Learning and Inference for Computer Vision

Master Mathématiques, Vision, et Apprentissage
12/01/2016, ENS Cachan

Iasonas Kokkinos and Nikos Paragios
CentraleSupelec

http://cvn.ecp.fr/personnel/iasonas/teaching.html

Organization - I

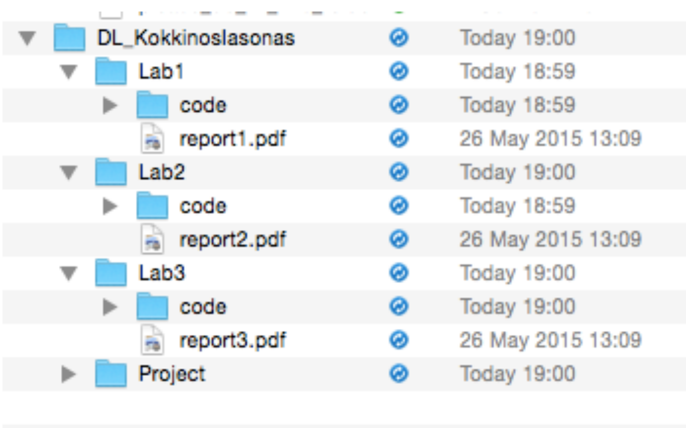
- Class webpage:
  - http://cvn.ecp.fr/personnel/iasonas/teaching.html

- 3 labs in Matlab (10 points)
  - Start with small preparatory exercises (synthetic data)
  - Most of the work: at home
- 1 Project (10 points)
  - Extension of labs to real data
- No `binome'
  - Code of conduct: if any copying is detected, everyone gets zero.
- Default: research project (20/20)
  - Study and implement a paper related to topics covered in class
  - Potential binome (but **harder**)

Organization - II

- Teaching assistants:
  - Maxim Berman, Siddhartha Chandra, Stefan Kinaeur, Riza Alp Guler
- Tonight: Everyone sends me this email
  - Subject: Subscribe MLCV15 [YourSchoolHere]
  - Text: name.familyname@institutional.email.fr
  - NO OTHER TEXT, NO PERSONAL EMAIL

- All deliverables: use Dropbox
  - Example
  - No more than 10Mb /user
  - Please do not remove items from my folder



Quick poll

- Raise your hand if..

  You have taken the 'deep learning' class

  You have taken the 'computer vision' class (2nd year, ECP only)

  You have taken the 'graphical model' class (Obozinski & Bach)

  You know what PCA (or, ACP) is

  You know what the EM algorithm is

  You know what the Lukas-Kanade algorithm is

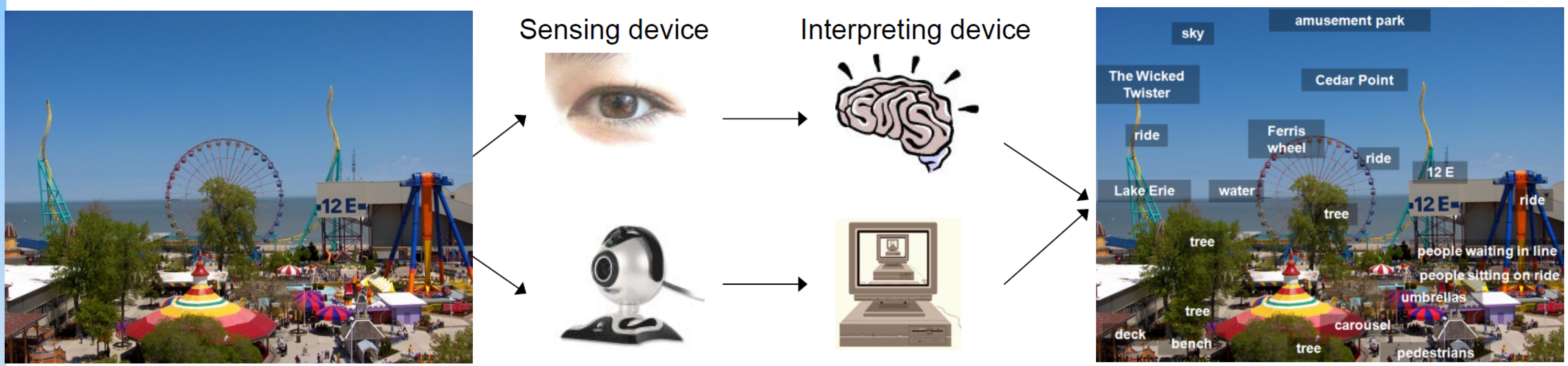
  You know what a Deformable Part Model is (Felzenszwalb et al)

  You know what an Active Shape/Appearance Model is

  You know what TRW-S, ADMM, LBP stand for

What is Computer Vision?

- Vision: Discovering what is where by looking at an image
- Computer Vision: Giving this ability to computers

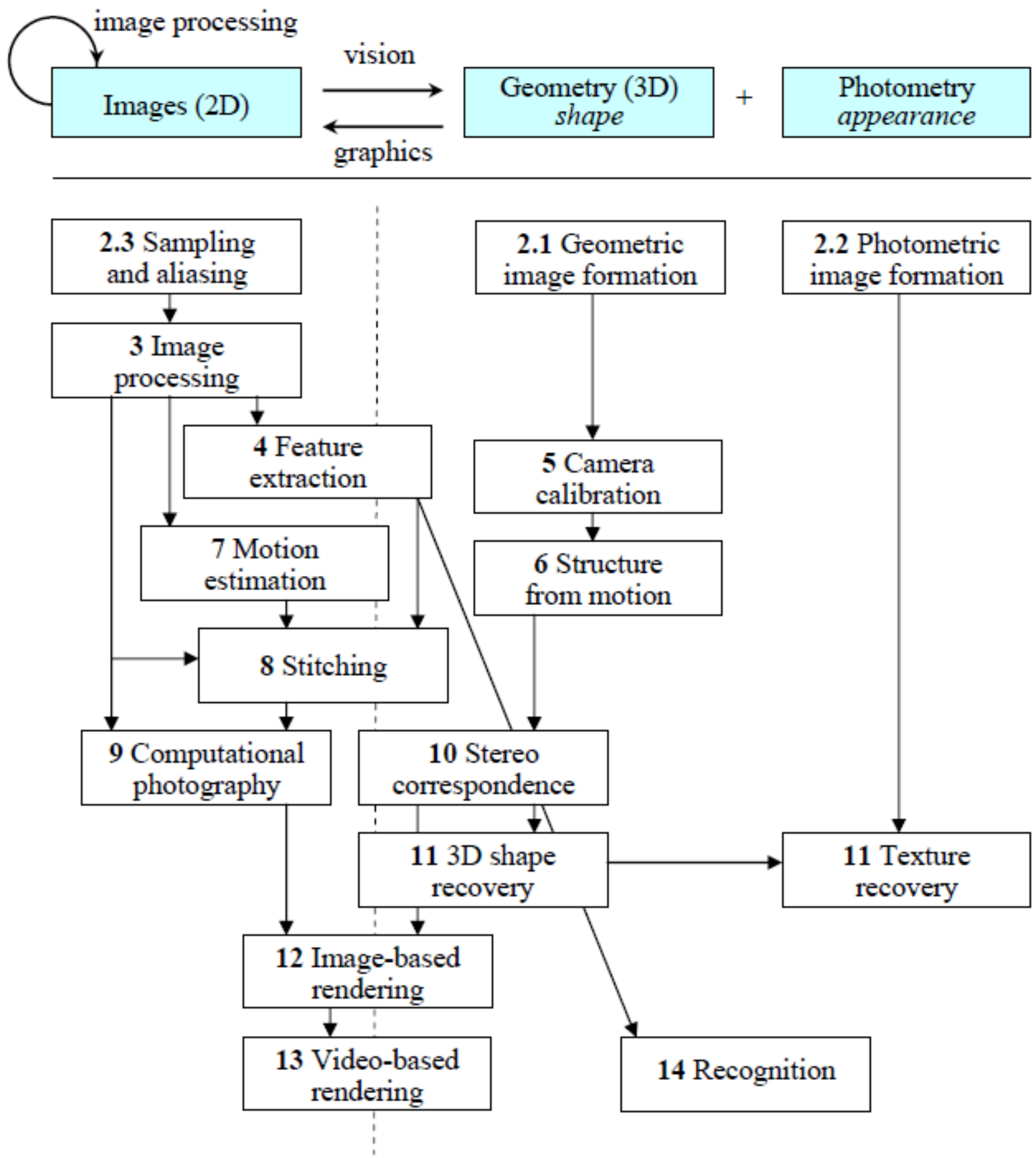


`An image is worth a thousand words'

