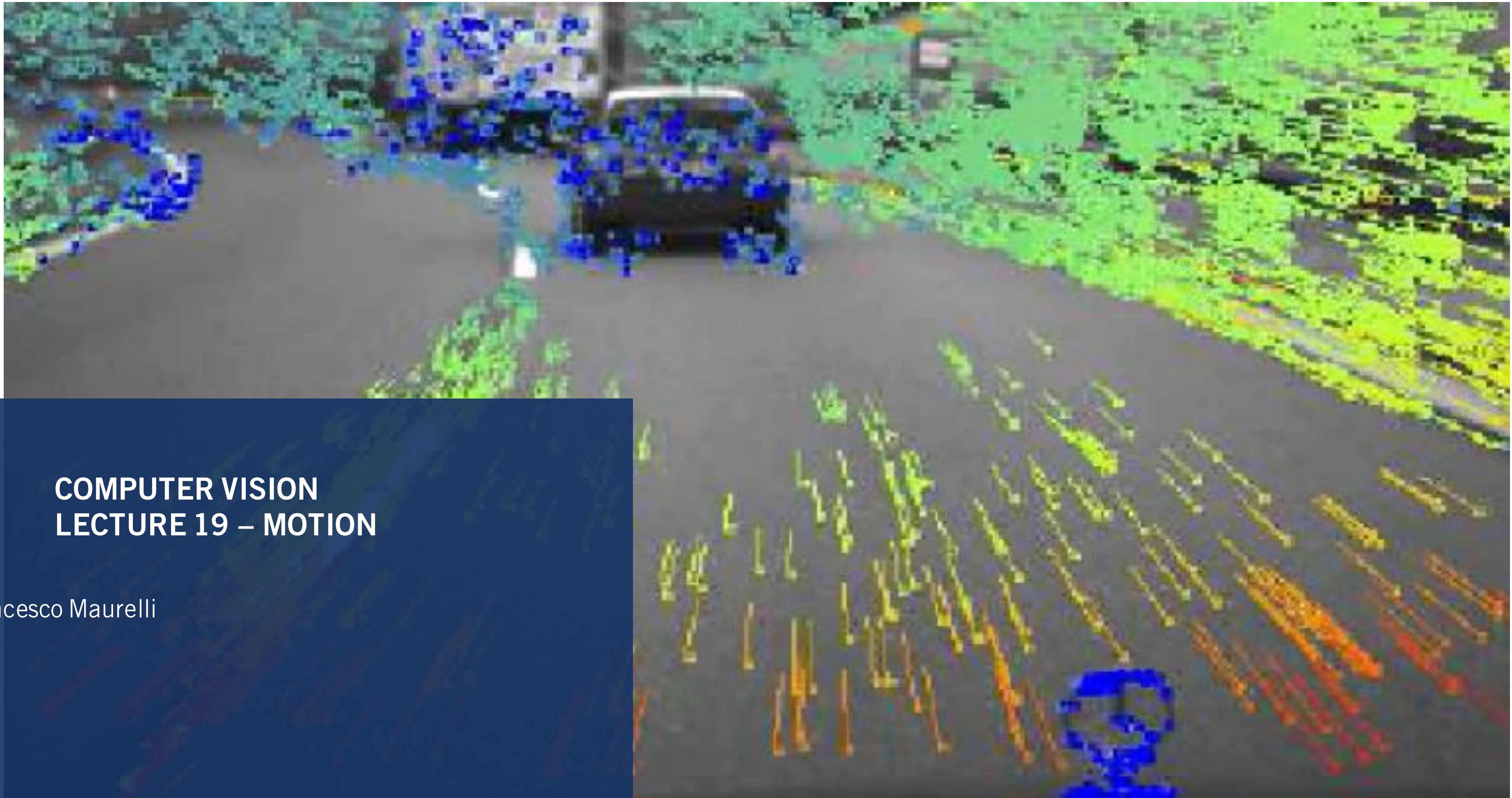


COMPUTER VISION
LECTURE 19 – MOTION

Prof. Dr. Francesco Maurelli

2018-11-09

WHAT WE WILL LEARN TODAY?

1. Optical flow
2. Lucas-Kanade method
3. Horn-Schunk method
4. Pyramids for large motion
5. Common fate
6. Applications

Reading: [Szeliski] Chapters: 8.4, 8.5

[Fleet & Weiss, 2005]
http://www.cs.toronto.edu/pub/jepson/teaching/vision/2503/opticalFlow.pdf

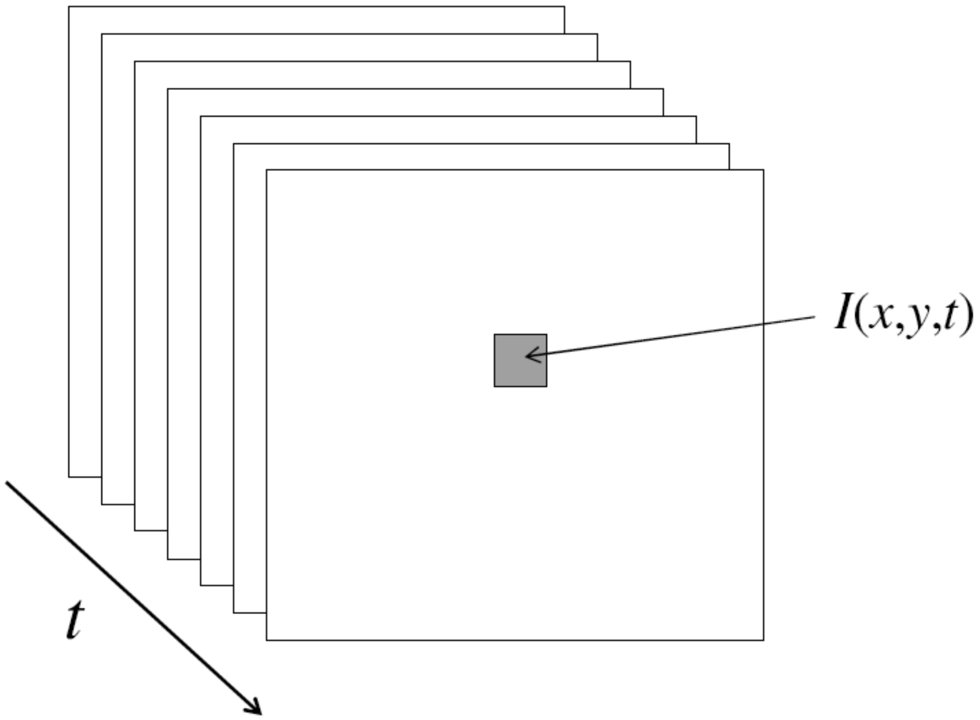
WHAT WE WILL LEARN TODAY?

1. Optical flow
2. Horn-Schunk method
3. Lucas-Kanade method
4. Pyramids for large motion
5. Common fate
6. Applications

Reading: [Szeliski] Chapters: 8.4, 8.5

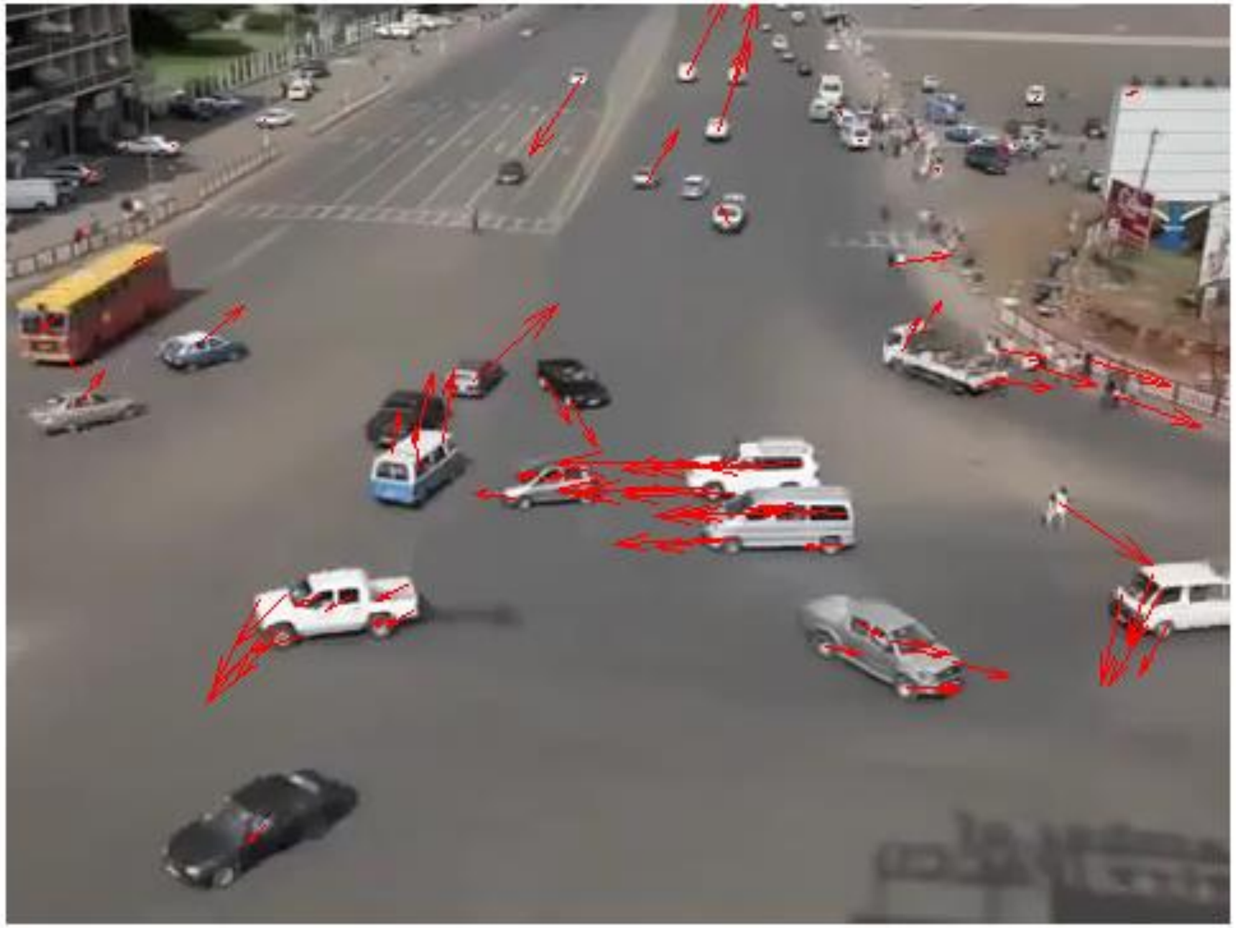
[Fleet & Weiss, 2005]
http://www.cs.toronto.edu/pub/jepson/teaching/vision/2503/opticalFlow.pdf

1. A video is a sequence of frames captured over time
2. Now our image data is a function of space (x, y) and time (t)



WHY IS MOTION USEFUL?

WHY IS MOTION USEFUL?



Definition: optical flow is the *apparent* motion of brightness patterns in the image

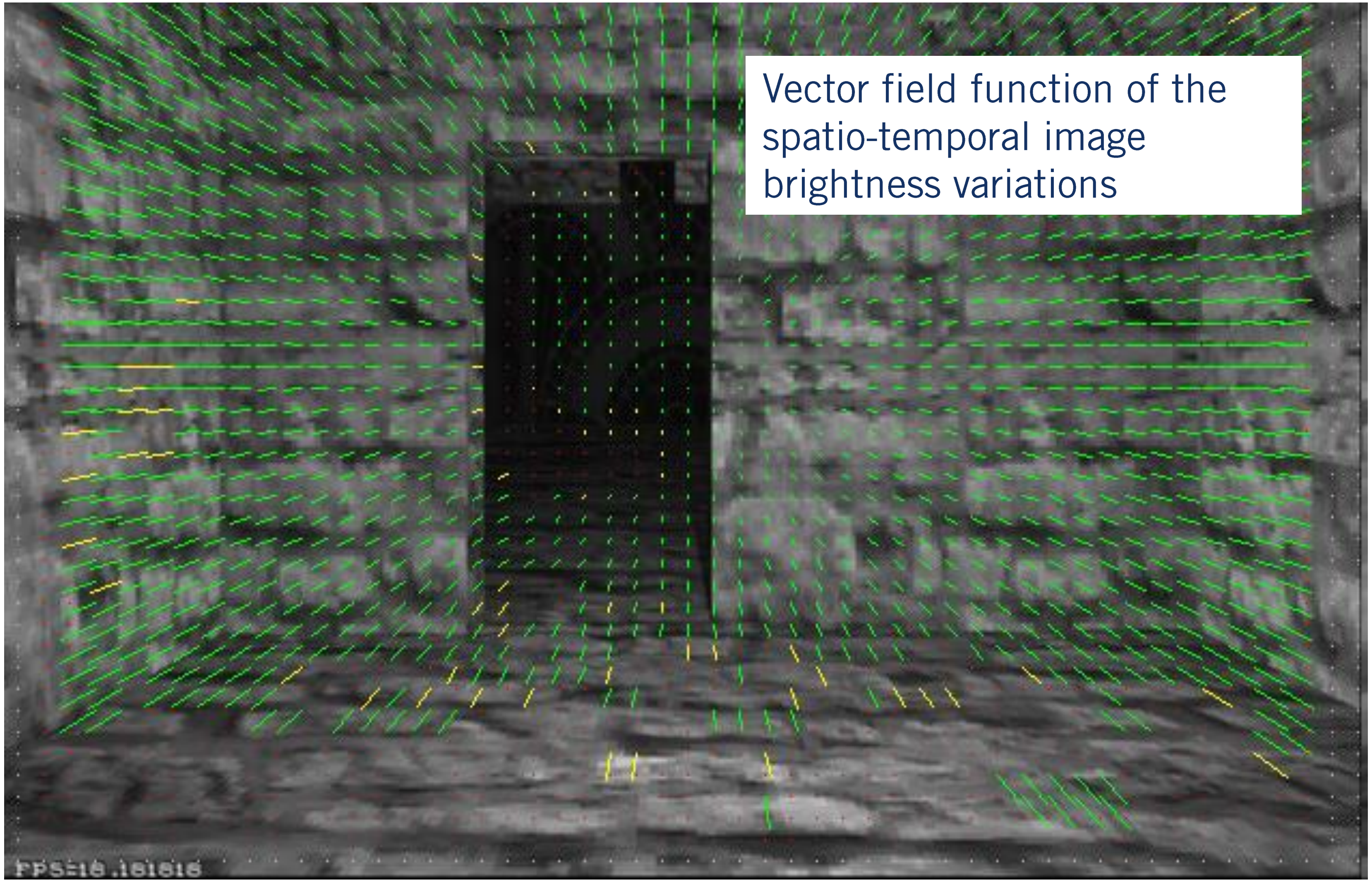
Note: apparent motion can be caused by lighting changes without any actual motion

