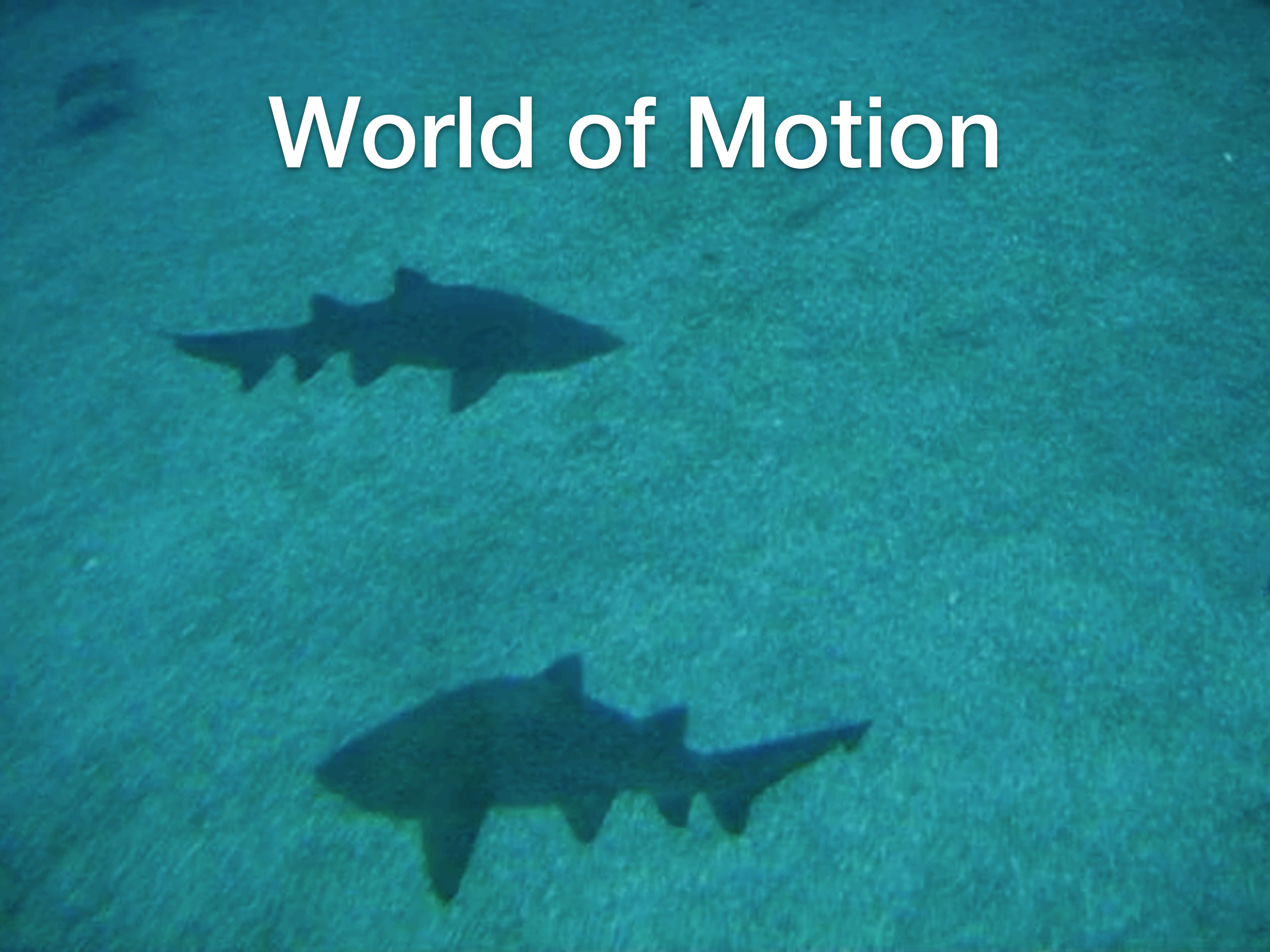
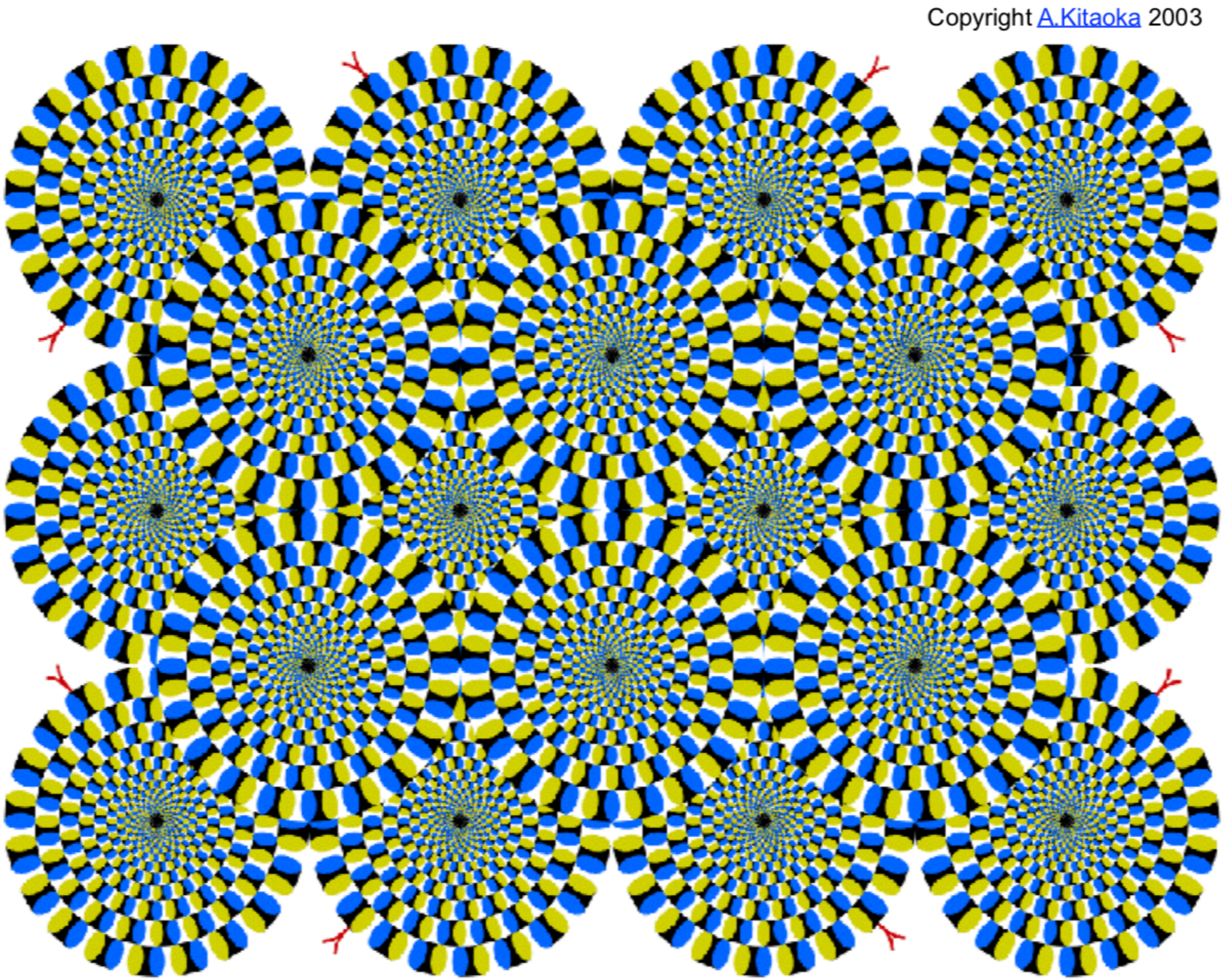
Motion and Flow

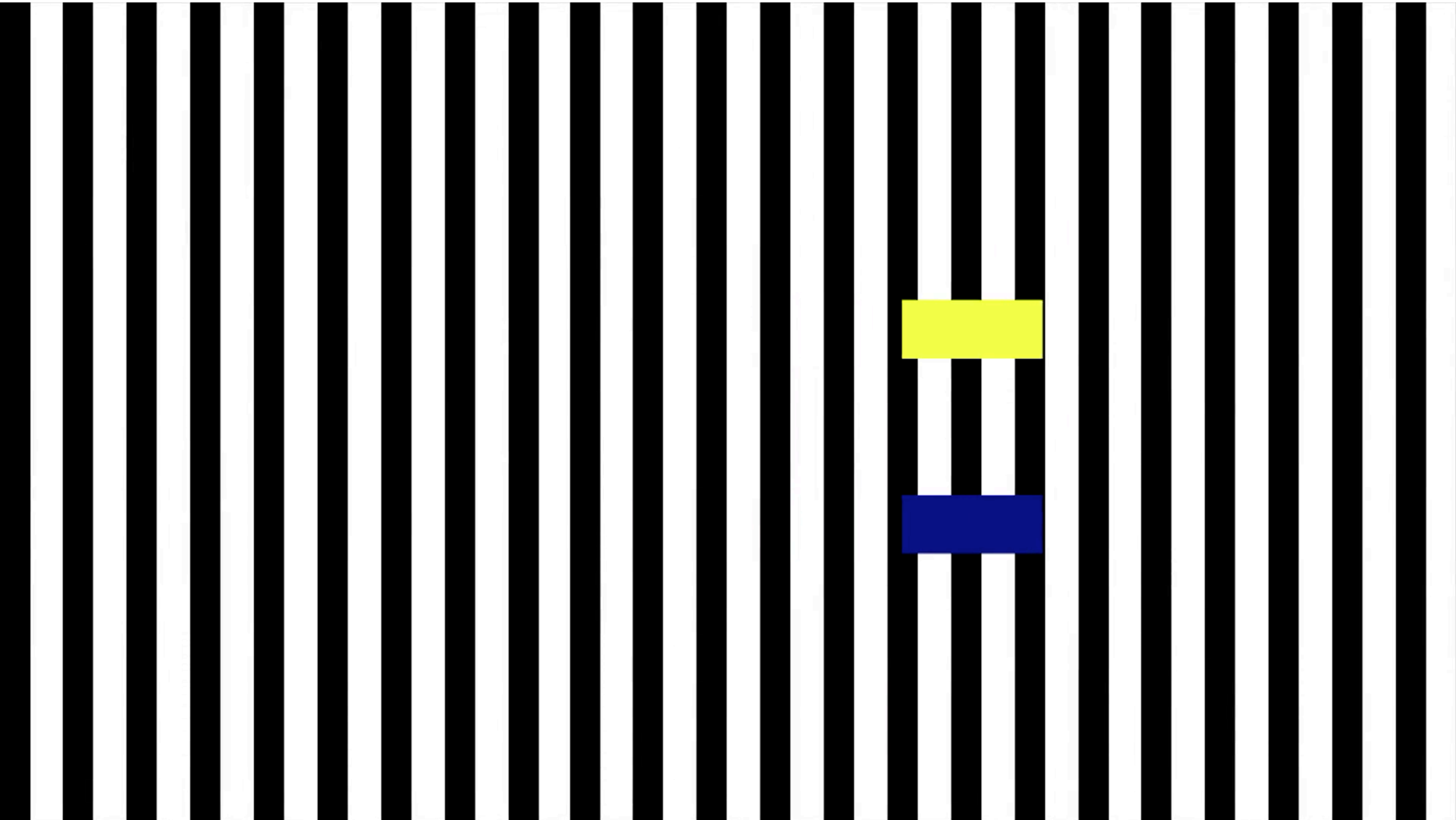
Computer Vision
Fall 2018
Columbia University


World of Motion

Illusionary Motion

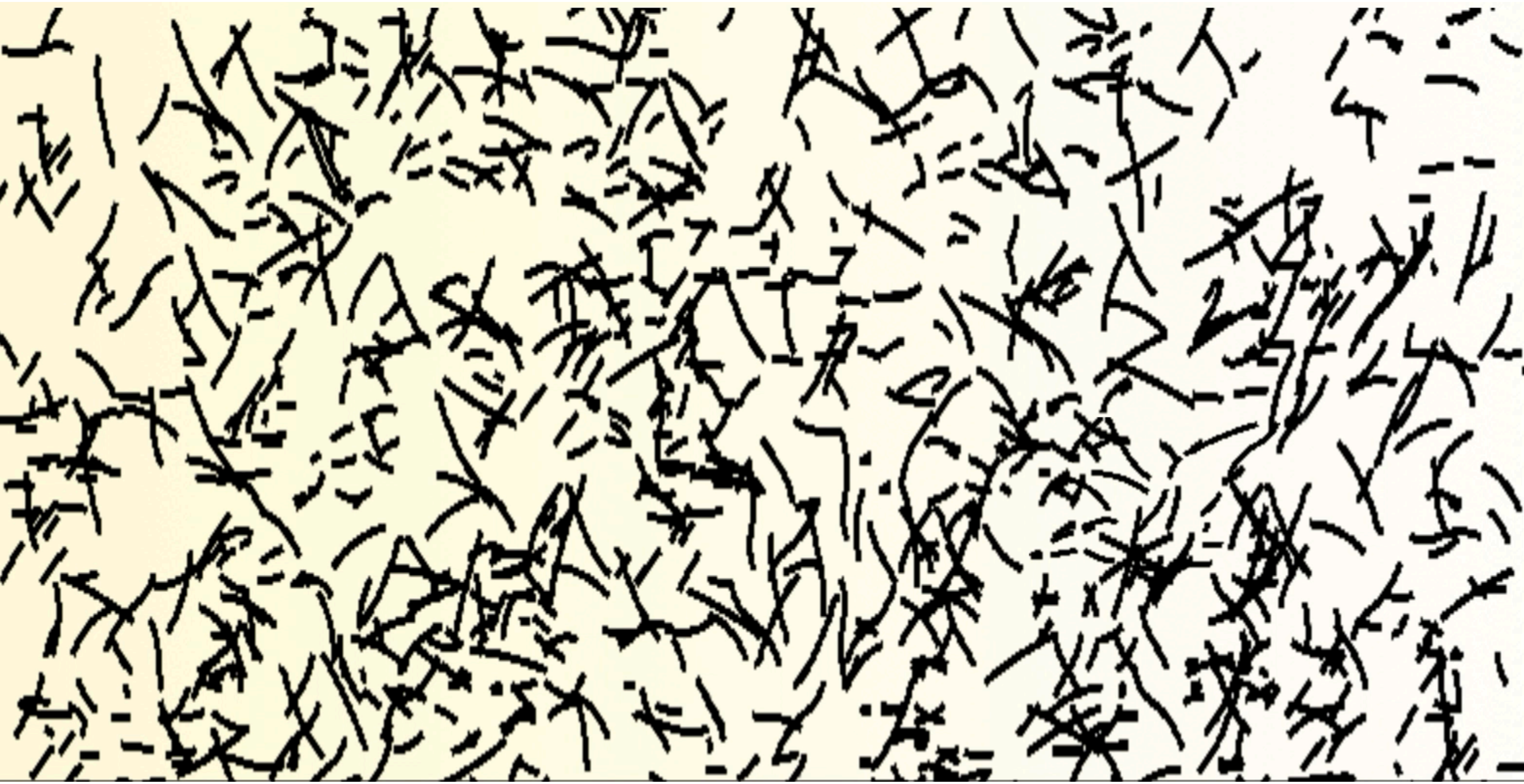


Stepping Feet Illusion



Sometimes motion is only cue

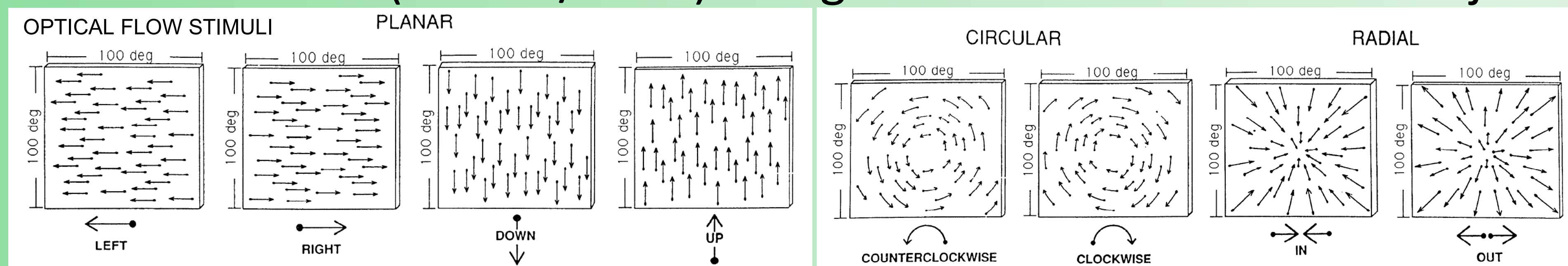
What do you see?

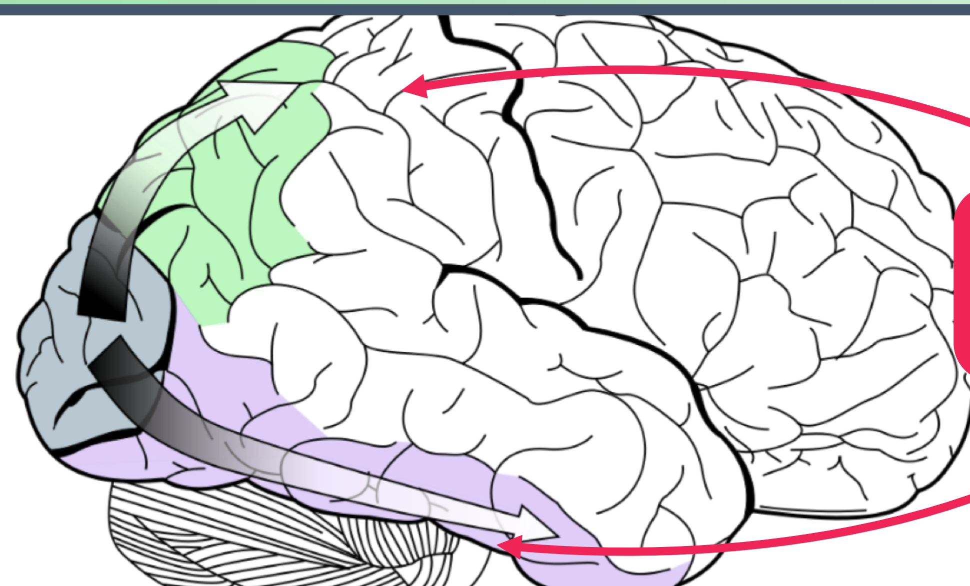


What do you see?

