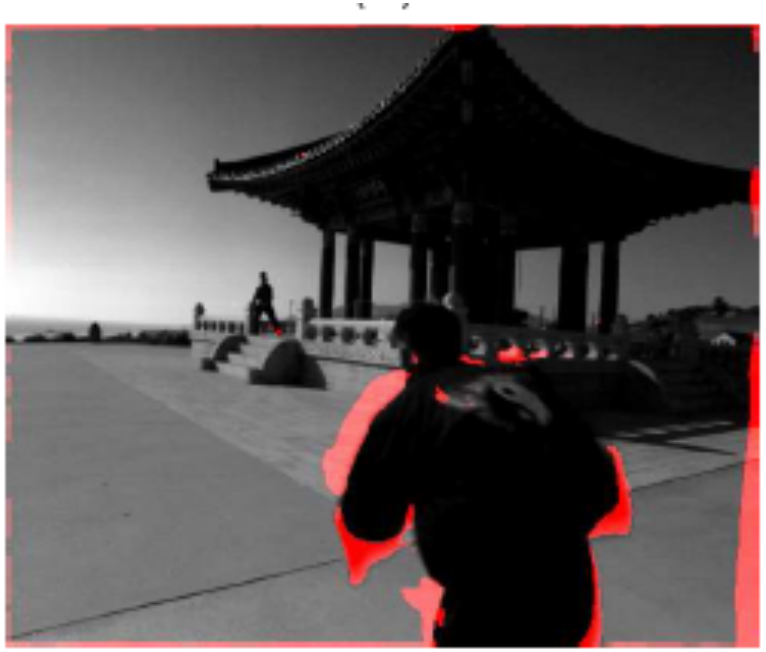
input video

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



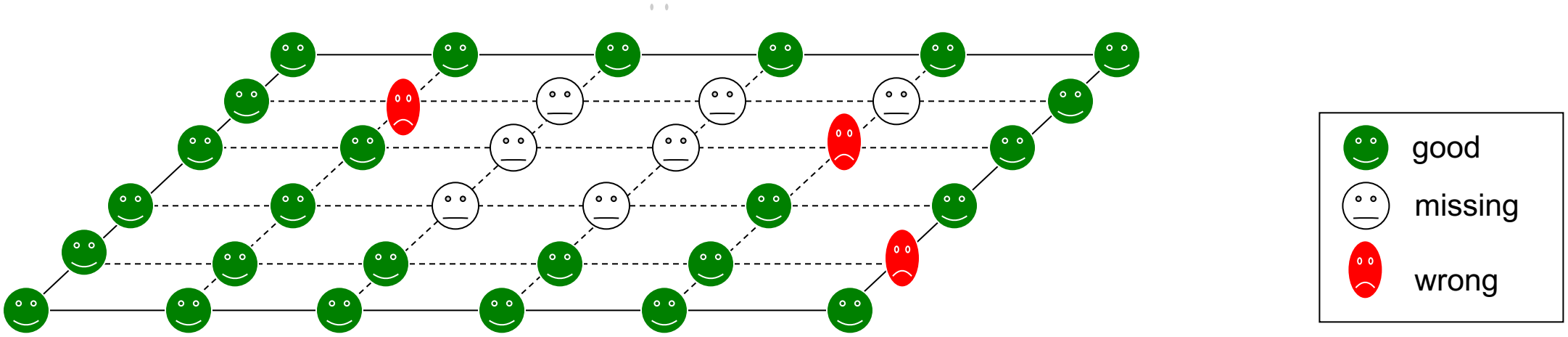
Occluded regions are marked in red

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: $E = \sum_i |\bar{x} - x_i|^2 \to \min$