- Numeric input
  - Sensor responses
- Symbolic output
  - Identities, locations and relations of objects

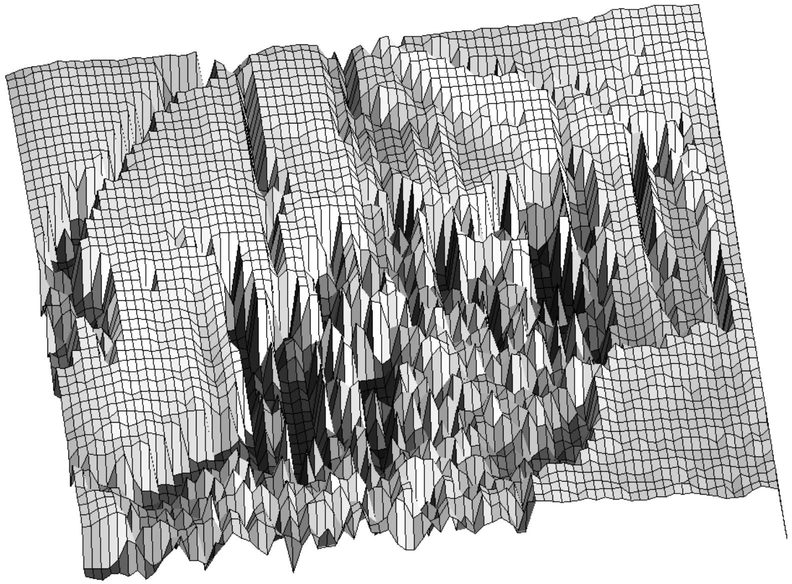
What is vision about?

R. Szelisky,
Introduction to
Computer Vision



Why can't computers see (yet)?

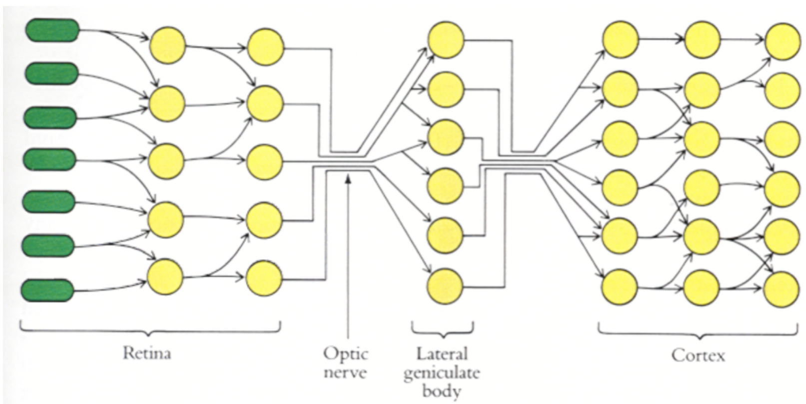
- Imagine describing `red' or `ugly' to a blind man
- Input to a computer: 2D/3D function

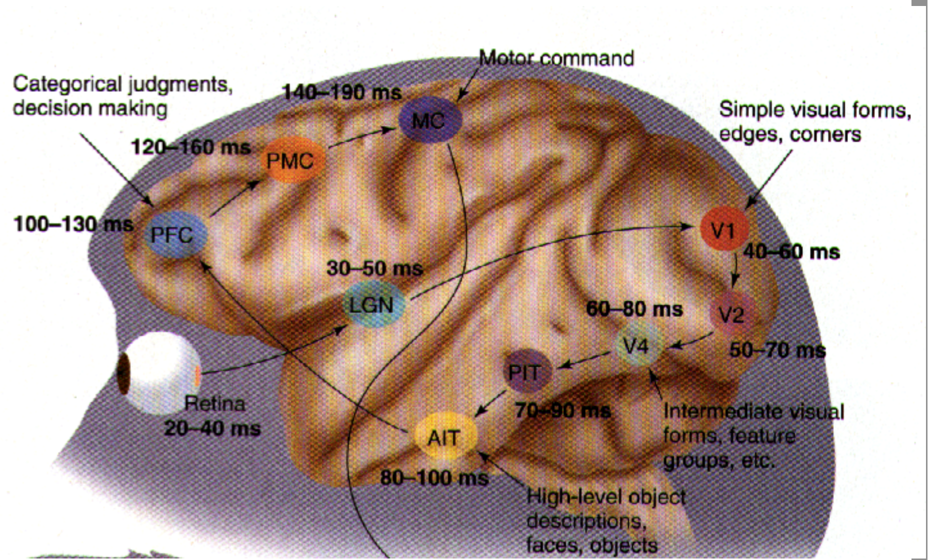


How do we solve vision?

We perform the vision task with amazing speed and accuracy
- But not effortessly:
  - Almost 50% of your brain is doing vision
    - Substantially more than what is involved in doing math!

Visual System





Challenges: What is an object?