Motivation: Separate visual pathways in nature

➔ Dorsal stream (‘where/how’) recognizes motion and locates objects





➔ **“Interconnection”**
e.g. in STS area

➔ Ventral (‘what’) stream performs object recognition

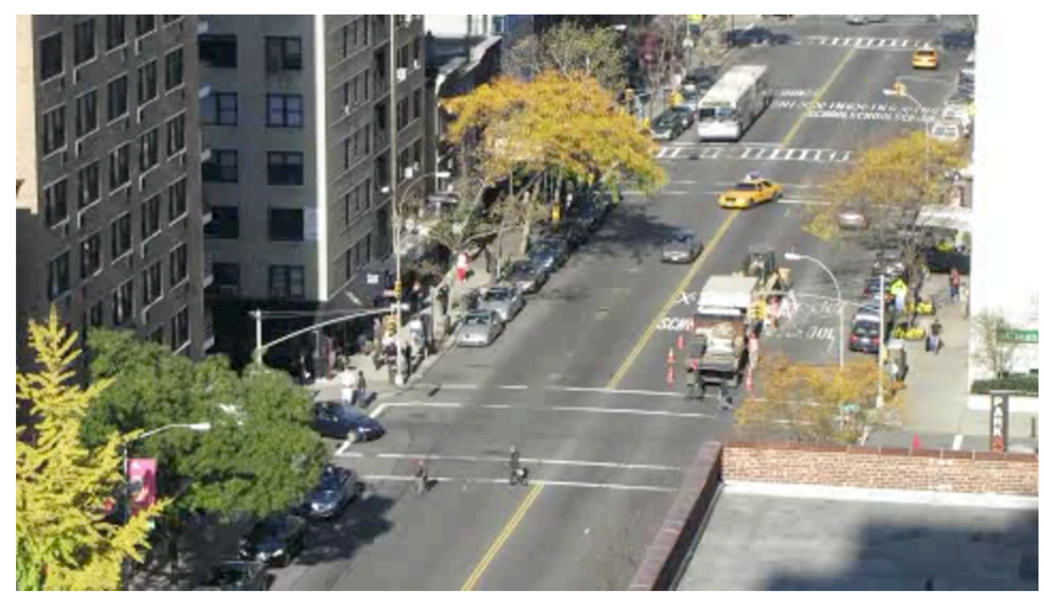
Sources: *“Sensitivity of MST neurons to optic flow stimuli. I. A continuum of response selectivity to large-field stimuli." Journal of neurophysiology 65.6 (1991).*
“A cortical representation of the local visual environment”, Nature. 392 (6676): 598–601, 2009
https://en.wikipedia.org/wiki/Two-streams_hypothesis

Sources of Motion

Static camera, moving scene

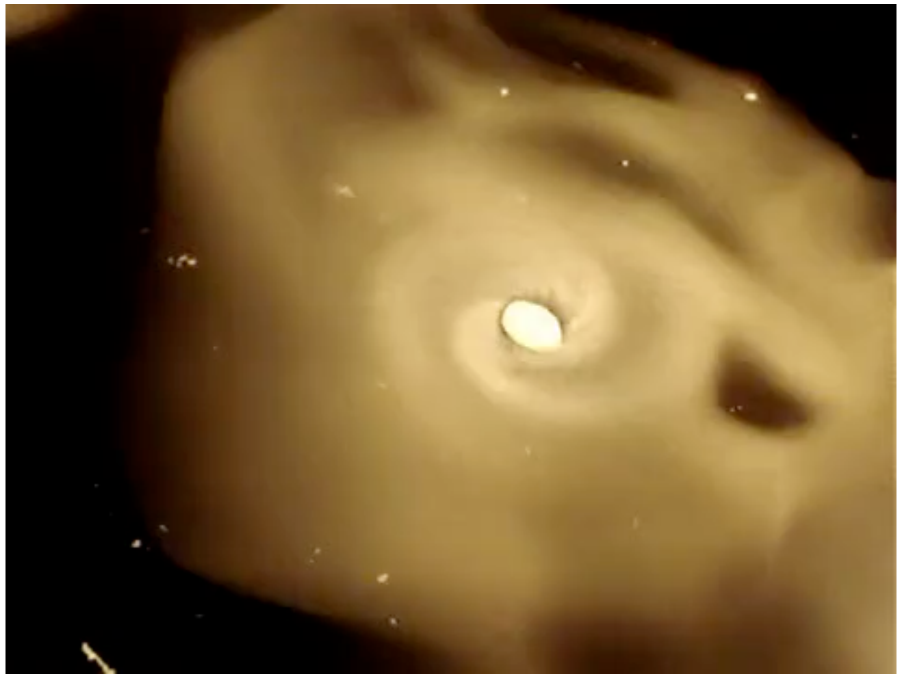
Moving camera, static scene

Moving camera and scene



Static camera, moving scene & light

Challenges



Representing Video

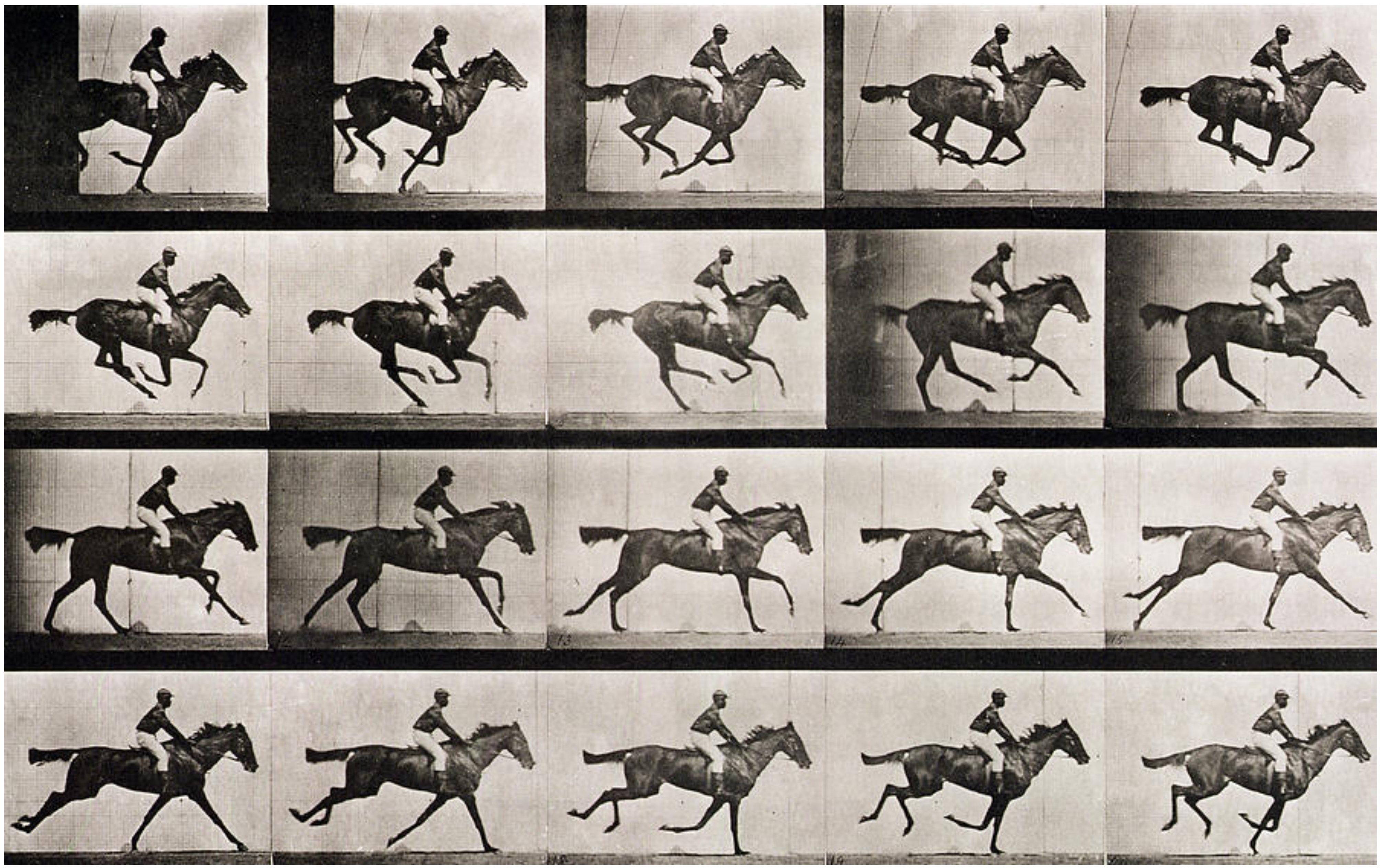
234	7	89	7	98	98	7	9	7	5
43	7	0	123	4	13	454	23	5	87
67	5	76	4	3	56	67	87	65	45
97	0	6	3	6	25	7	3	587	8
78	5	54	7	876	71	54	76	9	75
45	81	67	78	78	5	4	75	86	8
5	4	3	35	8	256	6	4	3	36
7	6	64	3	4	7	77	76	4	54
64	35	46	46	64	56	7	56	4	7
75	464	576	75	75	75	57	64	75	75

Height

Width

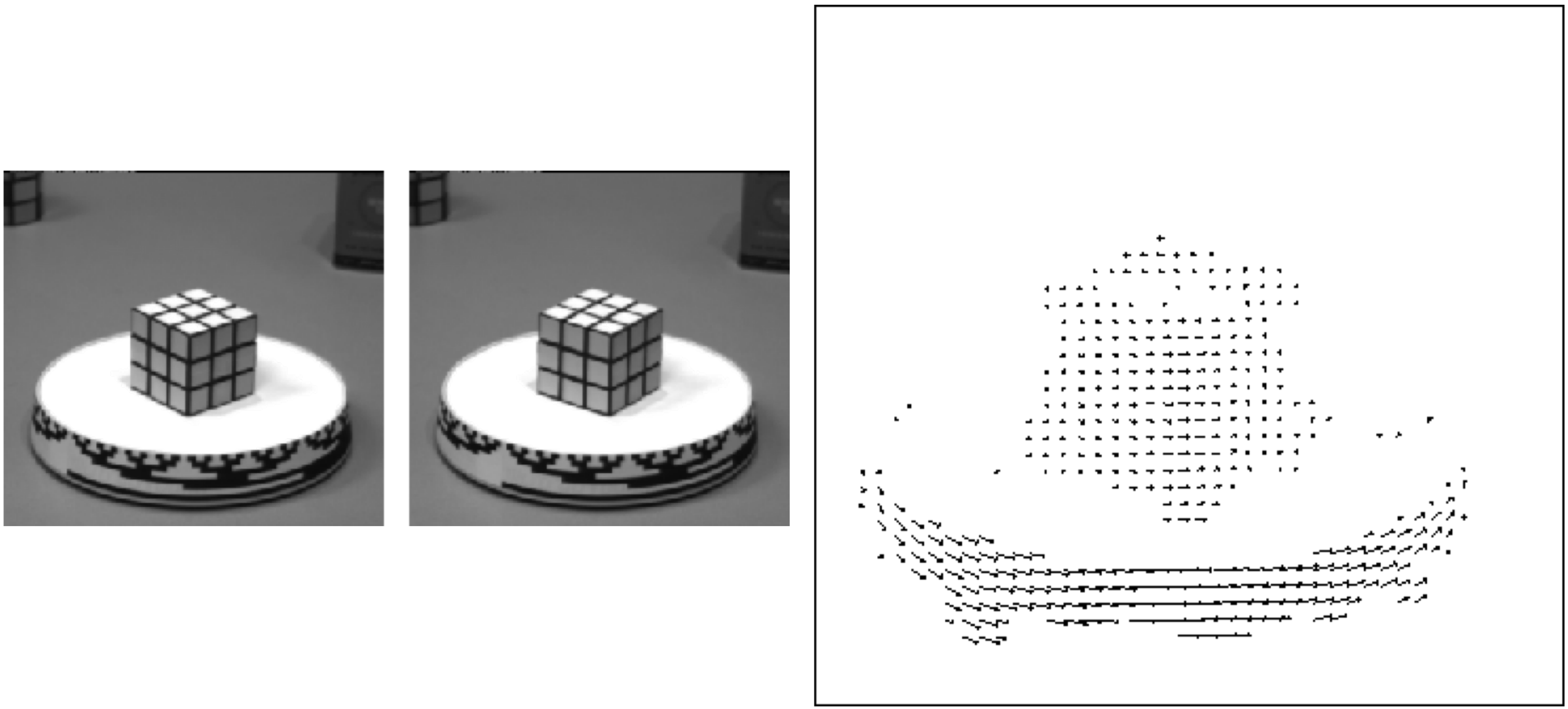
Time

Representing Video



Eadweard Muybridge's stop motion for studying animal location

Optical Flow

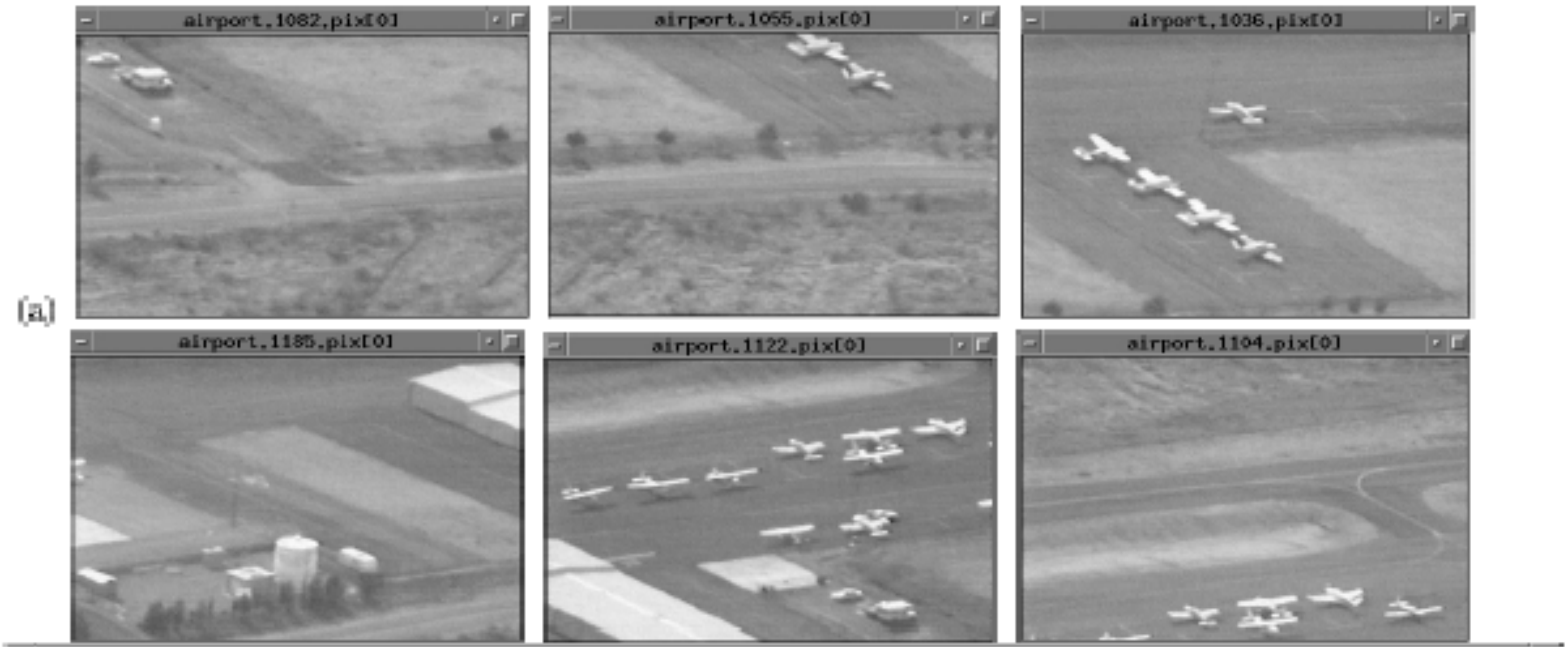


Will start by estimating motion of each pixel separately

Why estimate motion?

- Feature representation for DeepNets [coming up]
- Track object behavior
- Correct for camera jitter (stabilization)
- Align images (mosaics)
- 3D shape reconstruction
- Special effects

Mosaicing



(Michal Irani, Weizmann)

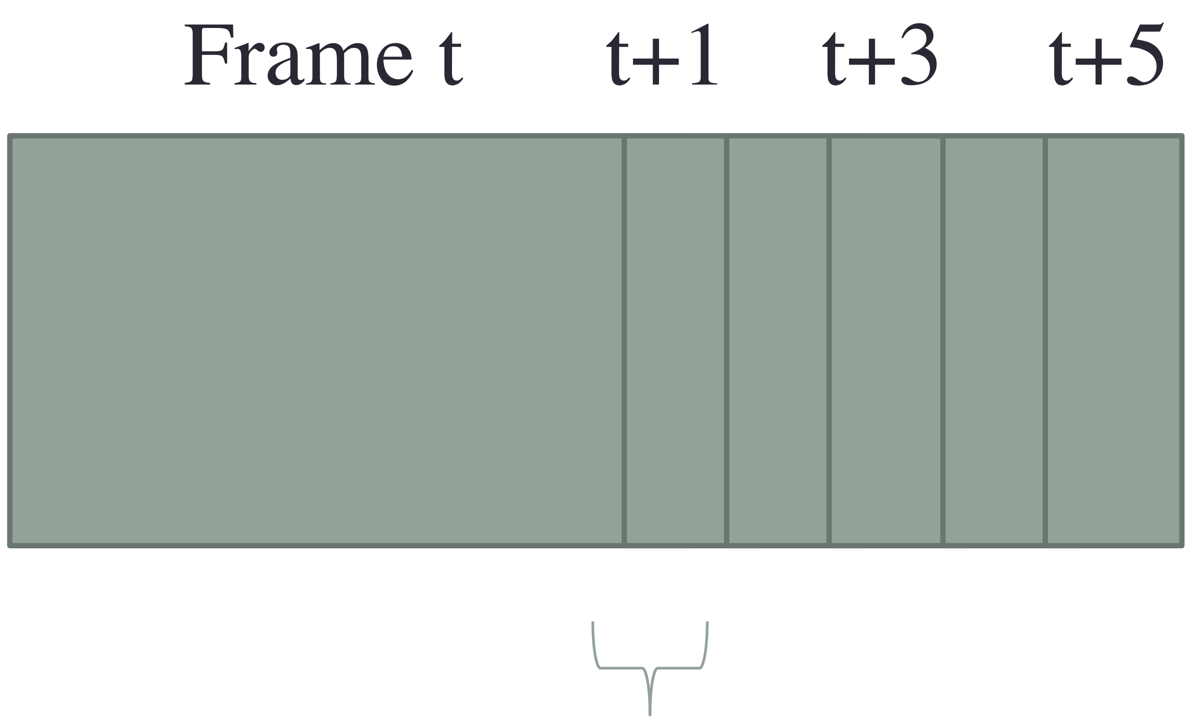
Mosaicing

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

(Michal Irani, Weizmann)

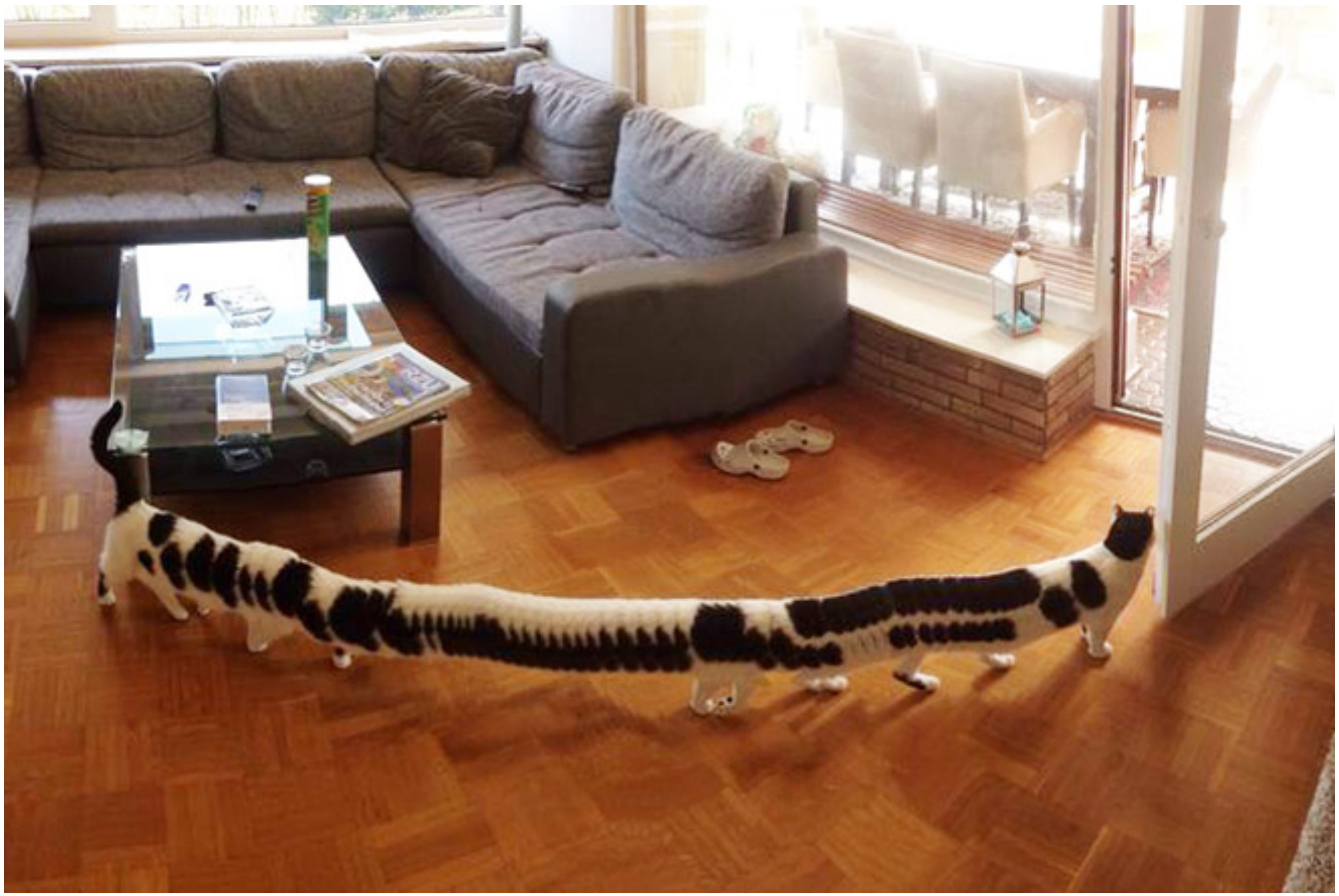
Mosaicing for Panoramas on Smartphones

Left to right sweep of video camera



Compare small overlap for efficiency

Mosaicing for Panoramas



Mosaicing for Panoramas

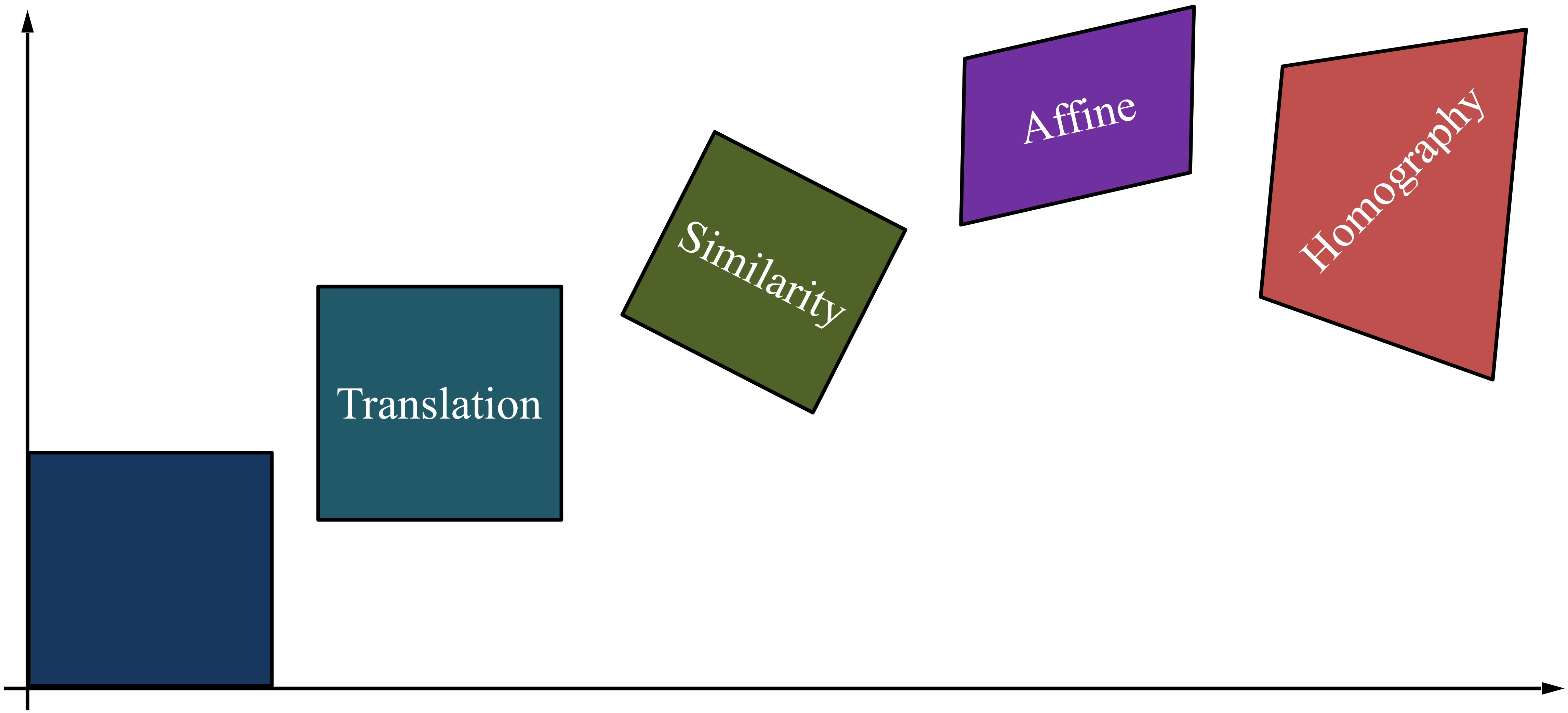
Mosaicing for Panoramas

Mosaicing for Panoramas

Anyone have funny failures from HW4?