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

GOAL

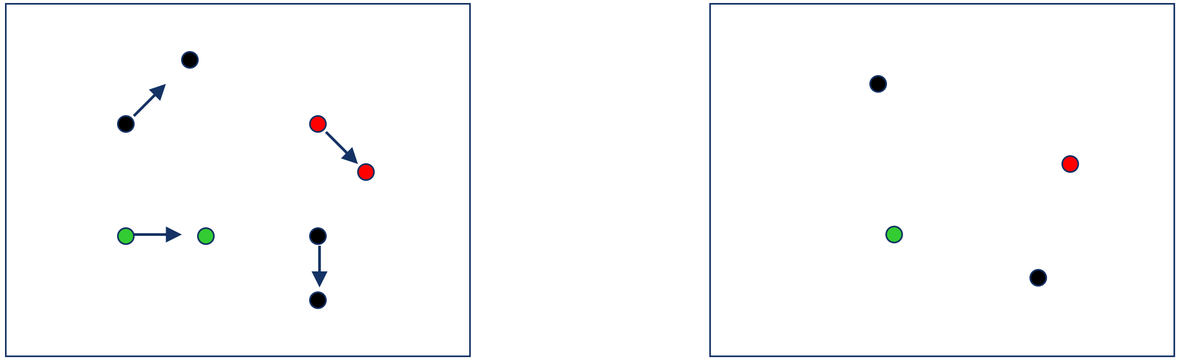
Recover image motion at each pixel from optical flow

OPTICAL FLOW



Picture courtesy of Selim Temizer - Learning and Intelligent Systems (LIS) Group, MIT

ESTIMATING OPTICAL FLOW



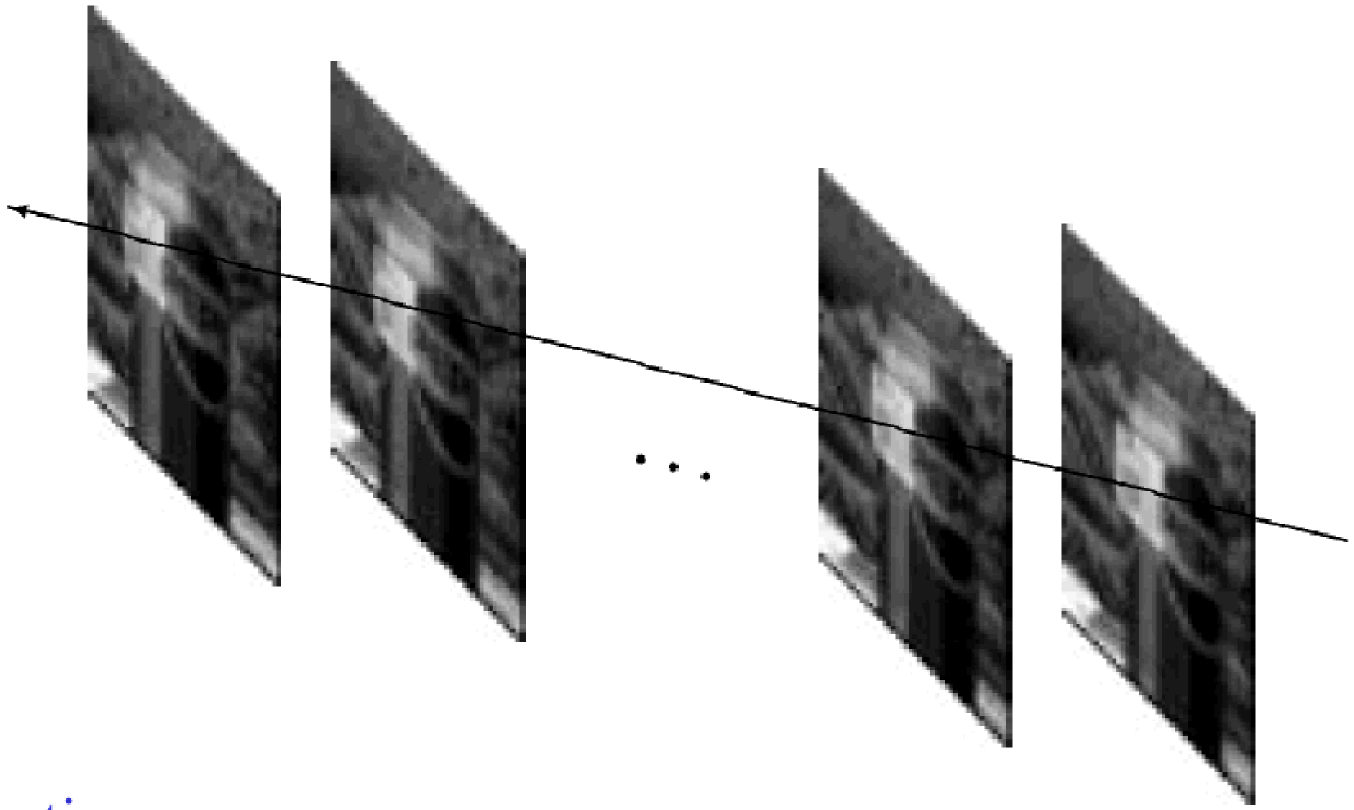
$I(x,y,t-1)$ $\quad$ $I(x,y,t)$

Given two subsequent frames, estimate the apparent motion field u(x,y), 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

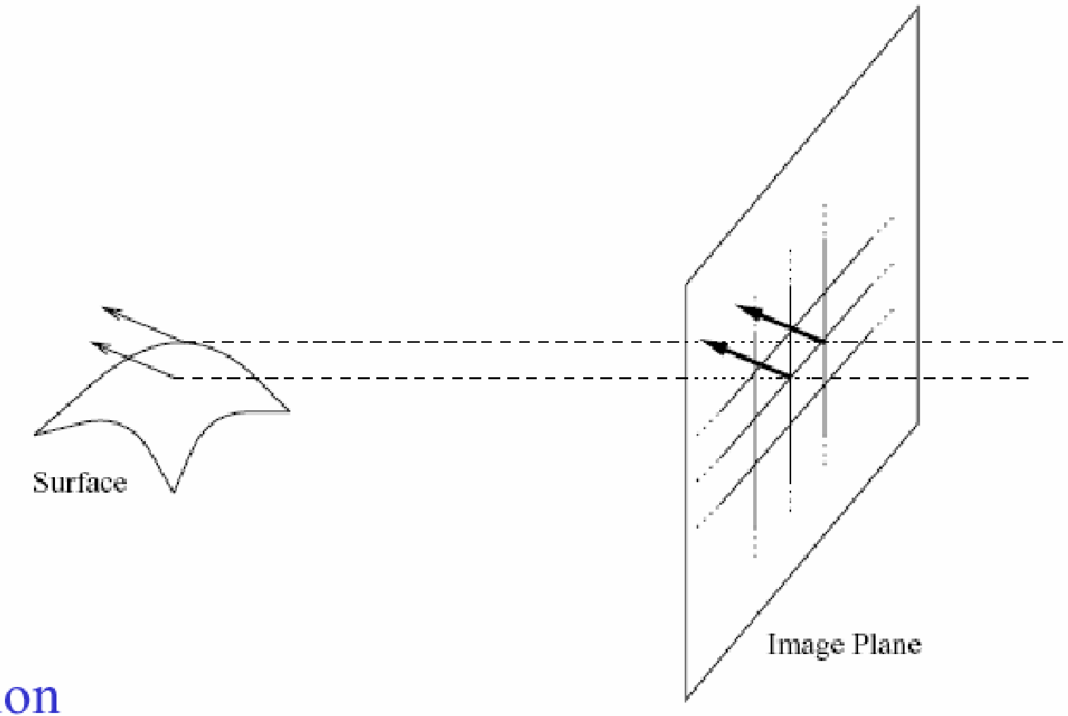
KEY ASSUMPTIONS: SMALL MOTIONS



Assumption:

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

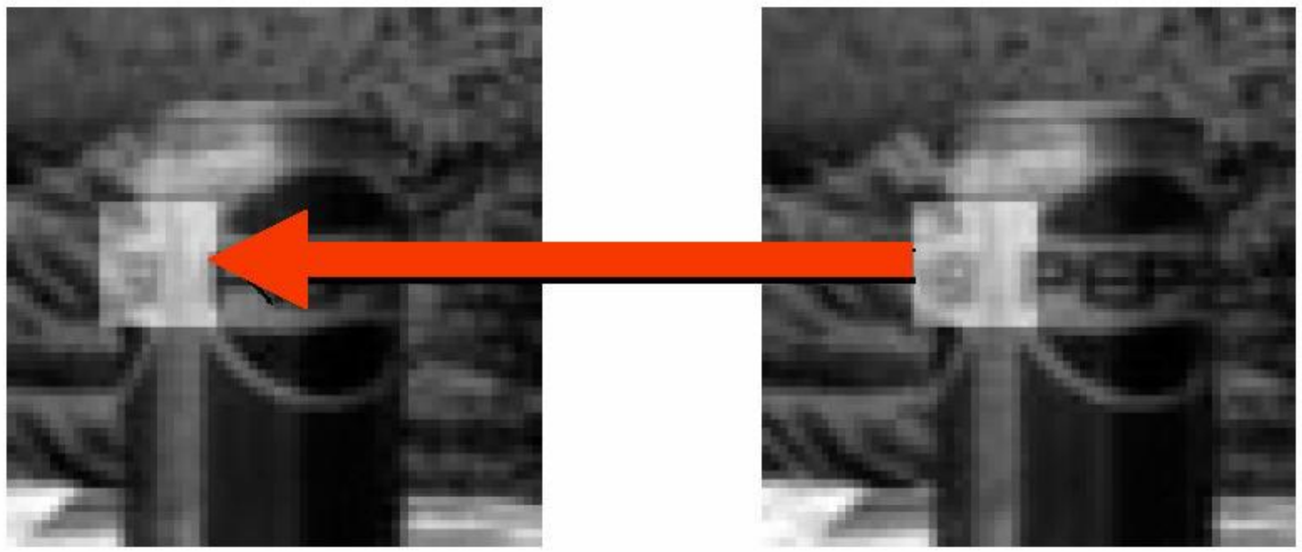
KEY ASSUMPTIONS: SPATIAL COHERENCE



Assumption

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

KEY ASSUMPTIONS: BRIGHTNESS CONSTANCY



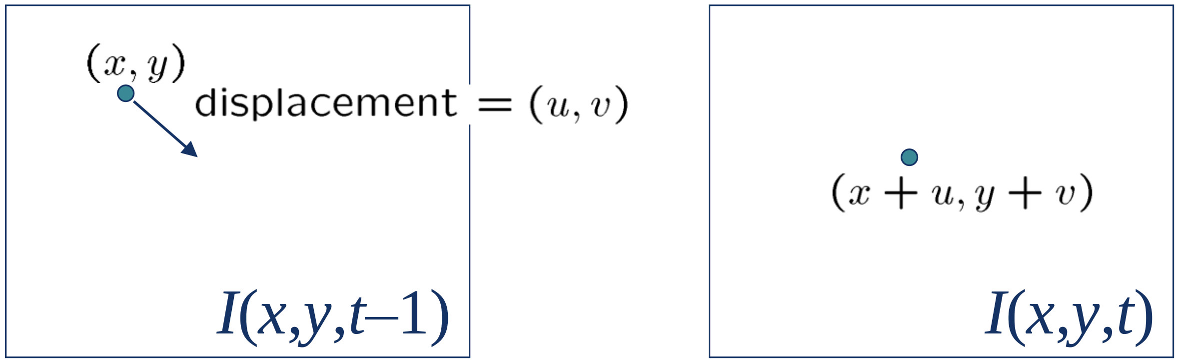
Assumption

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

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

(assumption)

JACOBS UNIVERSITY



1. 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:

Image derivative along x

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

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

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

THE BRIGHTNESS CONSTANCY CONSTRAINT

JACOBS UNIVERSITY

https://users.cs.cf.ac.uk/Dave.Marshall/Vision_lecture/node47.html

Optical Flow Constraint Equation

Let us suppose that the image intensity is given by $I(x,y,t)$, where the intensity is now a function of time, t, as well as of x and y.

At a point a small distance away, and a small time later, the intensity is

$$I(x+dx, y+dy, t+dt) = I(x,y,t) + \frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + \ldots, \qquad (61)$$

where the dots stand for higher order terms.

Now, suppose that part of an object is at a position (x,y) in the image at a time t, and that by a time dt later it has moved through a distance (dx,dy) in the image.

Furthermore, let us suppose that the intensity of that part of the object is just the same in our image before and afterwards.