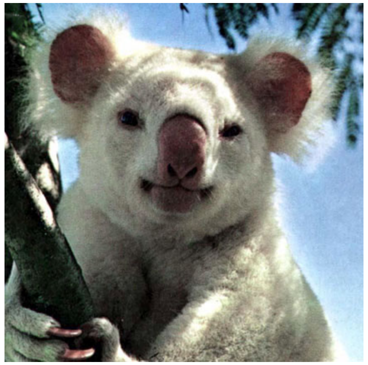
Challenges: Many Nuisance Parameters

Illumination

Object pose

Clutter

Occlusions



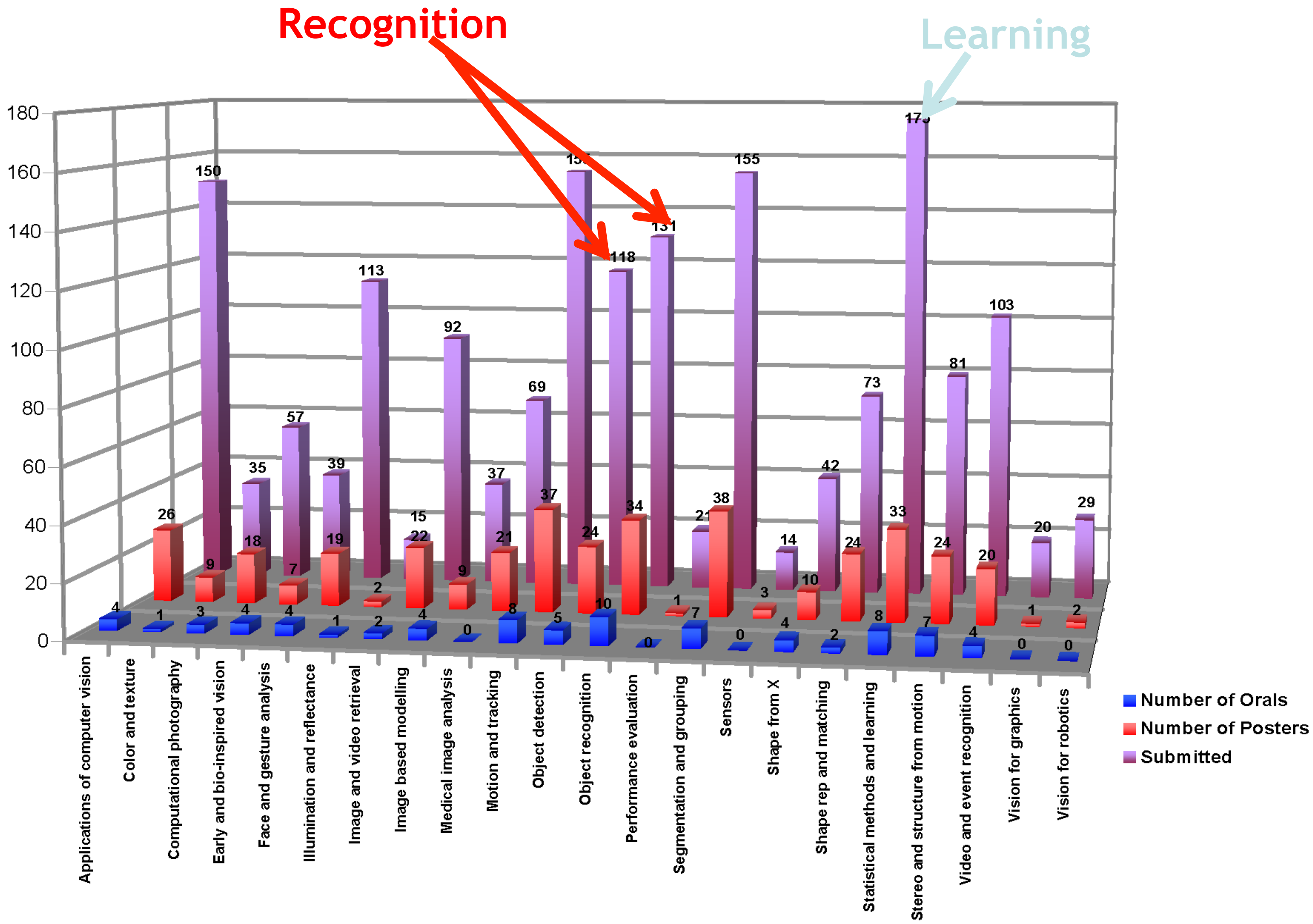
Intra-class appearance

Viewpoint

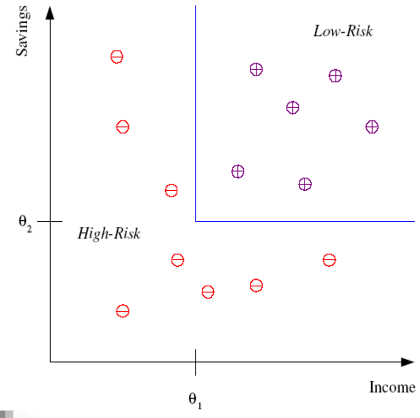
Challenges: Intra-Category Variation


~10,000 to 30,000

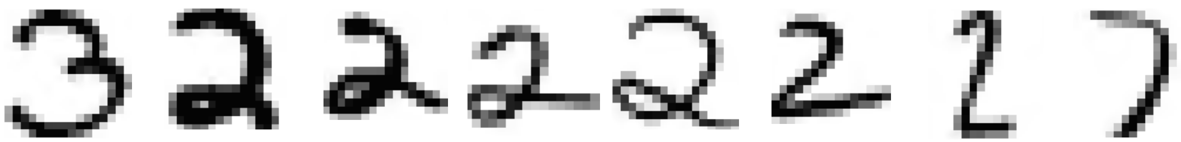
Submission Statistics from CVPR



Classification

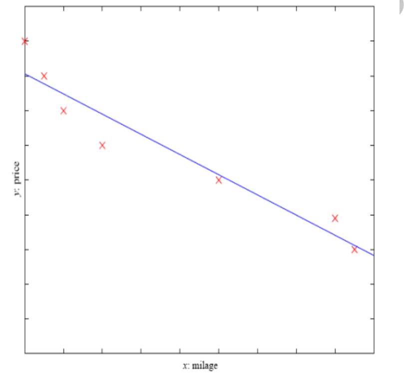


- Digit Recognition

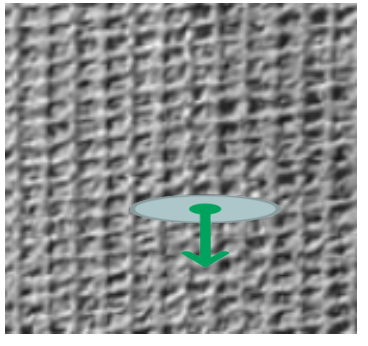


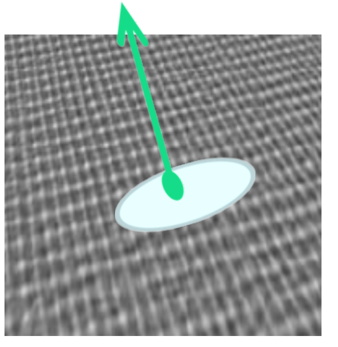
- Person Identification

Regression

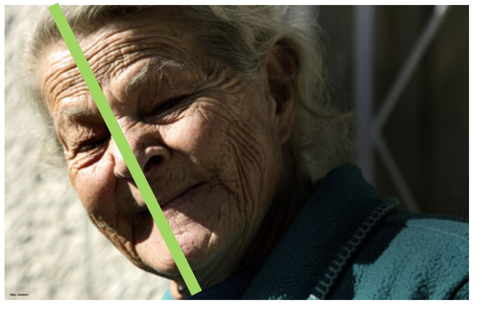


- Estimate surface normal

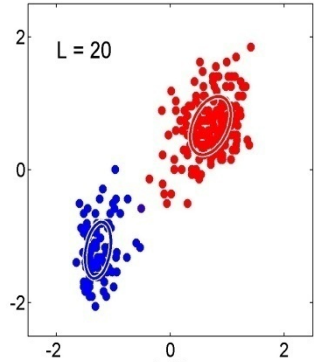




- Estimate face orientation/age

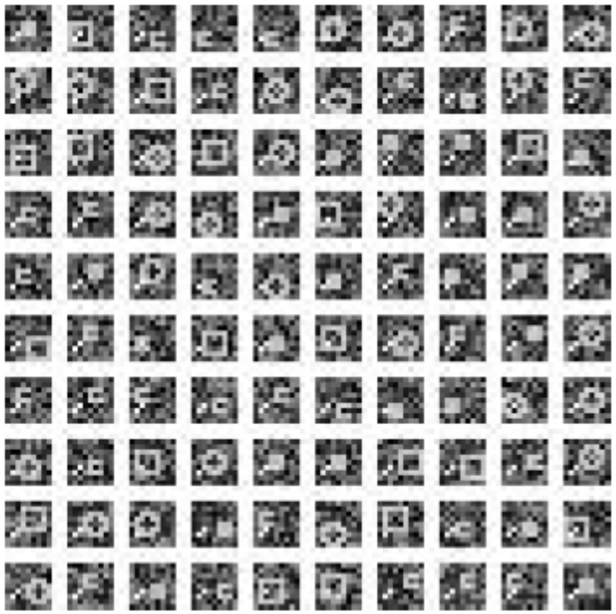


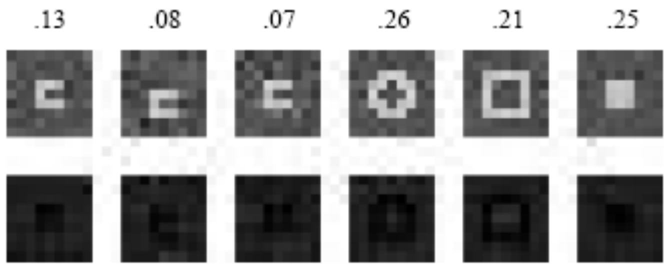
Clustering



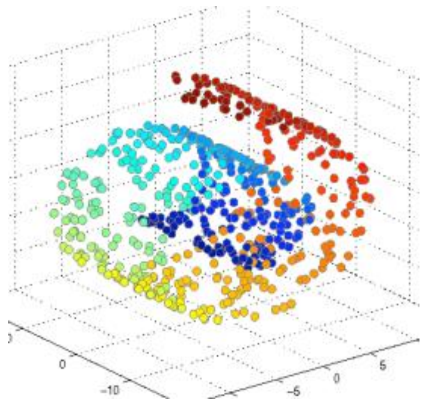
- Image Segmentation

- Multi-modal data modelling

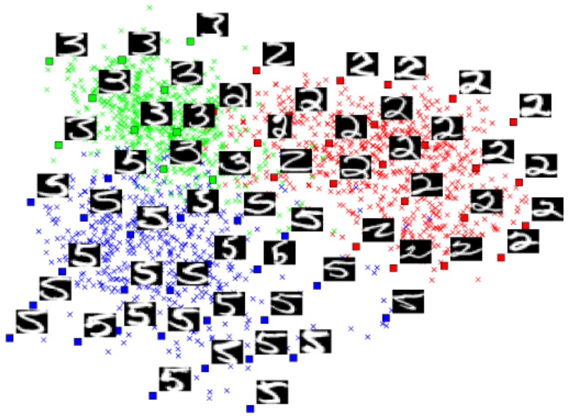




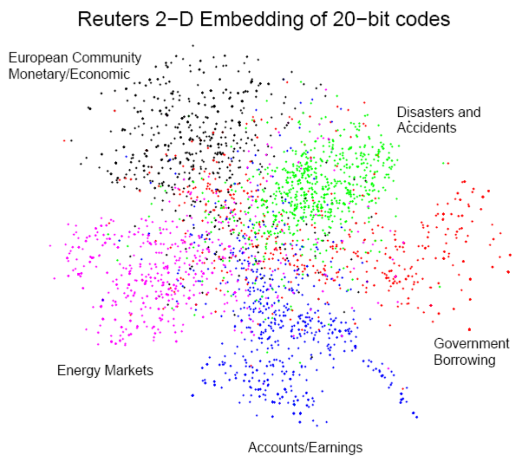
Dimensionality Reduction



- Visualization



- Compression
- Retrieval



Class outline

Lectures 1-2: Deformable Shape Models (ASMs/AAMs/..)