Parametric Motion Models

The previous models we used are too restricted to describe arbitrary motion

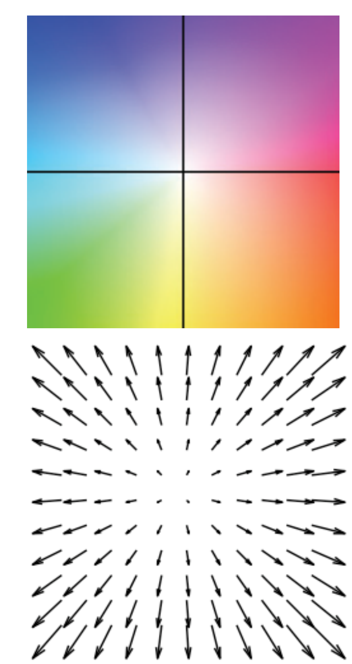


Optical Flow

- Optical flow field: assign a flow vector to each pixel
- Visualize: flow magnitude as saturation, orientation as hue

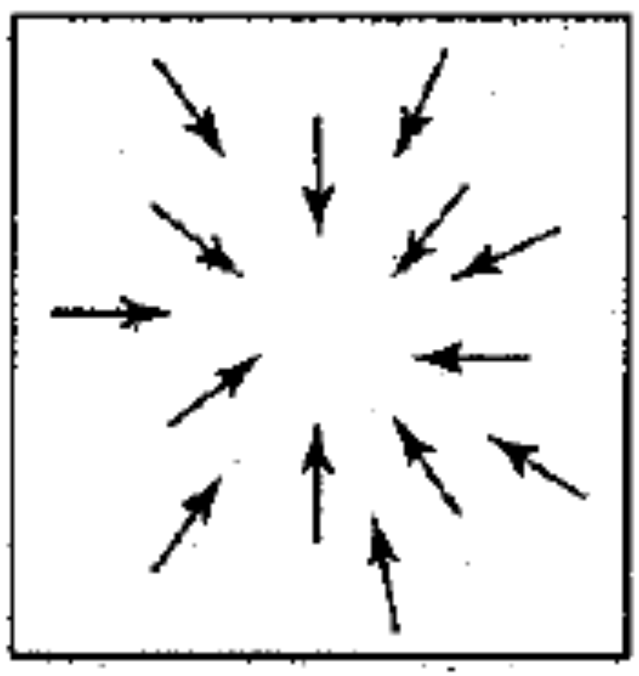
Input two frames

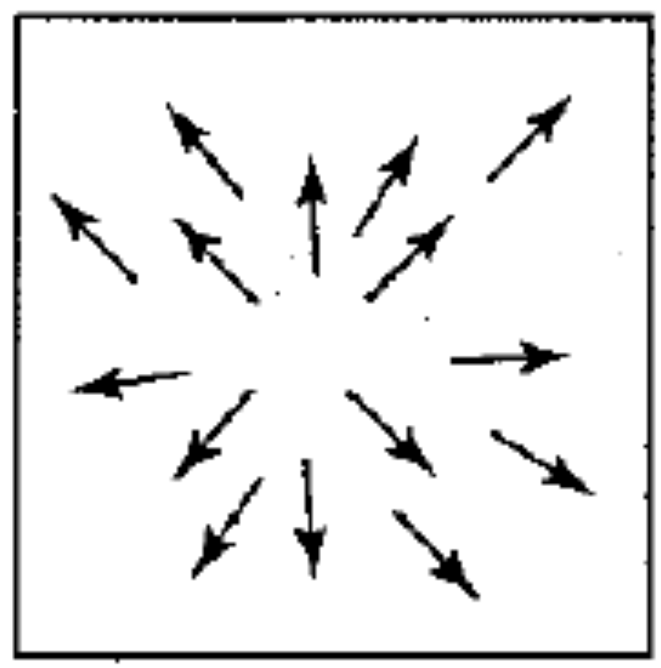
Ground-truth flow field

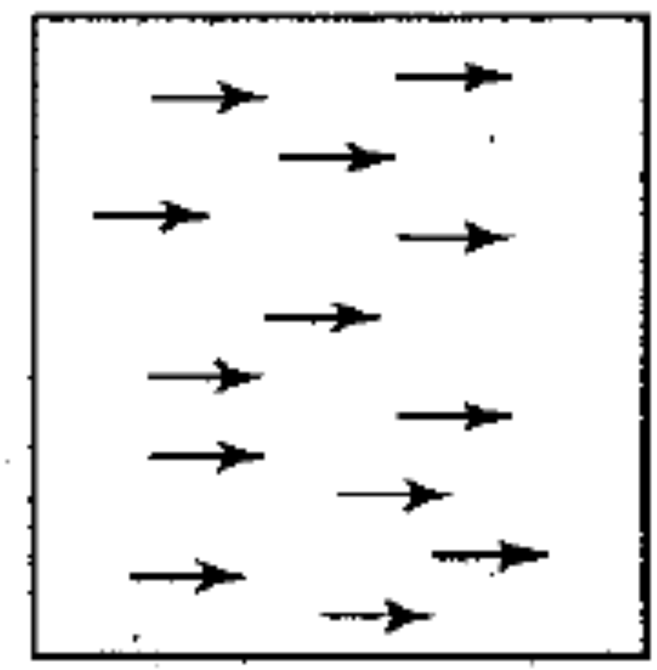


Visualization code [Baker et al. 2007]

Motion Fields

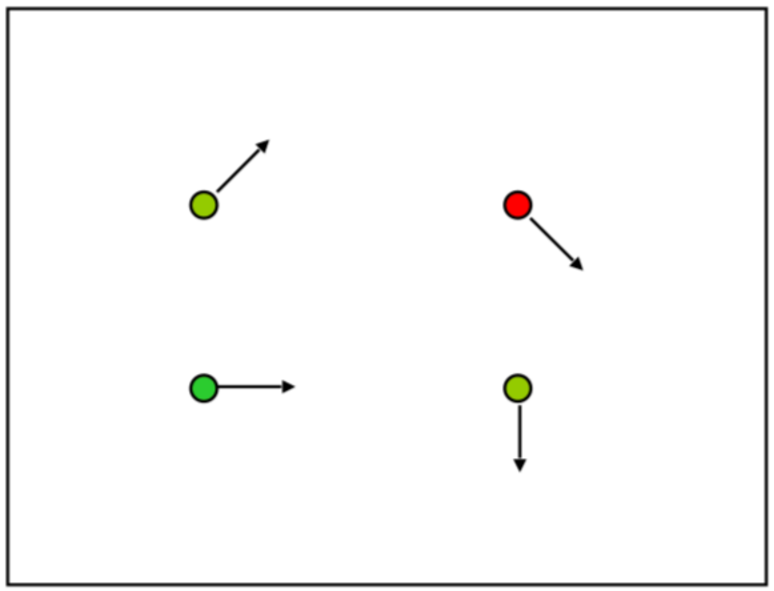




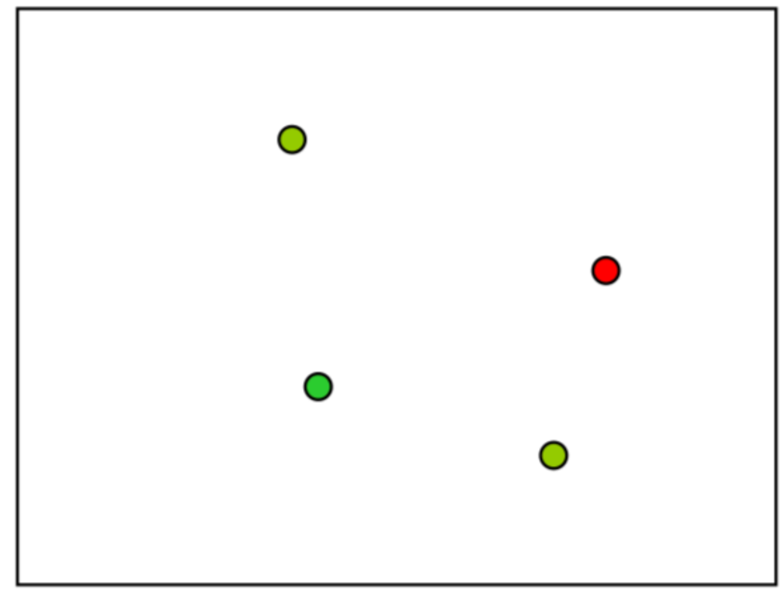


Zoom out **Zoom in** **Pan right to left**

Optical Flow



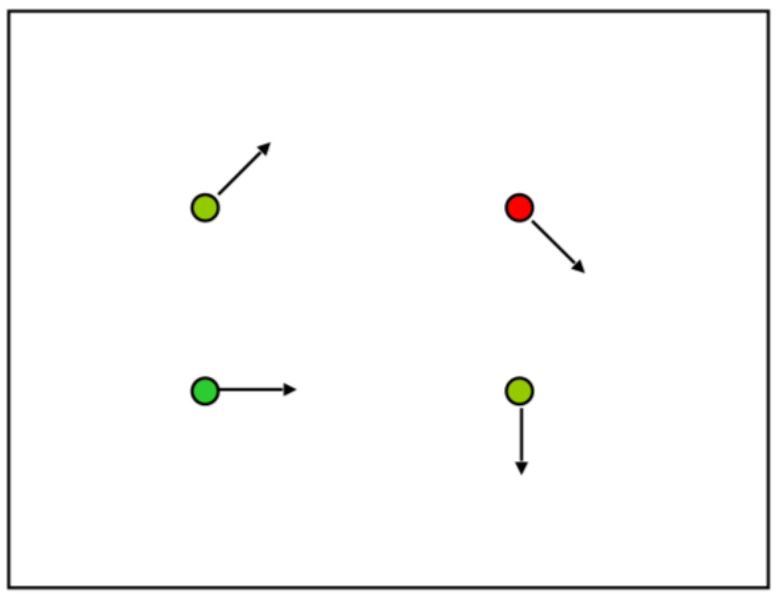
$H(x, y)$



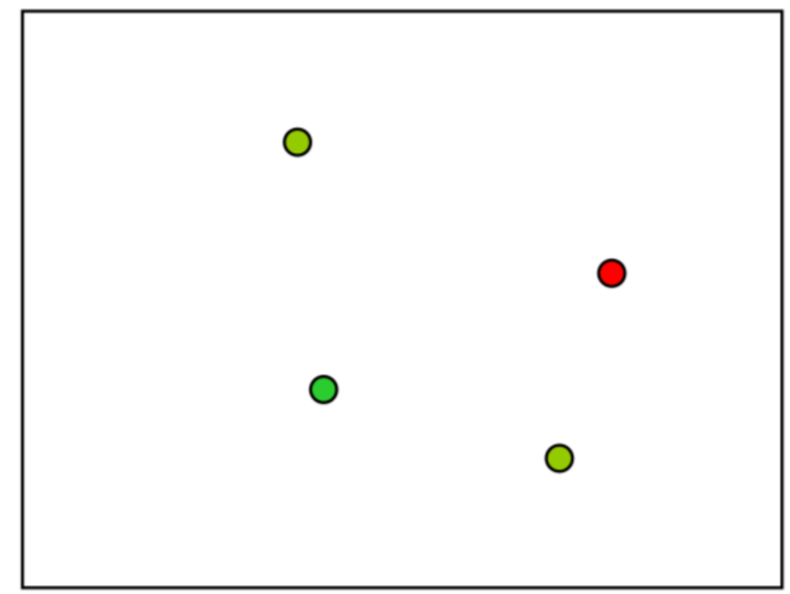
$I(x, y)$

How to estimate pixel motion from image H to image I?

Optical Flow



$H(x, y)$



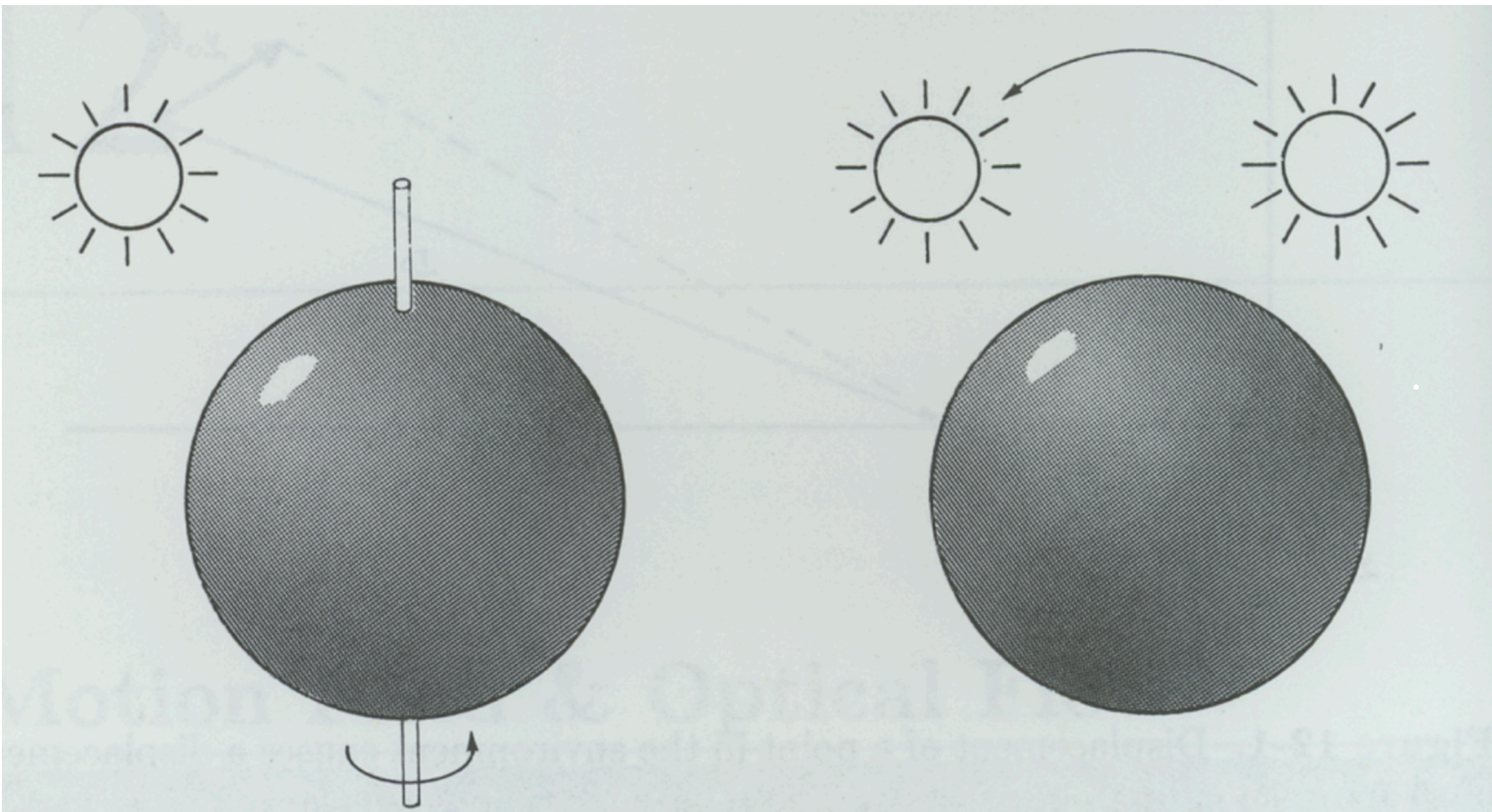
$I(x, y)$

How to estimate pixel motion from image H to image I?

Given pixel in H, look for **nearby** pixels in I that have same color.

This is called **color/brightness constancy assumption**

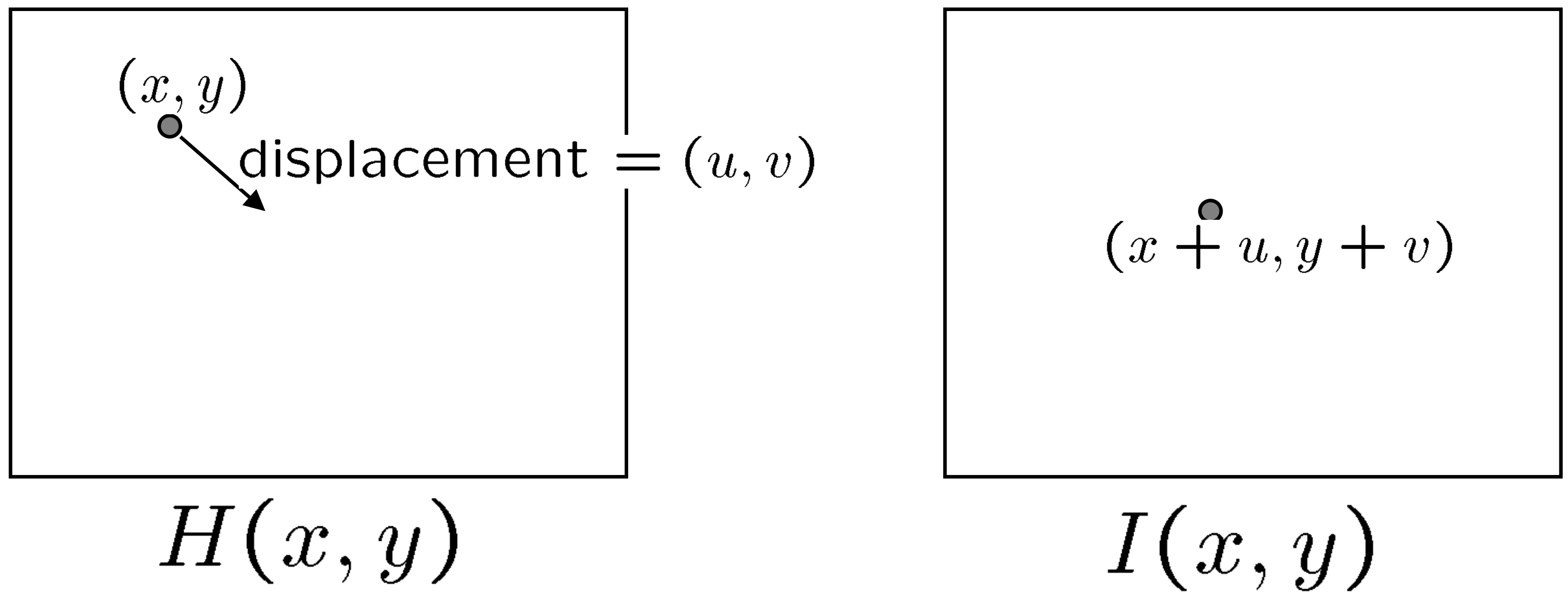
Apparent Motion versus Optical Flow



Motion field exists but no optical flow
(a)

No motion field but shading changes
(b)

Optical Flow Constraints



Let's look at these constraints more closely

- brightness constancy:

$$H(x,y) = I(x+u, y+v)$$

- small motion:

$$I(x+u, y+v) = I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v + \text{higher order terms}$$
$$\approx I(x,y) + \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v$$

Optical Flow Constraints

Combining these two equations

$$
\begin{aligned}
0 &= I(x+u, y+v) - H(x,y) \\
&\approx I(x,y) + I_x u + I_y v - H(x,y)
\end{aligned}
$$

shorthand: $I_x = \frac{\partial I}{\partial x}$

Optical Flow Constraints

Combining these two equations

$$
\begin{aligned}
0 &= I(x+u, y+v) - H(x,y) \\
&\approx I(x,y) + I_x u + I_y v - H(x,y) \\
&\approx (I(x,y) - H(x,y)) + I_x u + I_y v \\
&\approx I_t + I_x u + I_y v
\end{aligned}
$$

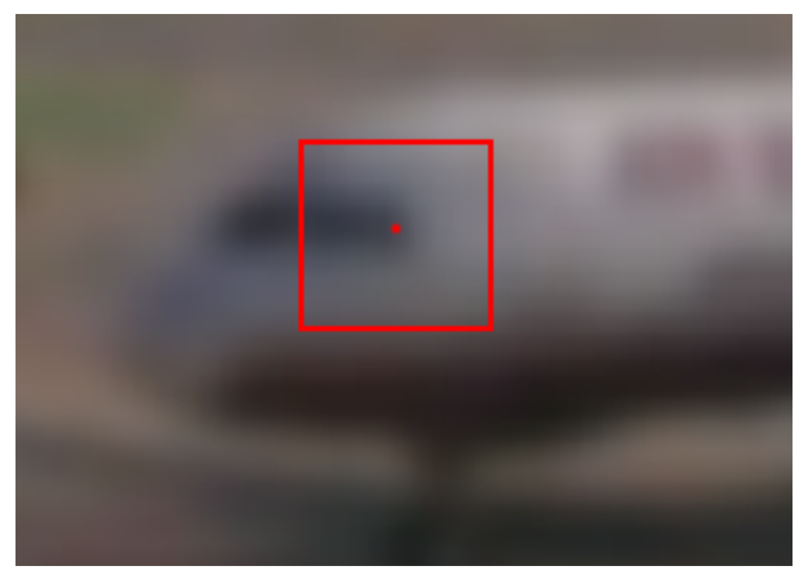
shorthand: $I_x = \frac{\partial I}{\partial x}$

How does this make sense?