Provided that we are justified in making this assumption, we then have that

$$I(x+dx, y+dy, t+dt) = I(x,y,t), \qquad (62)$$

and so

$$\frac{\partial I}{\partial x}dx + \frac{\partial I}{\partial y}dy + \frac{\partial I}{\partial t}dt + \ldots = 0. \qquad (63)$$

However, dividing through by dt, we have that

$$\frac{dx}{dt} = u, \qquad \frac{dy}{dt} = v, \qquad (64)$$

as these are the speeds the object is moving in the x and y directions respectively. Thus, in the limit that dt tends to zero, we have

$$-\frac{\partial I}{\partial t} = \frac{\partial I}{\partial x}u + \frac{\partial I}{\partial y}v, \qquad (65)$$

which is called the ***optical flow constraint equation***.

THE BARBER POLE ILLUSION

WHAT WE WILL LEARN TODAY?

1. Optical flow
2. Lucas-Kanade method
3. Horn-Schunk method
4. Pyramids for large motion
5. Common fate
6. Applications

Reading: [Szeliski] Chapters: 8.4, 8.5

[Fleet & Weiss, 2005]
http://www.cs.toronto.edu/pub/jepson/teaching/vision/2503/opticalFlow.pdf

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

How to get more equations for a pixel?

Spatial coherence constraint:

Assume the pixel's neighbors have the same (u,v)

If we use a 5x5 window, that gives us 25 equations per pixel

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

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

Overconstrained linear system:

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\times 2}{A} \; \underset{2\times 1}{d} = \underset{25\times 1}{b}$$

JACOBS UNIVERSITY

1. Overconstrained linear system

$$\begin{bmatrix} I_x(\mathbf{p}_1) & I_y(\mathbf{p}_1) \\ I_x(\mathbf{p}_2) & I_y(\mathbf{p}_2) \\ \vdots & \vdots \\ I_x(\mathbf{p}_{25}) & I_y(\mathbf{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathbf{p}_1) \\ I_t(\mathbf{p}_2) \\ \vdots \\ I_t(\mathbf{p}_{25}) \end{bmatrix} \qquad \underset{25\text{x}2}{A} \; \underset{2\text{x}1}{d} = \underset{25\text{x}1}{b}$$

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

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

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

- Optimal (u, v) satisfies Lucas-Kanade equation

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

When is This Solvable?

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

Does this remind anything to you?

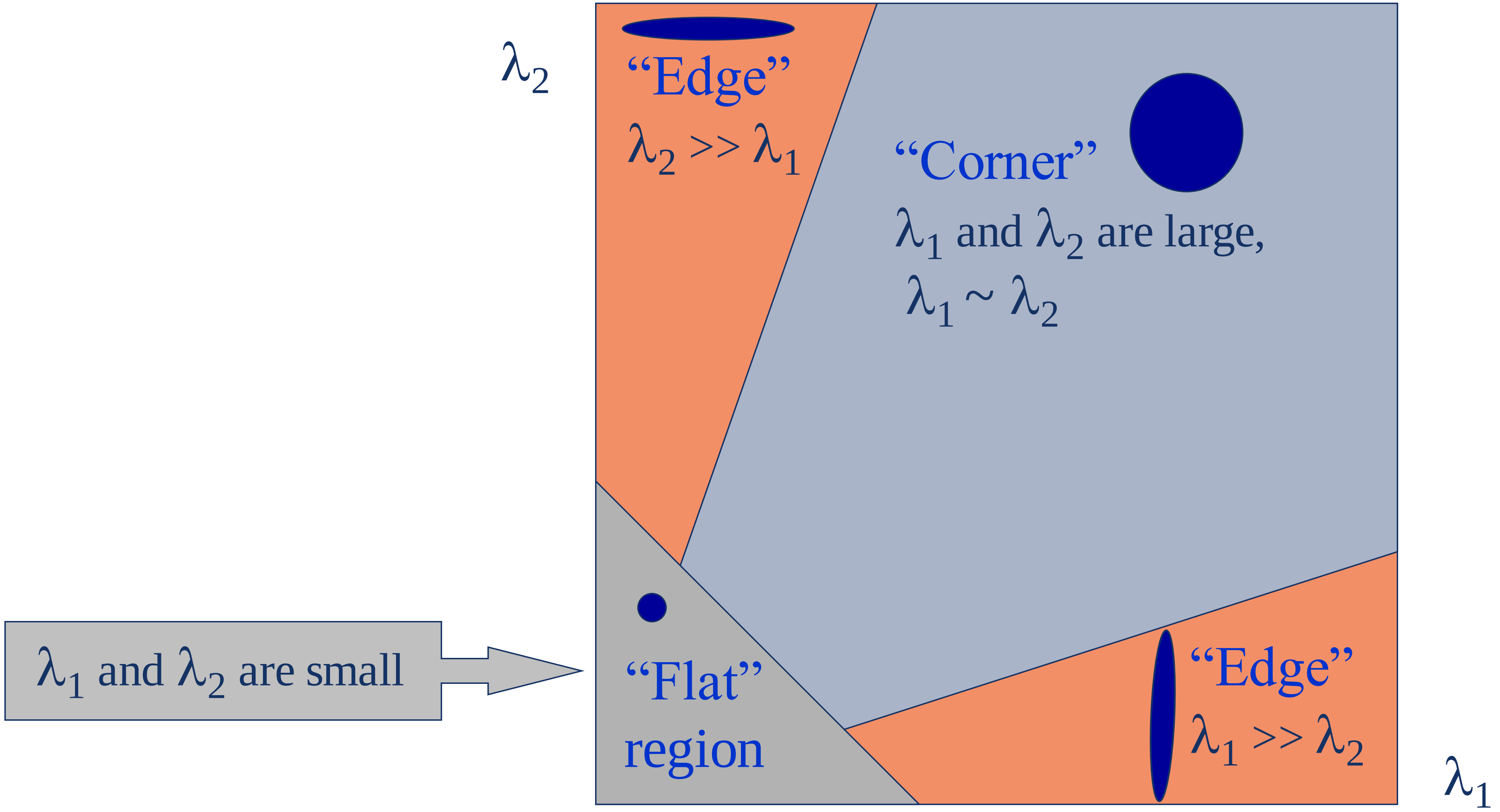
M = A^TA is the *second moment matrix* ! (Harris corner detector…)

$$A^TA = \begin{bmatrix} \sum I_x I_x & \sum I_x I_y \\ \sum I_x I_y & \sum I_y I_y \end{bmatrix} = \sum \begin{bmatrix} I_x \\ I_y \end{bmatrix} [I_x \; I_y] = \sum \nabla I (\nabla I)^T$$

Eigenvectors and eigenvalues of A^TA relate to edge direction and magnitude

- The eigenvector associated with the larger eigenvalue points in the direction of fastest intensity change
- The other eigenvector is orthogonal to it

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



EDGE

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

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

LOW-TEXTURE REGION

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

– gradients have small magnitude

– small λ_1, small λ_2

HIGH-TEXTURE REGION

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

- gradients are different, large magnitudes
- large λ_1, large λ_2

What are the potential causes of errors in this procedure?

- Suppose A^TA 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?

Recall our small motion assumption

$$0 = I(x+u, y+v) - I_{t-1}(x,y)$$

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

- This is not exact
  - To do better, we need to add higher order terms back in:

$$= I(x,y) + I_x u + I_y v + \text{higher order terms} - I_{t-1}(x,y)$$

- This is a polynomial root finding problem
  - Can solve using **Newton's method (out of scope for this class)**
  - Lukas-Kanade method does one iteration of Newton's method
    - Better results are obtained via more iterations

JACOBS UNIVERSITY

Iterative Lukas-Kanade Algorithm

1. Estimate velocity at each pixel by solving Lucas-Kanade equations
2. Warp I(t-1) towards I(t) using the estimated flow field

   - use image warping techniques

3. Repeat until convergence

WHEN DO THE OPTICAL FLOW ASSUMPTIONS FAIL?

In other words, in what situations does the displacement of pixel patches not represent physical movement of points in space?

Well, TV is based on illusory motion
– the set is stationary yet things seem to move

A uniform rotating sphere
– nothing seems to move, yet it is rotating

Changing directions or intensities of lighting can make things seem to move

4. Muscle movement can make some spots on a cheetah move opposite direction of motion.
– And infinitely more break downs of optical flow.

WHAT WE WILL LEARN TODAY?

1. Optical flow
2. Lucas-Kanade method
3. Horn-Schunk method
4. Pyramids for large motion
5. Common fate
6. Applications

Reading: [Szeliski] Chapters: 8.4, 8.5

[Fleet & Weiss, 2005]
http://www.cs.toronto.edu/pub/jepson/teaching/vision/2503/opticalFlow.pdf

HORN-SCHUNK METHOD FOR OPTICAL FLOW

The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 \left(\|\nabla u\|^2 + \|\nabla v\|^2 \right) \right] \mathrm{d}x \mathrm{d}y$$

HORN-SCHUNK METHOD FOR OPTICAL FLOW

The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[\boxed{(I_x u + I_y v + I_t)}^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \right] \mathrm{d}x\mathrm{d}y$$

The first part of the function is the brightness consistency.

HORN-SCHUNK METHOD FOR OPTICAL FLOW

The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 \left(\boxed{\|\nabla u\|^2 + \|\nabla v\|^2} \right) \right] \mathrm{d}x\mathrm{d}y$$

The second part is the smoothness constraint. It's trying to make sure that the changes between frames are small.

HORN-SCHUNK METHOD FOR OPTICAL FLOW

The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \boxed{\alpha^2}(\|\nabla u\|^2 + \|\nabla v\|^2)\right] \mathrm{d}x\mathrm{d}y$$

α is a regularization constant. Larger values of α lead to smoother flow.

HORN-SCHUNK METHOD FOR OPTICAL FLOW

The flow is formulated as a global energy function which is should be minimized:

$$E = \iint \left[(I_x u + I_y v + I_t)^2 + \alpha^2 (\|\nabla u\|^2 + \|\nabla v\|^2) \right] \mathrm{d}x\mathrm{d}y$$

By taking the derivative with respect to u and v, we get the following 2 equations:

$$I_x (I_x u + I_y v + I_t) - \alpha^2 \Delta u = 0$$
$$I_y (I_x u + I_y v + I_t) - \alpha^2 \Delta v = 0$$

By taking the derivative with respect to u and v, we get the following 2 equations:

$$I_x(I_x u + I_y v + I_t) - \alpha^2 \Delta u = 0$$
$$I_y(I_x u + I_y v + I_t) - \alpha^2 \Delta v = 0$$

Where $\Delta = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}$ is called the Lagrange operator.

In practice, it is measured using: $$\Delta u(x,y) = \overline{u}(x,y) - u(x,y)$$

where $\overline{u}(x,y)$ is the weighted average of u measured at (x,y).

Now we substitute in:

$$\Delta u(x,y) = \overline{u}(x,y) - u(x,y)$$

$$I_x(I_x u + I_y v + I_t) - \alpha^2 \Delta u = 0$$
$$I_y(I_x u + I_y v + I_t) - \alpha^2 \Delta v = 0$$

to get:

$$(I_x^2 + \alpha^2)u + I_x I_y v = \alpha^2 \overline{u} - I_x I_t$$
$$I_x I_y u + (I_y^2 + \alpha^2)v = \alpha^2 \overline{v} - I_y I_t$$

which is linear in u and v and can be solved for each pixel individually.

ITERATIVE HORN-SCHUNK

But since the solution depends on the neighboring values of the flow field, it must be repeated once the neighbors have been updated.

So instead, we can iteratively solve for u and v using:

$$u^{k+1} = \overline{u}^k - \frac{I_x(I_x\overline{u}^k + I_y\overline{v}^k + I_t)}{\alpha^2 + I_x^2 + I_y^2}$$

$$v^{k+1} = \overline{v}^k - \frac{I_y(I_x\overline{u}^k + I_y\overline{v}^k + I_t)}{\alpha^2 + I_x^2 + I_y^2}$$

WHAT WE WILL LEARN TODAY?