L1: $E = \sum_i |\bar{x} - x_i| \to \min$

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

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

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

Outliers influence the most

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

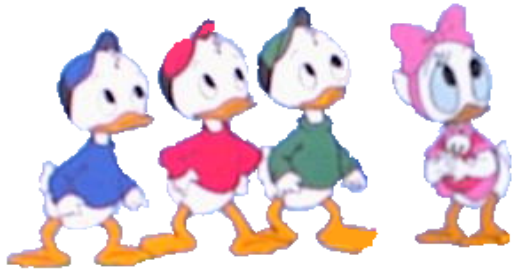
$E_{new} \cong E_{old} + \Delta$

equal for all x_i

Majority decides

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

Elections and robust statistics



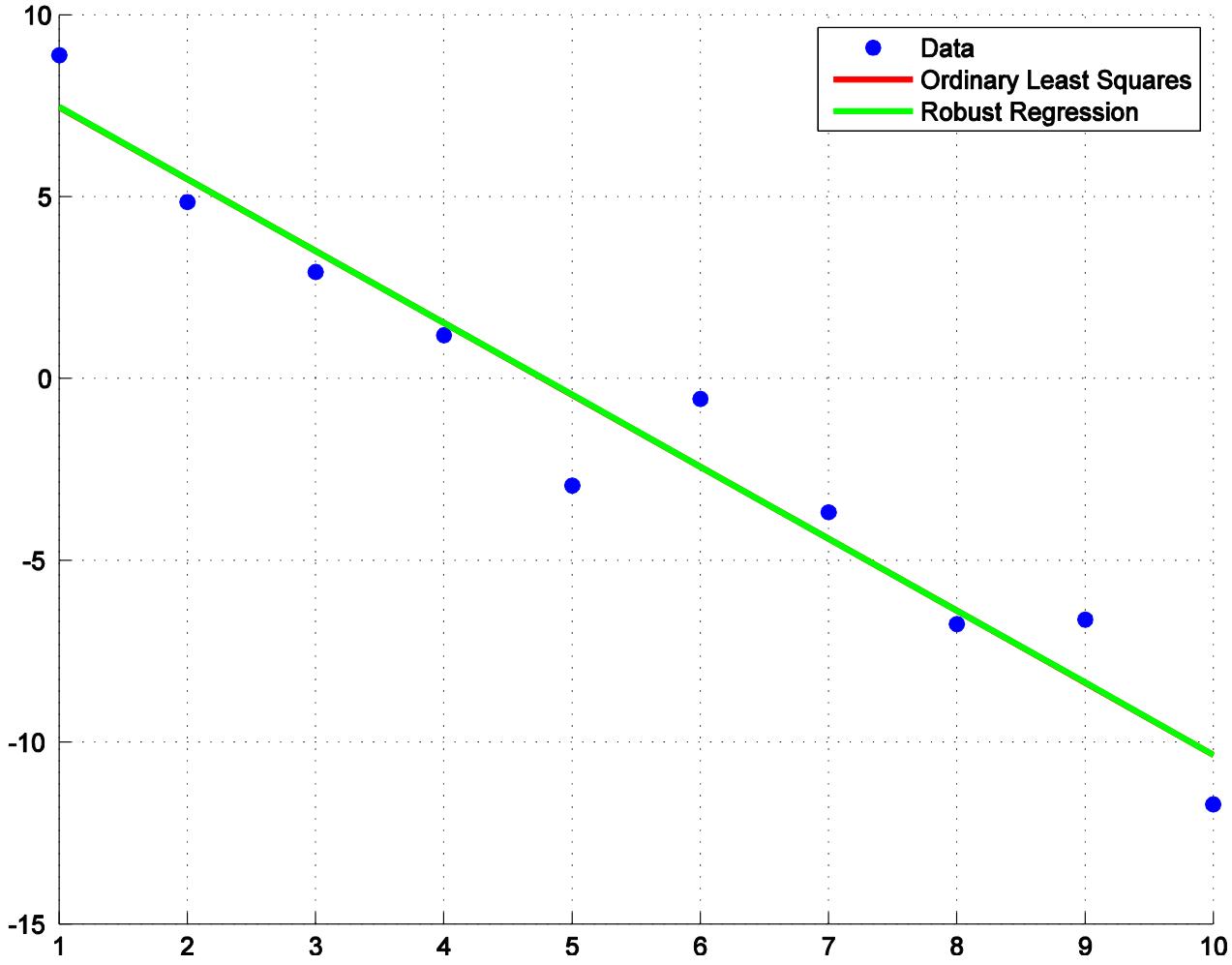
many ordinary people

a very rich man

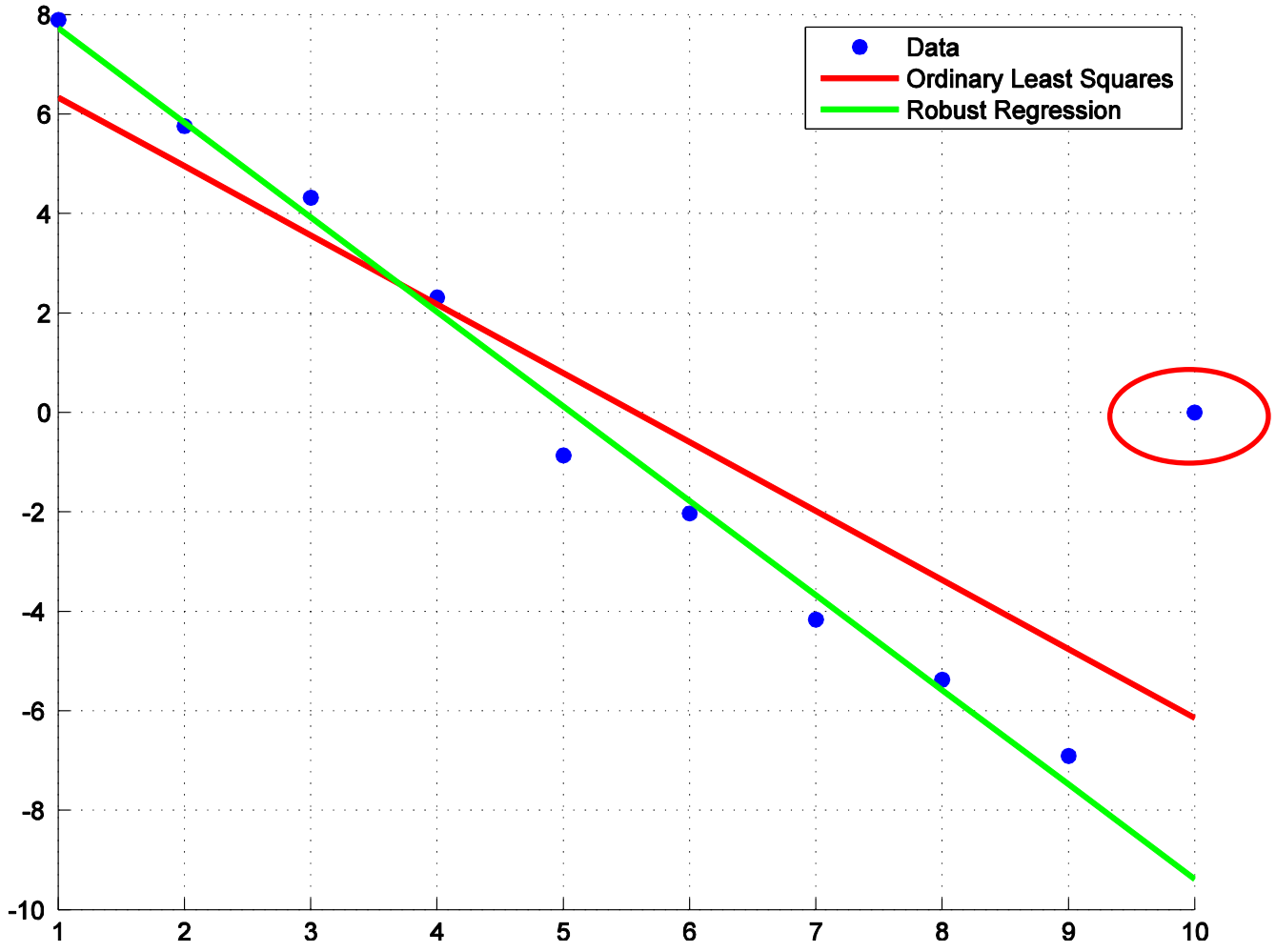
wealth

Oligarchy	Democracy
Votes proportional to the wealth	One vote per person
like in L2 norm minimization	like in L1 norm minimization