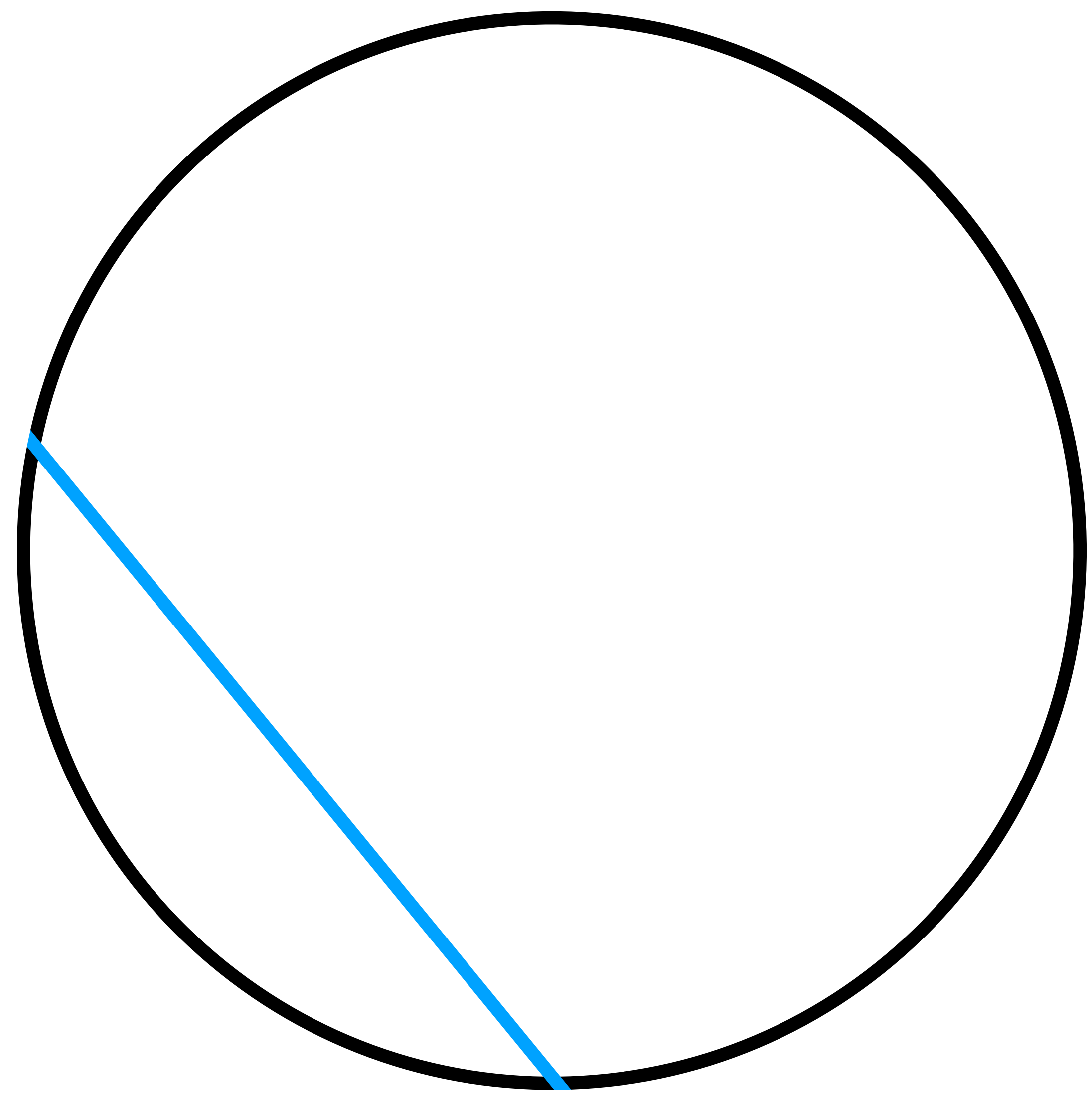
Brightness constancy constraint equation

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

What do the static image gradients
have to do with motion estimation?

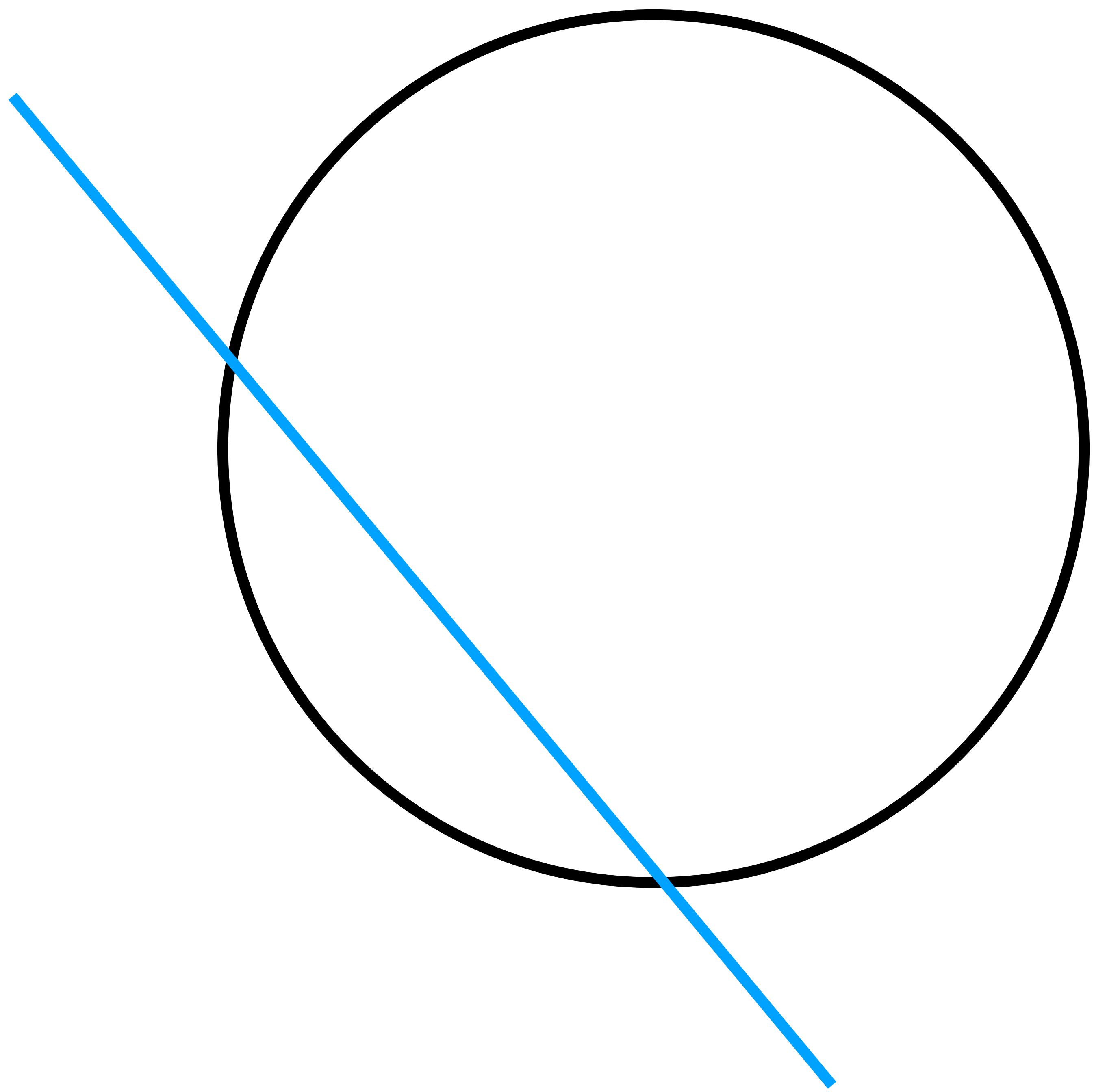


Aperture Problem



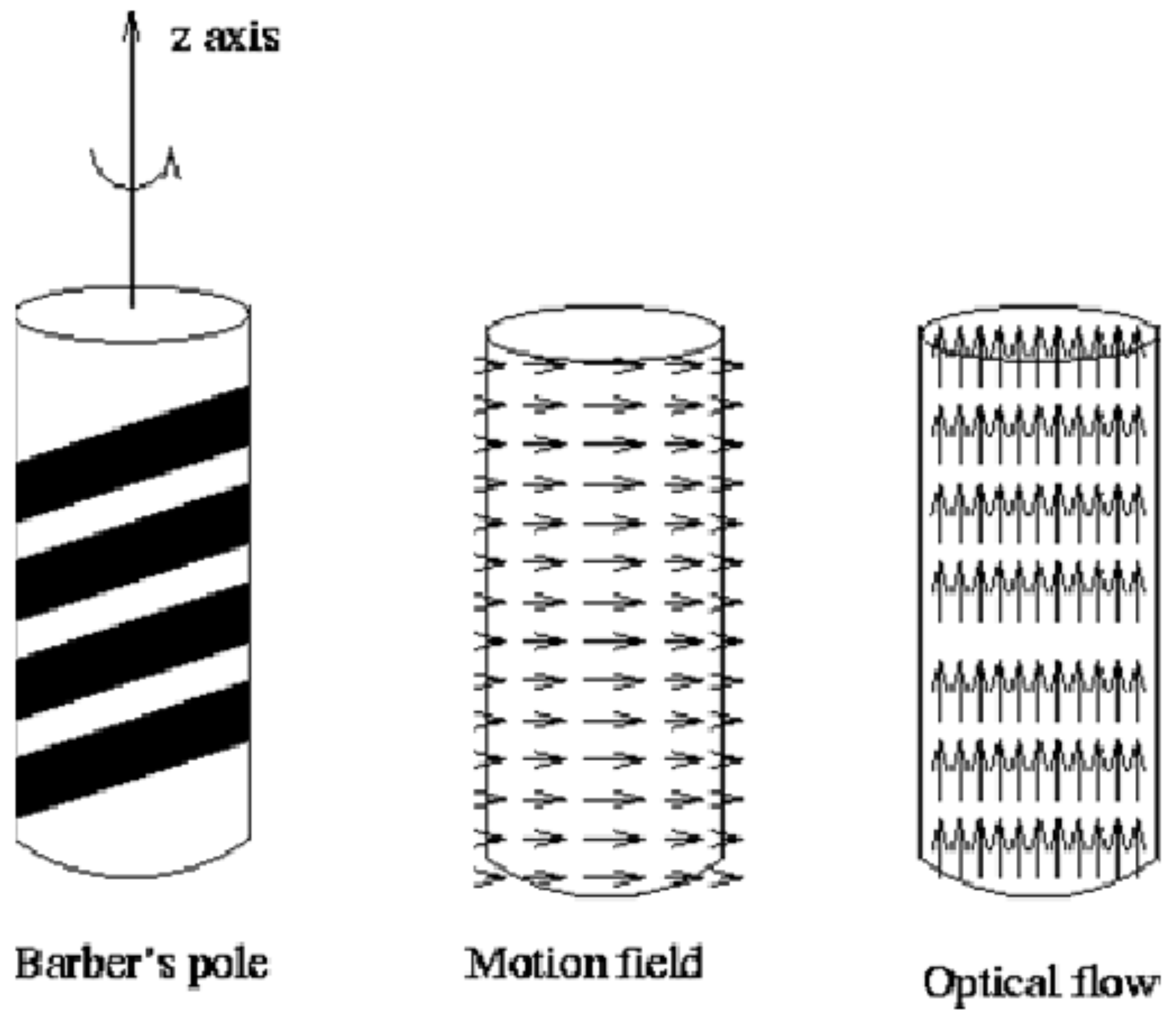
Which way did
the line move?

Aperture Problem



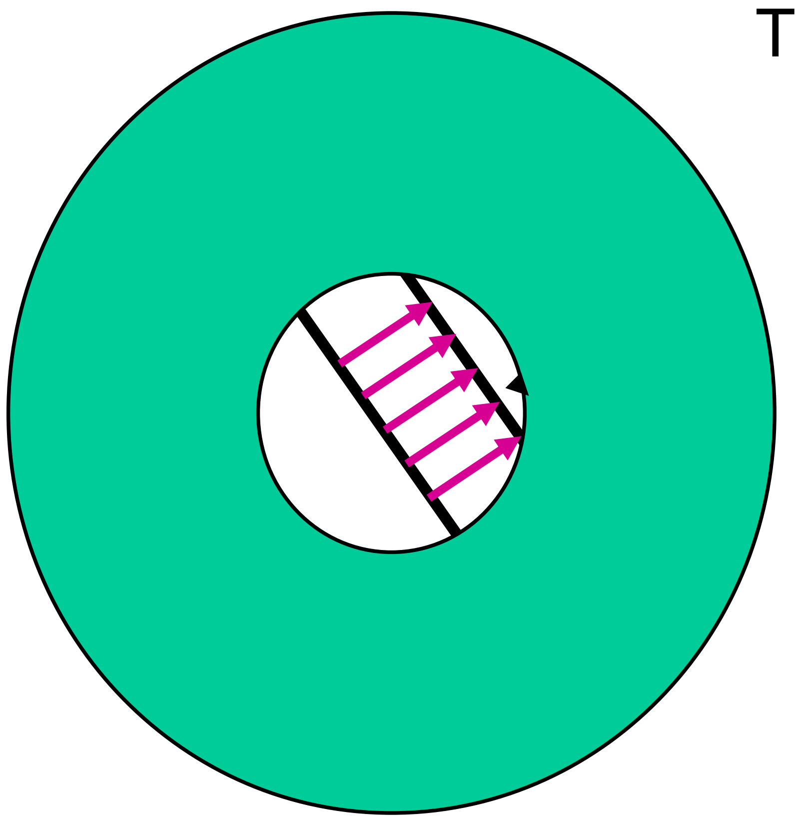
Which way did the line move?

The barber pole illusion



http://en.wikipedia.org/wiki/Barberpole_illusion

Aperture Problem



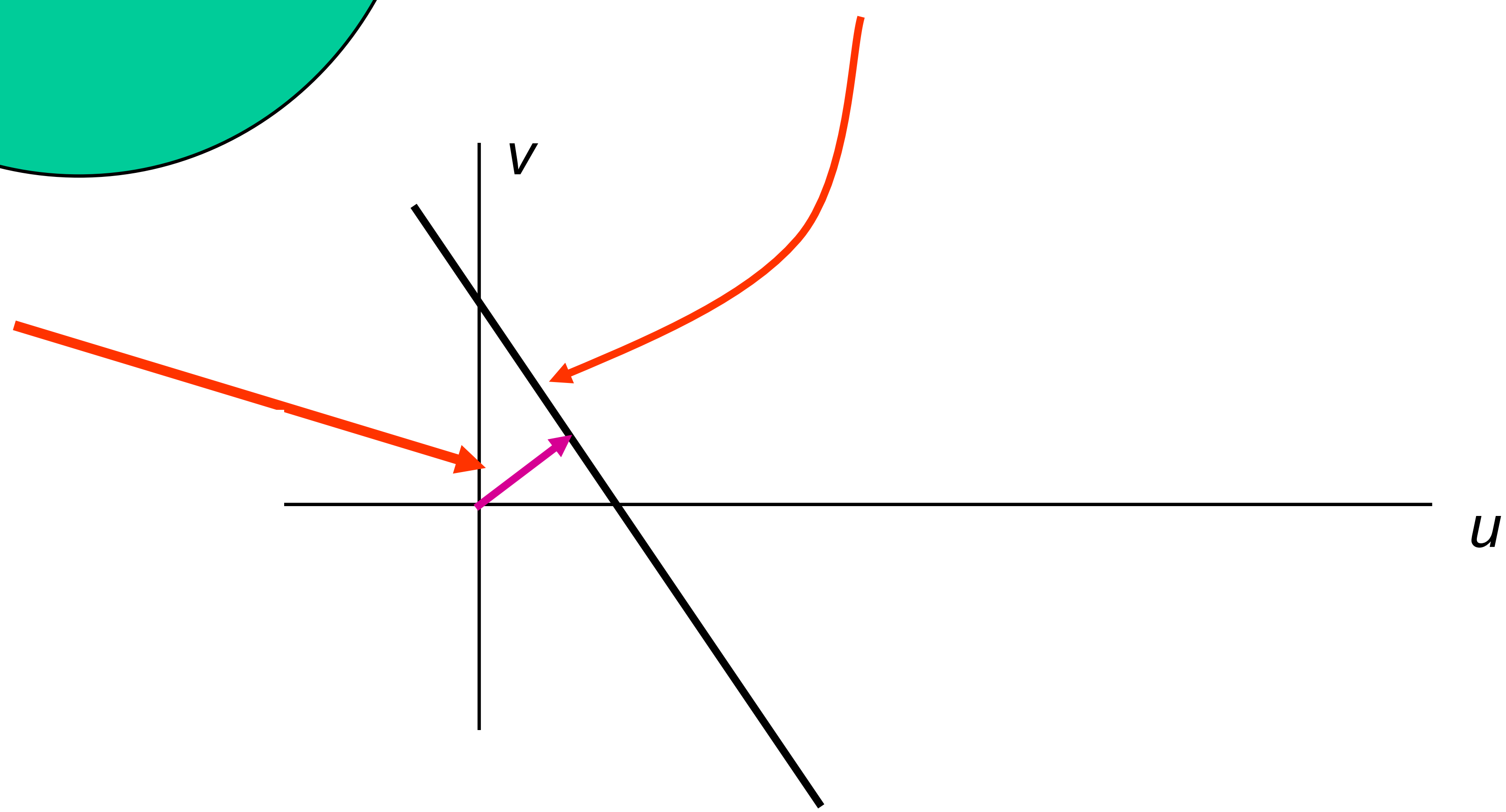
The gradient constraint:

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

$$\nabla I \bullet \vec{U} = 0$$

Defines a line in the *(u,v)* space

Normal Flow:



Solving the aperture problem

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

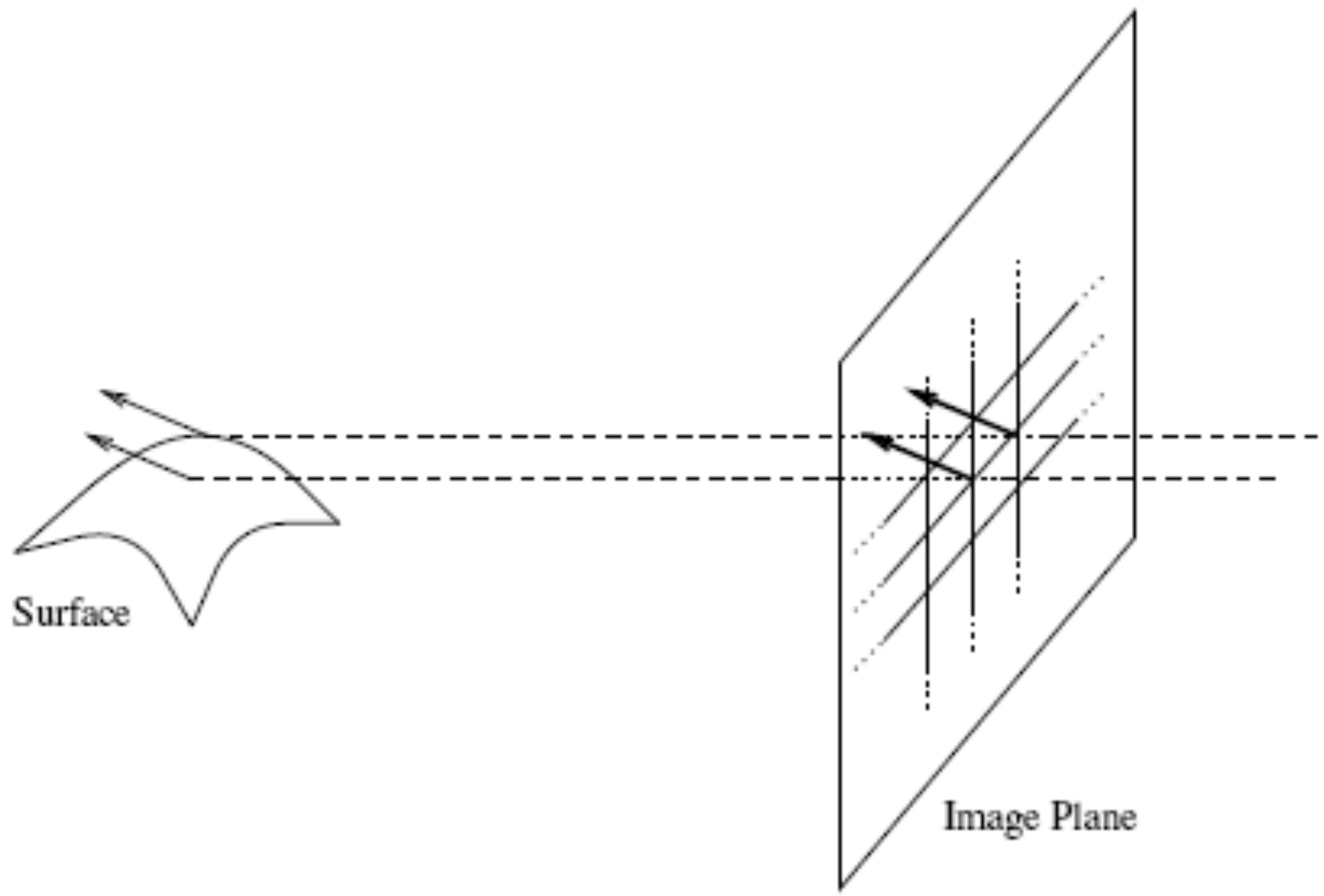


Figure 1.7: Spatial coherence assumption. Neighboring points in the image are assumed to belong to the same surface in the scene.