1. Optical flow
2. Lucas-Kanade method
3. Horn-Schunk method
4. Pyramids for large motion
5. Common fate
6. Applications

Key assumptions (Errors in Lucas-Kanade)

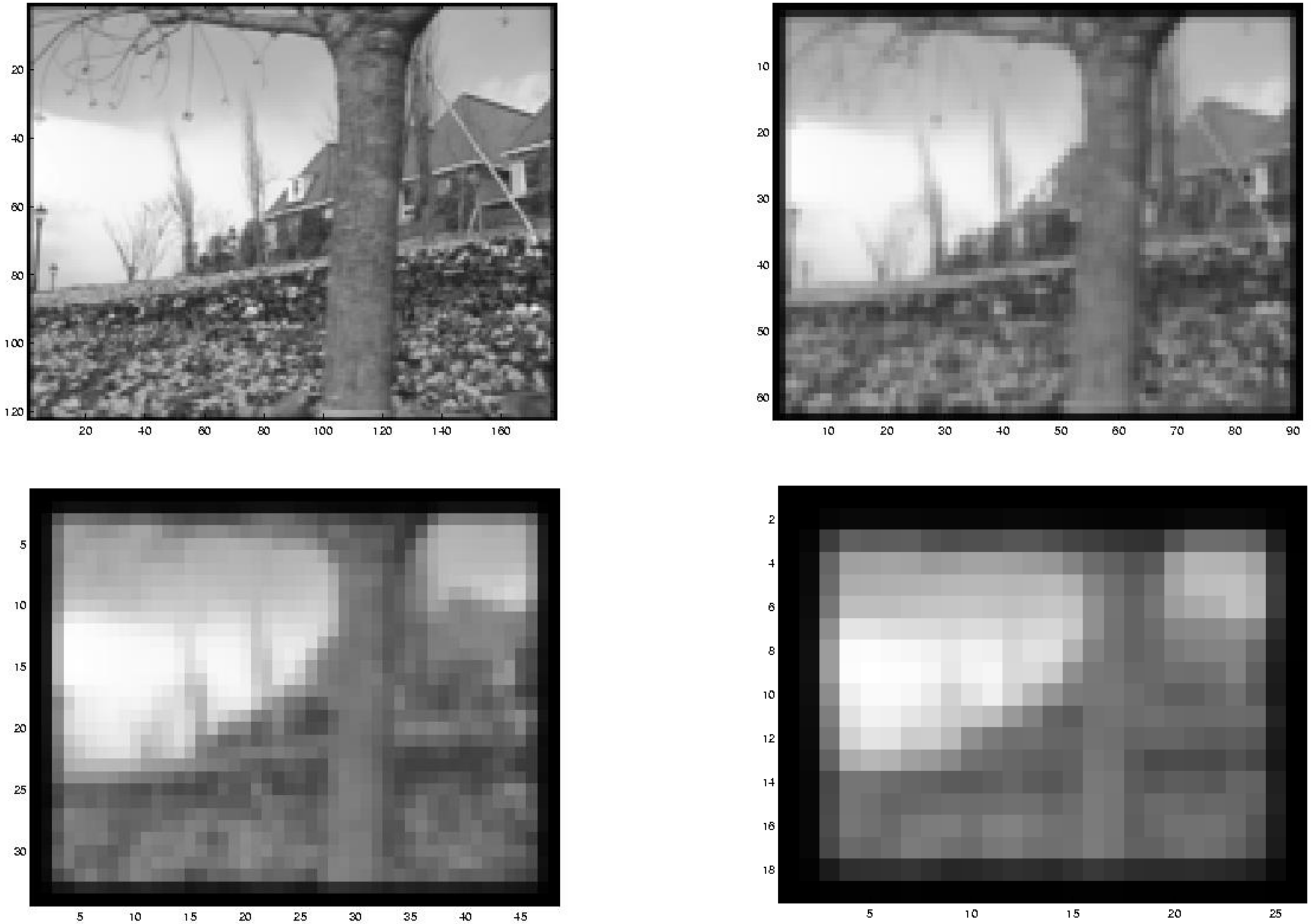
- **Small motion:** points do not move very far
- **Brightness constancy:** projection of the same point looks the same in every frame
- **Spatial coherence:** points move like their neighbors

REVISITING THE SMALL MOTION ASSUMPTION

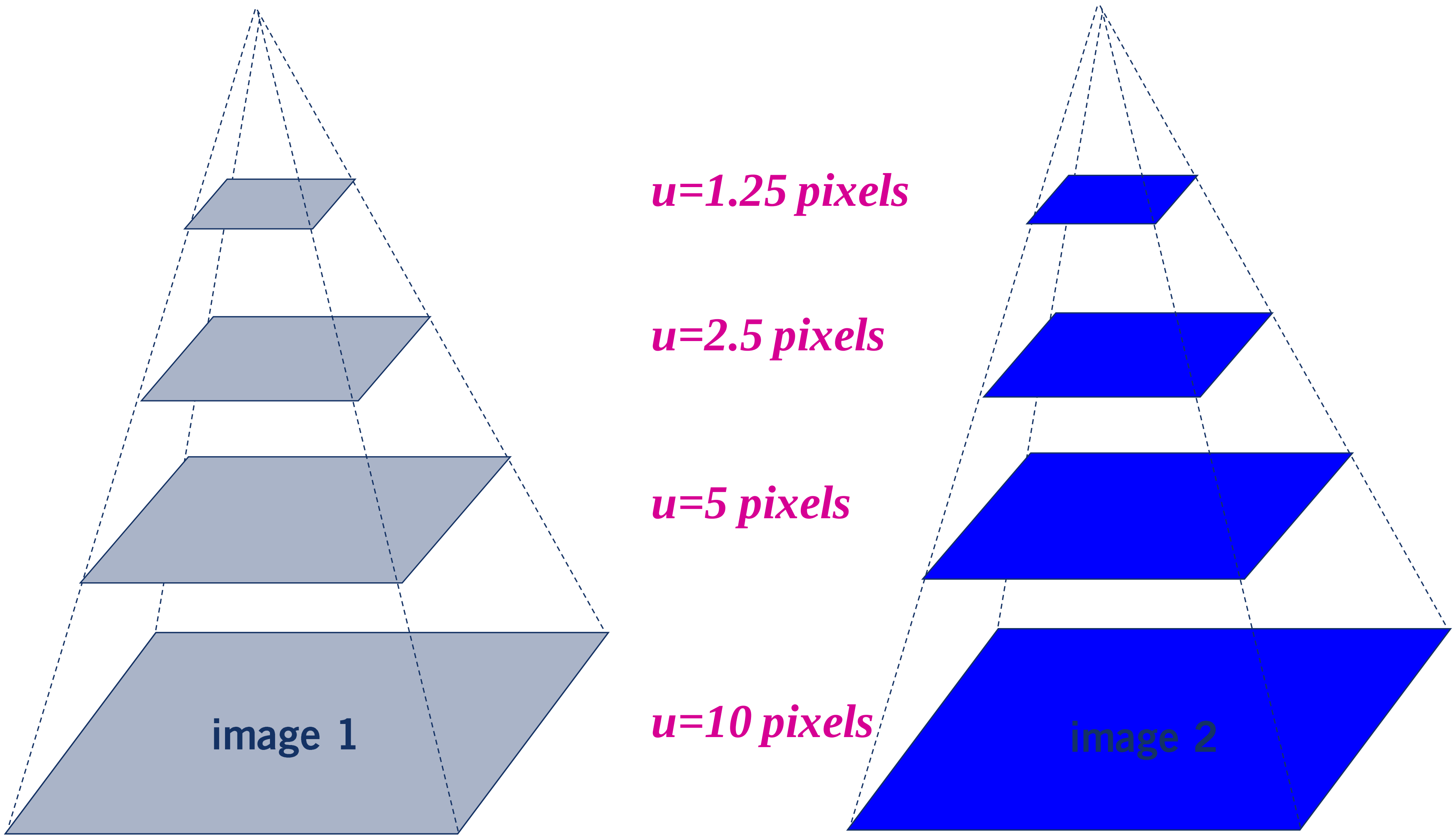
Is this motion small enough?

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

REDUCE THE RESOLUTION!



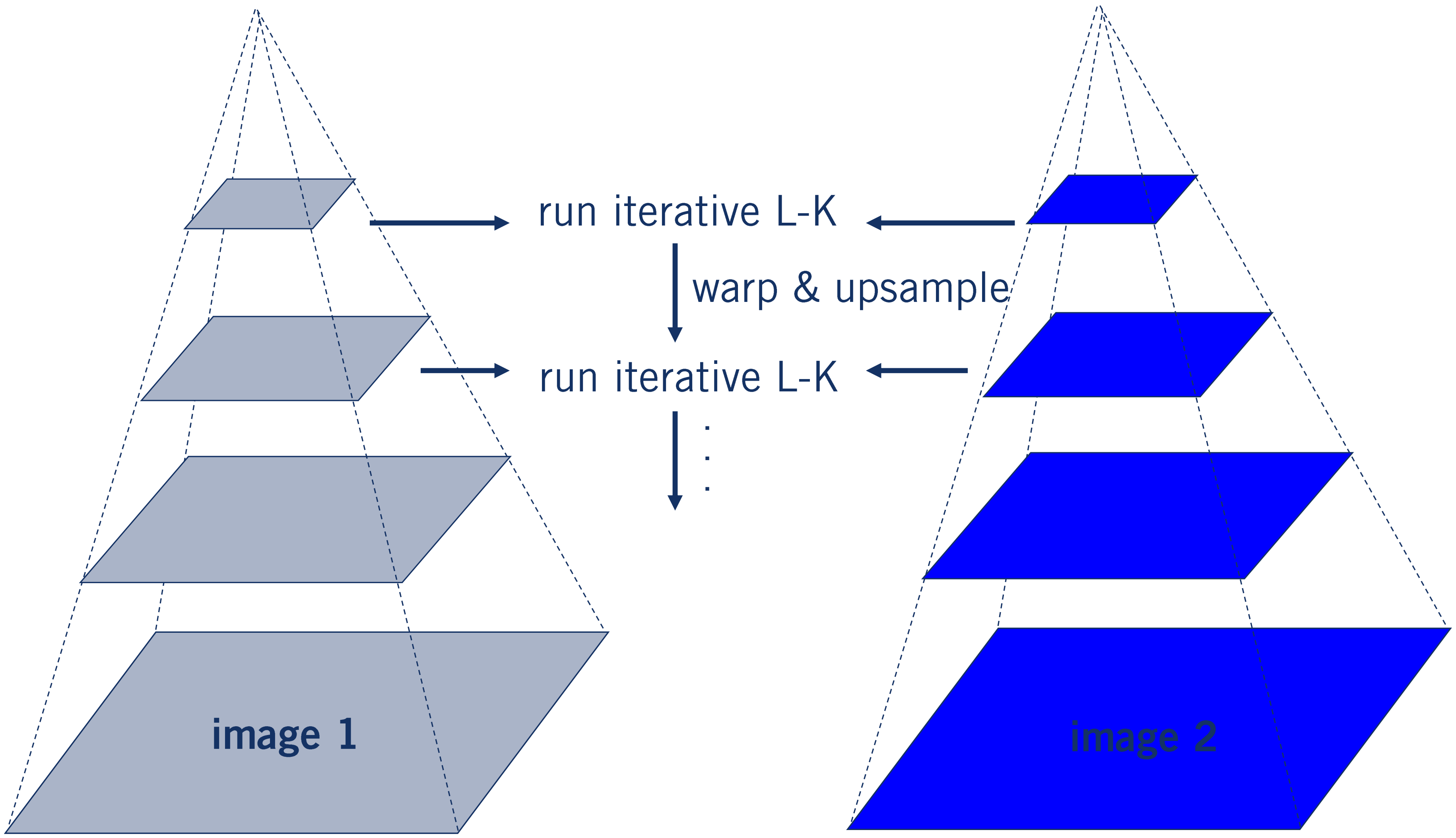
COARSE-TO-FINE OPTICAL FLOW ESTIMATION



Gaussian pyramid of image 1

Gaussian pyramid of image 2

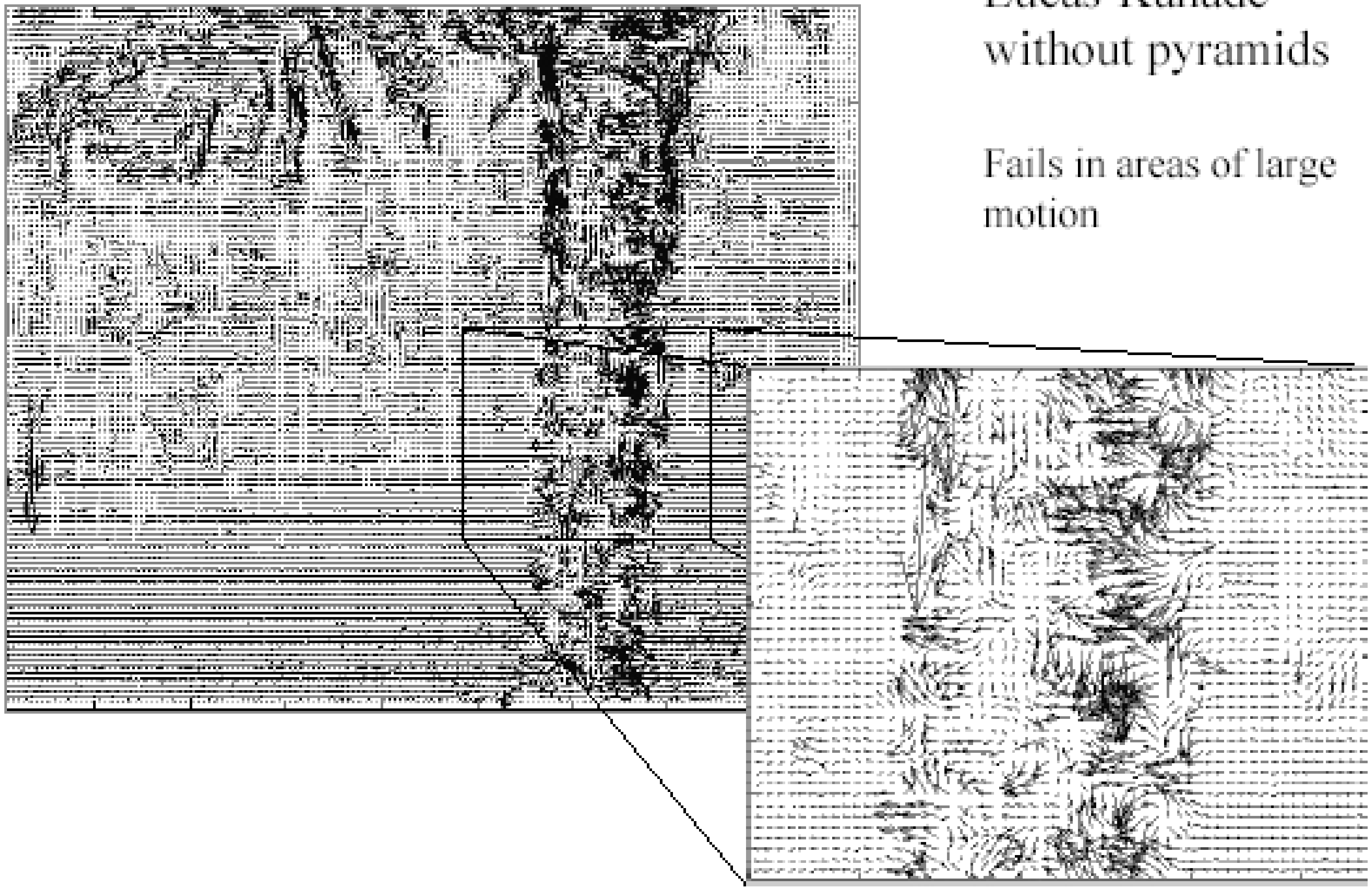
COARSE-TO-FINE OPTICAL FLOW ESTIMATION



Gaussian pyramid of image 1 (t)