L&O: EM, Factory Analysis/PCA, Mixture Modelling, G. Descent

Vision: Lukas-Kanade, Optical Flow, Faces, 3D modelling

Lectures 3-5: Deformable Part Models

L&O: Graphical Models, SVMs, Branch & Bound, ADMM, TRW-S

Vision: Detection, Pose Estimation, Action Recognition, Segmentation,..

Lecture 6: Hierarchical Models

L&O: A*, Reinforcement Learning

Vision: Scene Parsing, Facades, Detection,..

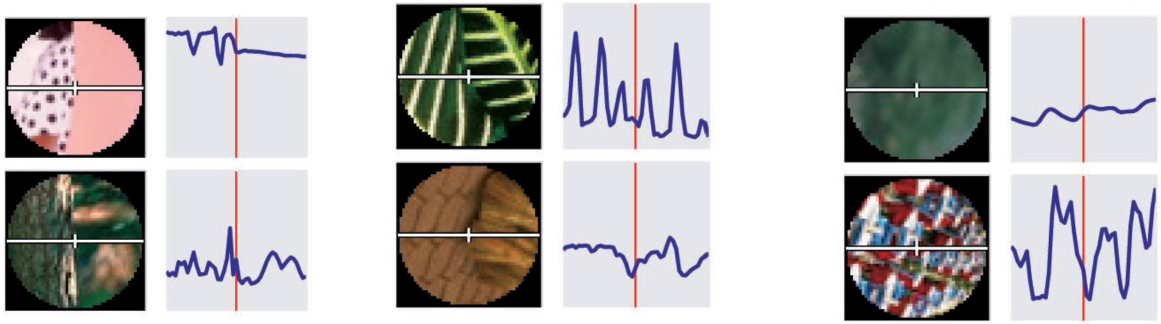
Lecture 7-8: Energy-based Models

L&O: Maximum Entropy, Structured Prediction, CNNs

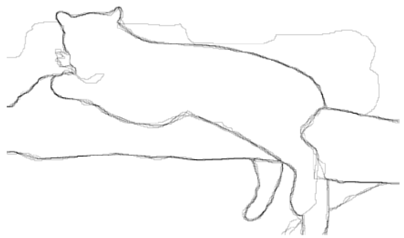
Vision: Optical flow, denoising, semantic segmentation, holistic vision

Learning & Vision Problem-I

- Boundary Detection



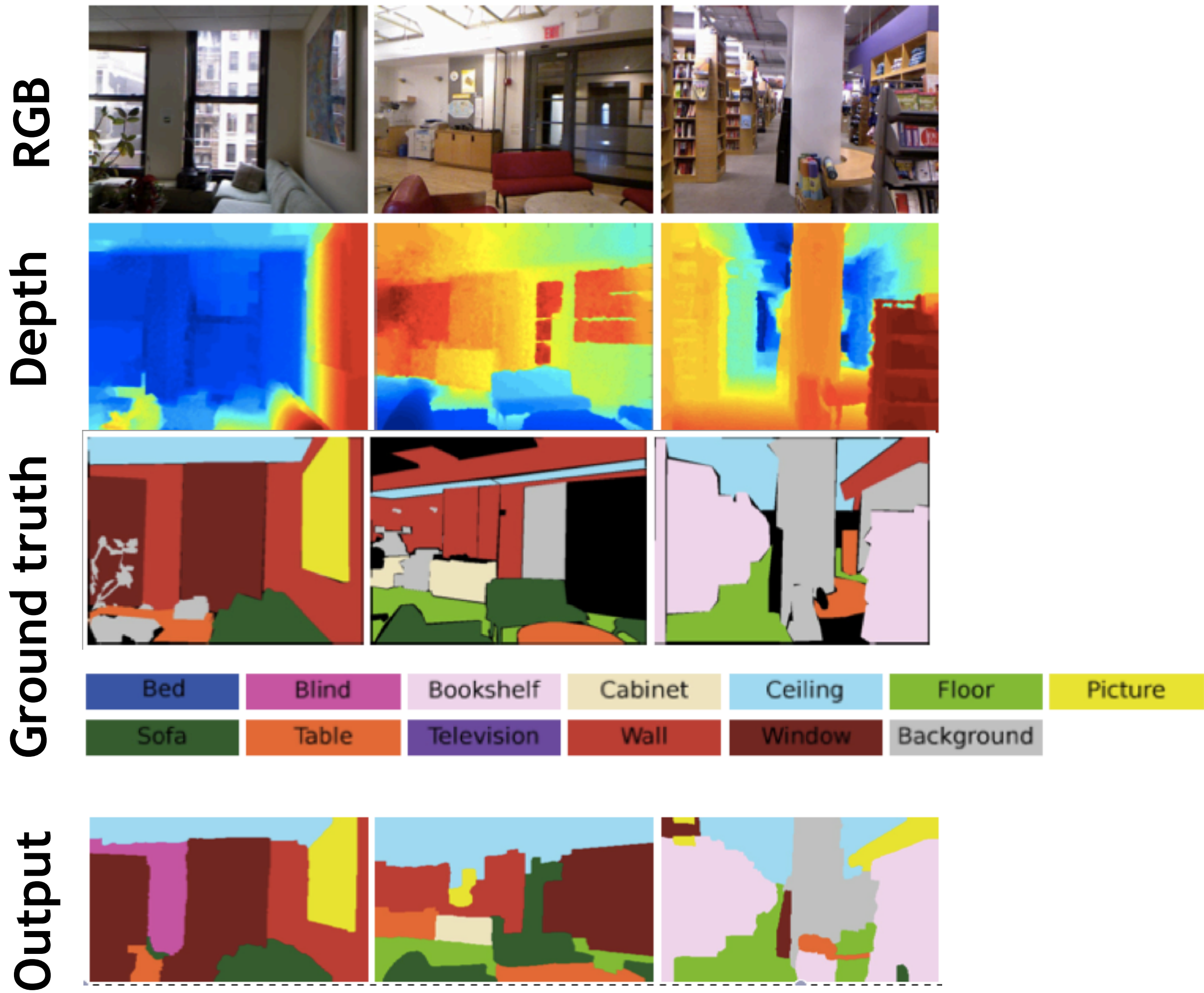
  - Boundary or non-boundary?
  - Input for segmentation, recognition, tracking….
- Abundance of vision approaches:
  - Filtering/Differential operators/Scale Space/Statistics/…
- Learning-based approach:



  - Use *labelled* training data.
    - 200 training/100 test images
  - Use all cues as inputs
  - Use decision trees/logisitic regression/boosting/… and *learn* how to combine the individual inputs.

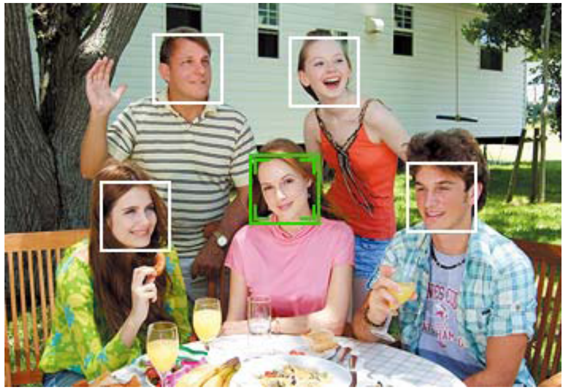
Learning & Vision Problem-II

- RGB-D Scene labelling

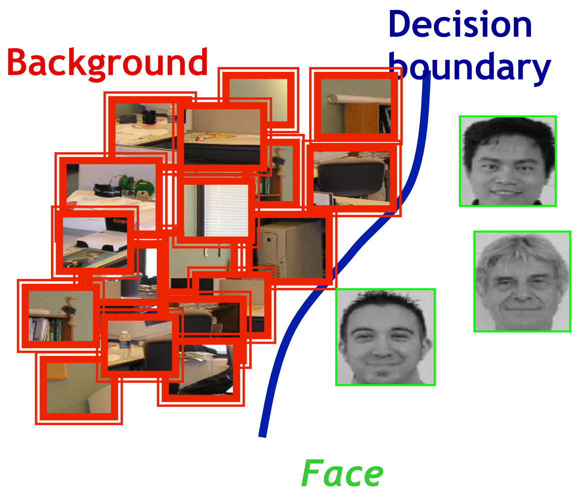


Learning & Vision Problem-III

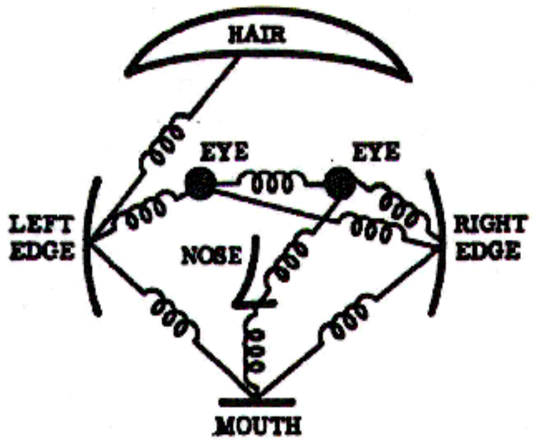
- Face Detection & Pose estimation



- Learn function that quantifies `faceness'