Figure by Michael Black

Solving the aperture problem

- How to get more equations for a pixel?
- **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

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

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

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

Slide credit: Steve Seitz

Solving the aperture problem

Problem: we have more equations than unknowns

$$\underset{25\times2}{A}\ \underset{2\times1}{d} = \underset{25\times1}{b} \longrightarrow \text{minimize } \|Ad - b\|^2$$

Slide credit: Steve Seitz

Solving the aperture problem

Problem: we have more equations than unknowns

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

Solution: solve least squares problem

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

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

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

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

Slide credit: Steve Seitz

Conditions for solvability

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

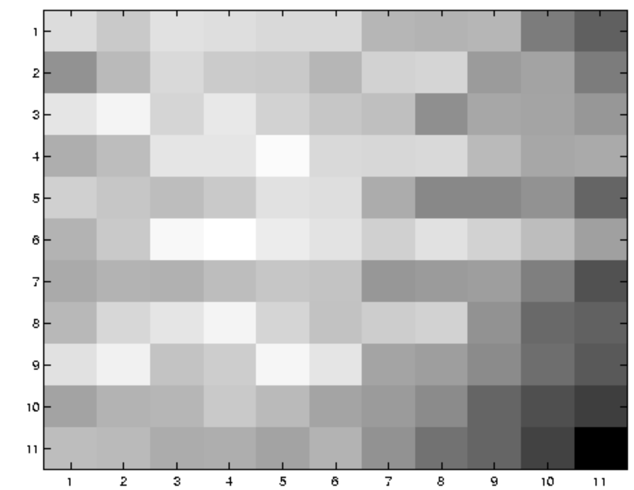
$$A^T A \qquad\qquad\qquad\qquad A^T b$$

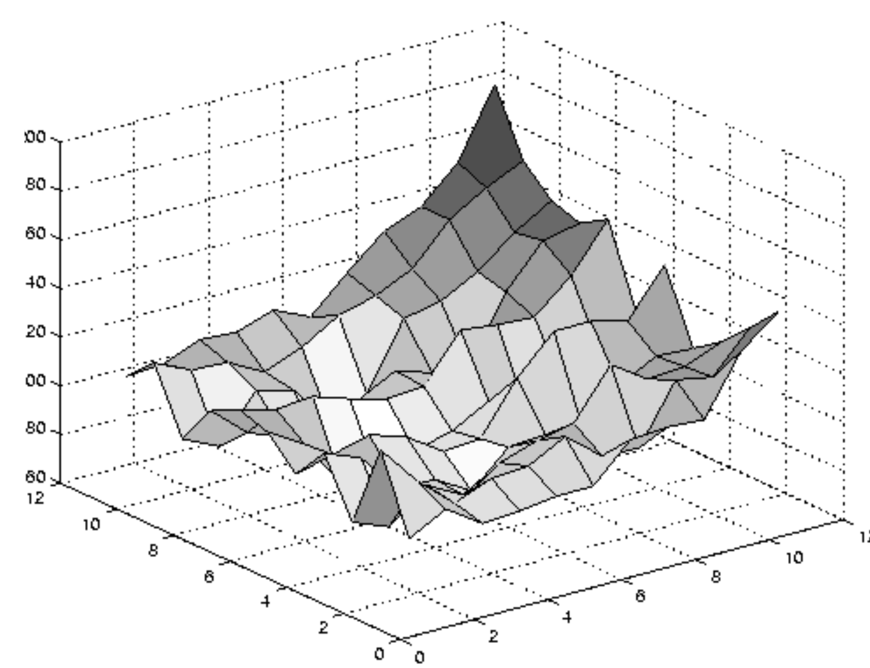
When is this solvable?

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

Where have we seen this before?

Low texture region



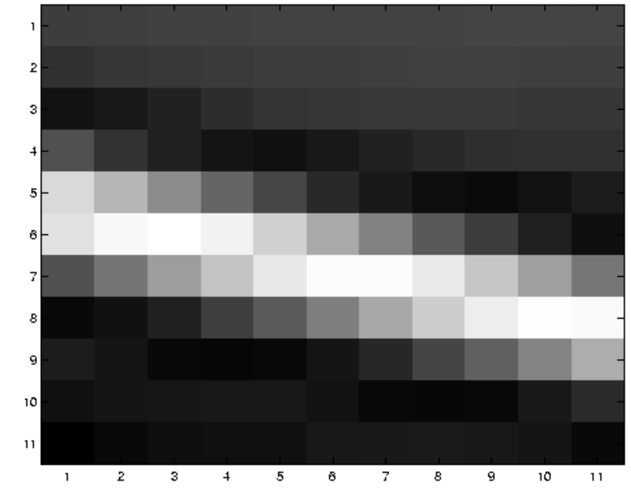


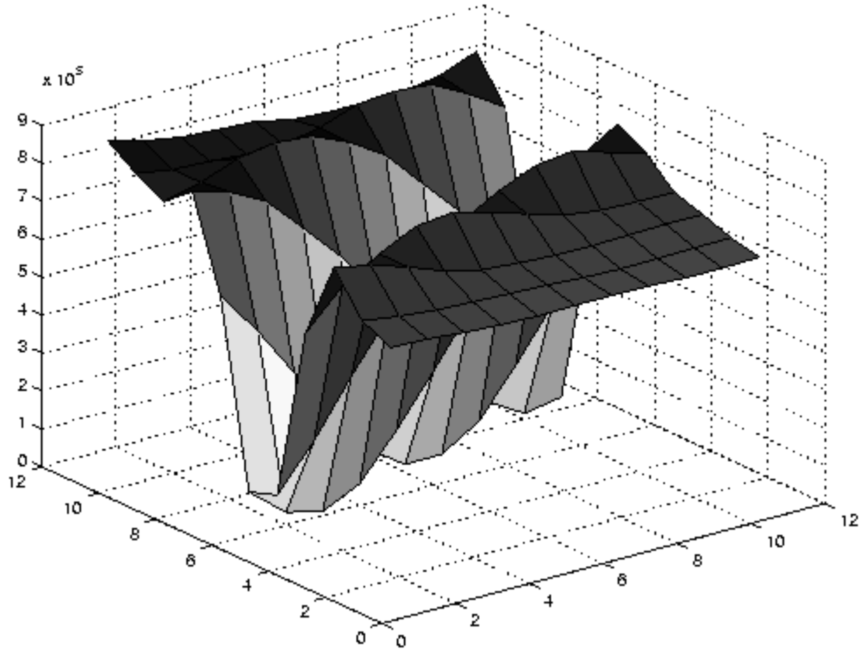
$$\sum \nabla I (\nabla I)^T$$

– gradients have small magnitude

– small λ_1, small λ_2

Edge

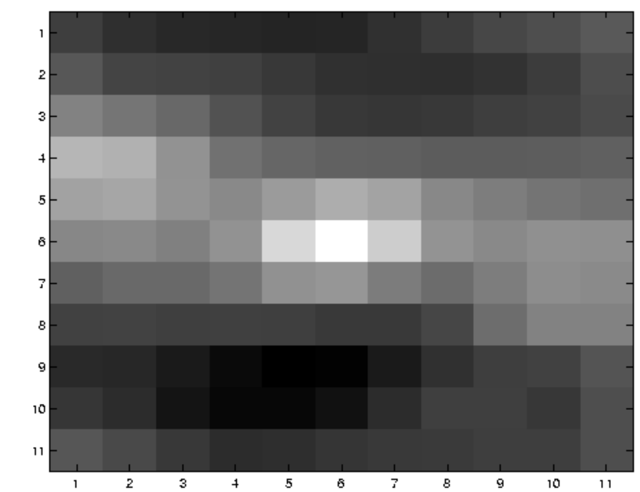


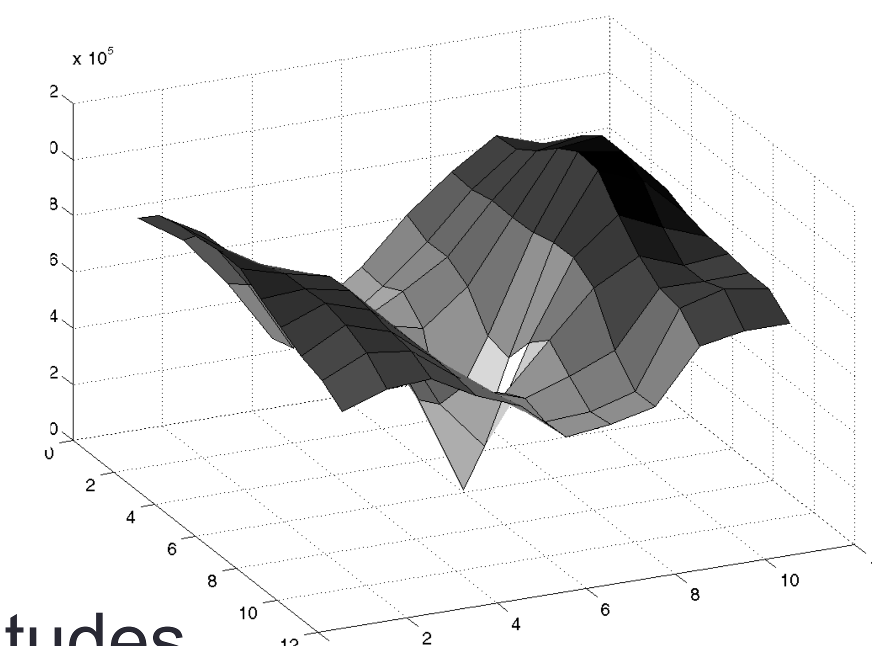


$$\sum \nabla I (\nabla I)^T$$

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

High textured region



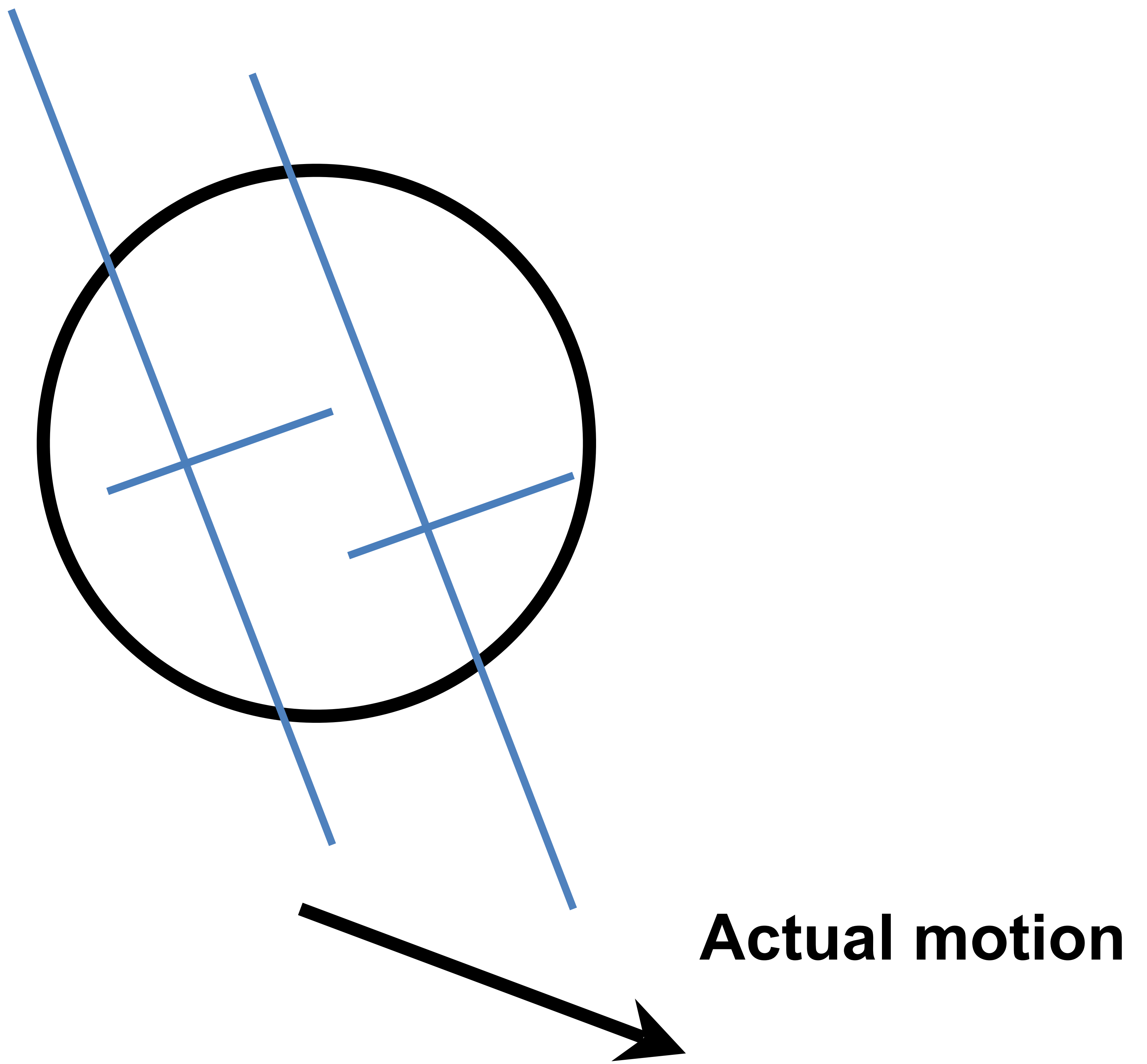


$$\sum \nabla I (\nabla I)^T$$

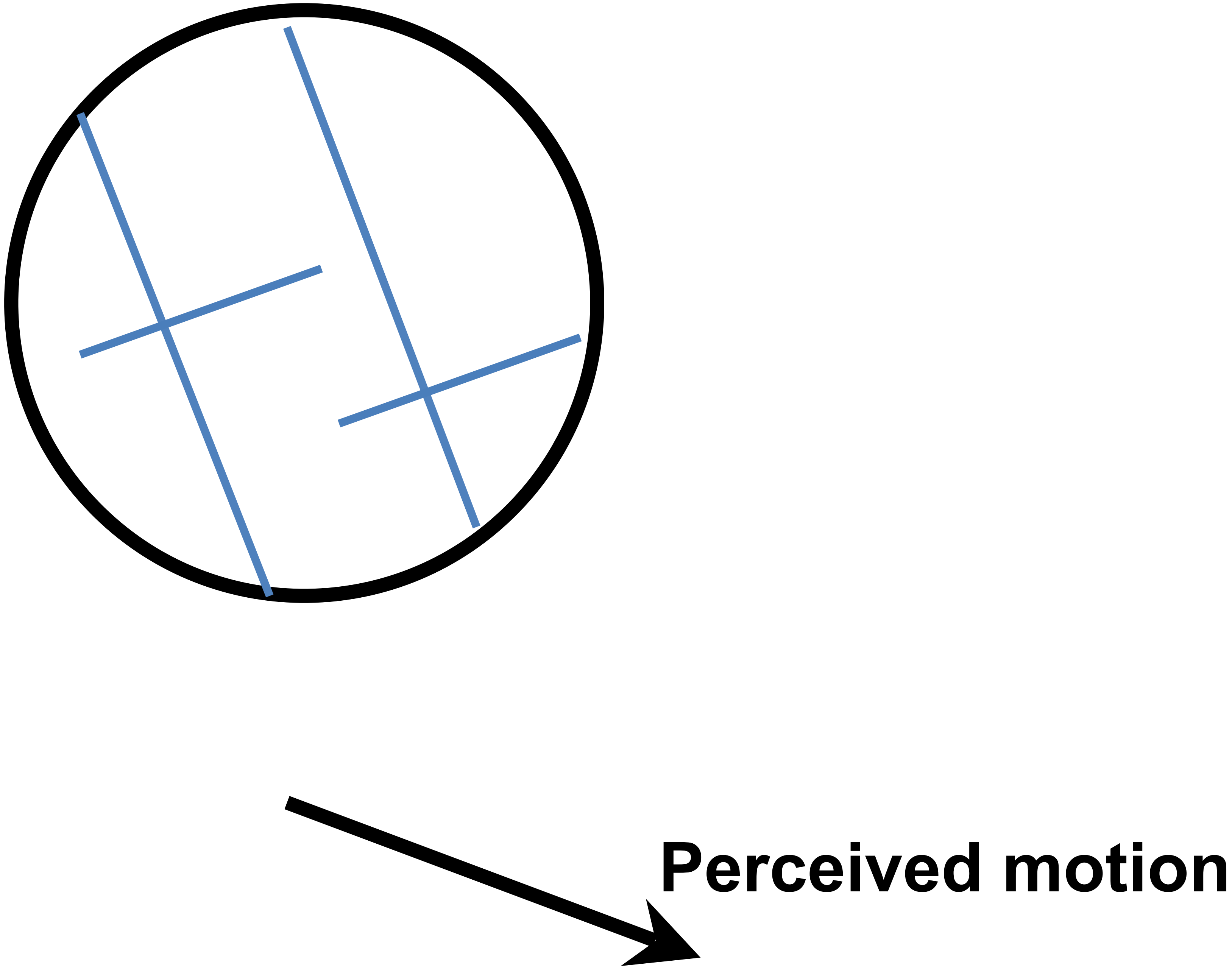
– gradients are different, large magnitudes

– large λ_1, large λ_2

The aperture problem resolved



The aperture problem resolved

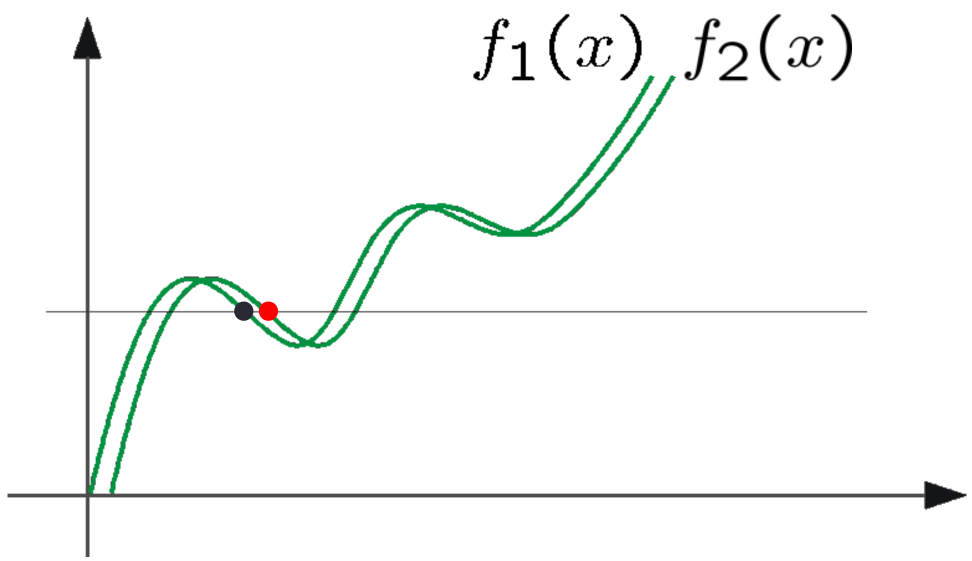


Revisiting the small motion assumption

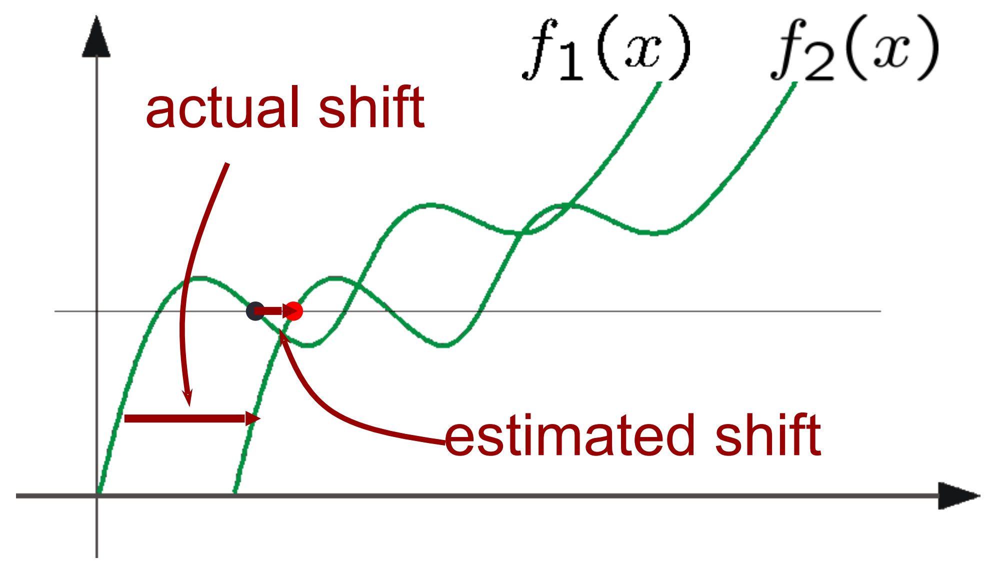
- Is this motion small enough?
  - Probably not—it's much larger than one pixel
  - How might we solve this problem?

