Introduction to Visual Computing

Prof. Feng Liu

Winter 2020

http://www.cs.pdx.edu/~fliu/courses/cs410/

02/06/2020

Last Time

- Image warping

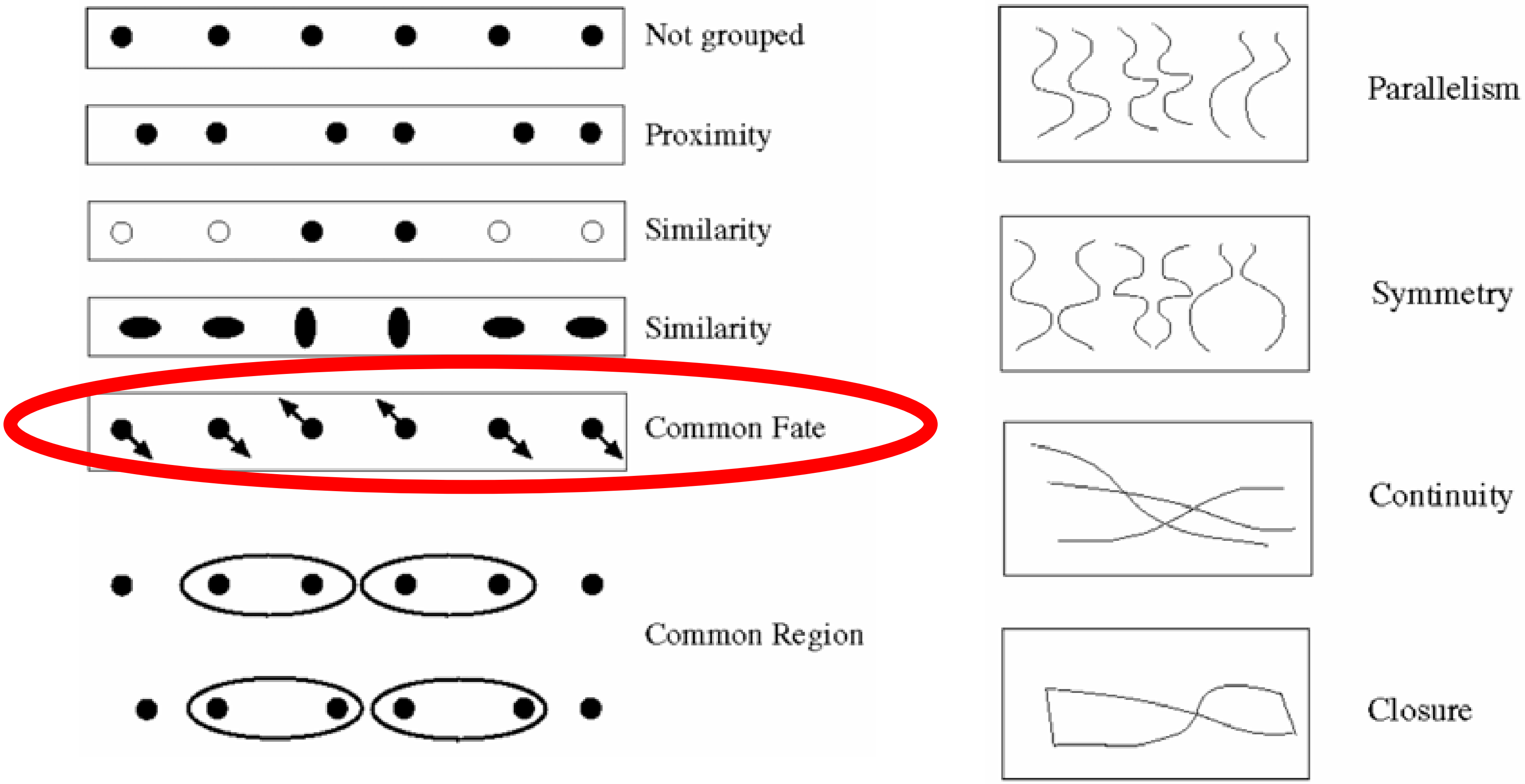
Today

- Motion estimation
  - Parametric motion
    - image alignment
  - Optical flow
  - Feature tracking

The slides for this topic are used from S. Lazebnik, Y-Y Chuang, M. Black, P. Anandan, S. Seitz, R. Szeliski, and M. Pollefeys.

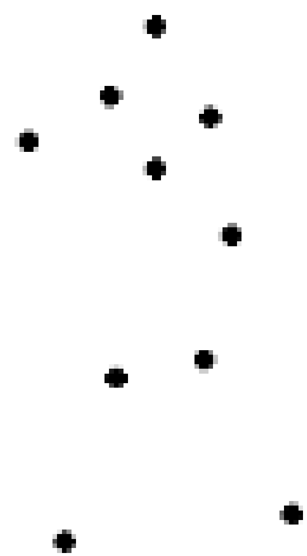
Motion and perceptual organization

- Sometimes, motion is the only cue



Motion and perceptual organization

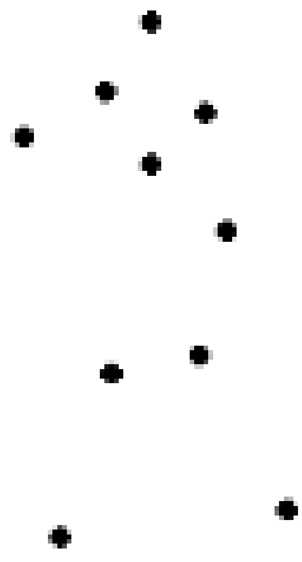
- Even “impoverished” motion data can evoke a strong percept



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

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.*

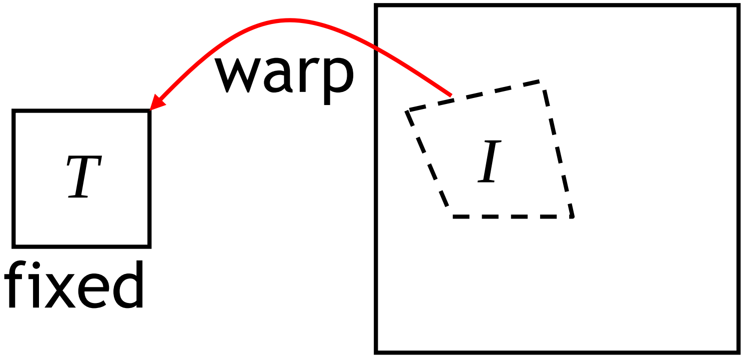
Uses of motion

- Estimating 3D structure
- Segmenting objects based on motion cues
- Learning and tracking dynamical models
- Recognizing events and activities