A Simple Example



A Simple Example

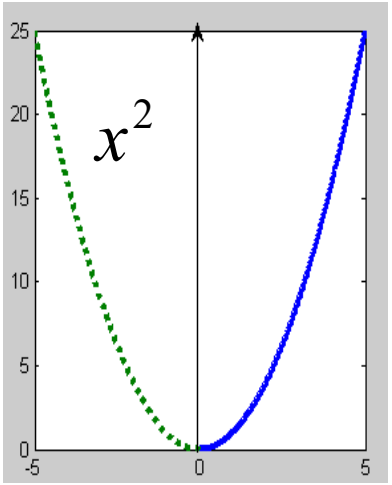


Combination of two flow constraints

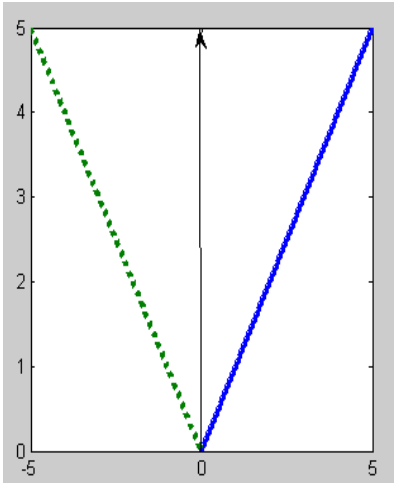
$$\min \int_{video} \phi\left(\left|I_{warped} - I\right|\right) + \alpha\phi\left(\left|\nabla I_{warped} - \nabla I\right|\right)$$

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

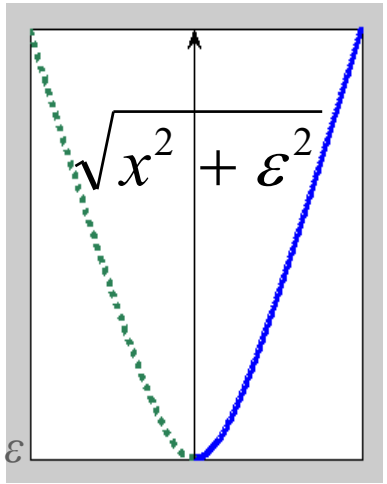
usual: L2



robust: L1



robust regularized



✓ 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

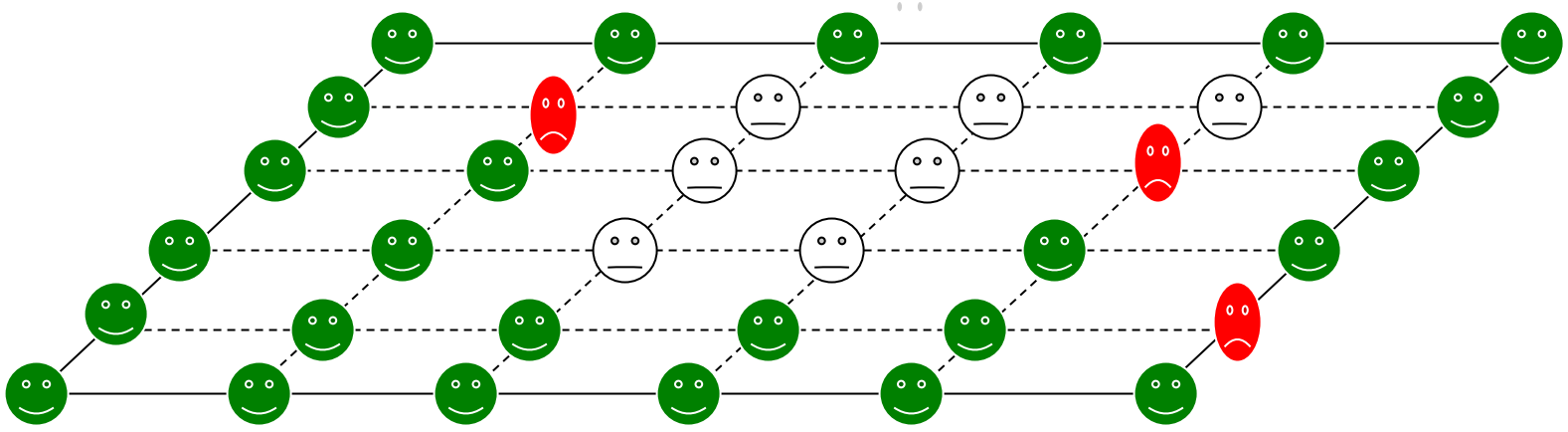
Spatial Propagation

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

Optical flow direction using only local constraints

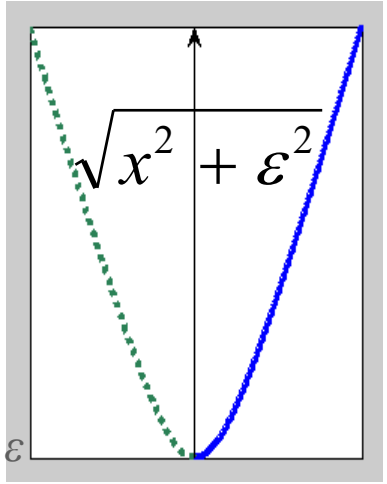
This constraint is not correct on motion boundaries
=> over-smoothing of the resulting flow

[Horn&Schunck 1981]

Robustness to flow discontinuities

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

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

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

Necessary condition:

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

Euler-Lagrange equation

Analogy:

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

Necessary condition

$$f'(x) = 0$$

Variational: Euler equation

单元单标量函数	$$E(u) = \int_0^{`} F(x, u, u', u'')\, dx$$
	$$\frac{\partial F}{\partial u} - \frac{d}{dx}\left(\frac{\partial F}{\partial u'}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial u''}\right) = 0$$
多元单标量函数	$$E(u) = \iint_\Omega F(x, y, u, u_x, u_y, u_{xx}, u_{yy})\, dxdy$$
	$$\frac{\partial F}{\partial u} - \frac{d}{dx}\left(\frac{\partial F}{\partial u_x}\right) - \frac{d}{dy}\left(\frac{\partial F}{\partial u_y}\right) + \frac{d^2}{dx^2}\left(\frac{\partial F}{\partial u_{xx}}\right) + \frac{d^2}{dy^2}\left(\frac{\partial F}{\partial u_{yy}}\right) = 0$$ $$= \mathbf{F_u} - \mathbf{div}\left(\mathbf{F_{u_x}}, \mathbf{F_{u_y}}\right) + \Delta\left(\mathbf{F_{u_{xx}}}, \mathbf{F_{u_{yy}}}\right) = 0$$
多元多标量函数 **Multi multivariable (scalar) function**	$$E[u, v] = \int_\Omega F(x, y, u, u_x, u_y, v, v_x, v_y)dxdy$$
	$$\begin{cases} \mathbf{F_u} - \mathbf{div}\left(\mathbf{F_{u_x}}, \mathbf{F_{u_y}}\right) = 0 \\ \mathbf{F_v} - \mathbf{div}\left(\mathbf{F_{v_x}}, \mathbf{F_{v_y}}\right) = 0 \end{cases}$$