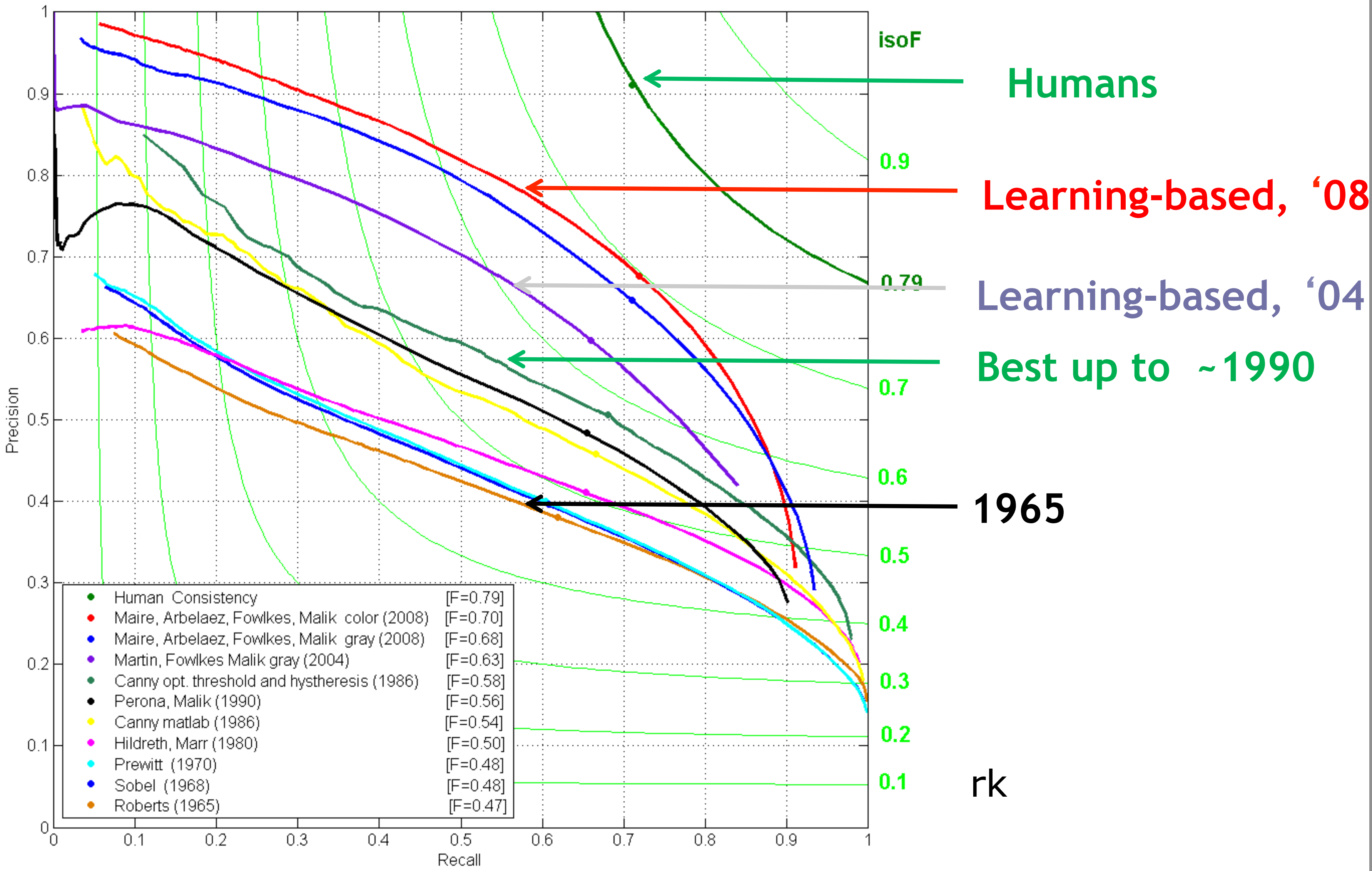


Binary classifier



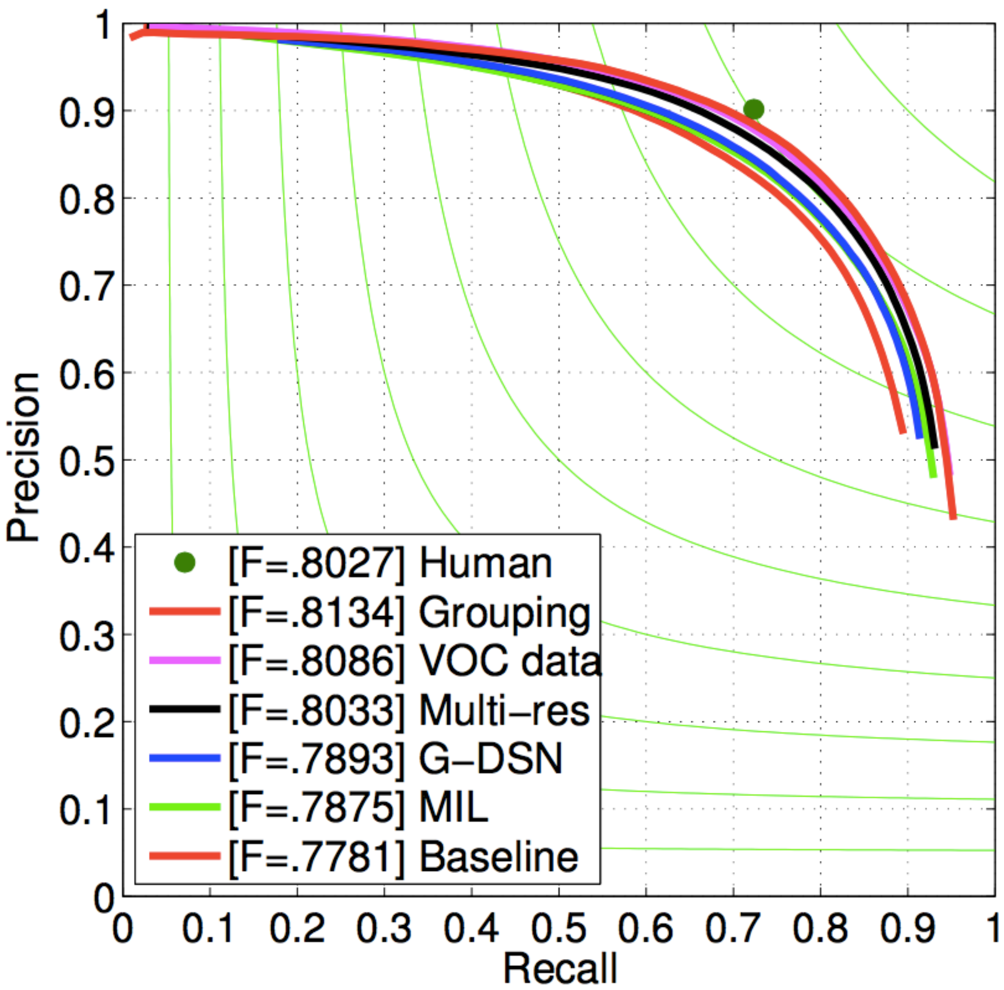
Energy of `physical system'