Image registration

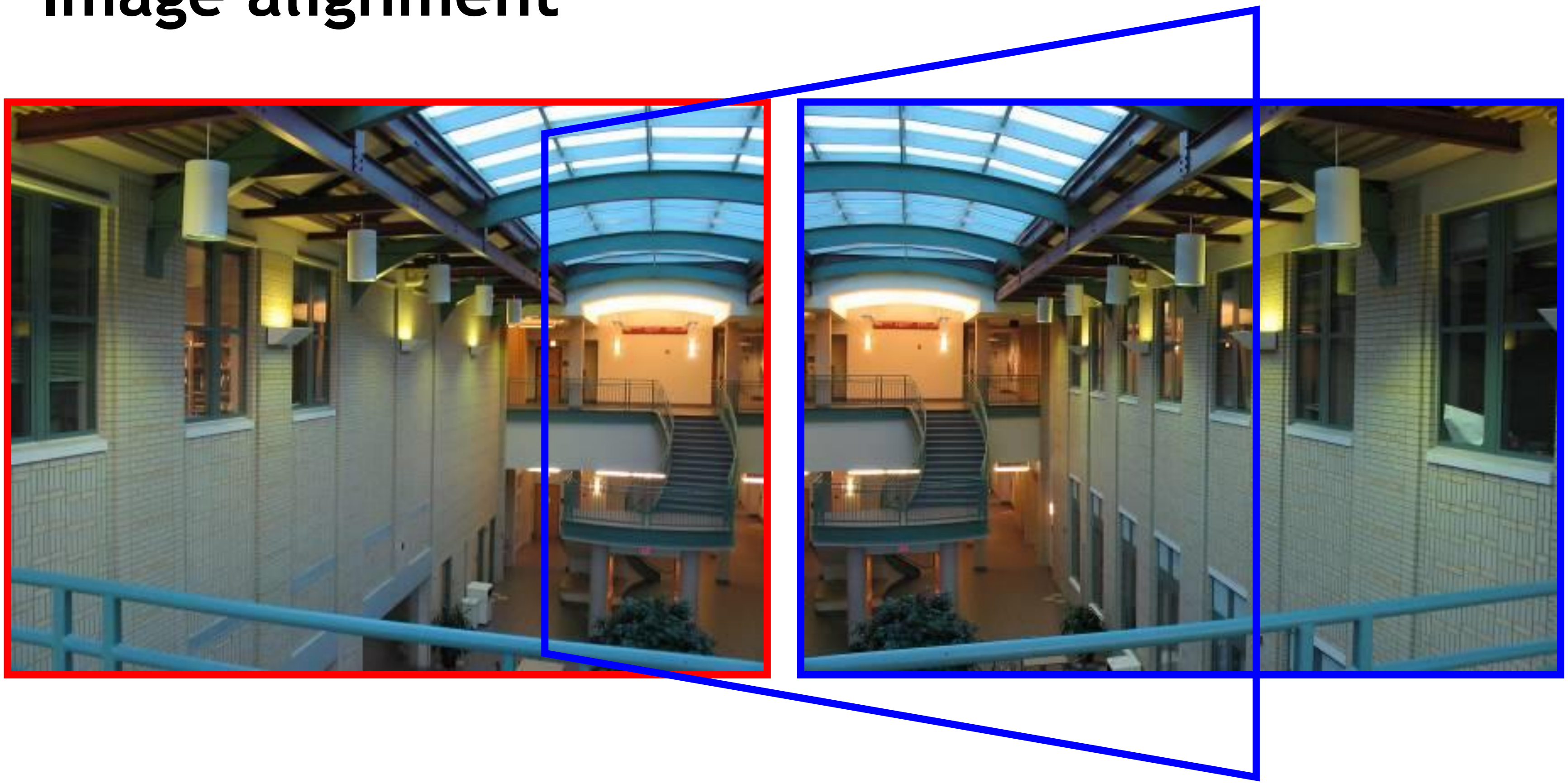
- Goal: register a template image $T(x)$ and an input image $I(x)$, where $x=(x,y)^T$. (warp I so that it matches T)

- Image alignment: $I(x)$ and $T(x)$ are two images
- Tracking: $T(x)$ is a small patch around a point p in the image at t. $I(x)$ is the image at time $t+1$.
- Optical flow: $T(x)$ and $I(x)$ are patches of images at t and $t+1$.