Summary

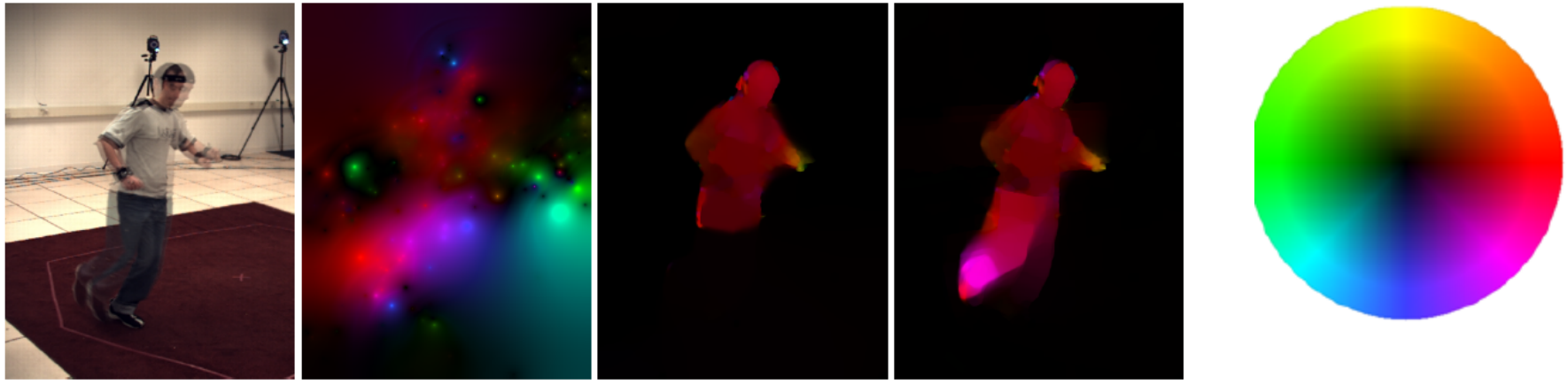
- Major contributions from Lucas, Tomasi, Kanade
  - Tracking feature points
  - Optical flow
  - Stereo
  - Structure from motion

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

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

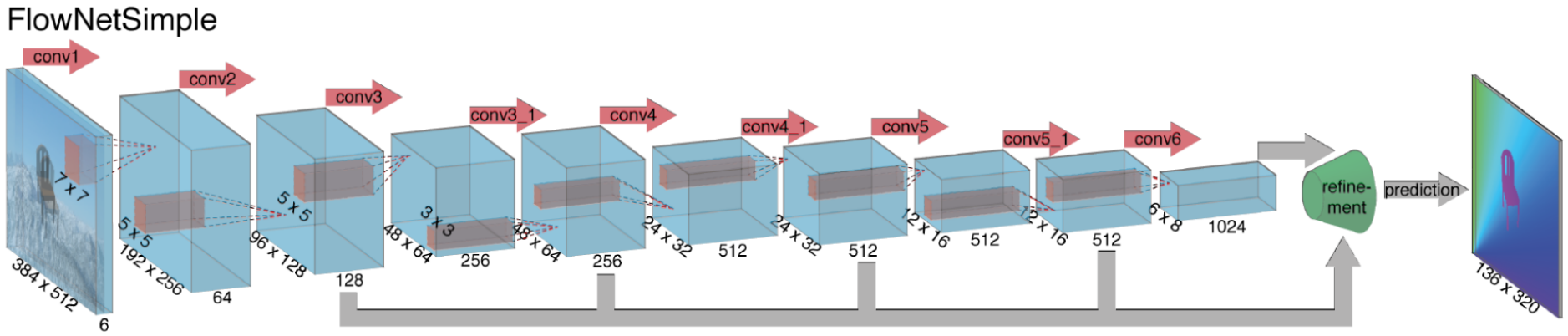
Large displacement optical flow, Brox et al., CVPR 2009