Optical Flow: Aliasing

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

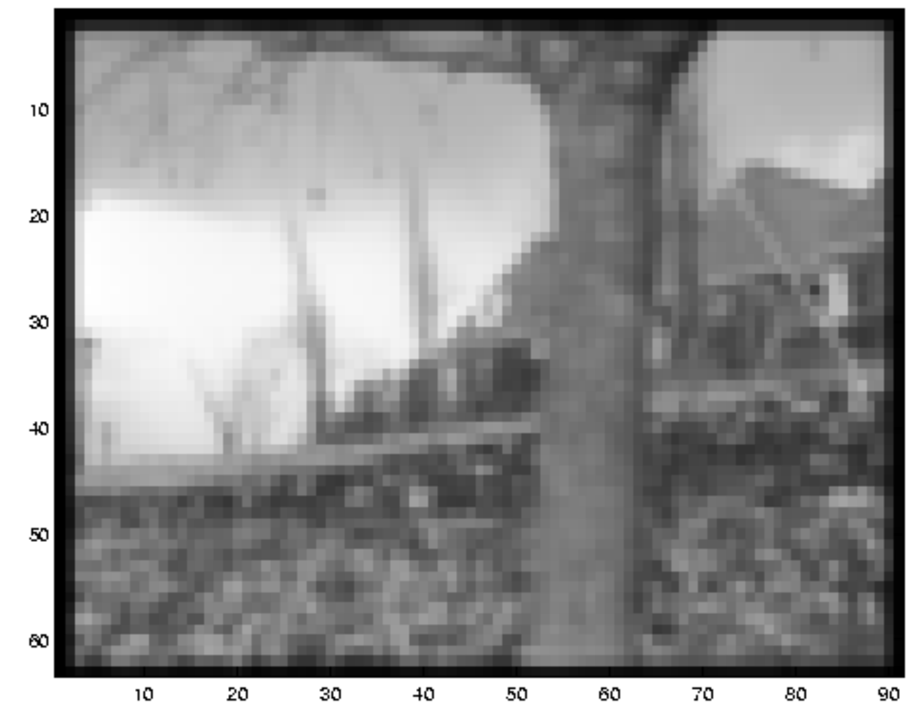


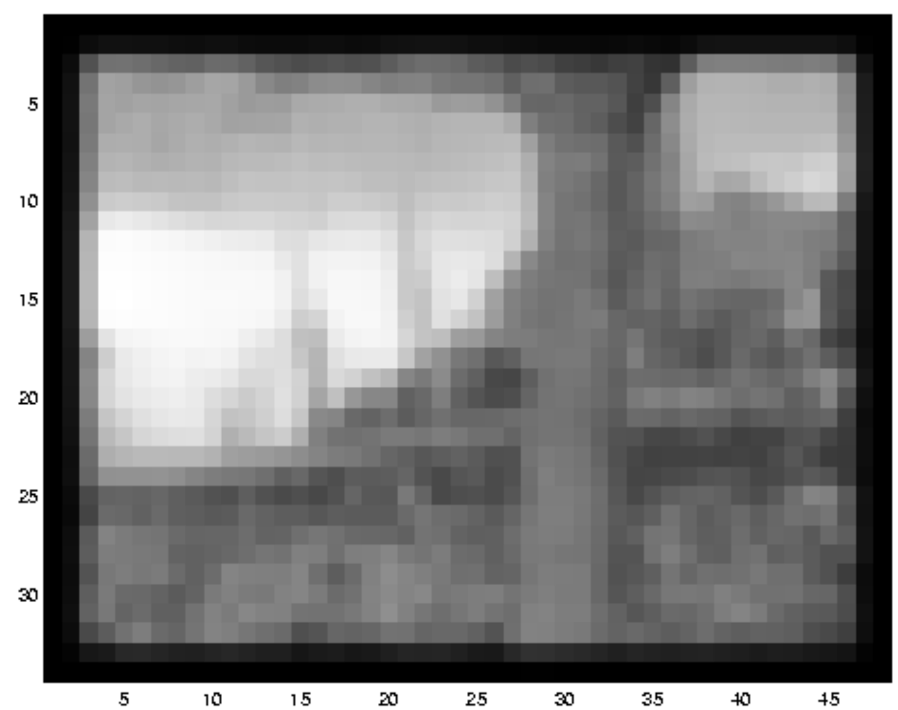
nearest match is correct (no aliasing)

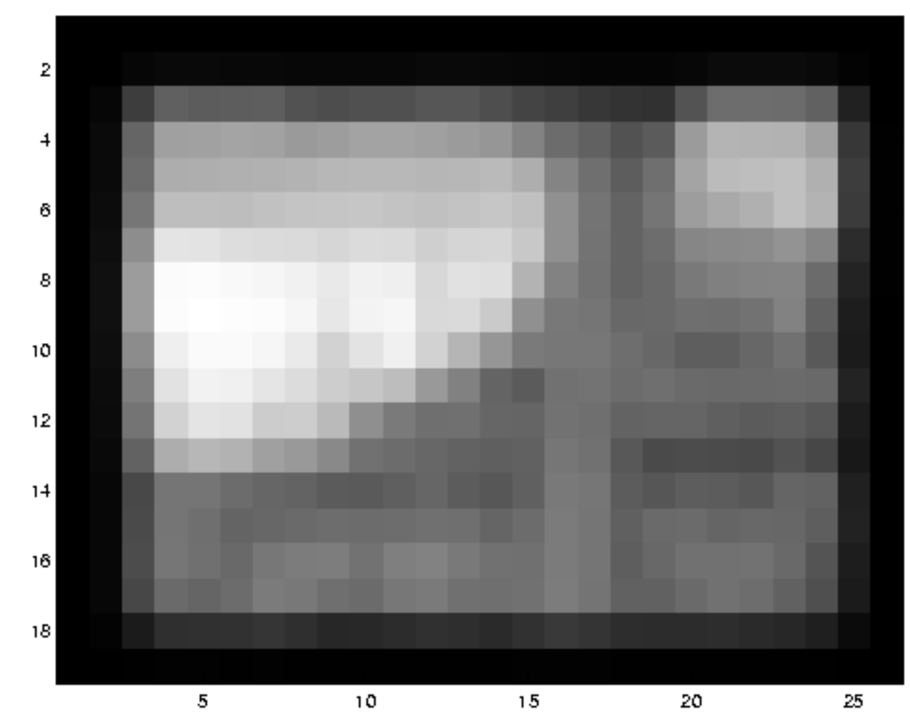


nearest match is incorrect (aliasing)