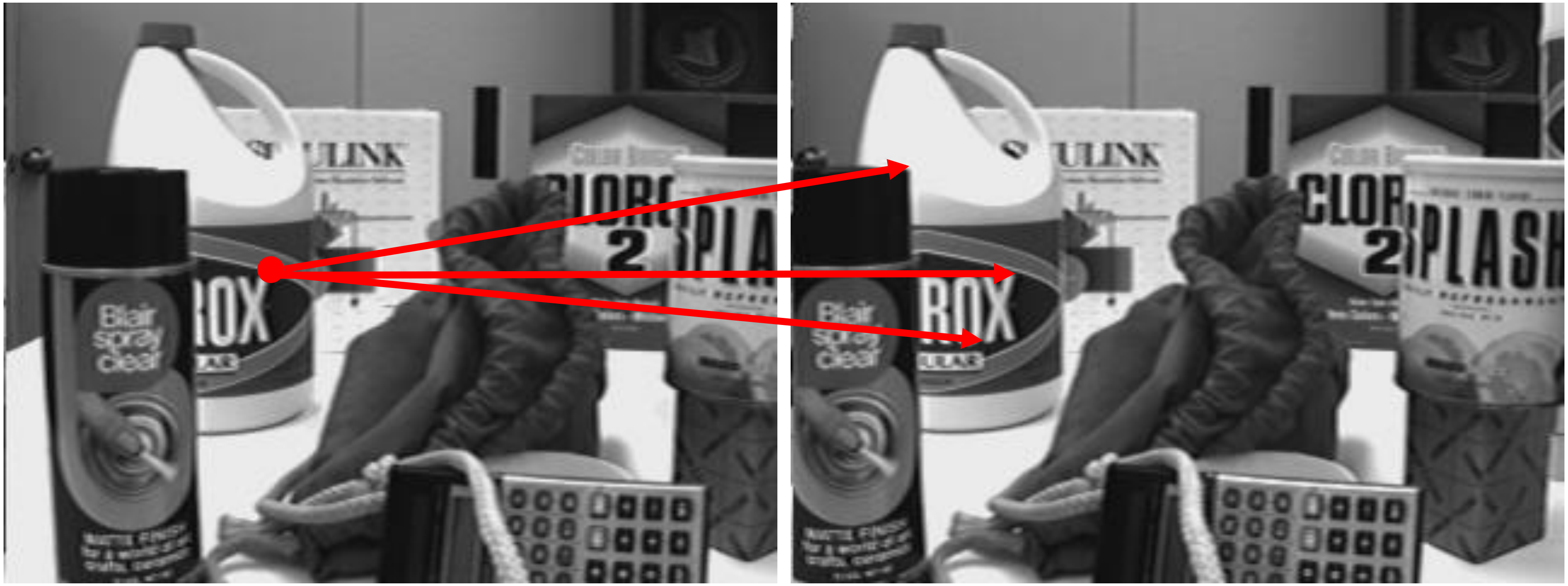


Parametric motion

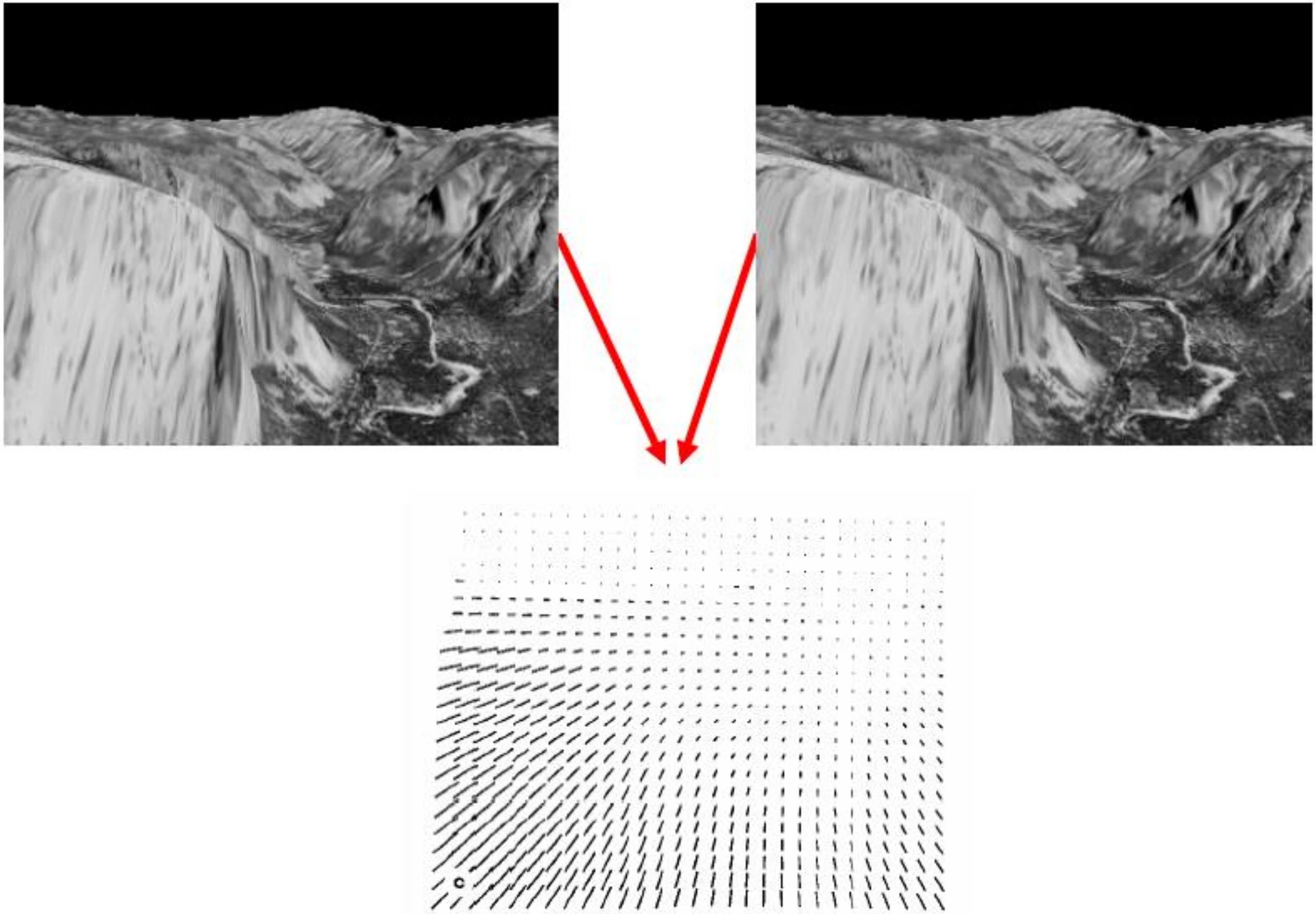
Image alignment



Tracking


ULINK
ROX
Blair spray clear
ULINK
CLOR
2
SPLASH
ROX
Blair spray clear

Optical flow



Three assumptions

- Brightness consistency
- Spatial coherence
- Temporal persistence

Brightness consistency

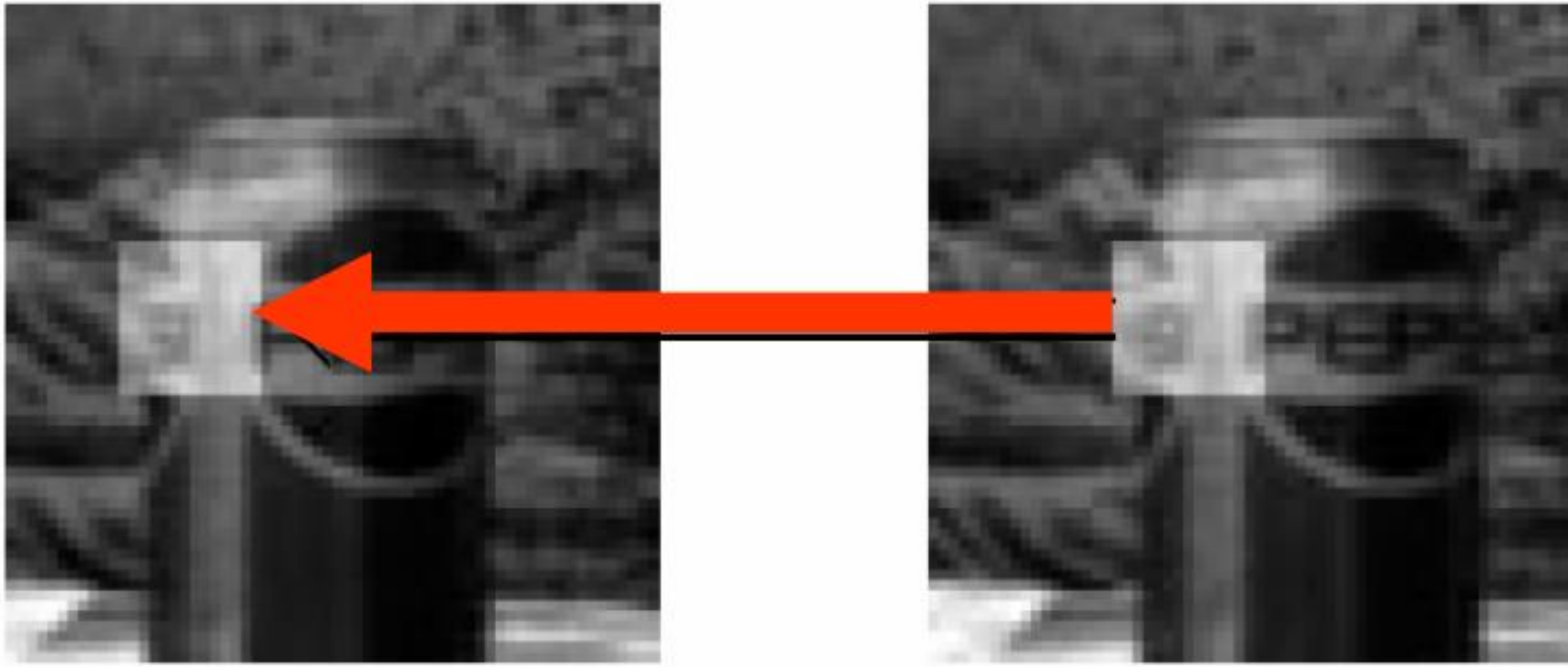
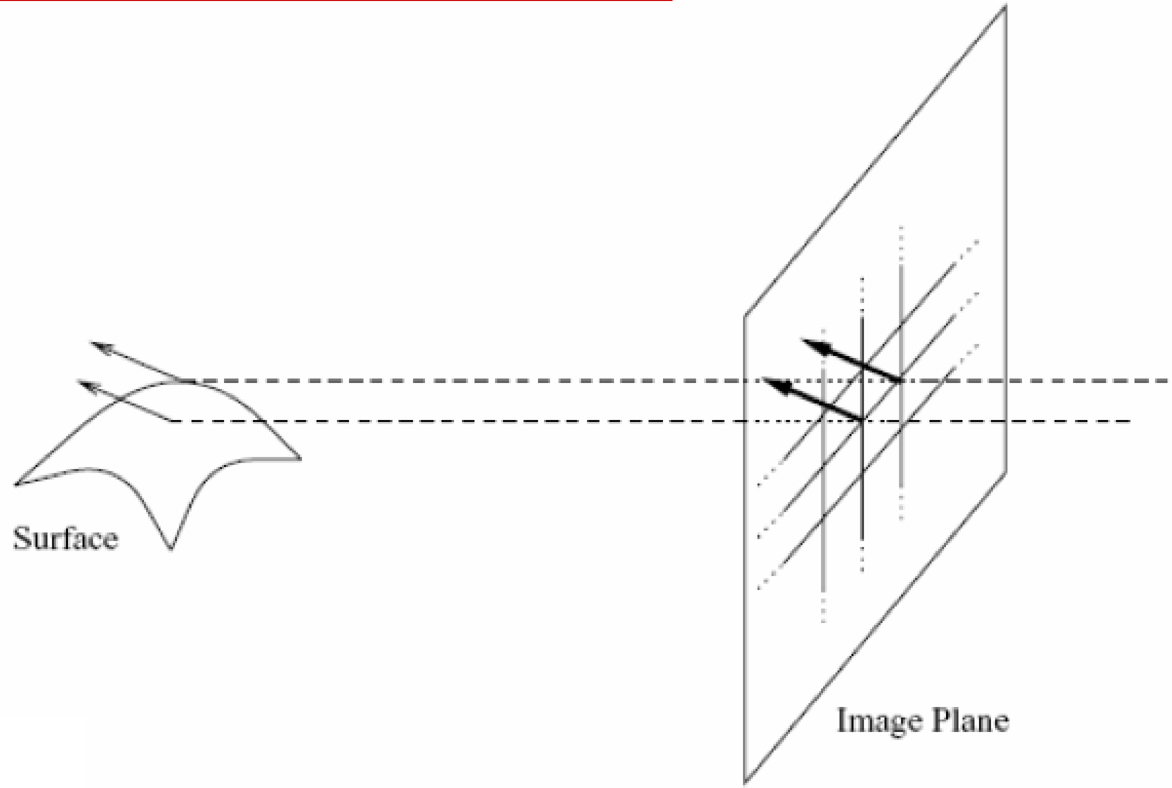