State-of-the-art optical flow, 2015

CNN encoder/decoder

Pair of input frames

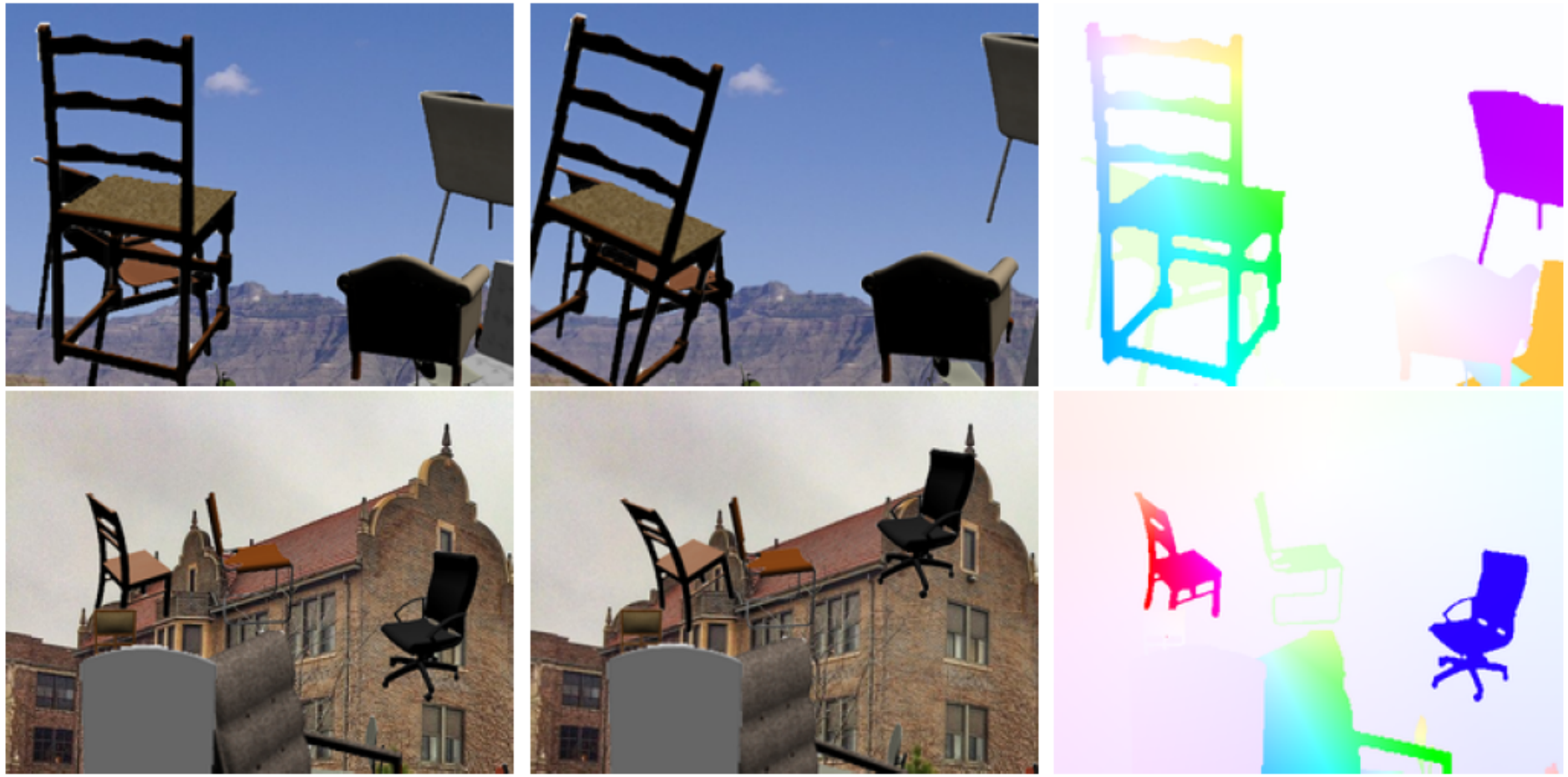
Upsample estimated flow back to input resolution