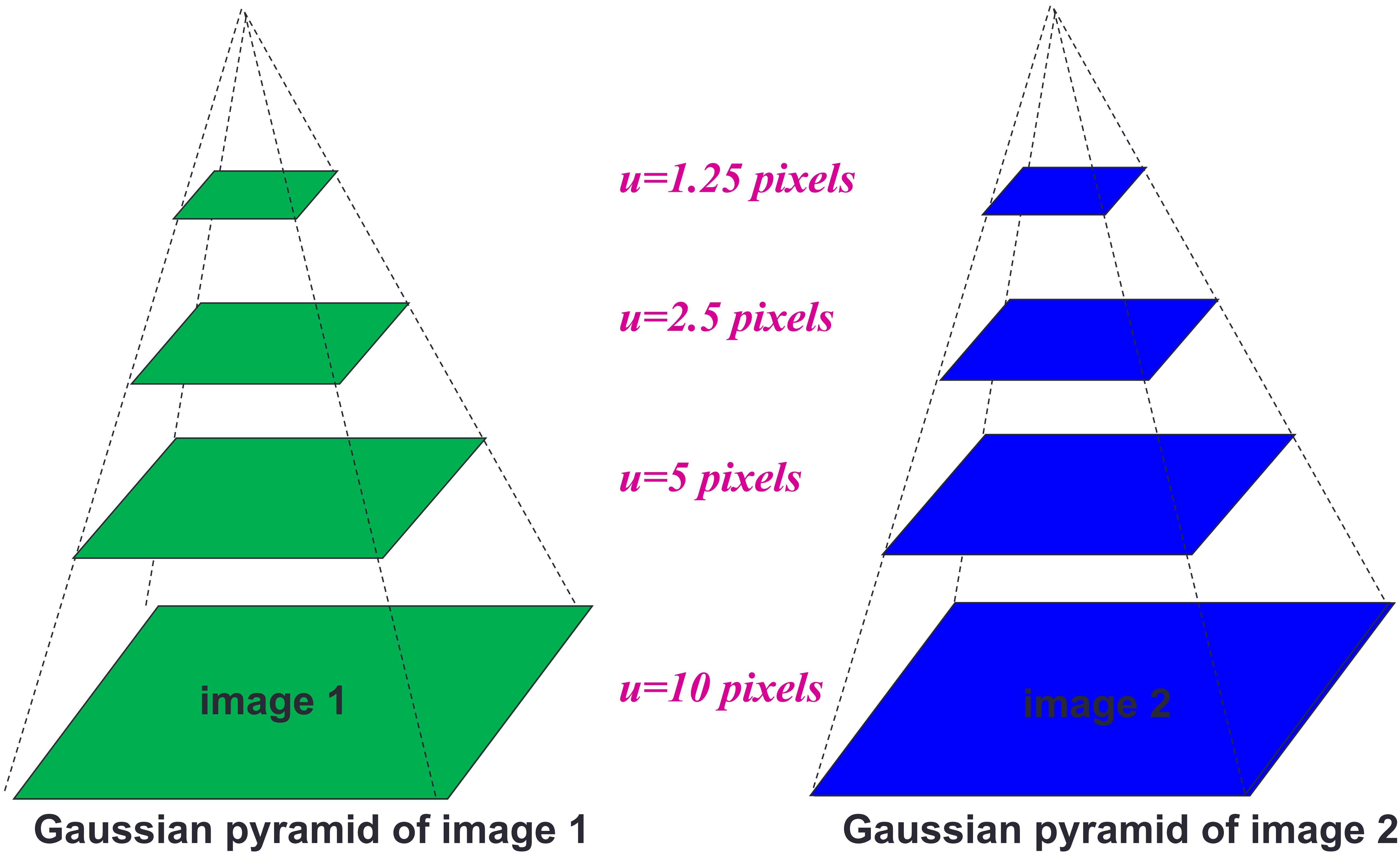
Reduce the resolution



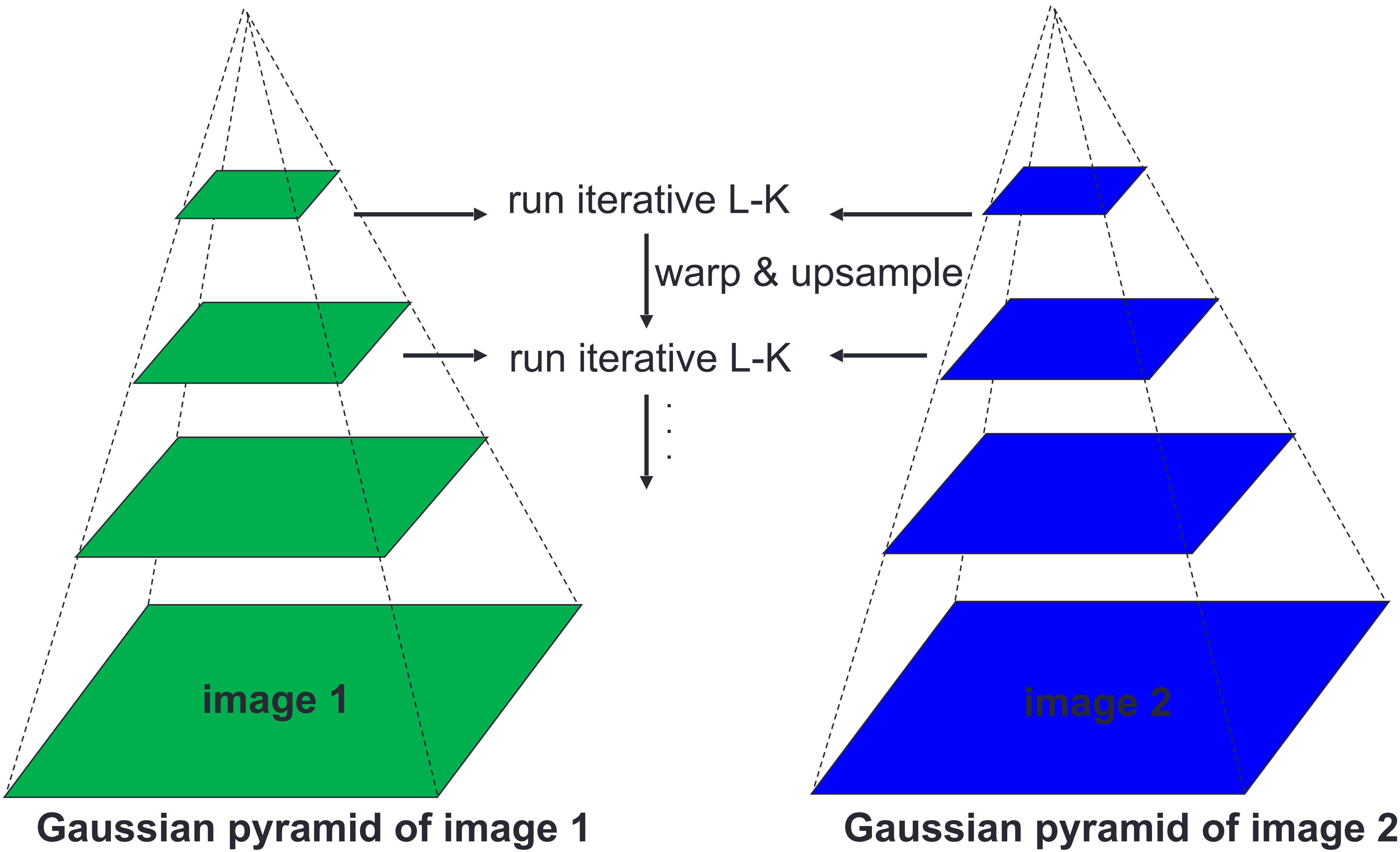




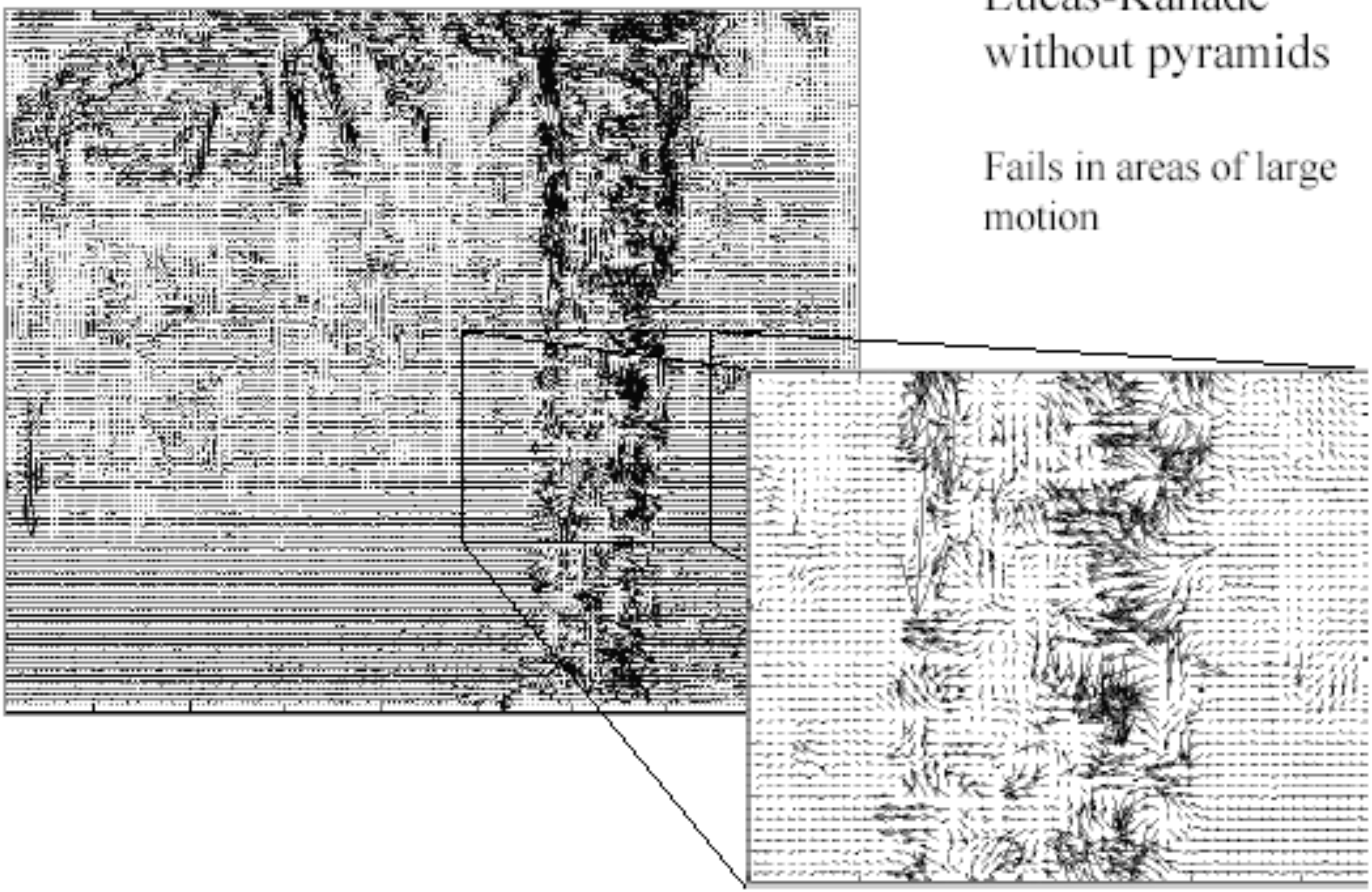
Coarse-to-fine optical flow estimation



Coarse-to-fine optical flow estimation



Optical Flow Results

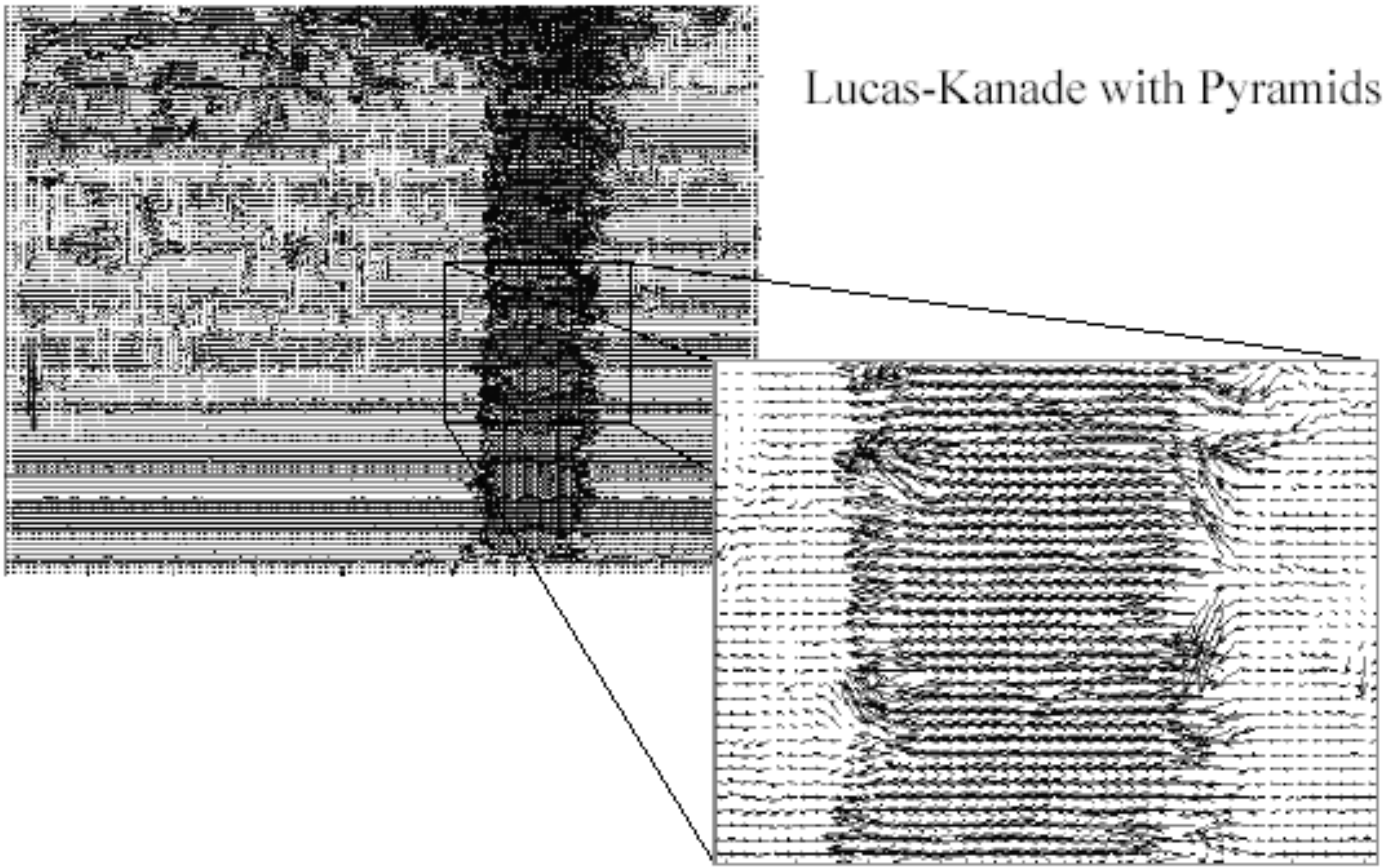


Lucas-Kanade without pyramids

Fails in areas of large motion

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003

Optical Flow Results



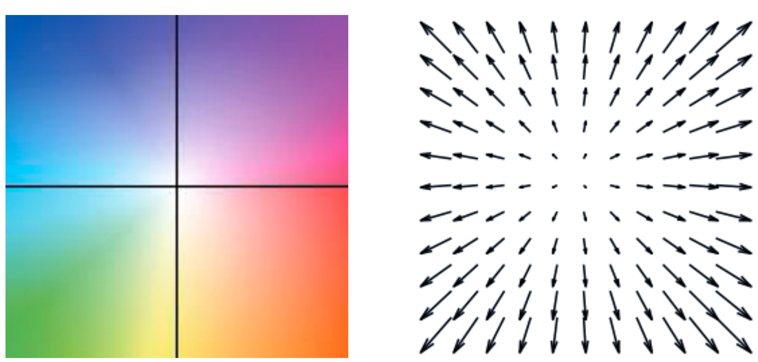
Lucas-Kanade with Pyramids

* From Khurram Hassan-Shafique CAP5415 Computer Vision 2003

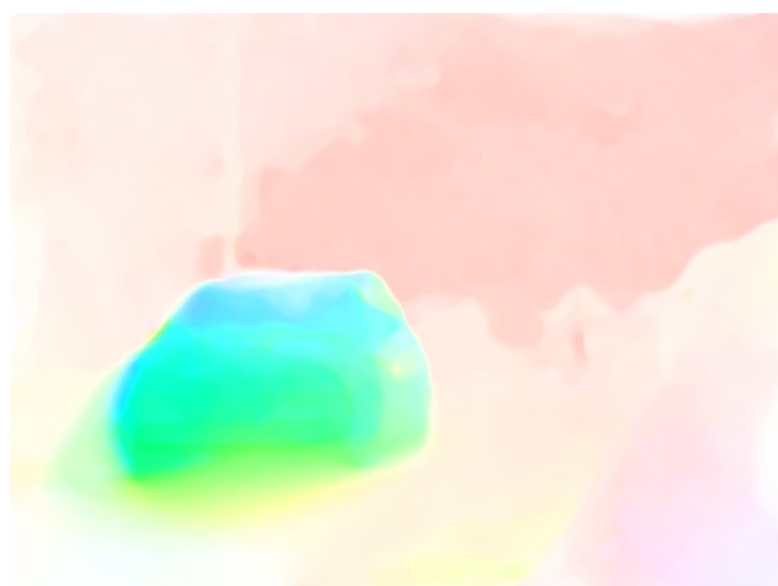
Optical Flow Results

Input two frames

Coarse-to-fine LK



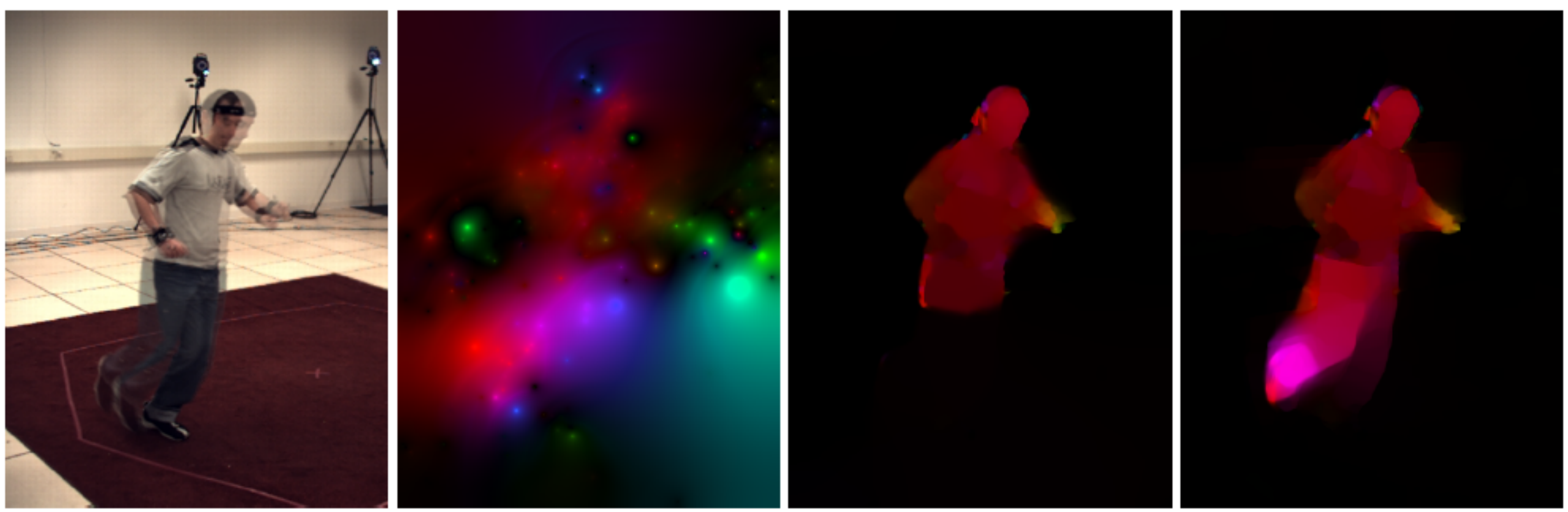
Flow visualization



Coarse-to-fine LK with median filtering

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)



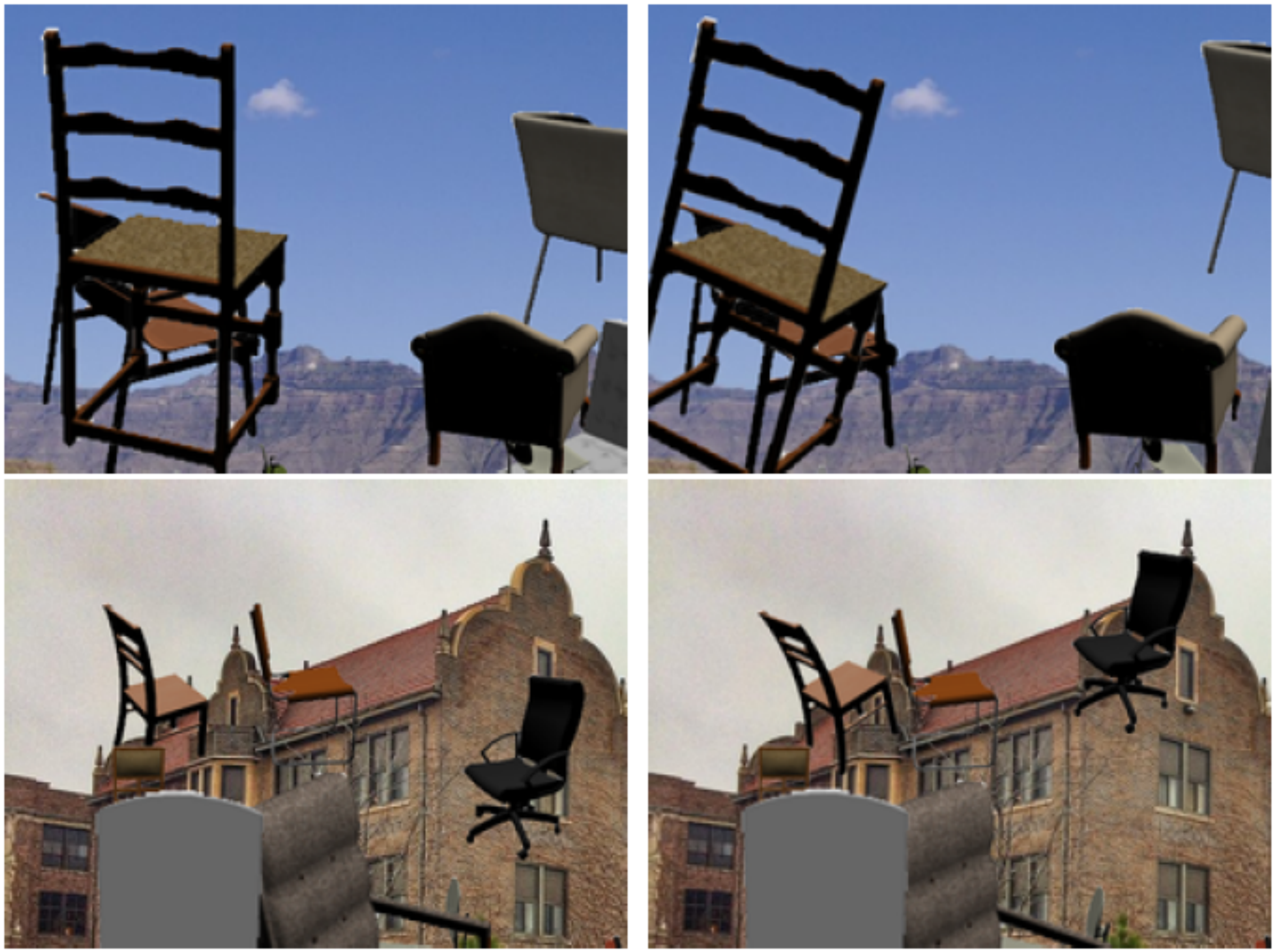
Region-based +Pixel-based +Keypoint-based

[Large displacement optical flow](), Brox et al., CVPR 2009

Learning optic flow

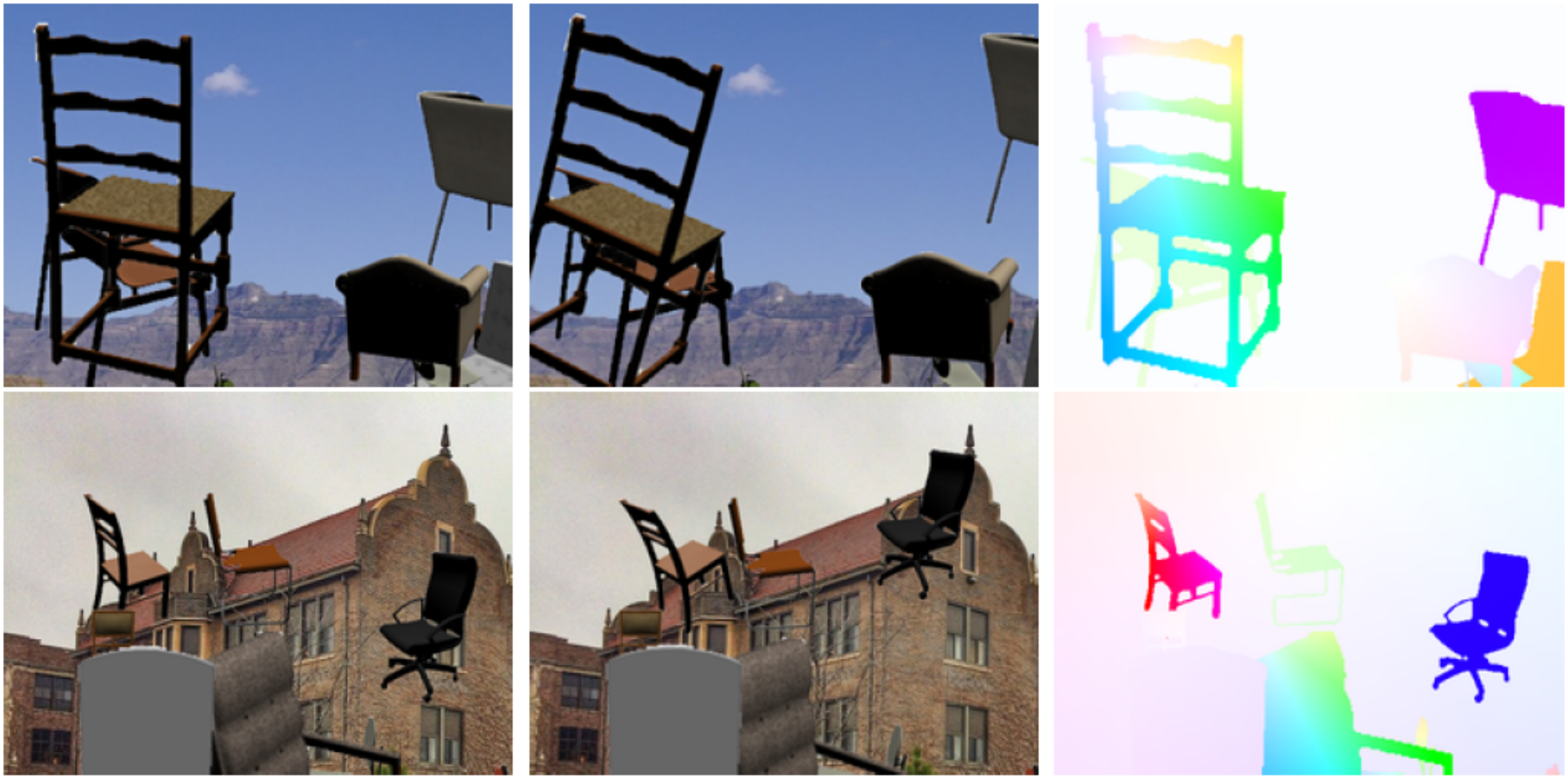
Where do you get ground truth data?

Learning optic flow



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

Learning optic flow



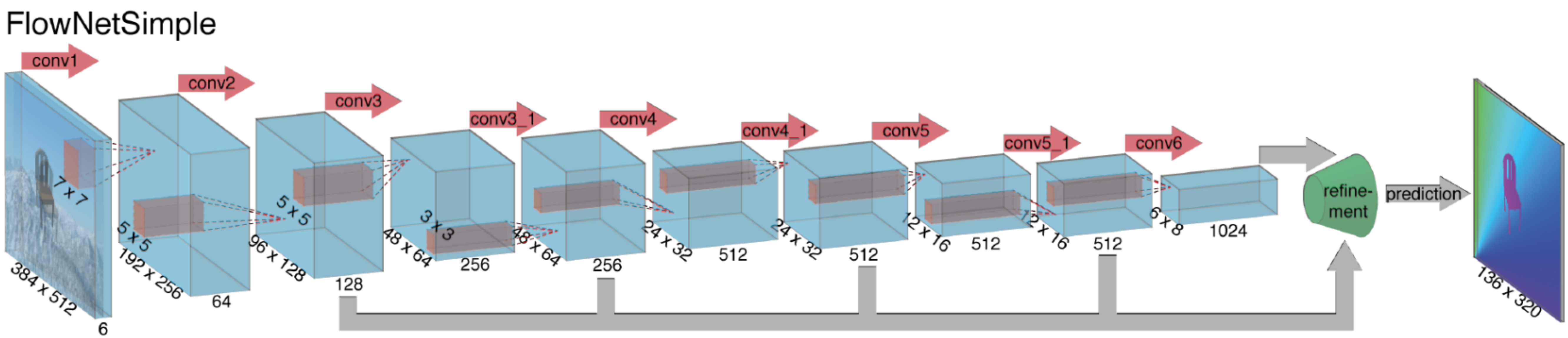
Fischer et al. 2015. https://arxiv.org/abs/1504.06852

Learning optic flow

CNN: input is pair of input frames

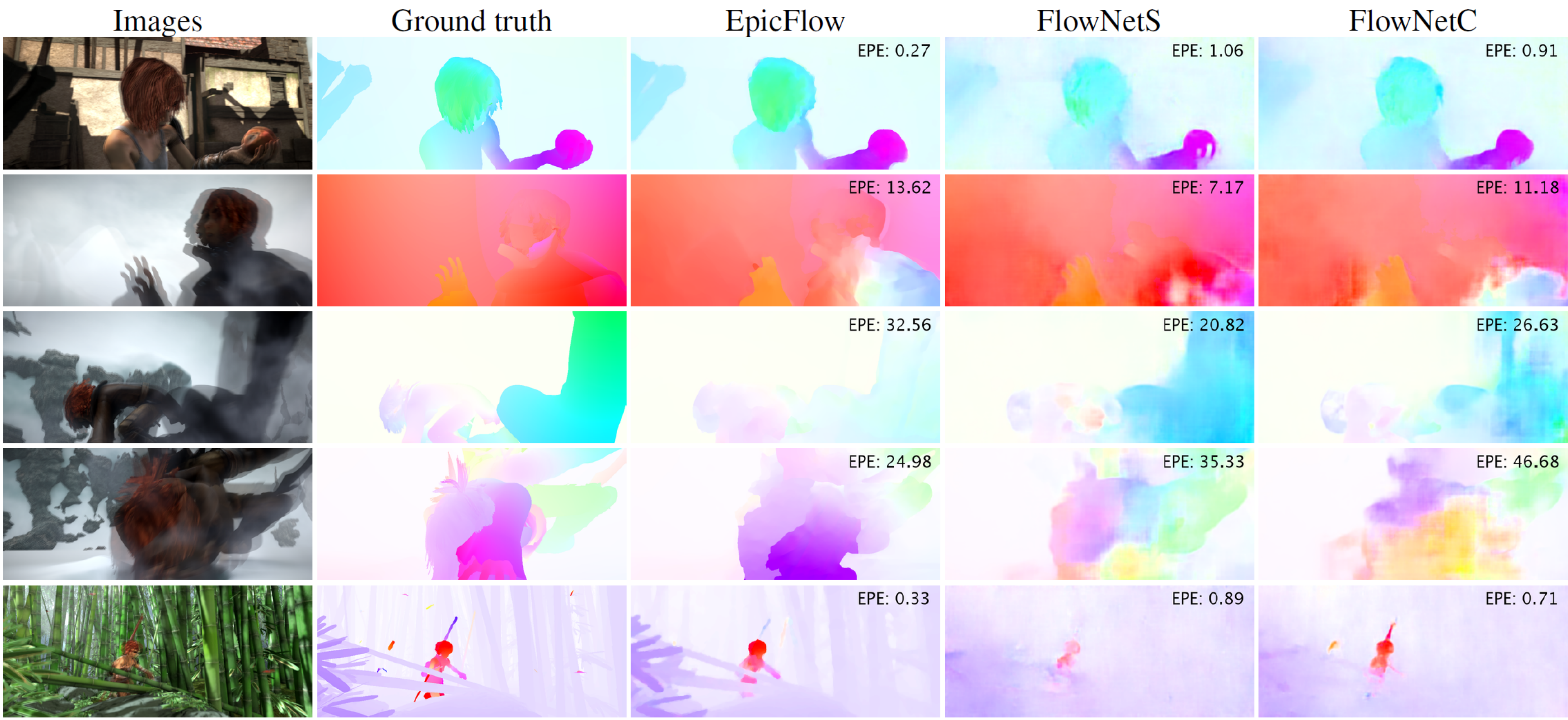
Upsample estimated flow back to input resolution