Progress during the previous 4 decades



rk

Progress during the previous 4 months



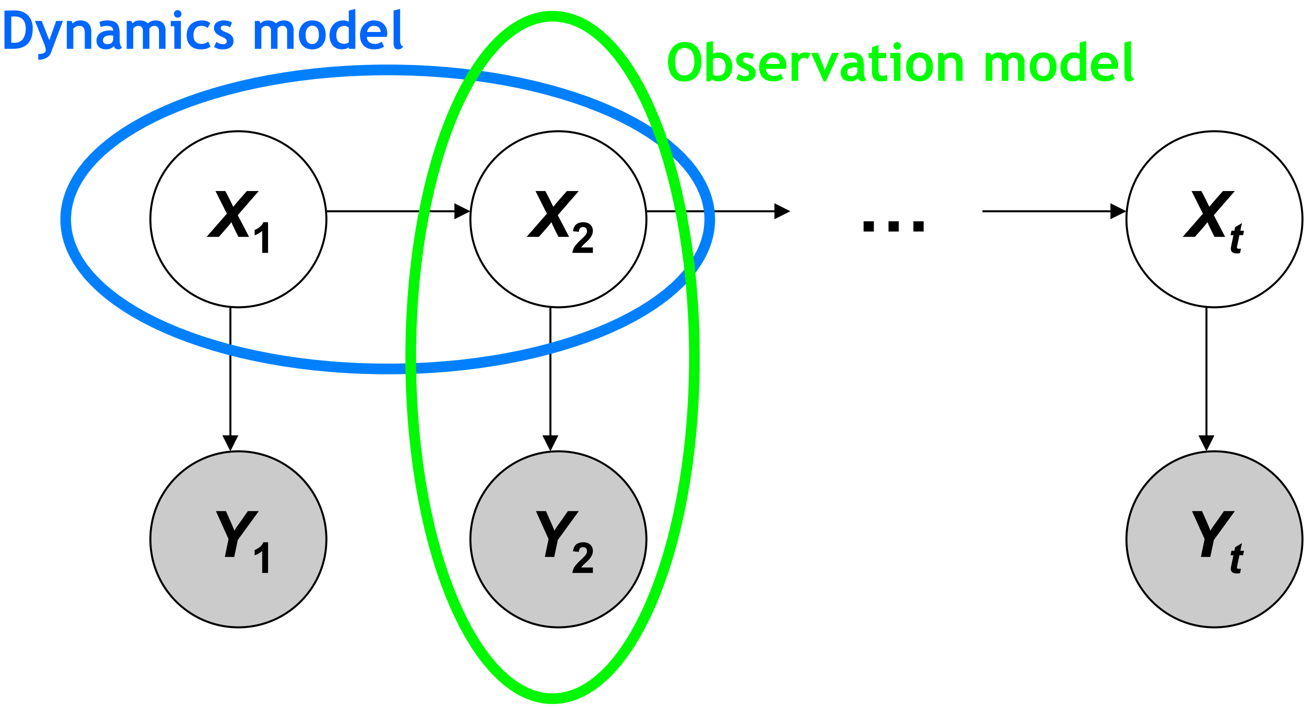
Method	ODS	OIS	AP
gPb-owt-ucm (Arbelaez et al., 2011)	0.726	0.757	0.696
SE-Var (Dollár & Zitnick, 2015)	0.746	0.767	0.803
DeepNets (Kivinen et al., 2014)	0.738	0.759	0.758
N4-Fields (Ganin & Lempitsky, 2014)	0.753	0.769	0.784
DeepEdge (Bertasius et al., 2015)	0.753	0.772	0.807
CSCNN (Hwang & Liu, 2015)	0.756	0.775	0.798
DeepContour (Shen et al., 2015)	0.756	0.773	0.797
HED-fusion (Xie & Tu, 2015)	0.790	0.808	0.811
HED-late merging (Xie & Tu, 2015)	0.788	0.808	0.840
Ours (DCNN + sPb)	0.8134	0.8308	0.866

earning', I. Kokkinos, arxiv 2015

Class objectives

- Treatment of a broad range of learning techniques
  - No single technique works for all problems
- Hands-on experience through computer vision applications
- Main asset: coupling of theory with applications

Coupling of theory with applications

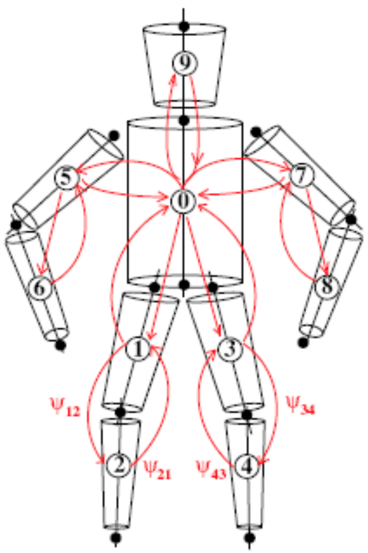


$$P(X_t|X_{t-1}, X_{t-2}, X_1) = P(X_t|X_{t-1})$$

$$P(Y_t|X_1, \ldots X_T, Y_1, \ldots Y_{t-1}, Y_{t+1}, \ldots Y_T) = P(Y_t|X_t)$$

Coupling of theory with applications

- **Inference on Markov Random Fields**
  - **Main problem: high-dimensional state space**



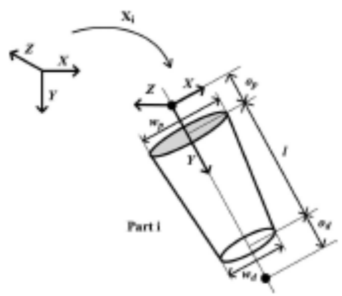


Figure 2: *Parameterization of part i.*

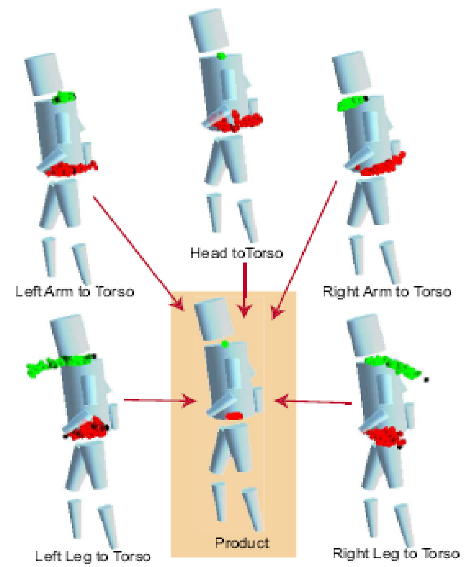
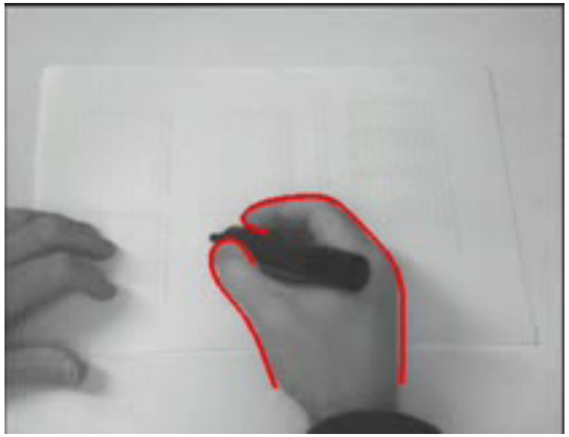


Figure 4: *Message Product.* The head, upper arms, and upper legs send messages to the torso. Samples from these messages are illustrated by showing the predicted torso location with green balls. The distribution over the orientation of the torso is illustrated by showing a red ball at the distal end of the torso for each sample. While any single message respresents uncertain information about the torso pose, the product of these messages tightly constrains the torso position and orientation.

$$m_{i,j}(c_j) = \sum_{c_i} \Phi_i(c_i)\Psi_{i,j}(c_i, c_j) \prod_{k\in\{\mathcal{N}(i)\backslash j\}} m_{k,i}(c_i)$$