Gaussian pyramid of image 2 (t+1)

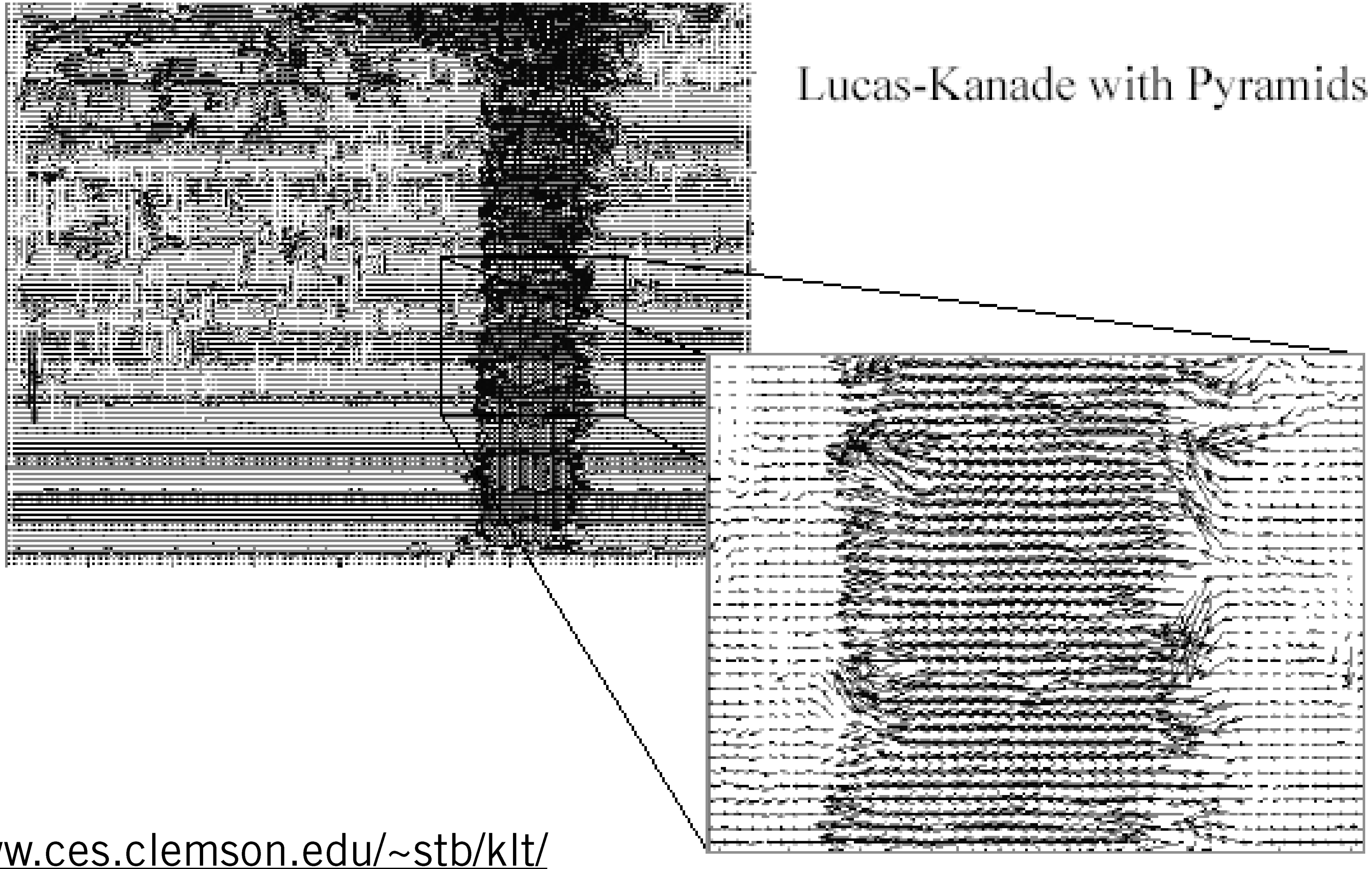
OPTICAL FLOW RESULTS



Lucas-Kanade
without pyramids

Fails in areas of large
motion

OPTICAL FLOW RESULTS



http://www.ces.clemson.edu/~stb/klt/
OpenCV

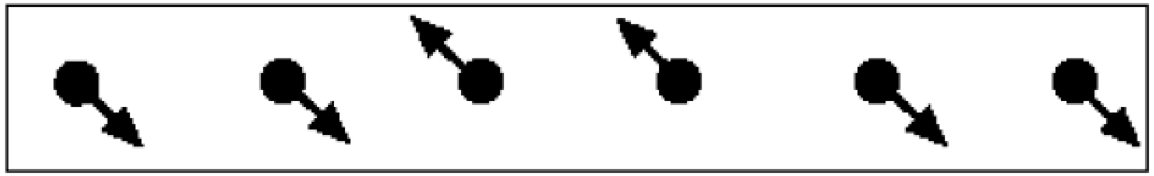
WHAT WE WILL LEARN TODAY?

1. Optical flow
2. Lucas-Kanade method
3. Horn-Schunk method
4. Pyramids for large motion
5. Common fate
6. Applications

Key assumptions (Errors in Lucas-Kanade)

- **Small motion:** points do not move very far
- **Brightness constancy:** projection of the same point looks the same in every frame
- **Spatial coherence:** points move like their neighbors

REMINDER: GESTALT – COMMON FATE



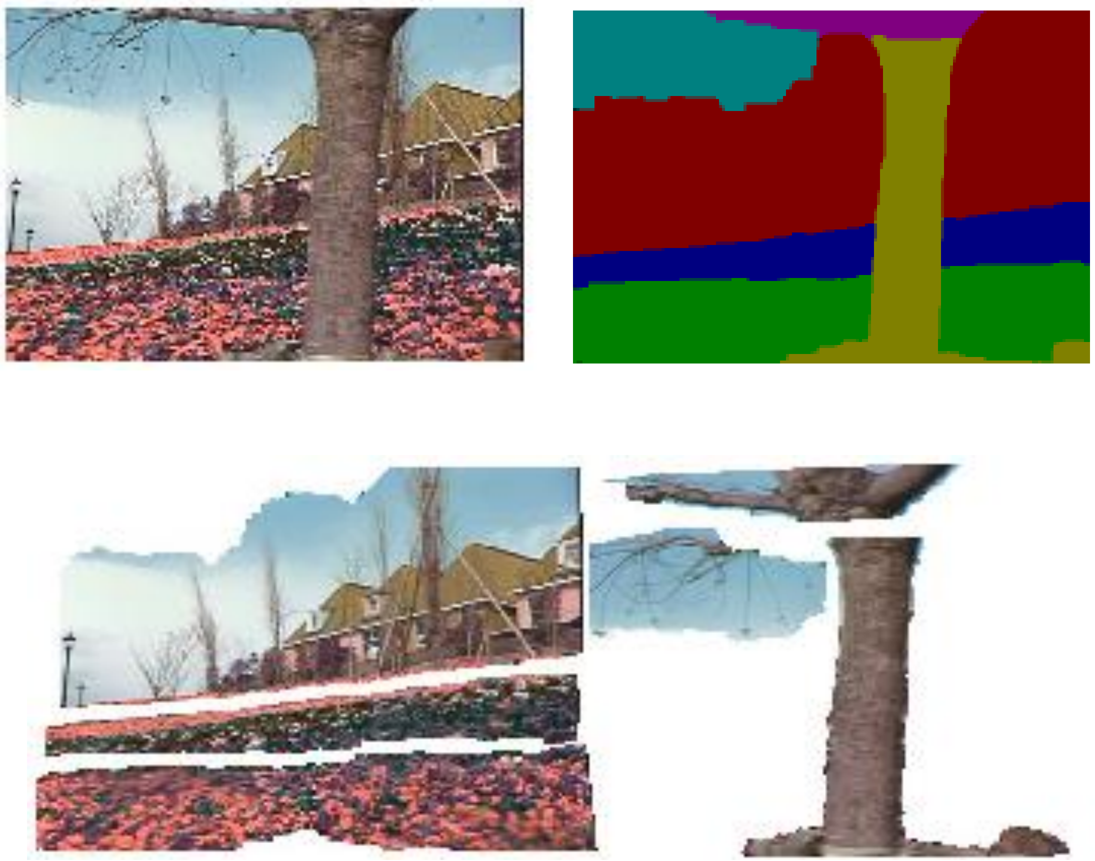
Common Fate

How do we represent the motion in this scene?

MOTION SEGMENTATION

J. Wang and E. Adelson. Layered Representation for Motion Analysis. *CVPR 1993*.

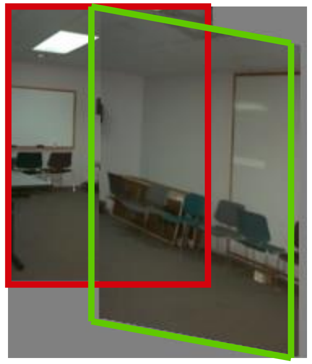
Break image sequence into “layers” each of which has a coherent (affine) motion



AFFINE MOTION

$$u(x, y) = a_1 + a_2 x + a_3 y$$

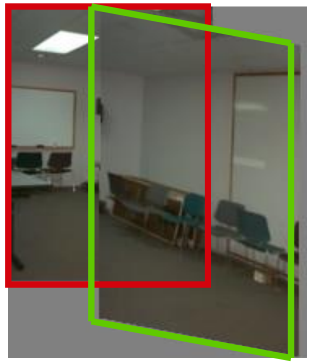
$$v(x, y) = a_4 + a_5 x + a_6 y$$



Substituting into the brightness constancy equation:

$$\boxed{I_x \cdot u + I_y \cdot v + I_t \approx 0}$$

AFFINE MOTION



$$u(x,y) = a_1 + a_2 x + a_3 y$$
$$v(x,y) = a_4 + a_5 x + a_6 y$$

Substituting into the brightness constancy equation:

$$\boxed{I_x(a_1 + a_2 x + a_3 y) + I_y(a_4 + a_5 x + a_6 y) + I_t \approx 0}$$

- Each pixel provides 1 linear constraint in 6 unknowns
- Least squares minimization:

$$\boxed{Err(\vec{a}) = \sum \left[I_x(a_1 + a_2 x + a_3 y) + I_y(a_4 + a_5 x + a_6 y) + I_t \right]^2}$$

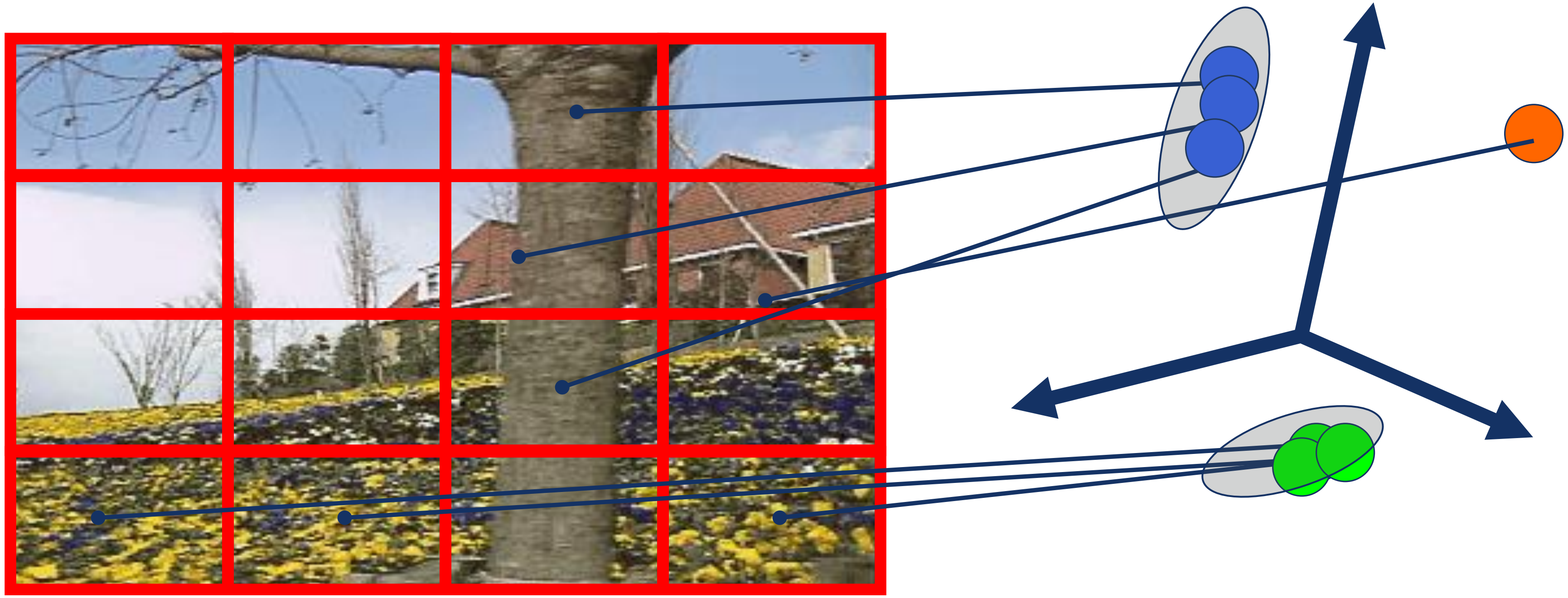
JACOBS UNIVERSITY

1. Obtain a set of initial affine motion hypotheses

- Divide the image into blocks and estimate affine motion parameters in each block by least squares