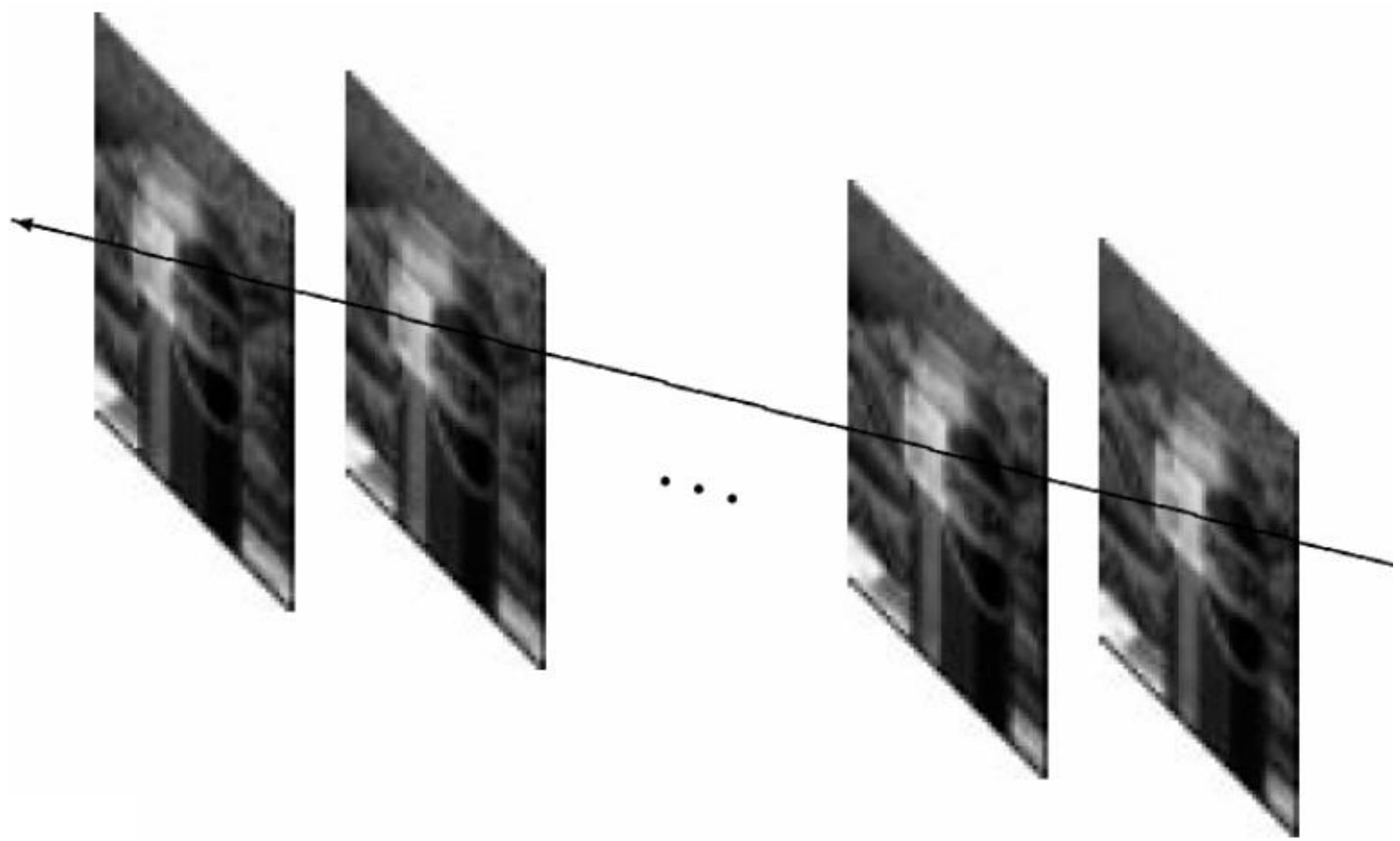
Image measurement (e.g. brightness) in a small region remain the same although their location may change.

Spatial coherence



- Neighboring points in the scene typically belong to the same surface and hence typically have similar motions.
- Since they also project to nearby pixels in the image, we expect spatial coherence in image flow.

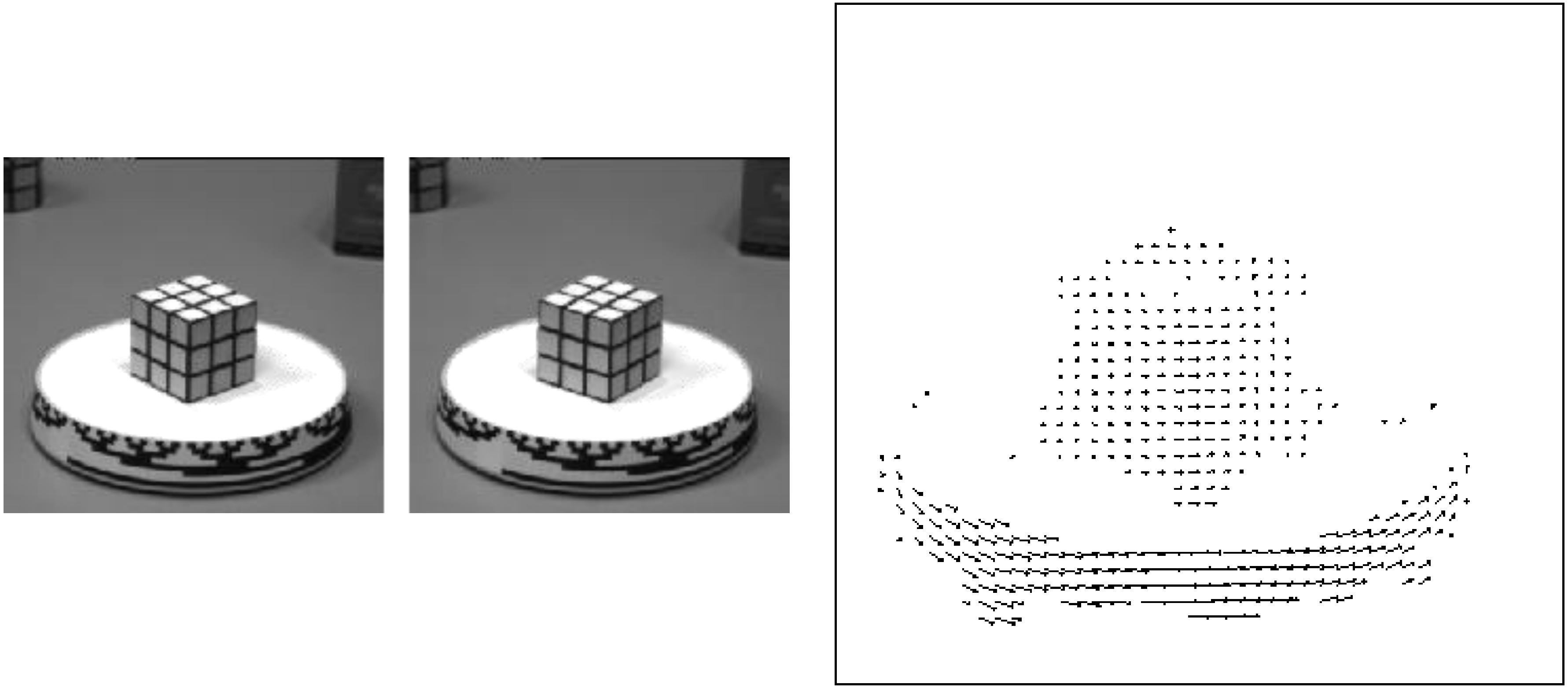
Temporal persistence



The image motion of a surface patch changes gradually over time.