Coupling of theory with applications

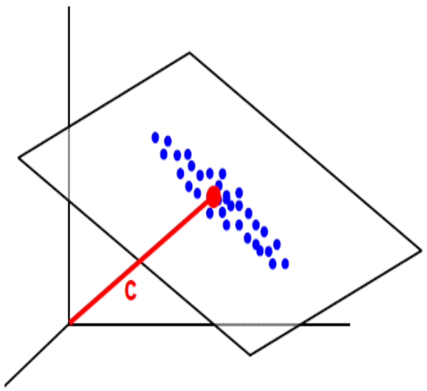


http://www.robots.ox.ac.uk/~misard/condensation.html

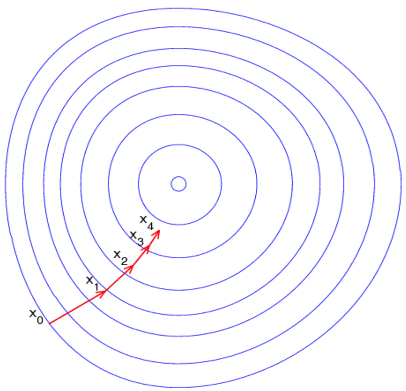
Coupling of theory with applications

PCA + Newton-Raphson

PCA



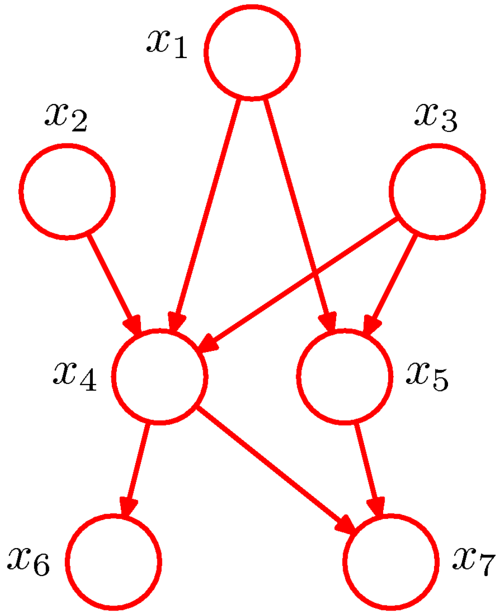
Newton-Raphson

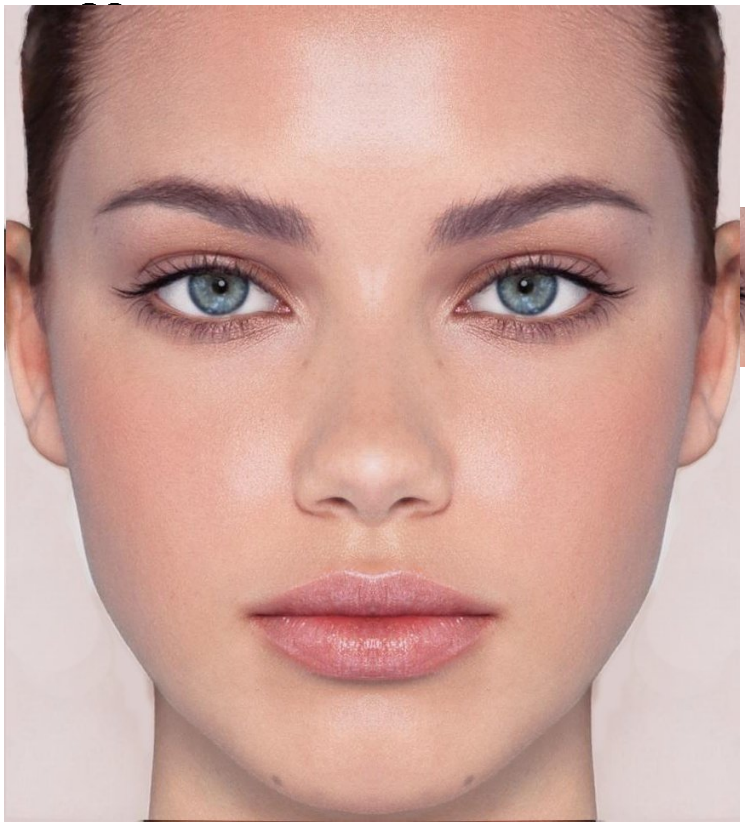


Coupling of theory with applications

Graphical Models + Object Detection

Bayesian Network

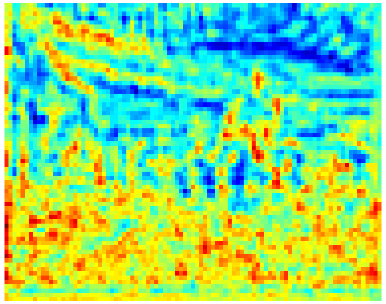




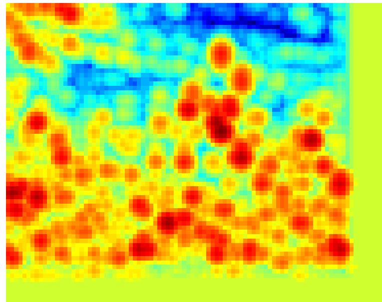
Coupling of theory with applications

Dynamic Programming + Detection

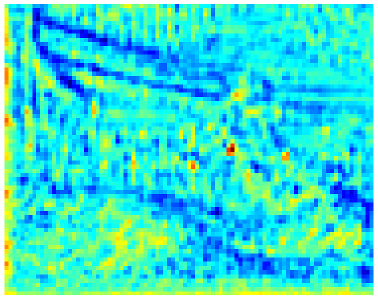
$$U_p(x') = \langle \mathbf{w}_p, \mathbf{H}(x') \rangle \qquad \max_{x'} [U_p(x') + B_p(x, x')]$$