  Eliminate hypotheses with high residual error
- Map into motion parameter space
- Perform k-means clustering on affine motion parameters

  Merge clusters that are close and retain the largest clusters to obtain a smaller set of hypotheses to describe all the motions in the scene



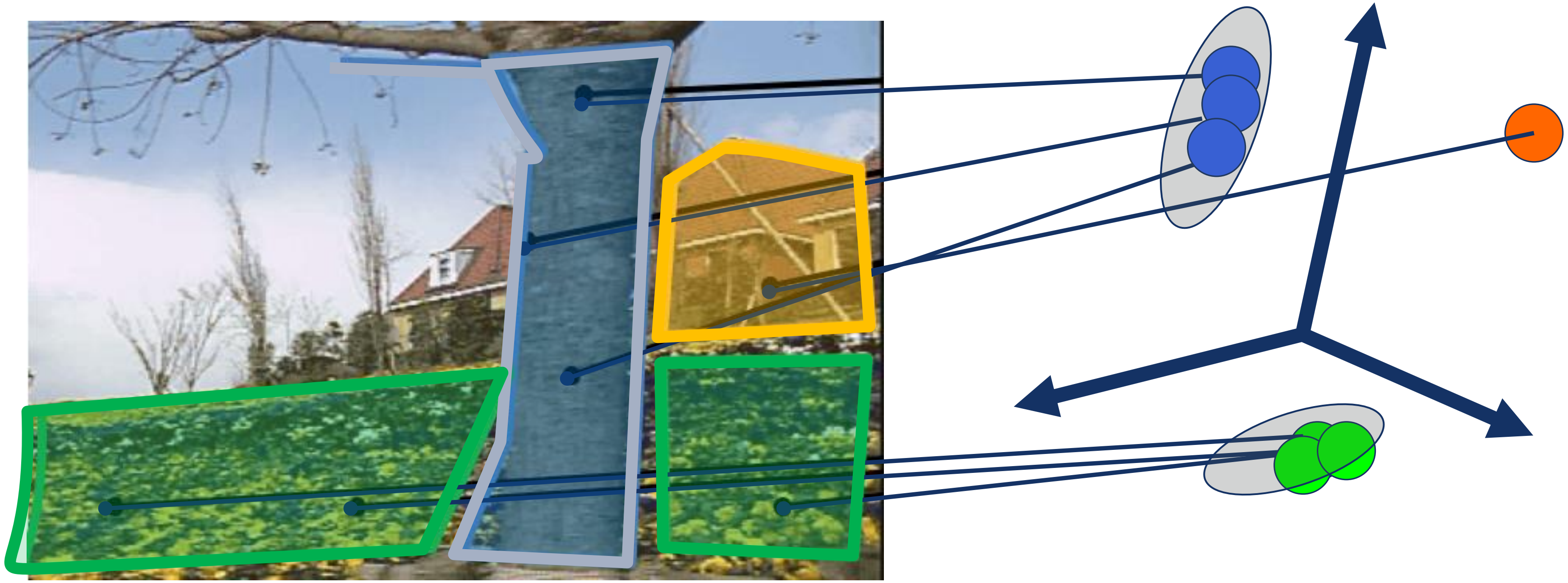
JACOBS UNIVERSITY

1. Obtain a set of initial affine motion hypotheses

- Divide the image into blocks and estimate affine motion parameters in each block by least squares

  Eliminate hypotheses with high residual error
- Map into motion parameter space
- Perform k-means clustering on affine motion parameters

  Merge clusters that are close and retain the largest clusters to obtain a smaller set of hypotheses to describe all the motions in the scene



HOW DO WE ESTIMATE THE LAYERS?

1. Obtain a set of initial affine motion hypotheses

- Divide the image into blocks and estimate affine motion parameters in each block by least squares

  Eliminate hypotheses with high residual error
- Map into motion parameter space
- Perform k-means clustering on affine motion parameters

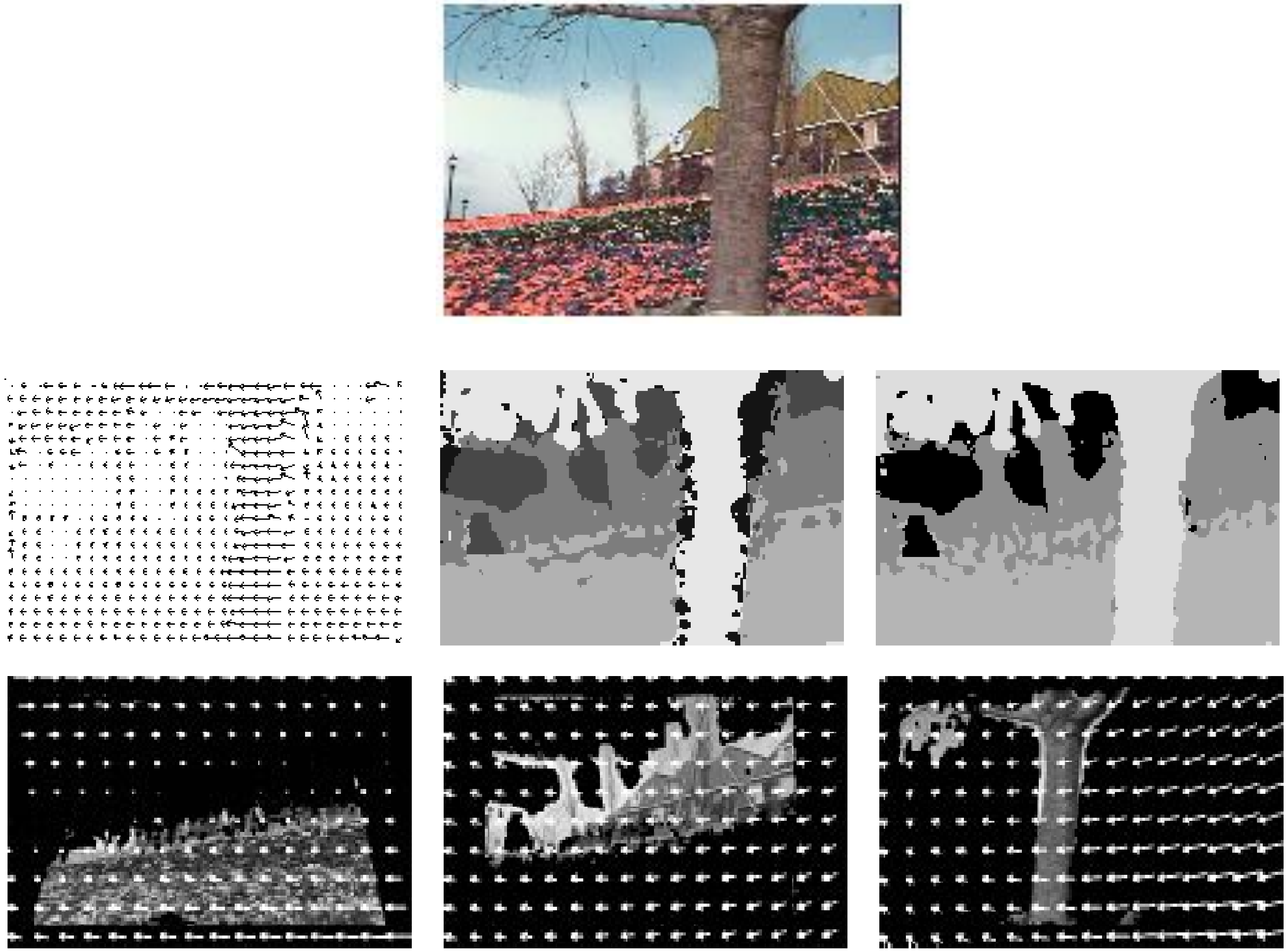
  Merge clusters that are close and retain the largest clusters to obtain a smaller set of hypotheses to describe all the motions in the scene

2. Iterate until convergence

- Assign each pixel to best hypothesis

  Pixels with high residual error remain unassigned
- Perform region filtering to enforce spatial constraints
- Re-estimate affine motions in each region

EXAMPLE RESULT



J. Wang and E. Adelson. Layered Representation for Motion Analysis. *CVPR 1993*.

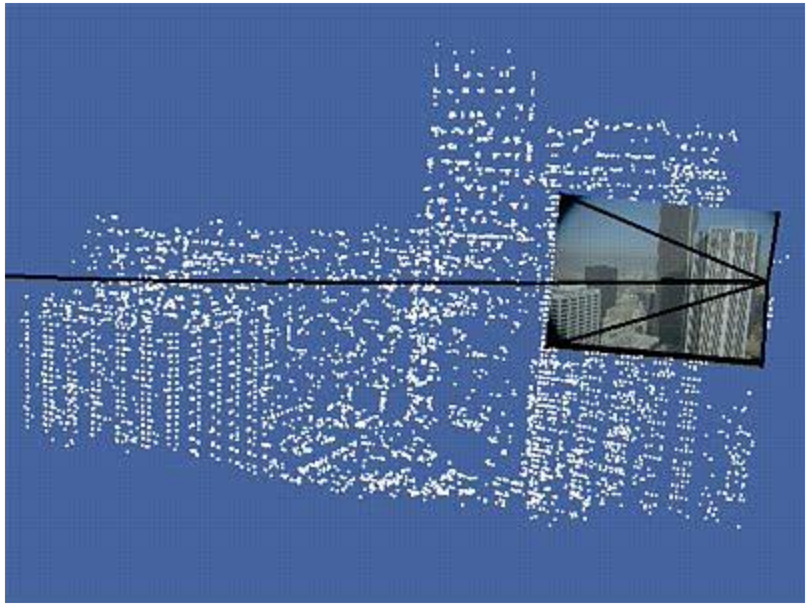
WHAT WE WILL LEARN TODAY?

1. Optical flow
2. Lucas-Kanade method
3. Horn-Schunk method
4. Pyramids for large motion
5. Common fate
6. Applications

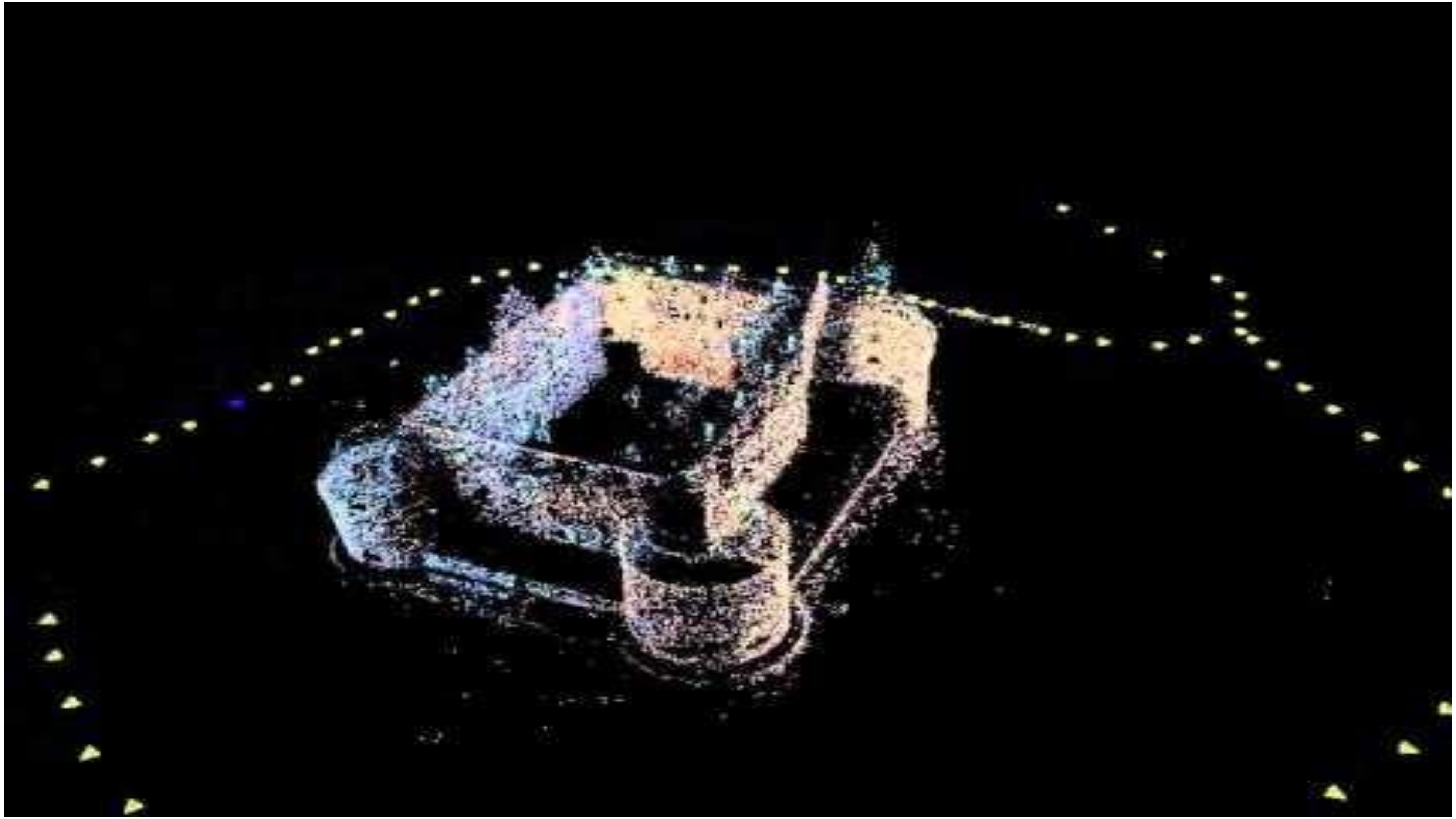
USES OF MOTION

- Tracking features
- Segmenting objects based on motion cues
- Learning dynamical models
- Improving video quality
    - Motion stabilization
    - Super resolution
- Tracking objects
- Recognizing events and activities