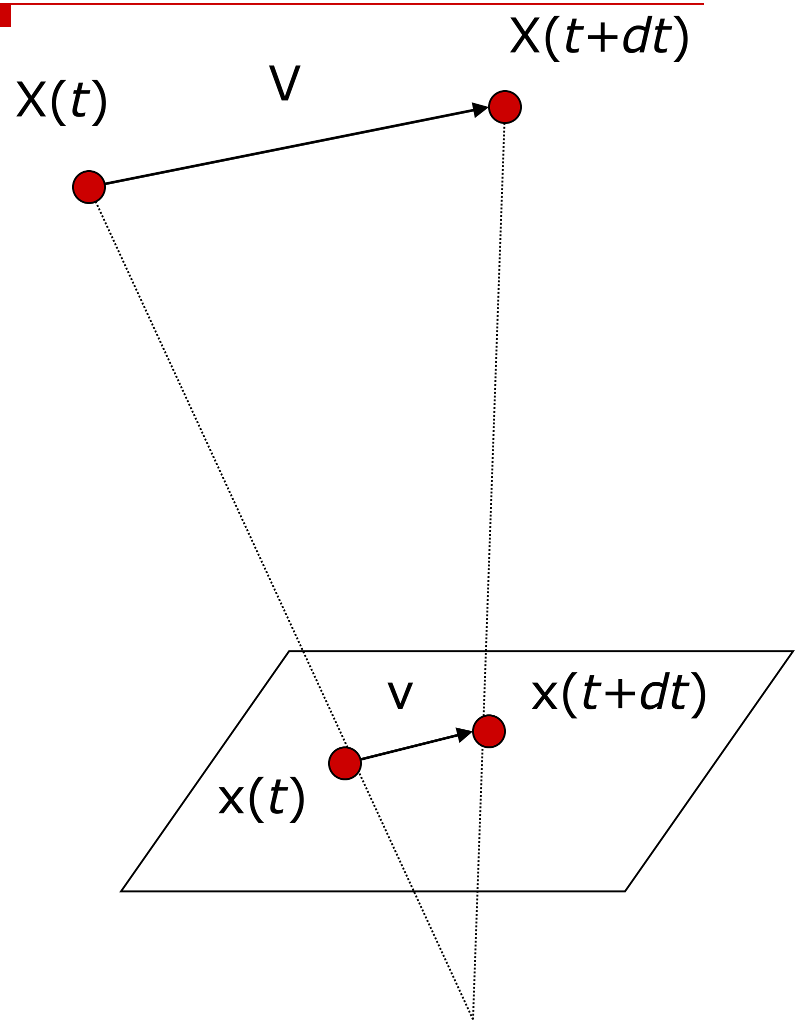
Motion field

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



Motion field and parallax

- **X**(t) is a moving 3D point
- Velocity of scene point: **V** = d**X**/dt
- **x**(t) = (x(t),y(t)) is the projection of **X** in the image
- Apparent velocity **v** in the image: given by components $v_x = dx/dt$ and $v_y = dy/dt$
- These components are known as the *motion field* of the image



Motion field and parallax

To find image velocity v, differentiate x=(x,y) with respect to t (using quotient rule):

$$x = f\frac{X}{Z} \qquad v_x = f\frac{ZV_x - V_z X}{Z^2} = \frac{fV_x - V_z x}{Z}$$

$$y = f\frac{Y}{Z} \qquad v_y = \frac{fV_y - V_z y}{Z}$$

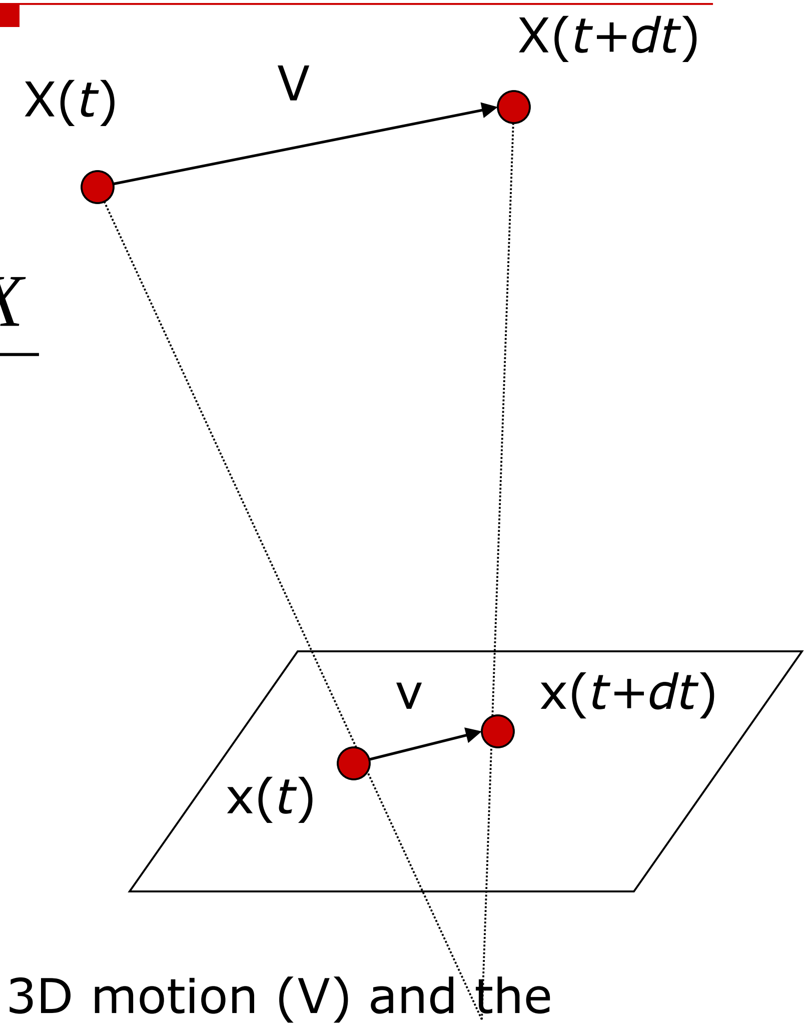


Image motion is a function of both the 3D motion (V) and the depth of the 3D point (Z)

Motion field and parallax

- Pure translation: V is constant everywhere

$$v_x = \frac{fV_x - V_z x}{Z}$$

$$v_y = \frac{fV_y - V_z y}{Z}$$

$$\mathbf{v} = \frac{1}{Z}(\mathbf{v}_0 - V_z \mathbf{x}), \qquad \mathbf{v}_0 = \left(fV_x, fV_y\right)$$

Motion field and parallax

- Pure translation: **V** is constant everywhere

$$\mathbf{v} = \frac{1}{Z}(\mathbf{v}_0 - V_z\mathbf{x}), \qquad \mathbf{v}_0 = \left(fV_x, fV_y\right)$$

- The length of the motion vectors is inversely proportional to the depth *Z*
- V_z is nonzero:
  - Every motion vector points toward (or away from) the vanishing point of the translation direction

Motion field and parallax

- Pure translation: **V** is constant everywhere

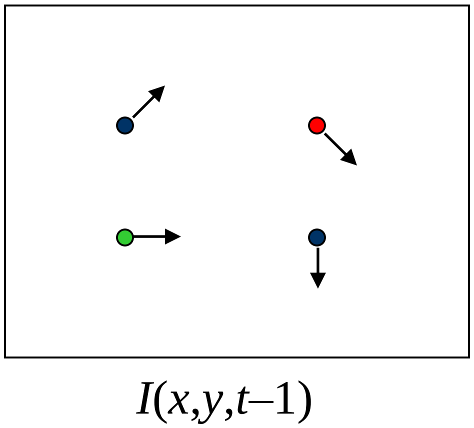
$$\mathbf{v} = \frac{1}{Z}(\mathbf{v}_0 - V_z \mathbf{x}), \qquad \mathbf{v}_0 = \left(f V_x, f V_y\right)$$

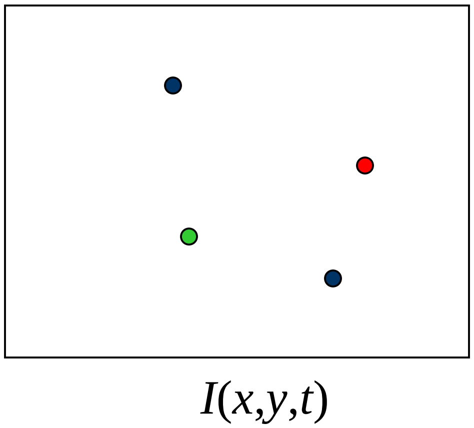
- The length of the motion vectors is inversely proportional to the depth *Z*
- V_z is nonzero:
  - Every motion vector points toward (or away from) the vanishing point of the translation direction
- V_z is zero:
  - Motion is parallel to the image plane, all the motion vectors are parallel

Optical flow

- 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
  - Think of a uniform rotating sphere under fixed lighting vs. a stationary sphere under moving illumination

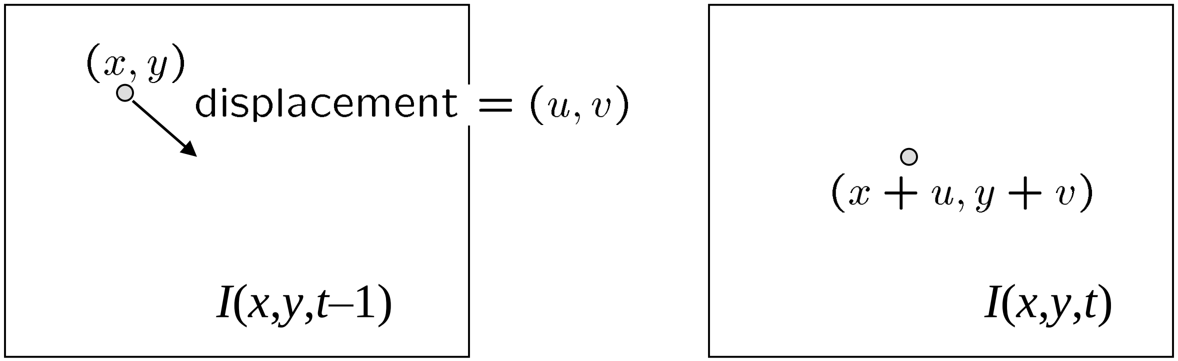
Estimating optical flow





- Given two subsequent frames, estimate the apparent motion field $u(x,y)$ and $v(x,y)$ between them
- Key assumptions
  - Brightness constancy: projection of the same point looks the same in every frame
  - Small motion: points do not move very far
  - Spatial coherence: points move like their neighbors