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

State-of-the-art optical flow, 2015

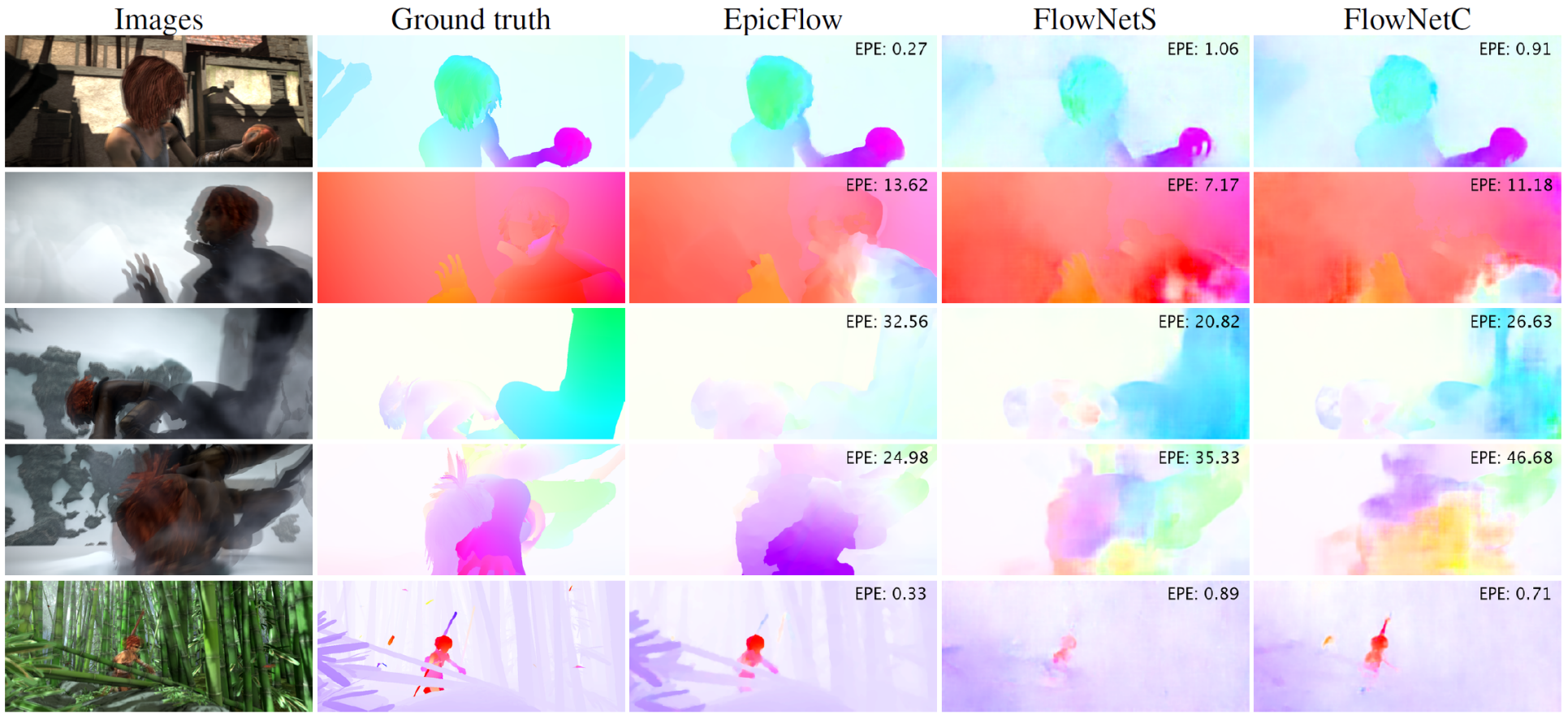
Synthetic Training data



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

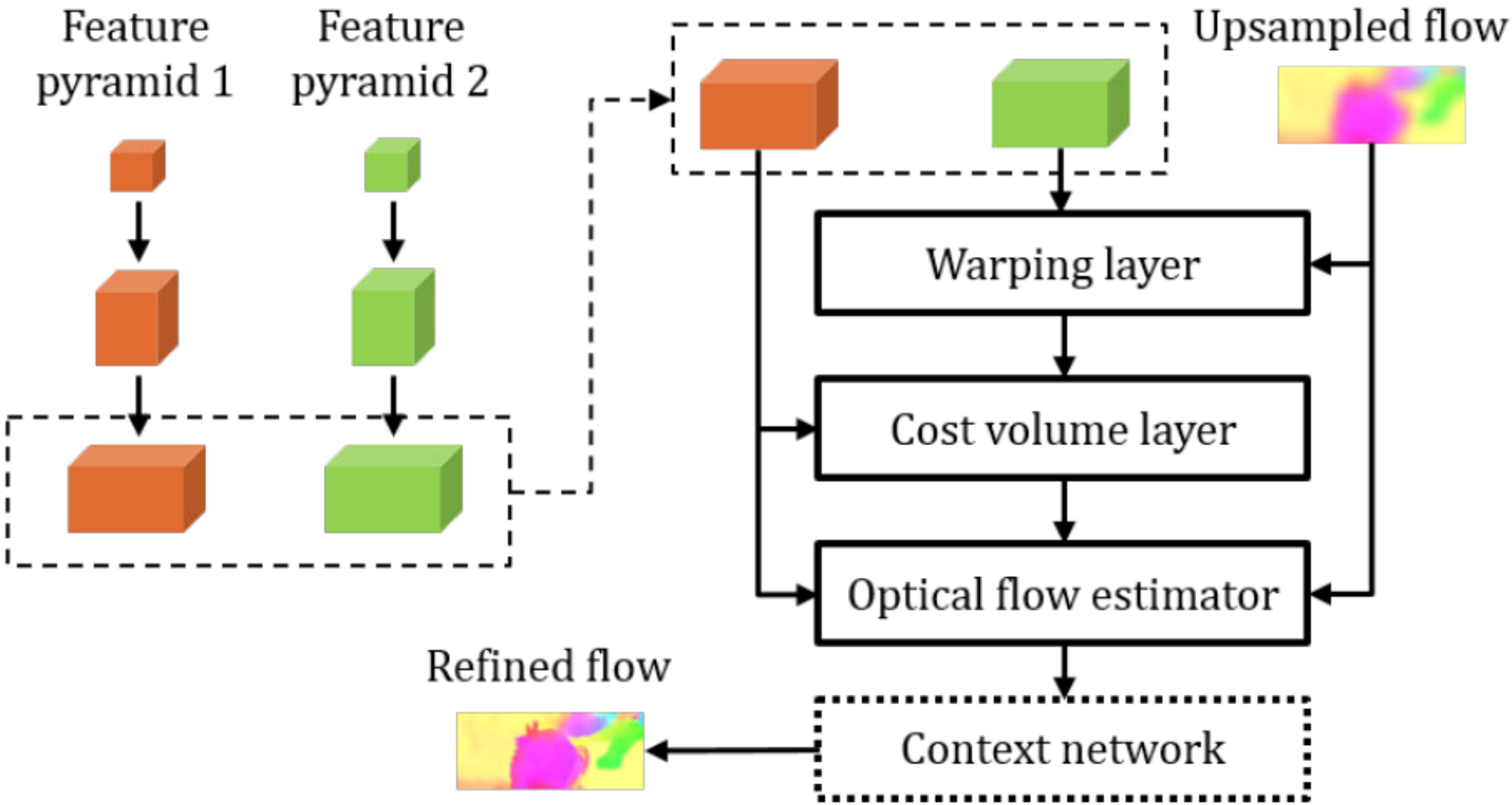