ESTIMATING 3D STRUCTURE



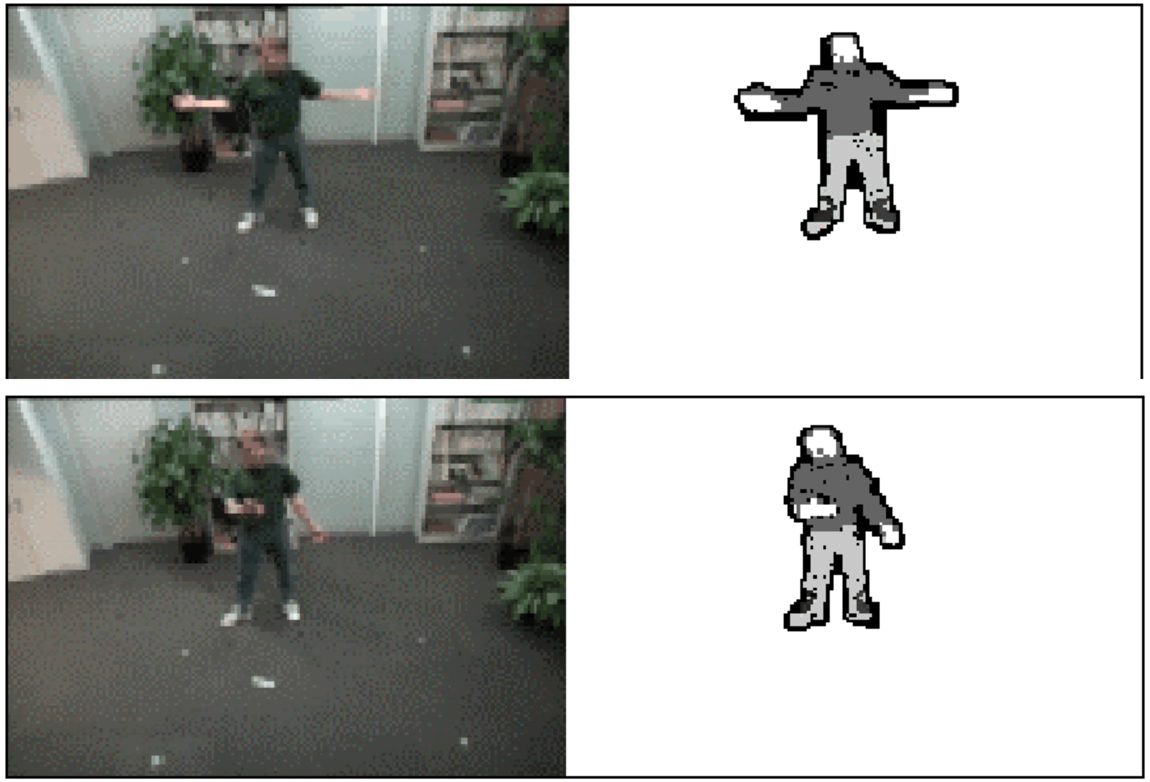
ESTIMATING 3D STRUCTURE



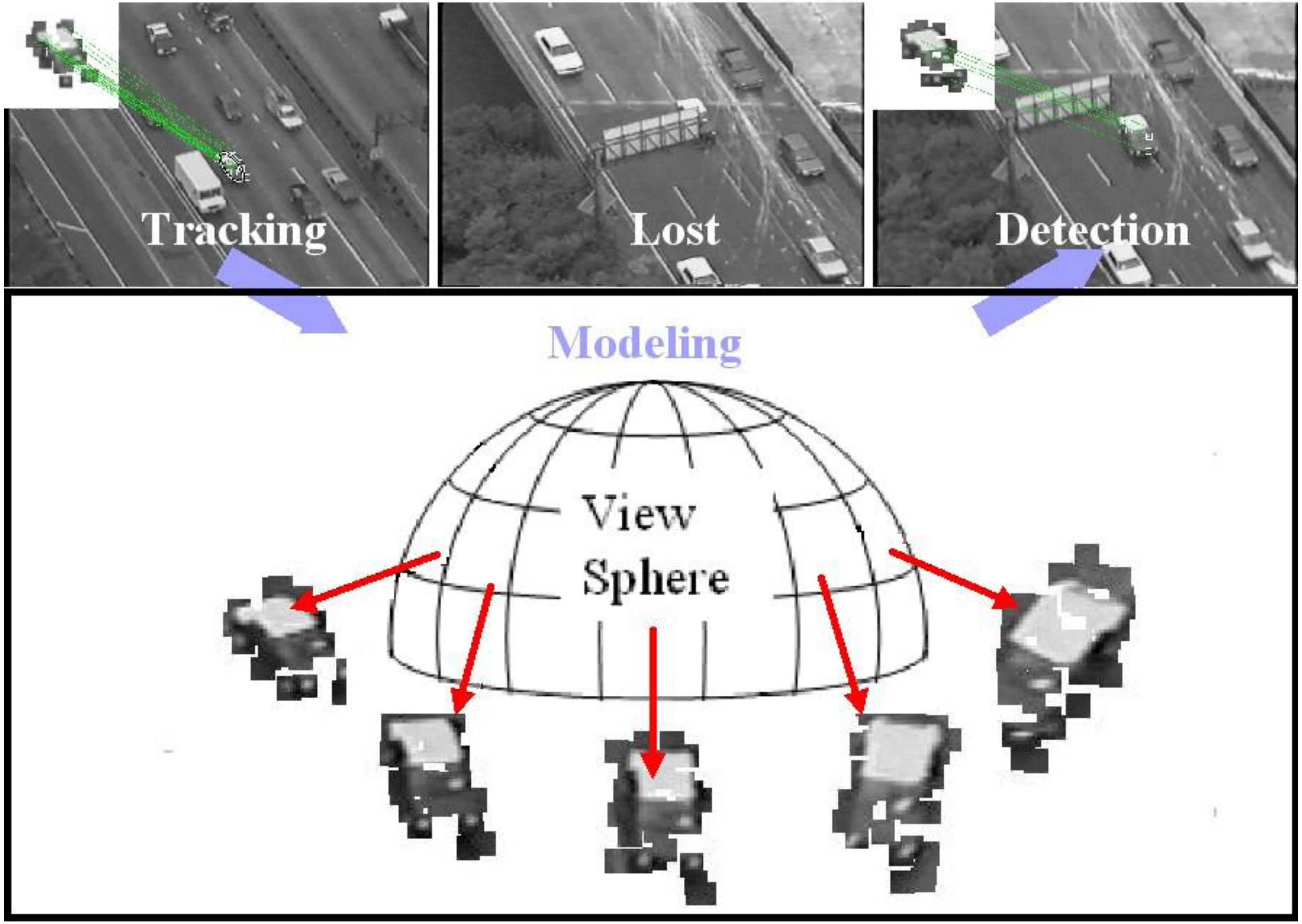
JACOBS UNIVERSITY

Background subtraction

- A static camera is observing a scene
- Goal: separate the static *background* from the moving *foreground*



TRACKING OBJECTS

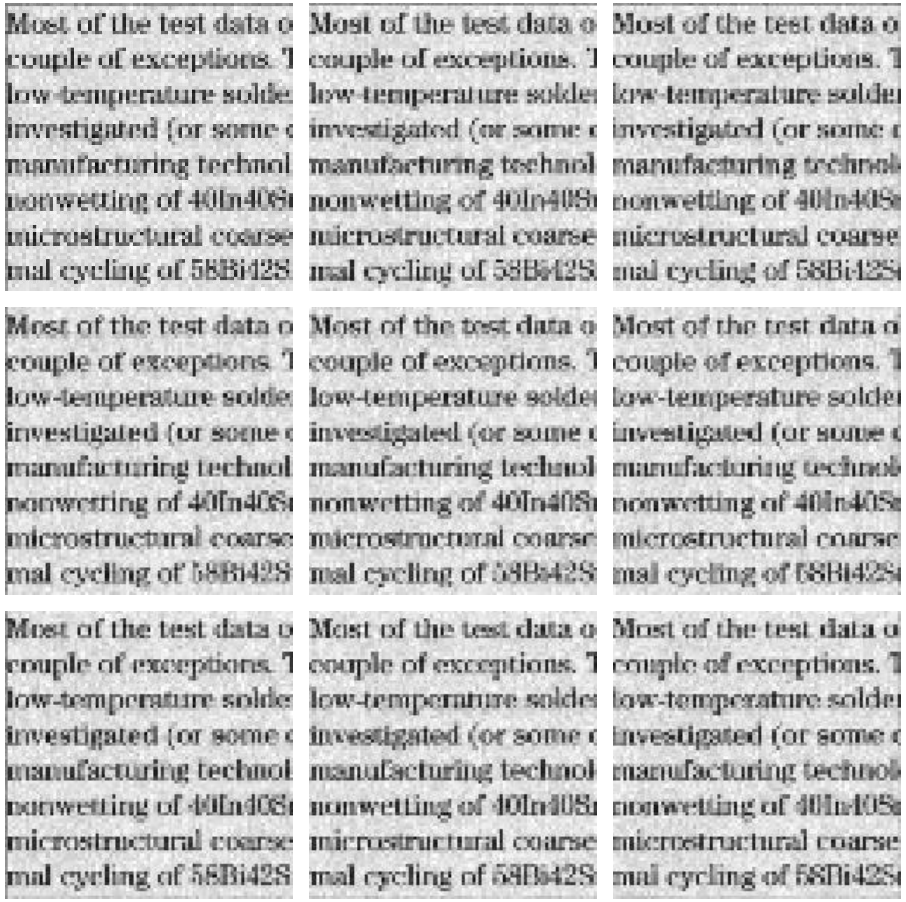


Z.Yin and R.Collins, "On-the-fly Object Modeling while Tracking," *IEEE Computer Vision and Pattern Recognition (CVPR '07),* Minneapolis, MN, June 2007.