The brightness constancy constraint



- Brightness Constancy Equation:

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

- Linearizing the right side using Taylor expansion:

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

Hence, $$I_x \cdot u + I_y \cdot v + I_t \approx 0$$

The brightness constancy constraint

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

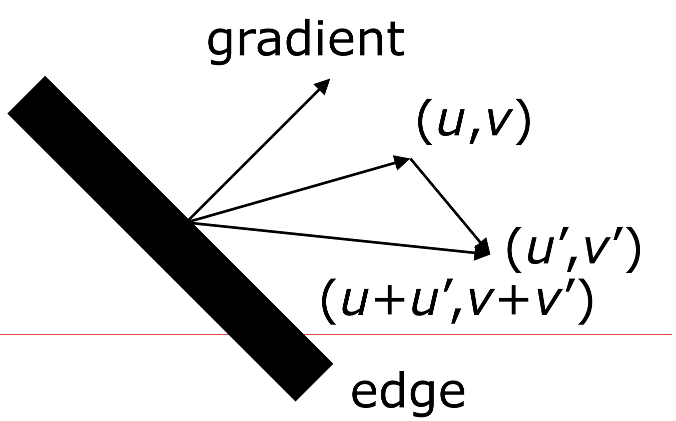
- How many equations and unknowns per pixel?
  - One equation, two unknowns
- Intuitively, what does this constraint mean? $\nabla I \cdot (u, v) + I_t = 0$
- The component of the flow perpendicular to the gradient (i.e., parallel to the edge) is unknown

The brightness constancy constraint

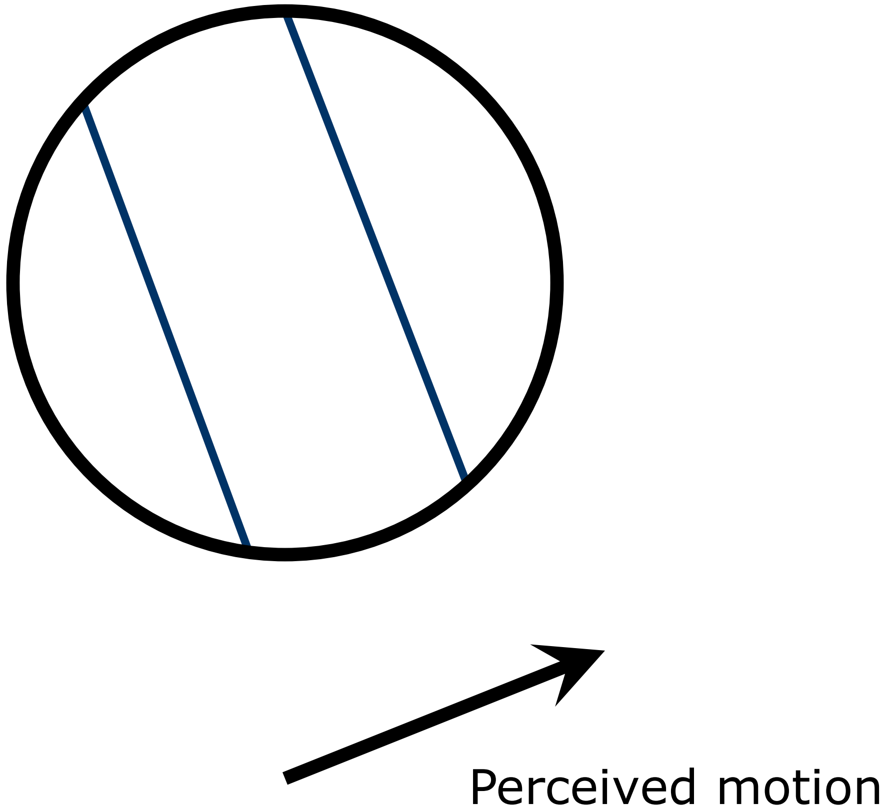
$$I_x \cdot u + I_y \cdot v + I_t = 0$$

- How many equations and unknowns per pixel?
  - One equation, two unknowns
- Intuitively, what does this constraint mean? $\nabla I \cdot (u,v) + I_t = 0$
- The component of the flow perpendicular to the gradient (i.e., parallel to the edge) is unknown

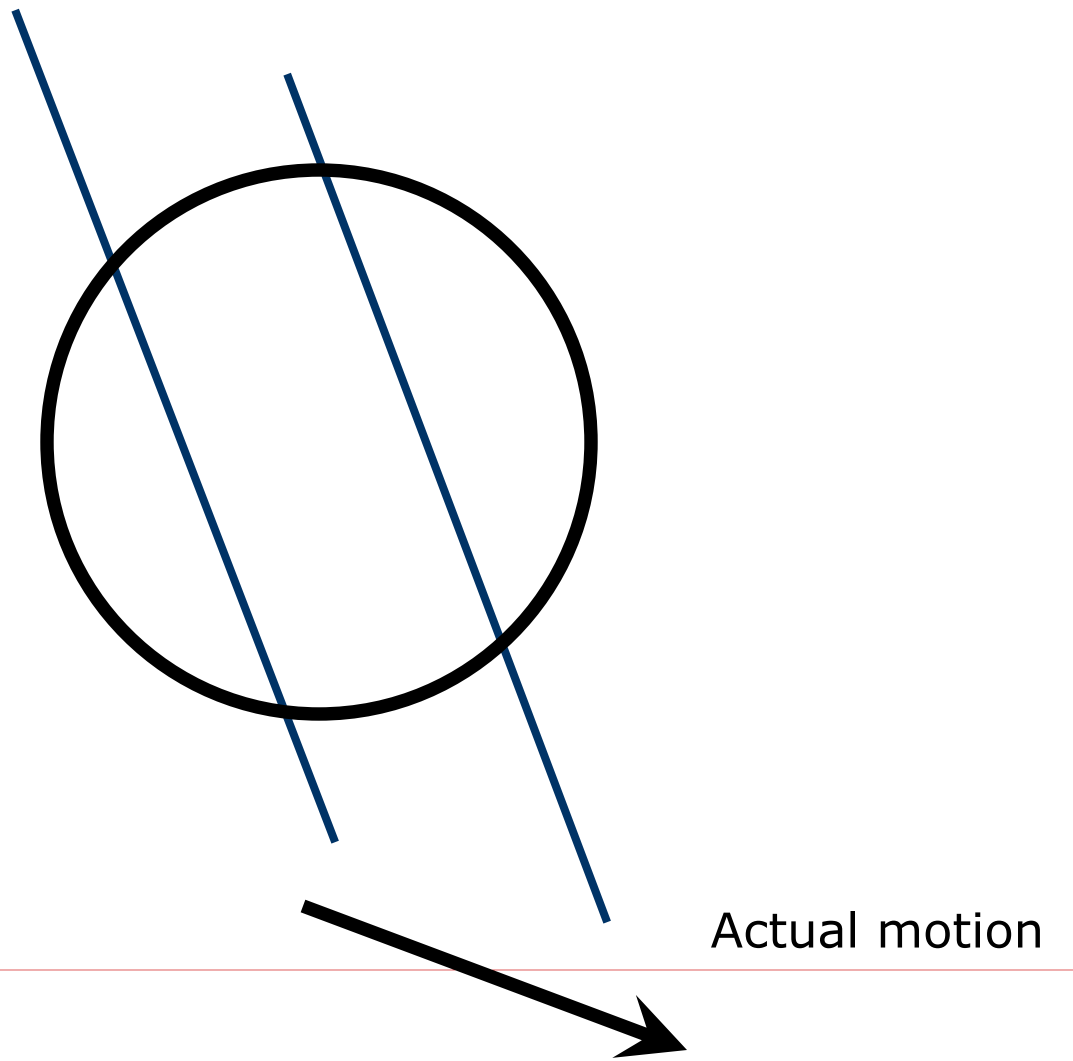
If (u, v) satisfies the equation,
so does ($u+u'$, $v+v'$) if $\nabla I \cdot (u', v') = 0$