State-of-the-art optical flow, 2015

Results on Sintel

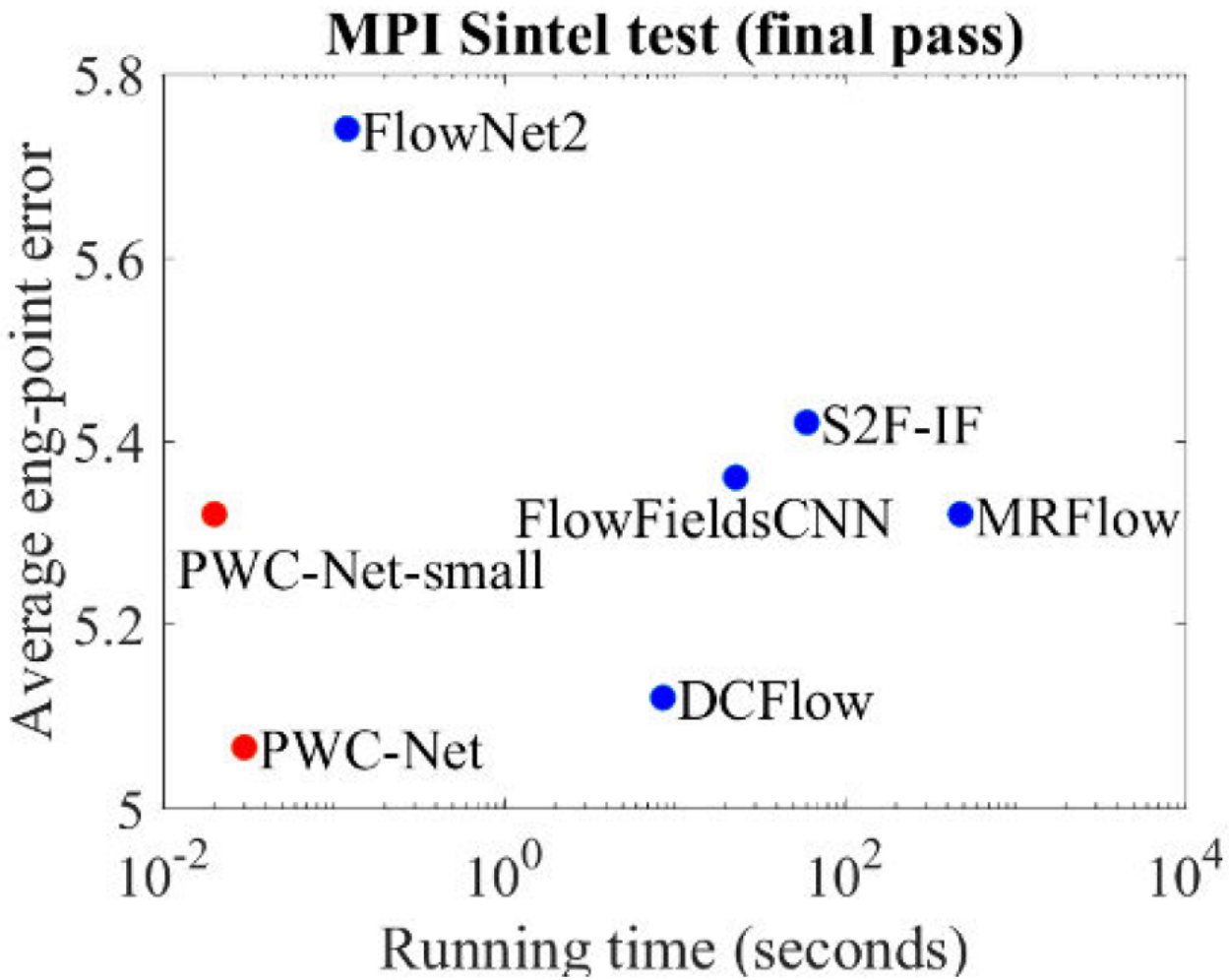


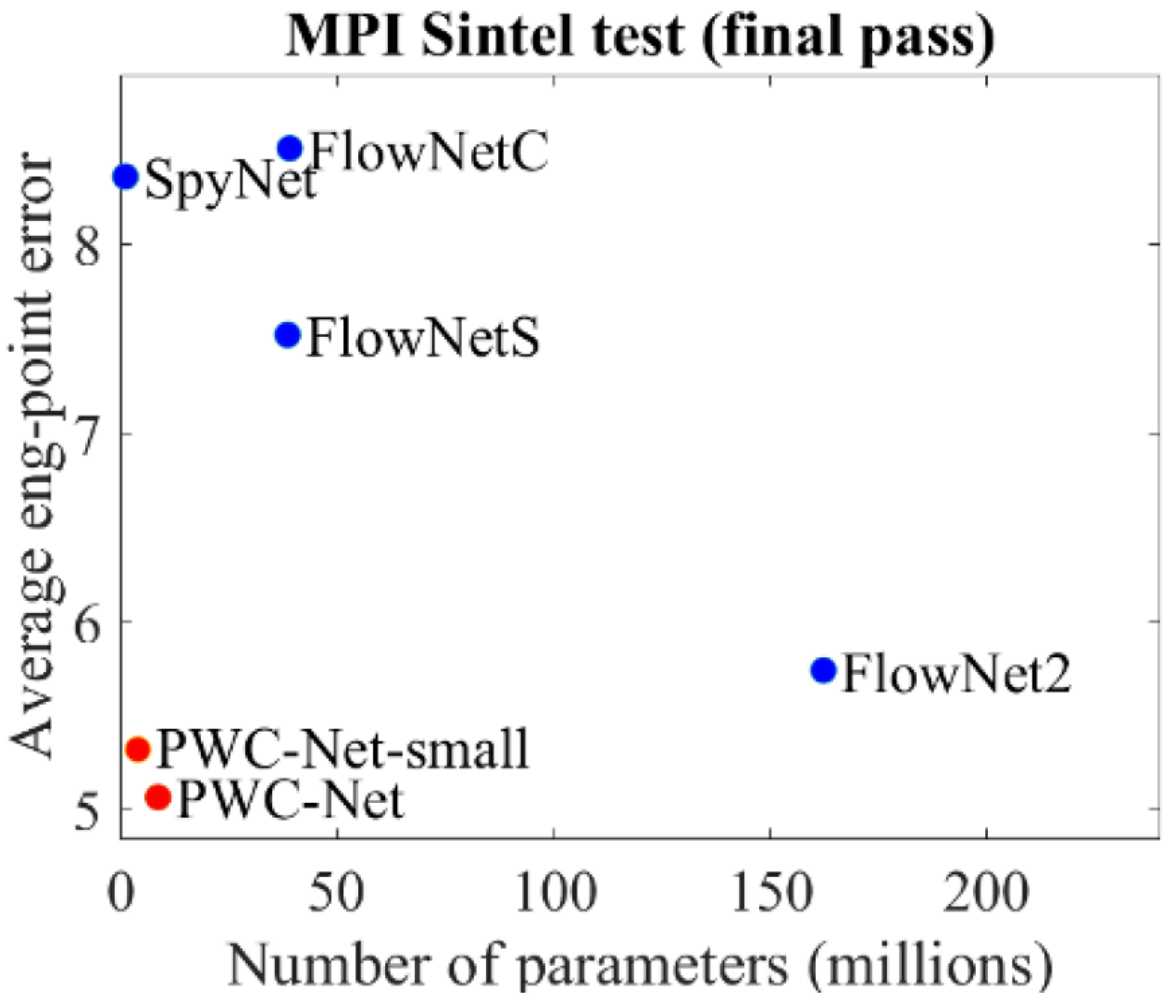
Fischer et al. 2015. https://arxiv.org/abs/1504.06852

State-of-the-art optical flow, 2018 (CVPR)



State-of-the-art optical flow, 2018 (CVPR)