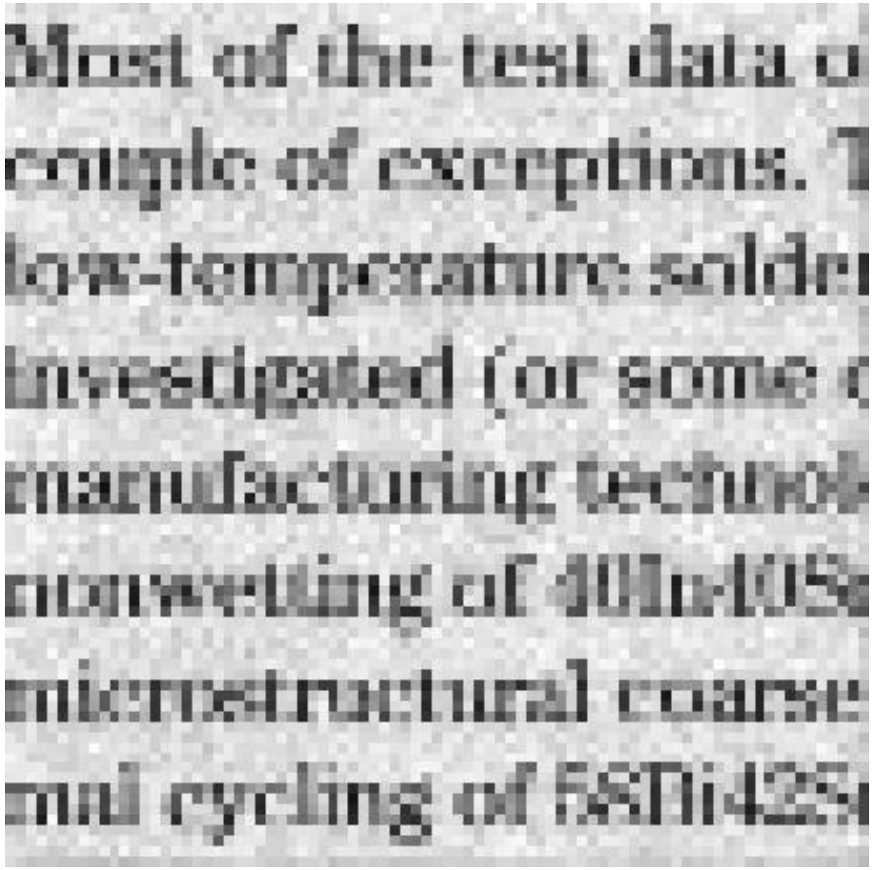
SYNTHESIZING DYNAMIC TEXTURES

SUPER-RESOLUTION

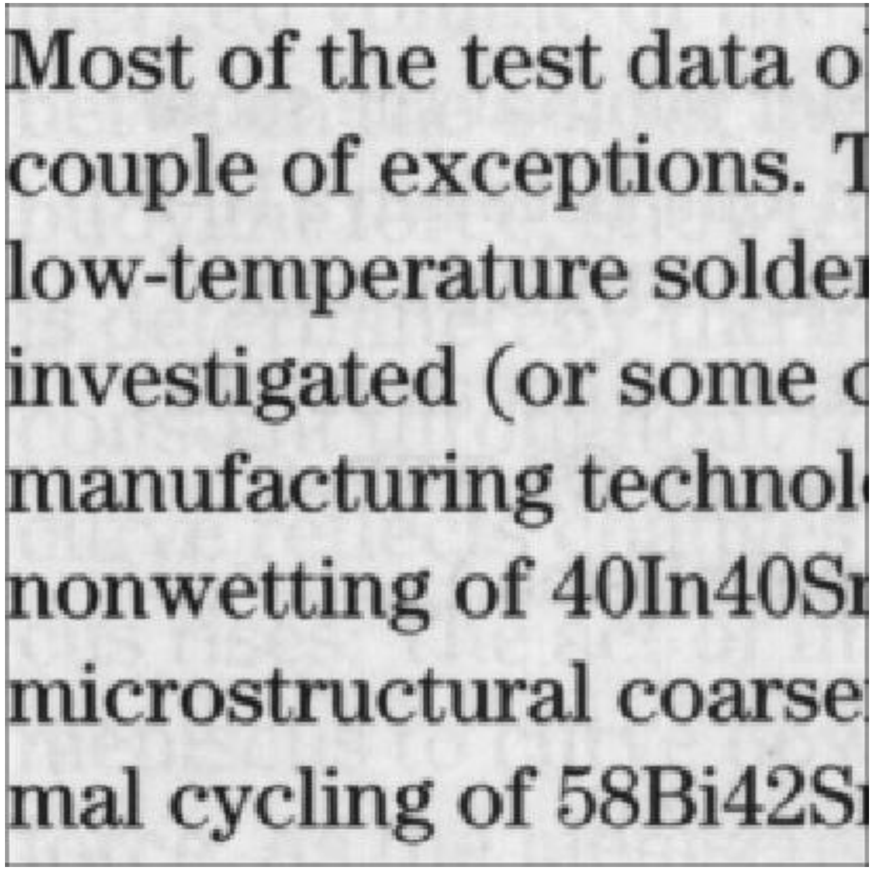
Example: A set of low quality images

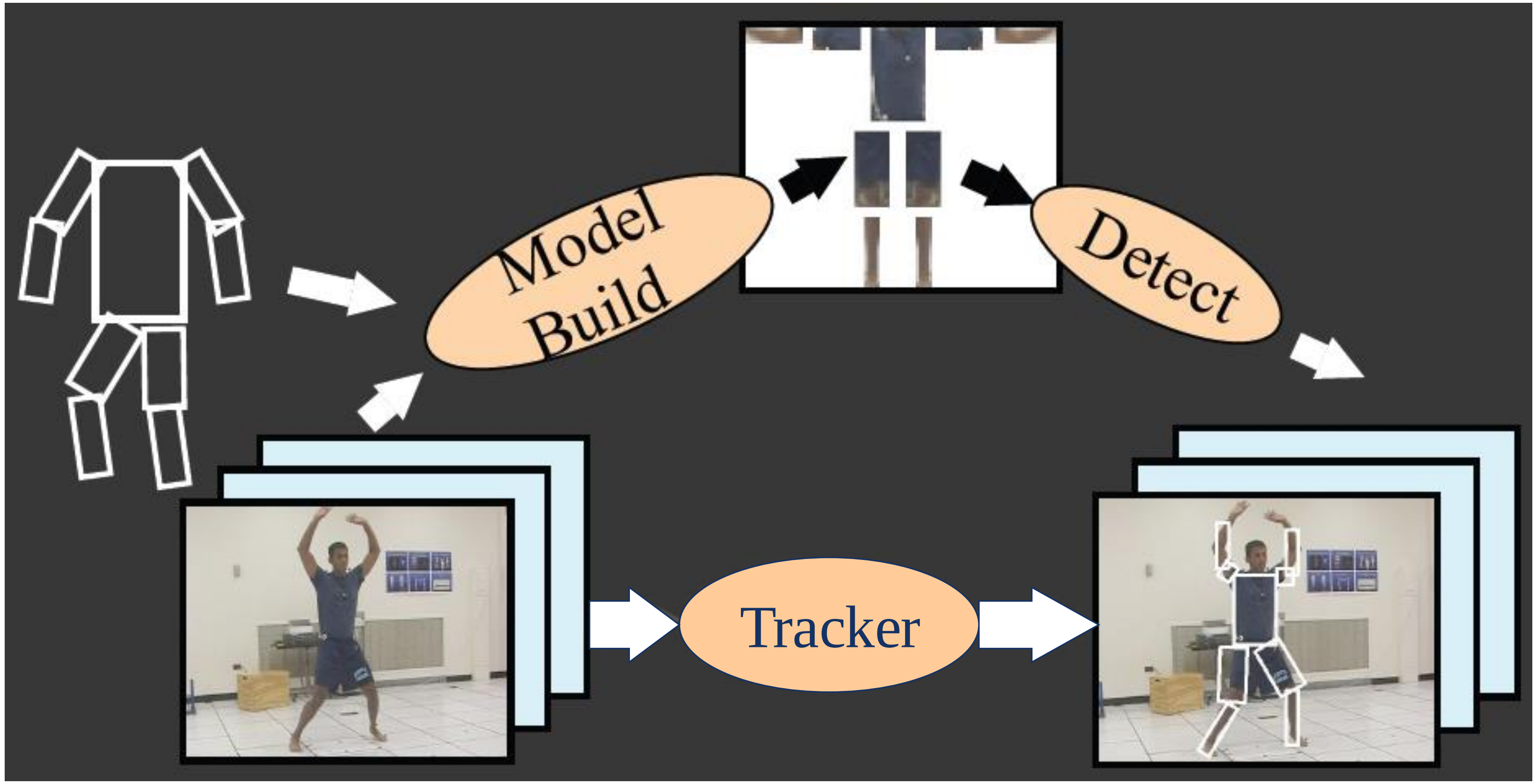


Each of these images looks like this:


Most of the test data o
couple of exceptions. T
low-temperature solder
investigated (or some c
manufacturing technol
nonwetting of 40In40S
microstructural coarse
and cycling of 58Bi42S

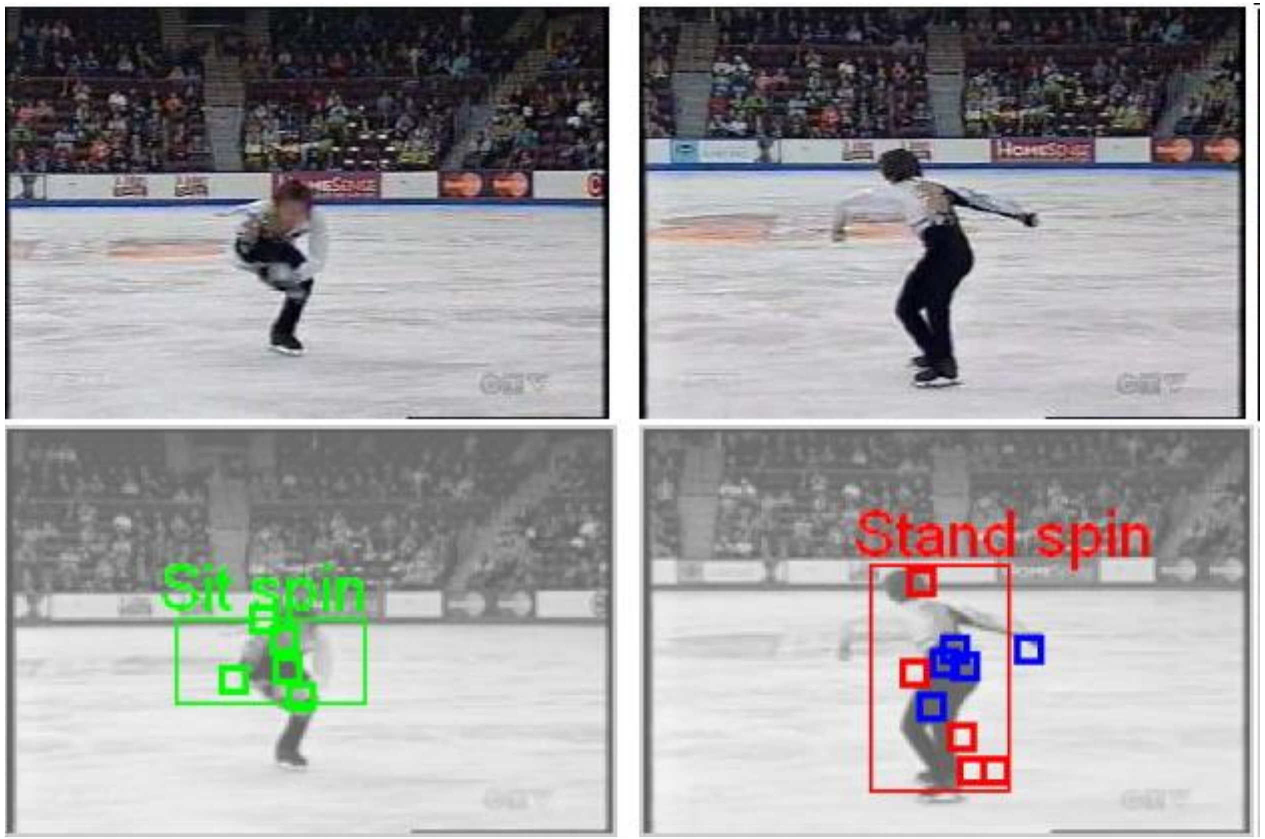
The recovery result:


Most of the test data o
couple of exceptions. T
low-temperature solde
investigated (or some c
manufacturing technol
nonwetting of 40In40S
microstructural coarse
mal cycling of 58Bi42S



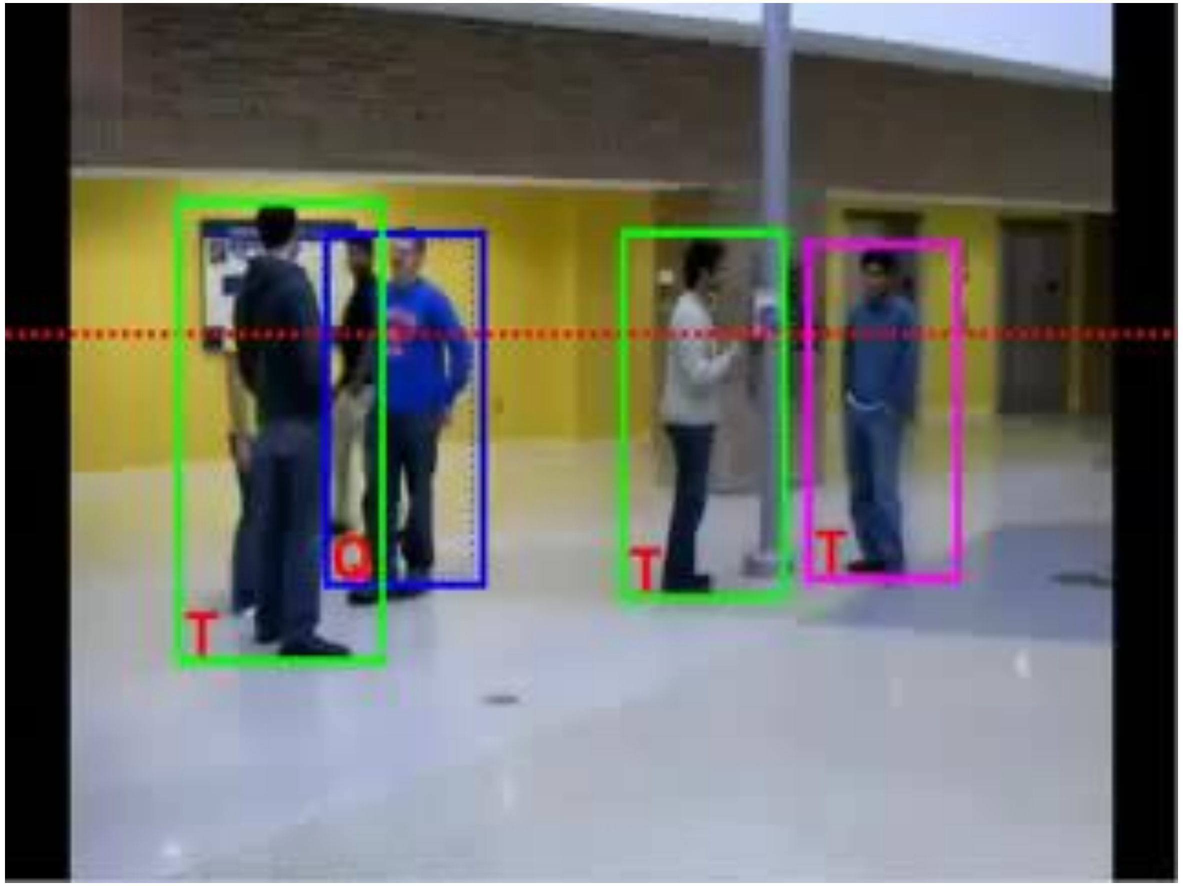
D. Ramanan, D. Forsyth, and A. Zisserman. Tracking People by Learning their Appearance. PAMI 2007.

RECOGNIZING EVENTS AND ACTIVITIES



Juan Carlos Niebles, Hongcheng Wang and Li Fei-Fei, **Unsupervised Learning of Human Action Categories Using Spatial-Temporal Words**, *(BMVC)*, Edinburgh, 2006.

Crossing – Talking – Queuing – Dancing – jogging



W. Choi & K. Shahid & S. Savarese WMC 2010

WHAT WE HAVE LEARNED TODAY?

JACOBS UNIVERSITY

1. Optical flow
2. Lucas-Kanade method
3. Horn-Schunk method
4. Pyramids for large motion
5. Common fate
6. Applications

Reading: [Szeliski] Chapters: 8.4, 8.5

[Fleet & Weiss, 2005]
http://www.cs.toronto.edu/pub/jepson/teaching/vision/2503/opticalFlow.pdf