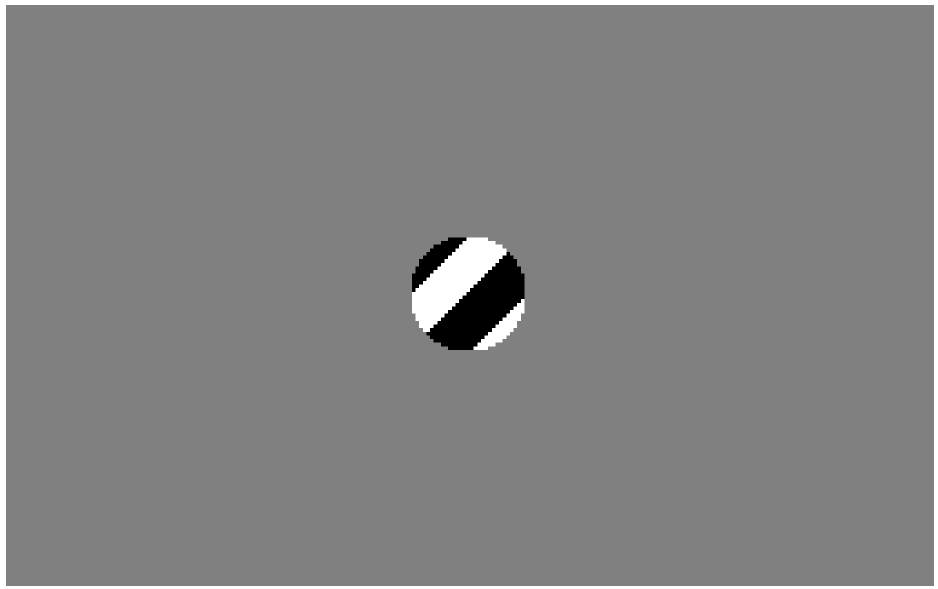
The aperture problem



The aperture problem

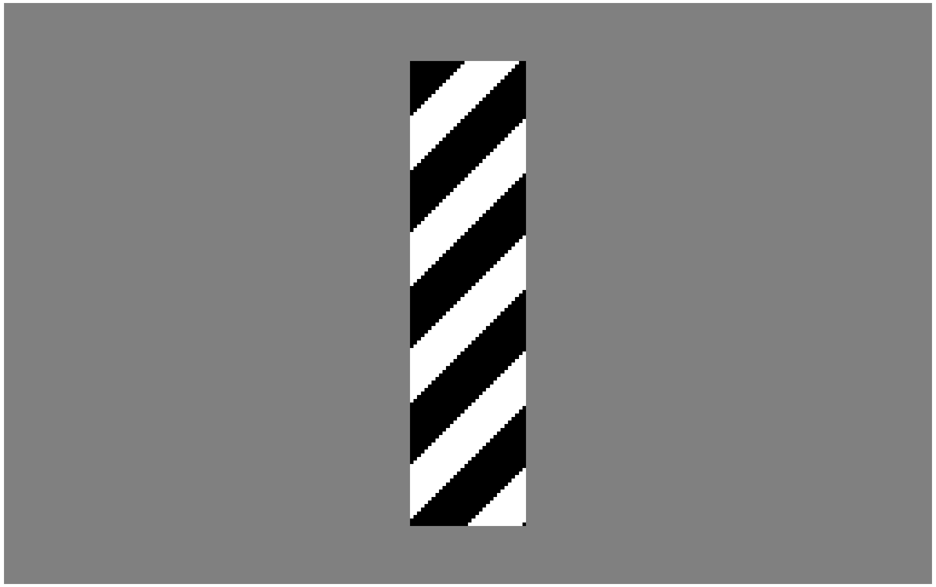


The barber pole illusion



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

The barber pole illusion



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

The barber pole illusion

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

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

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

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

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.

Solving the aperture problem

- Least squares problem:

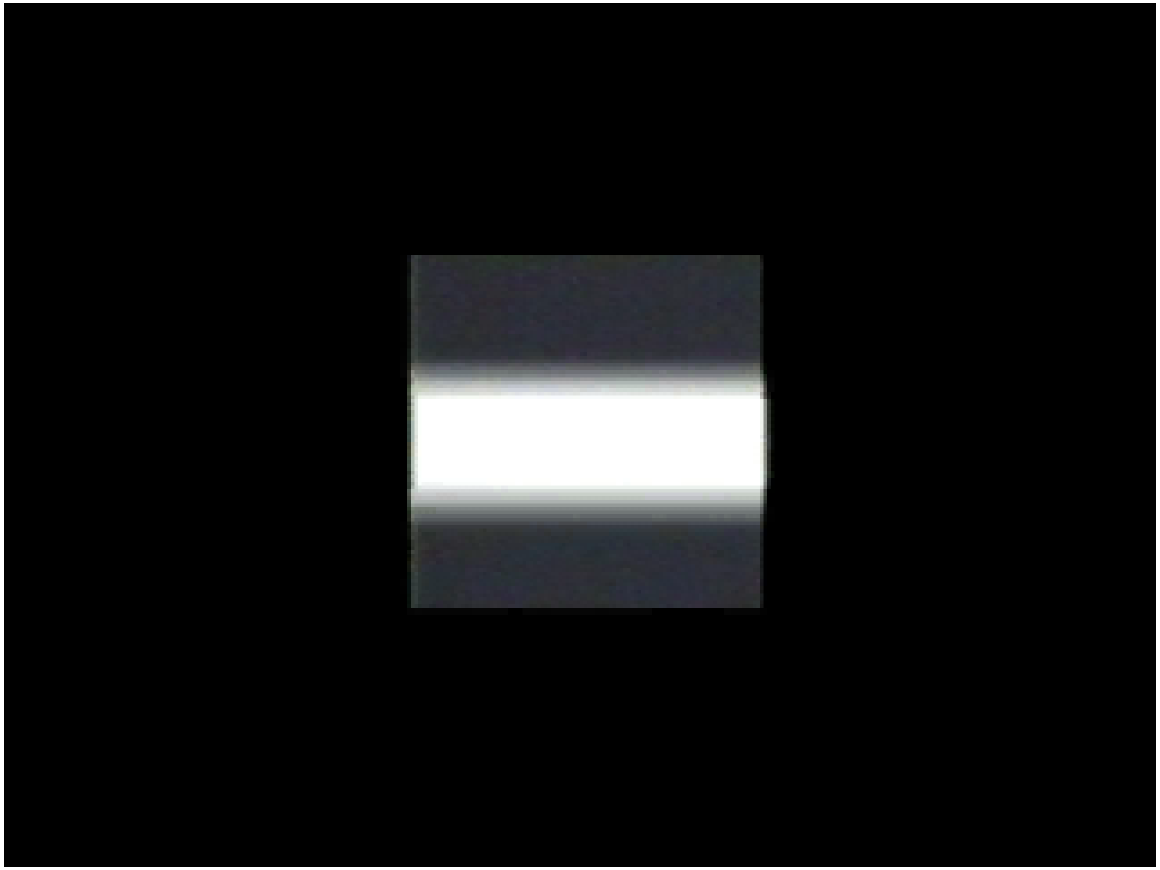
$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\times2}{A} \; \underset{2\times1}{d} = \underset{25\times1}{b}$$

- When is this system solvable?
  - What if the window contains just a single straight edge?

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

- “Bad” case: single straight edge



Lucas-Kanade flow

- Least squares problem:

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

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

Solution given by $(A^T A)\ d = A^T b$

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

The summations are over all pixels in the window

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.