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



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

Results

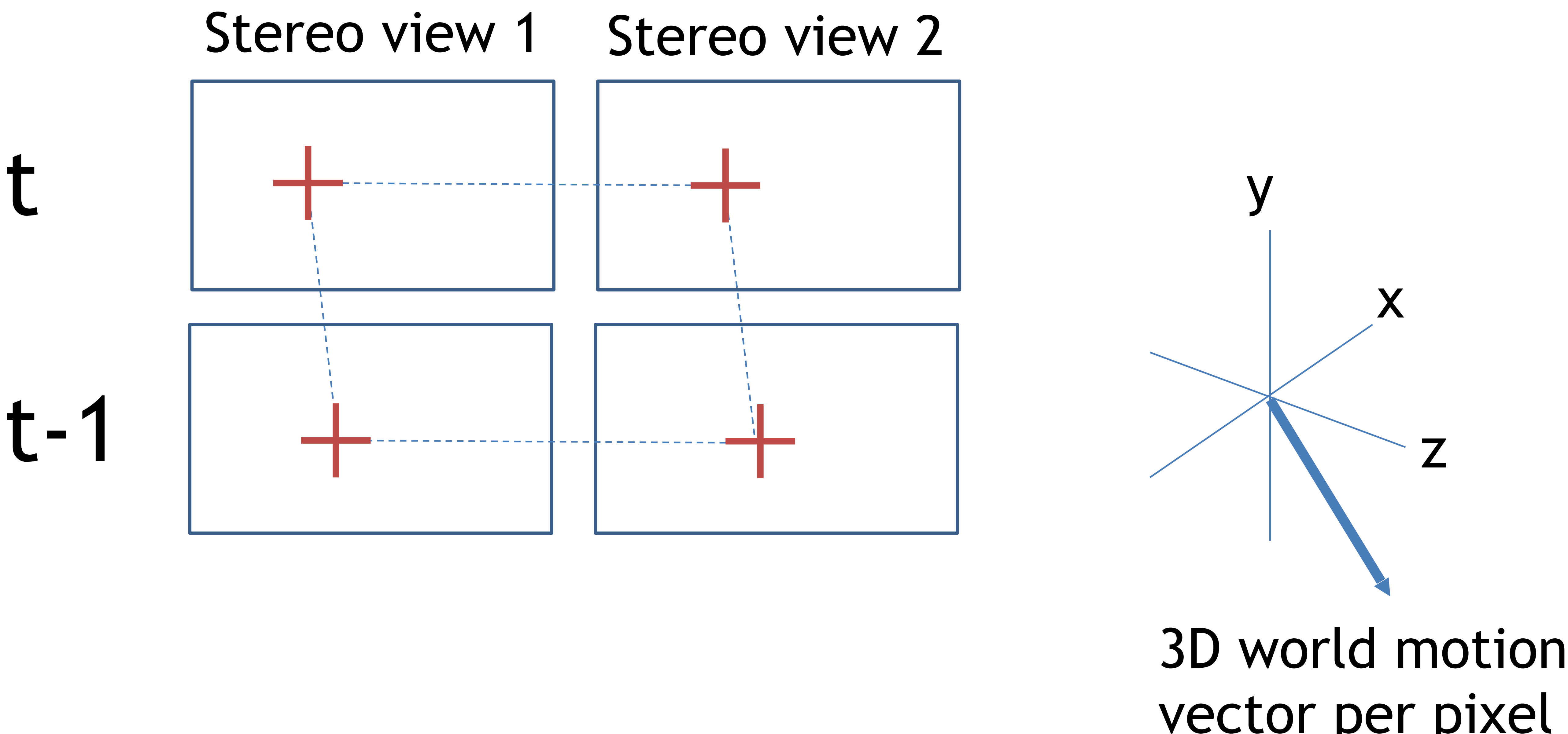


Fischer et al. 2015. https://arxiv.org/abs/1504.06852

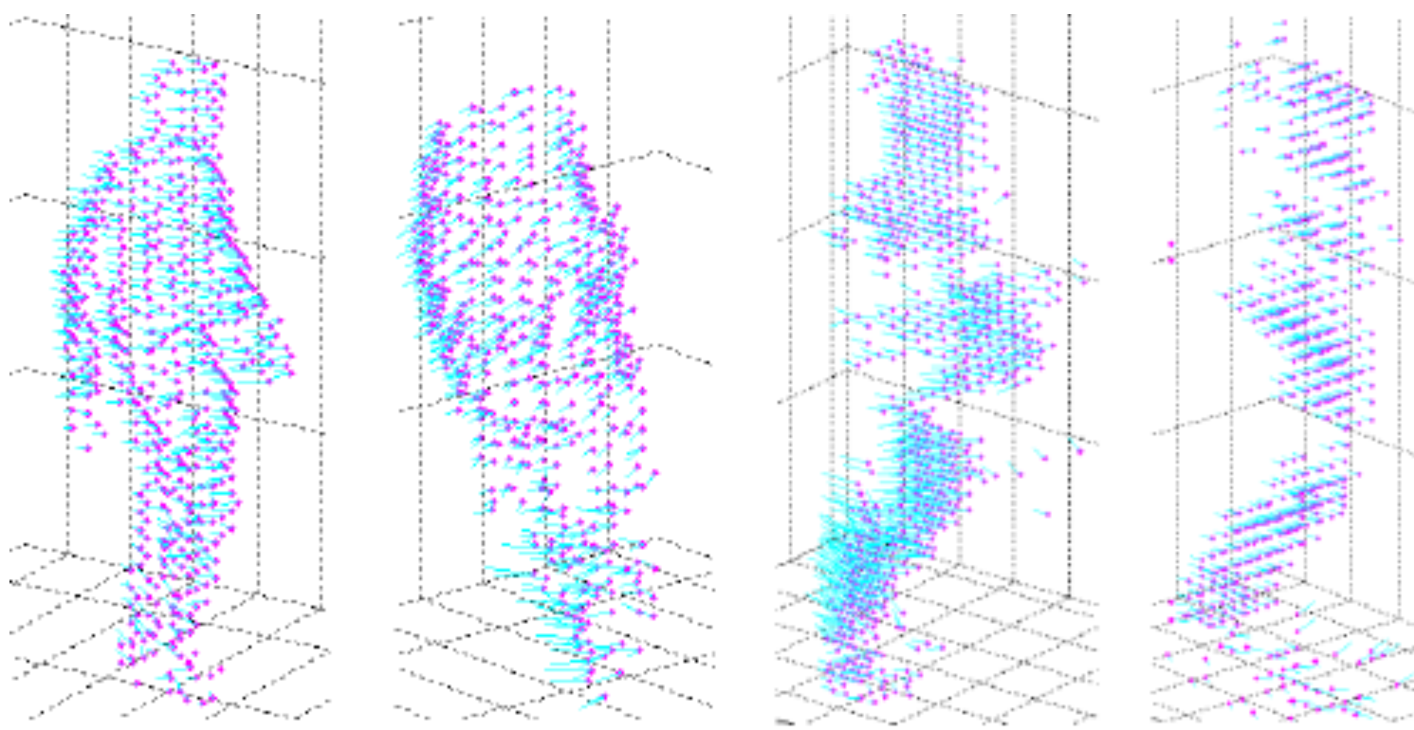
Can we do more? *Scene flow*

Combine spatial stereo & temporal constraints

Recover 3D vectors of world motion



Scene flow example for human motion


(a) (b) (c) (d)

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

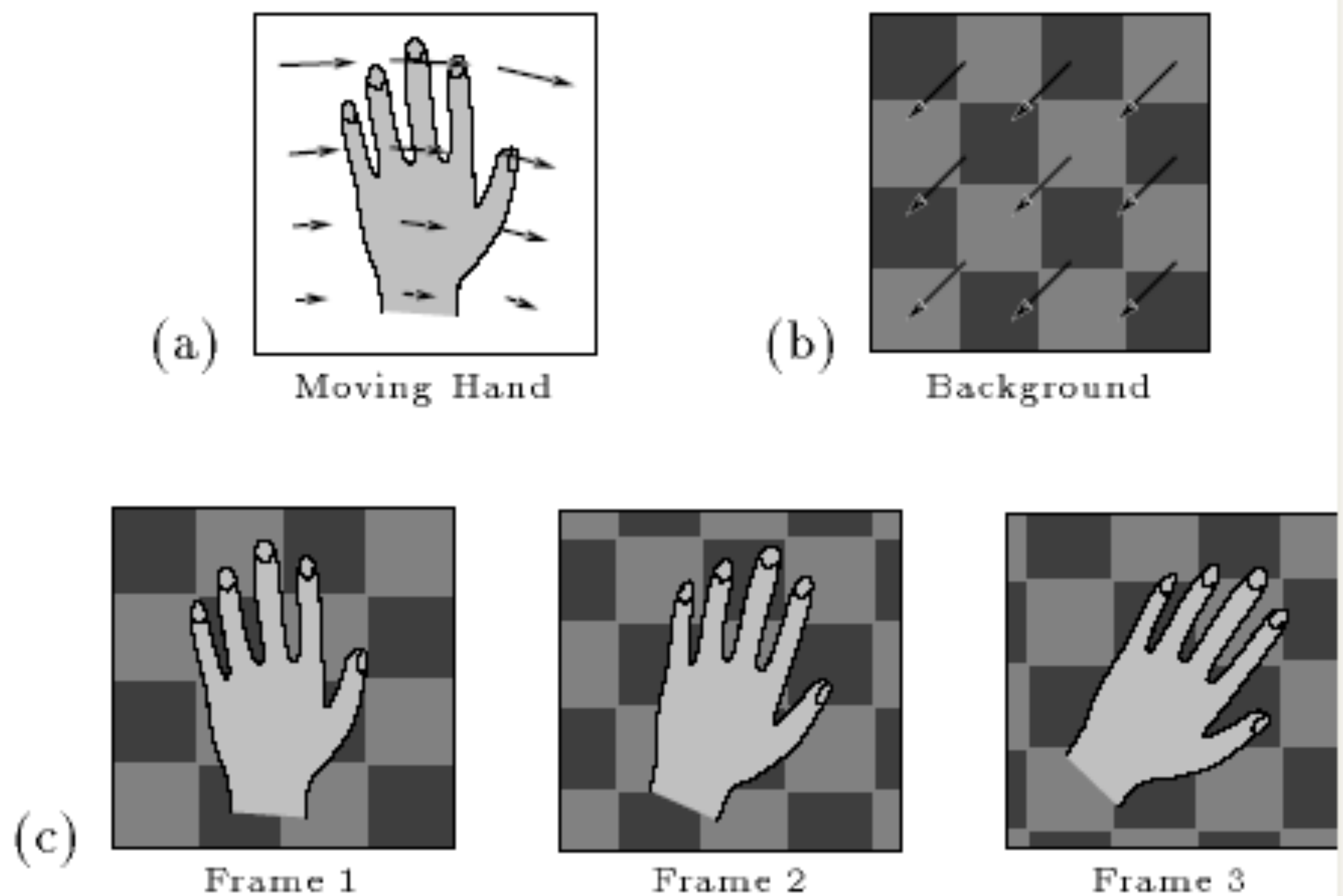
Scene Flow

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]

Layered model

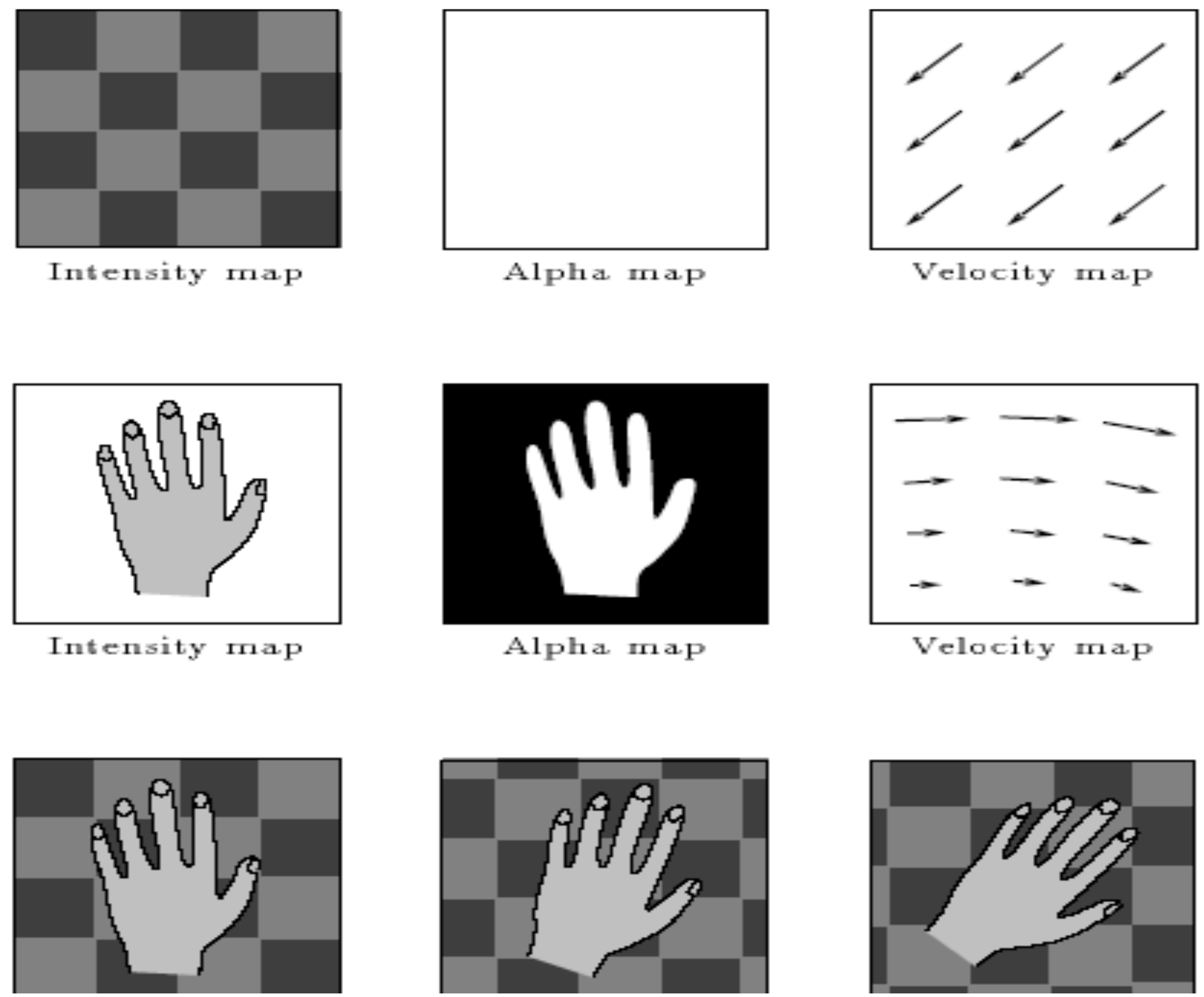


Mathematical formalism

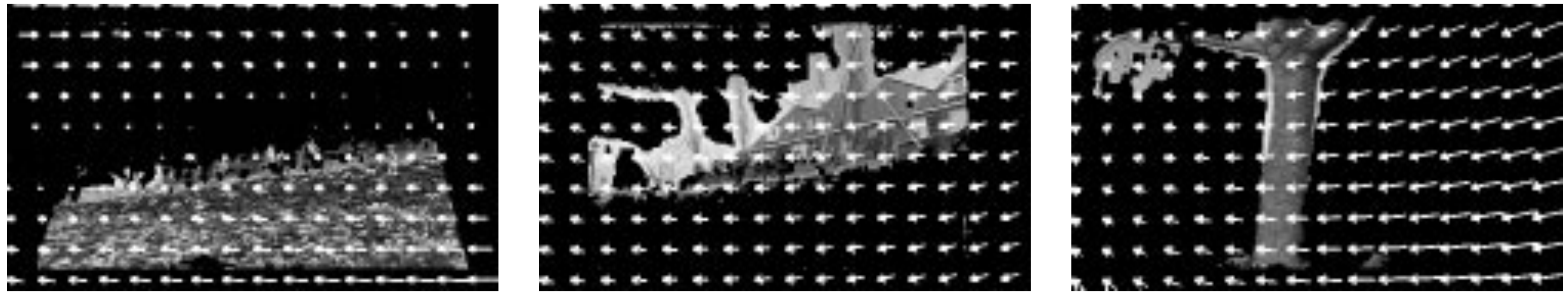
Layer 0 (BG)

Layer 1

Alpha composite



$$I_i(x, y) = \alpha_i(x, y)L_i(x, y) + (1 - \alpha_i(x, y))I_{i-1}(x, y)$$



(a) (b) (c)

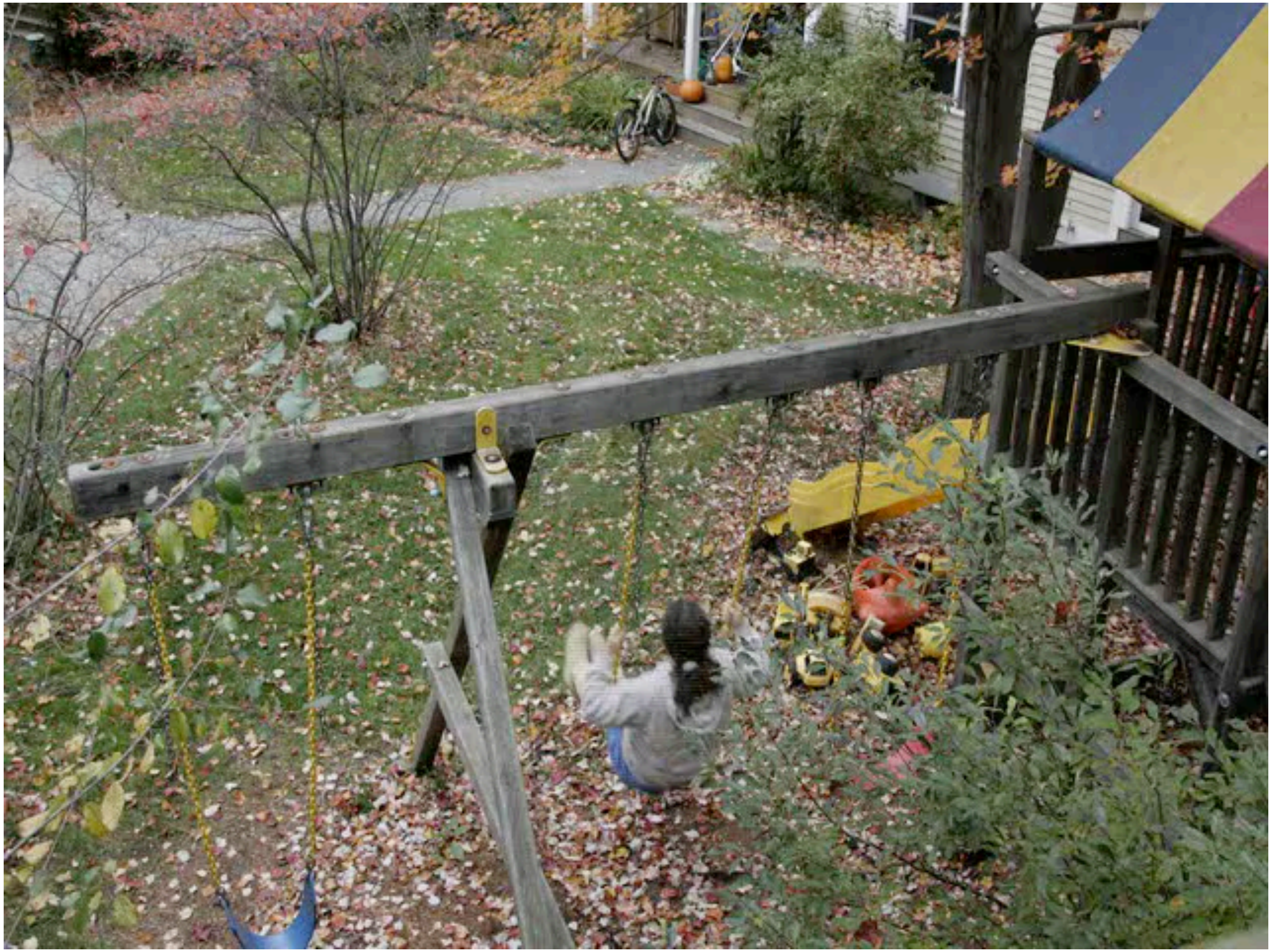
Figure 12: The layers corresponding to the tree, the flower bed, and the house shown in figures (a-c), respectively. The affine flow field for each layer is superimposed.

(a) (b) (c)

Figure 13: Frames 0, 15, and 30 as reconstructed from the layered representation shown in figures (a-c), respectively.

(a) (b) (c)

Figure 14: The sequence reconstructed without the tree layer shown in figures (a-c), respectively.

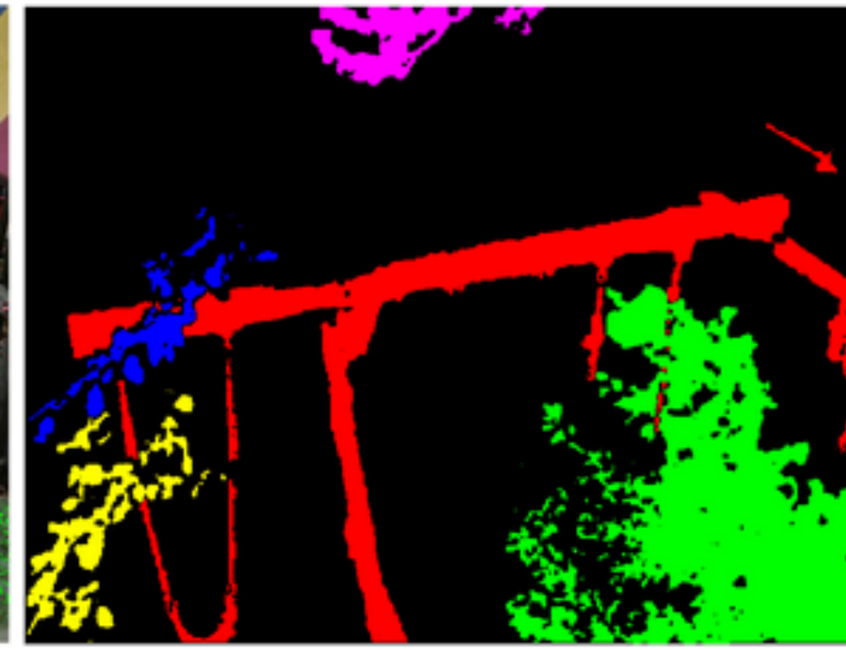


Motion Magnification

Motion Magnification

(a) Registered input frame

(b) Clustered trajectories of tracked features



(c) Layers of related motion and appearance

(d) Motion magnified, showing holes

(e) After texture in-painting to fill holes

(f) After user's modification to segmentation map in (c)

Motion Magnification

Motion Magnification