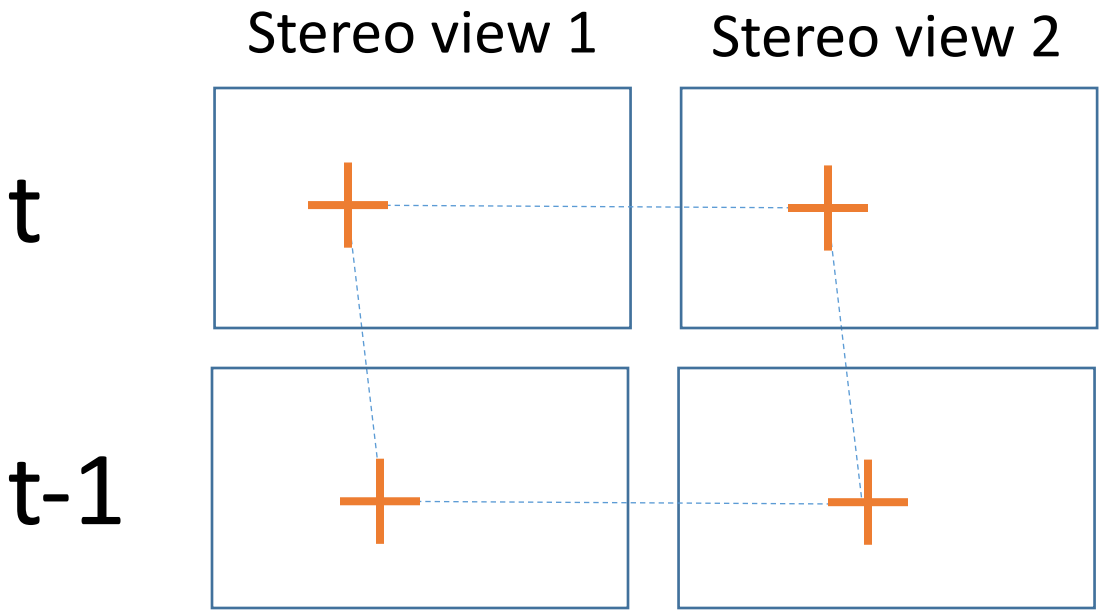




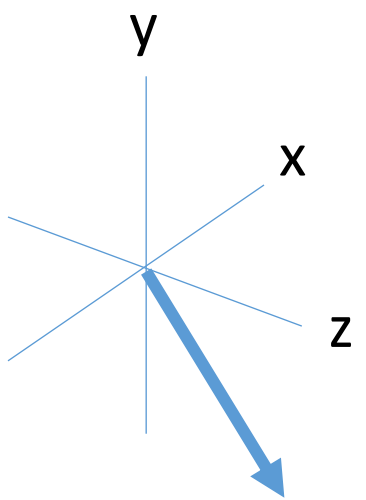
Can we do more? *Scene flow*

Combine spatial stereo & temporal constraints

Recover 3D vectors of world motion

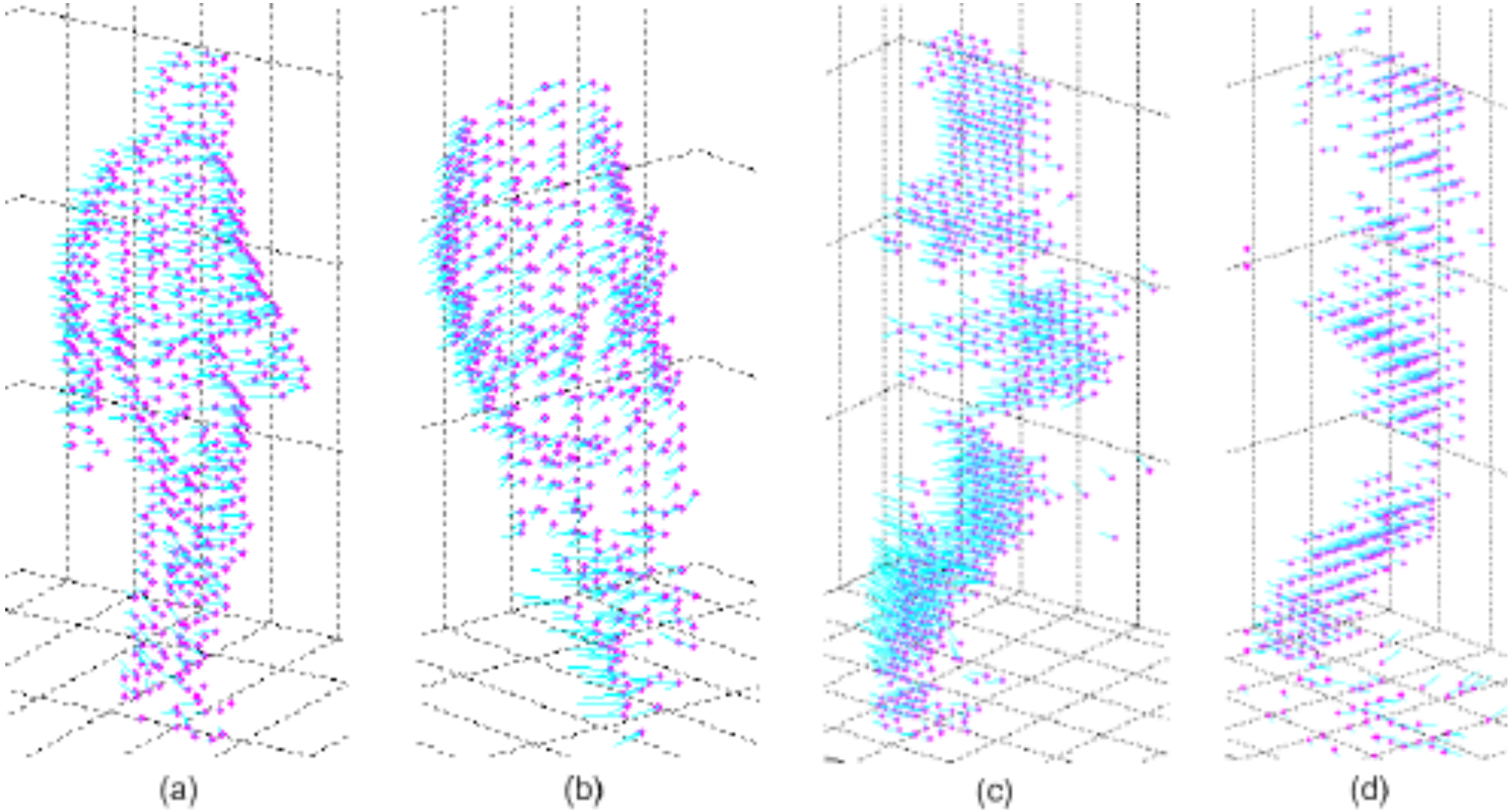


Epipolar constraints
across space and time!



3D world motion
vector per pixel

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]

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/