Lucas-Kanade flow

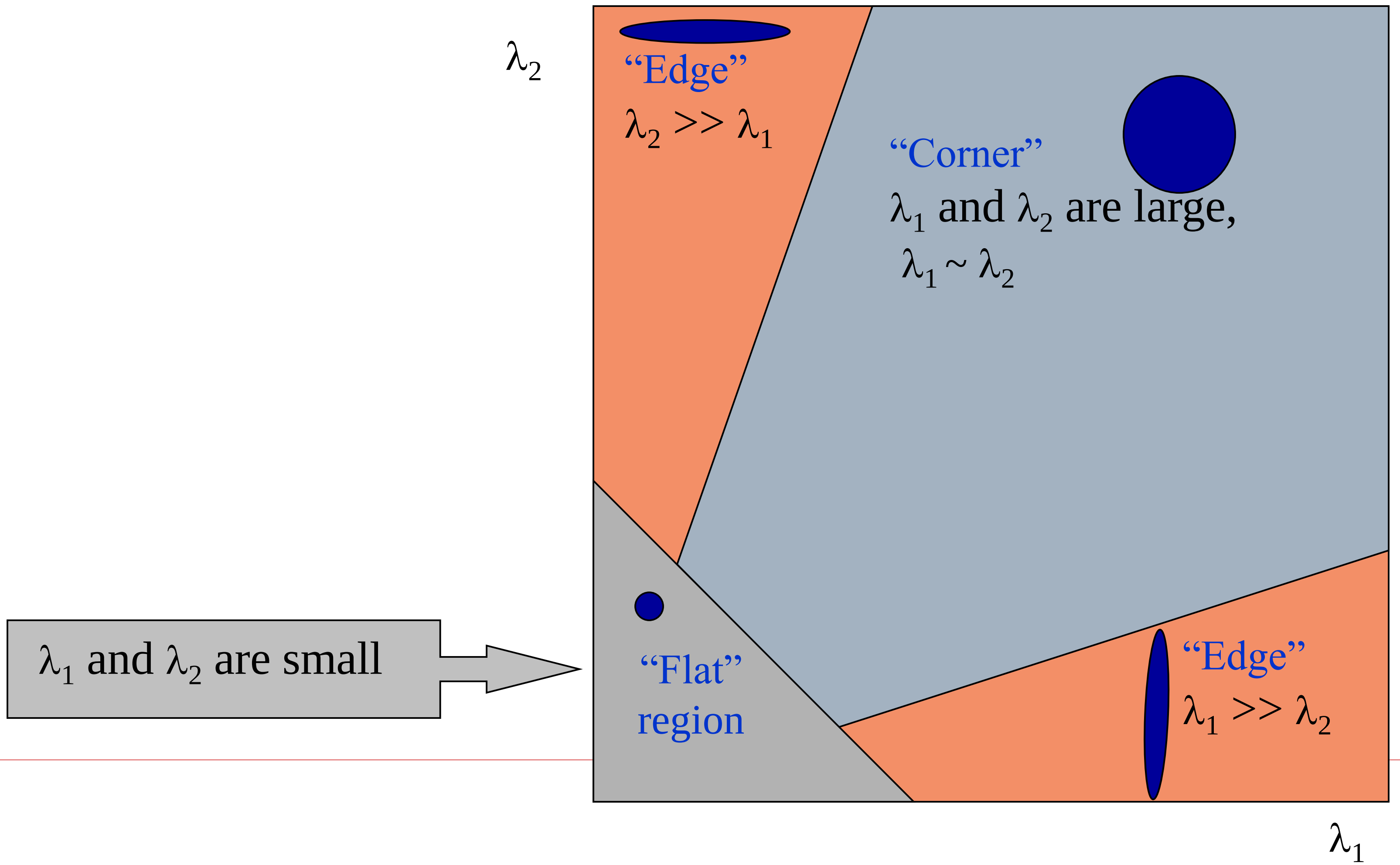
$$\begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} \sum I_x I_t \\ \sum I_y I_t \end{bmatrix}$$

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

- Recall the Harris corner detector: $M = A^T A$ is the *second moment matrix*
- We can figure out whether the system is solvable by looking at the eigenvalues of the second moment matrix
  - The eigenvectors and eigenvalues of M relate to edge direction and magnitude
  - The eigenvector associated with the larger eigenvalue points in the direction of fastest intensity change, and the other eigenvector is orthogonal to it

Interpreting the eigenvalues

Classification of image points using eigenvalues of the second moment matrix:



Uniform region

– gradients have small magnitude
– small λ_1, small λ_2
– system is ill-conditioned

Edge

- gradients have one dominant direction
- large λ_1, small λ_2
- system is ill-conditioned

High-texture or corner region

– gradients have different directions, large magnitude
– large λ_1, large λ_2
– system is well-conditioned

Errors in Lucas-Kanade

- The motion is large (larger than a pixel)
  - Iterative refinement
  - Coarse-to-fine estimation
  - Exhaustive neighborhood search (feature matching)
- A point does not move like its neighbors
  - Motion segmentation
- Brightness constancy does not hold
  - Exhaustive neighborhood search with normalized correlation

Feature tracking

- So far, we have only considered optical flow estimation in a pair of images
- If we have more than two images, we can compute the optical flow from each frame to the next
- Given a point in the first image, we can in principle reconstruct its path by simply “following the arrows”

Tracking challenges

- Ambiguity of optical flow
  - Need to find good features to track
- Large motions, changes in appearance, occlusions, disocclusions
  - Need mechanism for deleting, adding new features
- Drift – errors may accumulate over time
  - Need to know when to terminate a track

Tracking over many frames

- Select features in first frame
- For each frame:
  - Update positions of tracked features
    - Discrete search or Lucas-Kanade (or a combination of the two)
  - Terminate inconsistent tracks
    - Compute similarity with corresponding feature in the previous frame or in the first frame where it's visible
  - Find more features to track

Shi-Tomasi feature tracker

- Find good features using eigenvalues of second-moment matrix
  - Key idea: “good” features to track are the ones whose motion can be estimated reliably
- From frame to frame, track with Lucas-Kanade
  - This amounts to assuming a translation model for frame-to-frame feature movement
- Check consistency of tracks by *affine* registration to the first observed instance of the feature
  - Affine model is more accurate for larger displacements
  - Comparing to the first frame helps to minimize drift

J. Shi and C. Tomasi. Good Features to Track. CVPR 1994.

Tracking example

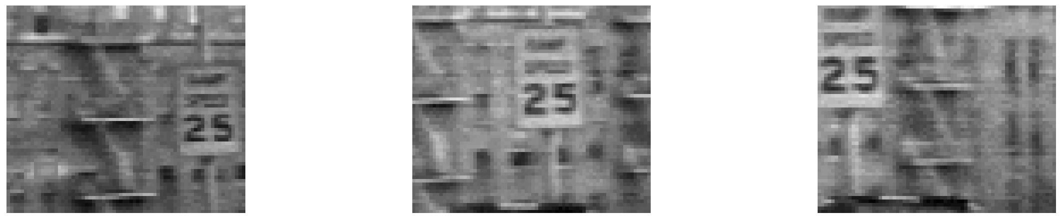


Figure 1: Three frame details from Woody Allen's *Manhattan*. The details are from the 1st, 11th, and 21st frames of a subsequence from the movie.

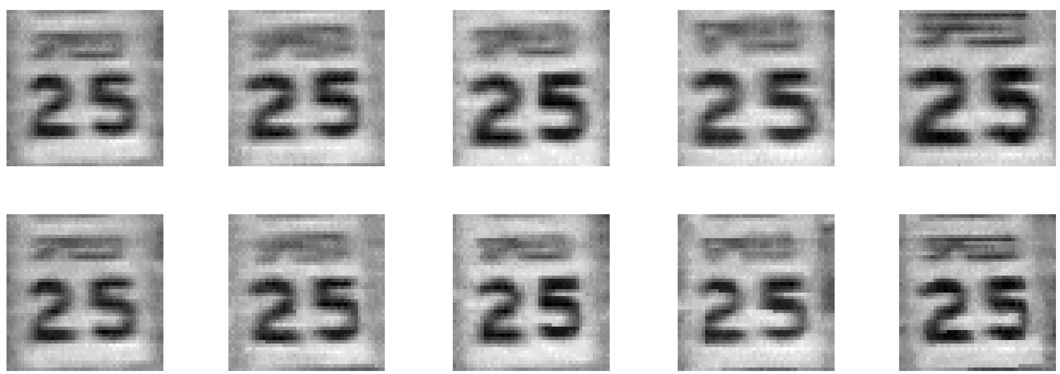


Figure 2: The traffic sign windows from frames 1,6,11,16,21 as tracked (top), and warped by the computed deformation matrices (bottom).

J. Shi and C. Tomasi. Good Features to Track. CVPR 1994.

Next Time

- Camera and calibration