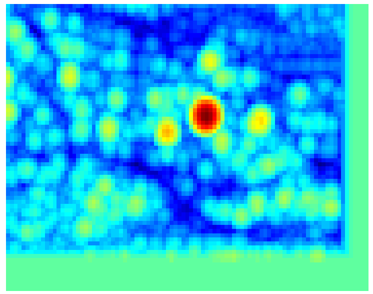
$p = 1$

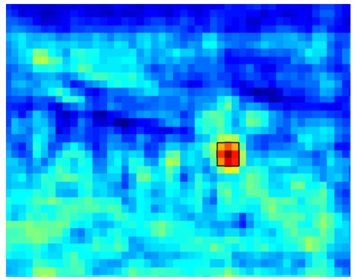


$\vdots$



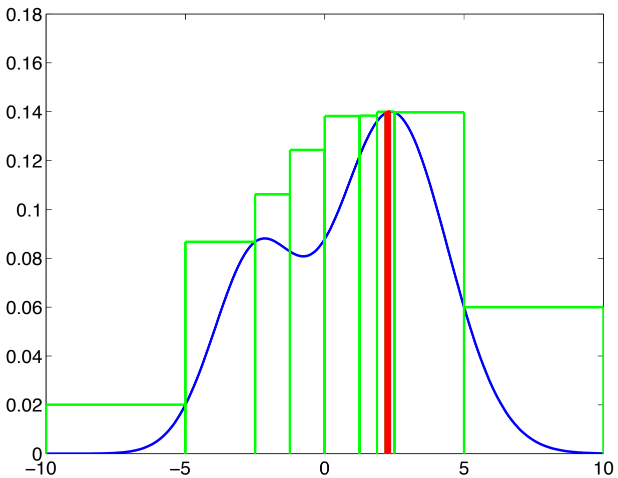
$p = P$





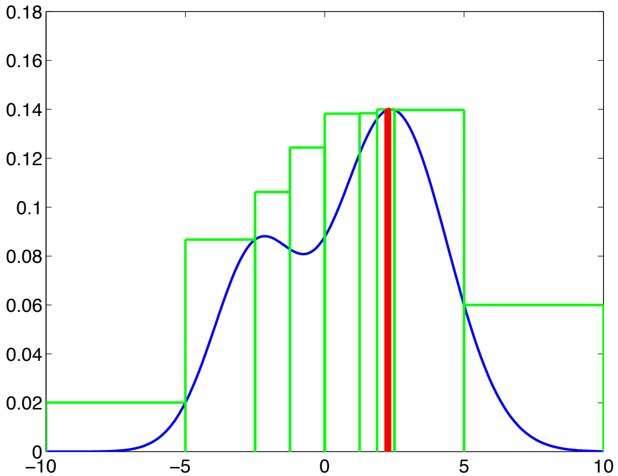
Coupling of theory with applications

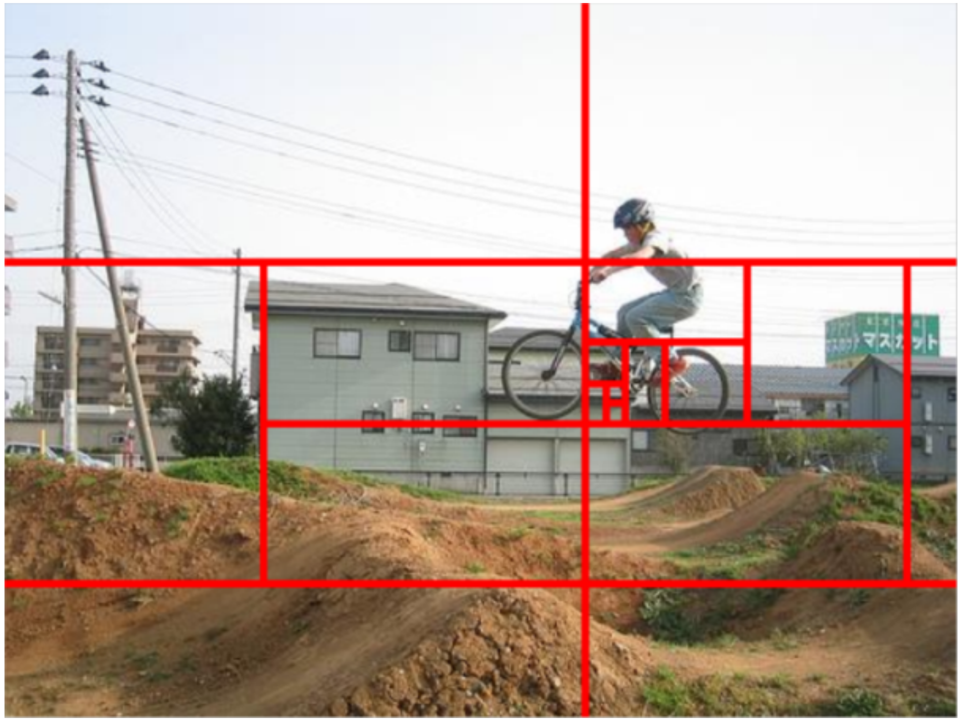
Branch-and-Bound + Detection



Coupling of theory with applications

Branch-and-Bound + Detection

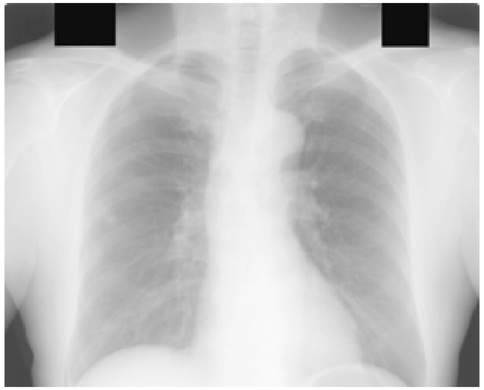




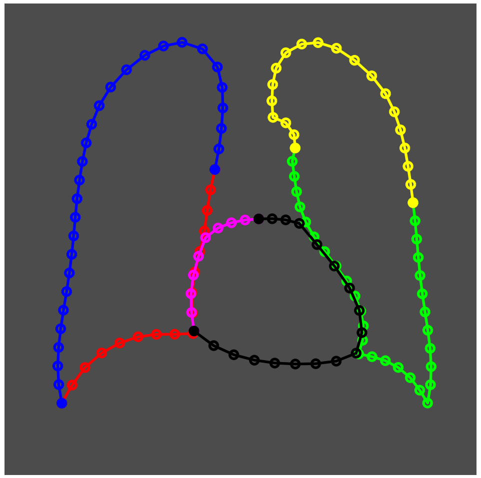
Coupling of theory with applications

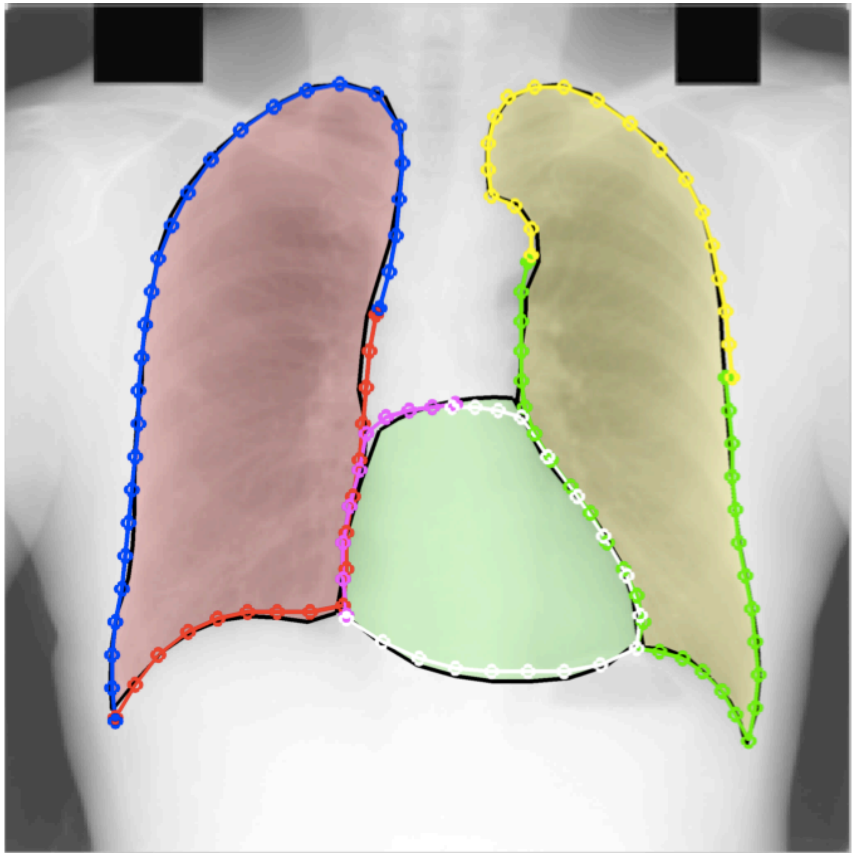
ADMM + MRFs for Shape Segmentation

Input image



The model

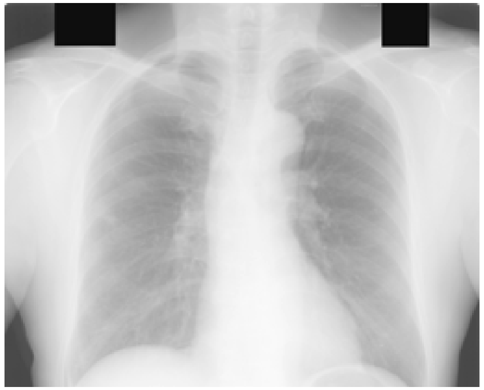




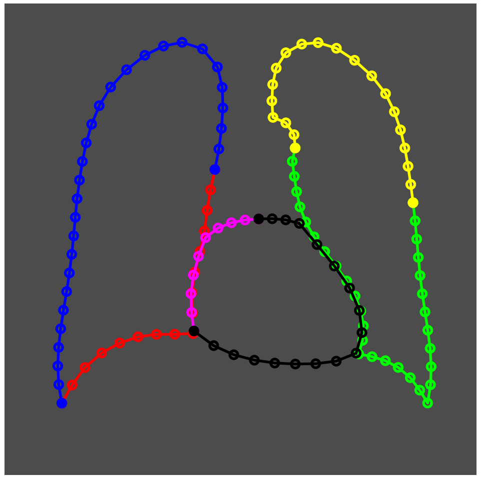
Coupling of theory with applications

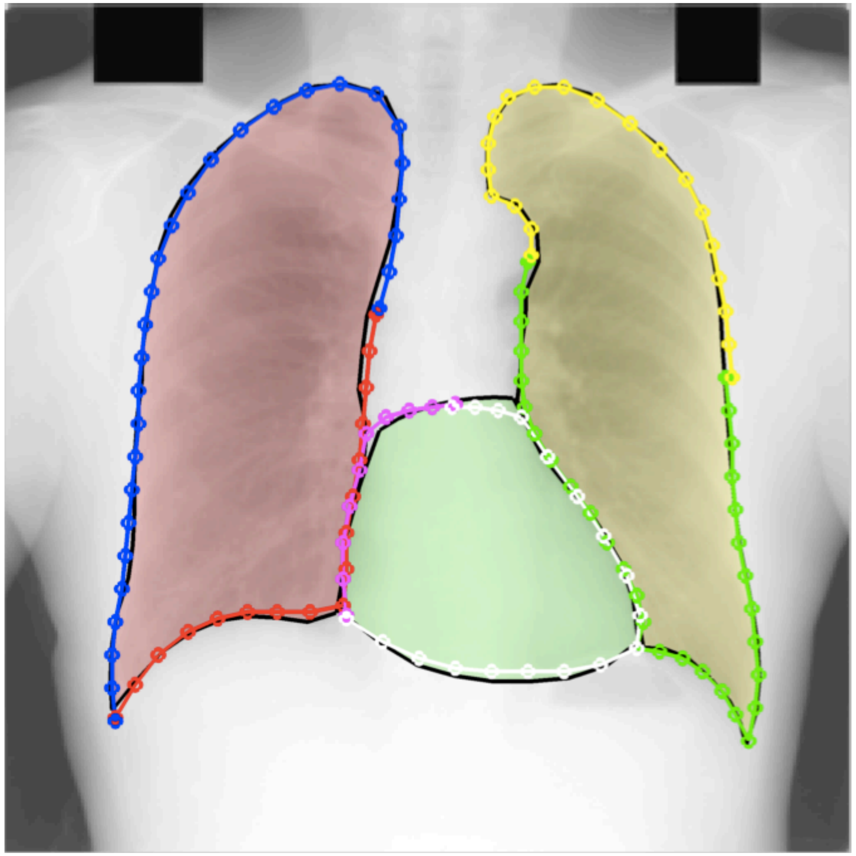
ADMM + MRFs for Shape Segmentation

Input image



The model





Bayes' theorem

Sum Rule	$p(X) = \sum_{Y} p(X, Y)$
Product Rule	$p(X, Y) = p(Y\|X)p(X)$

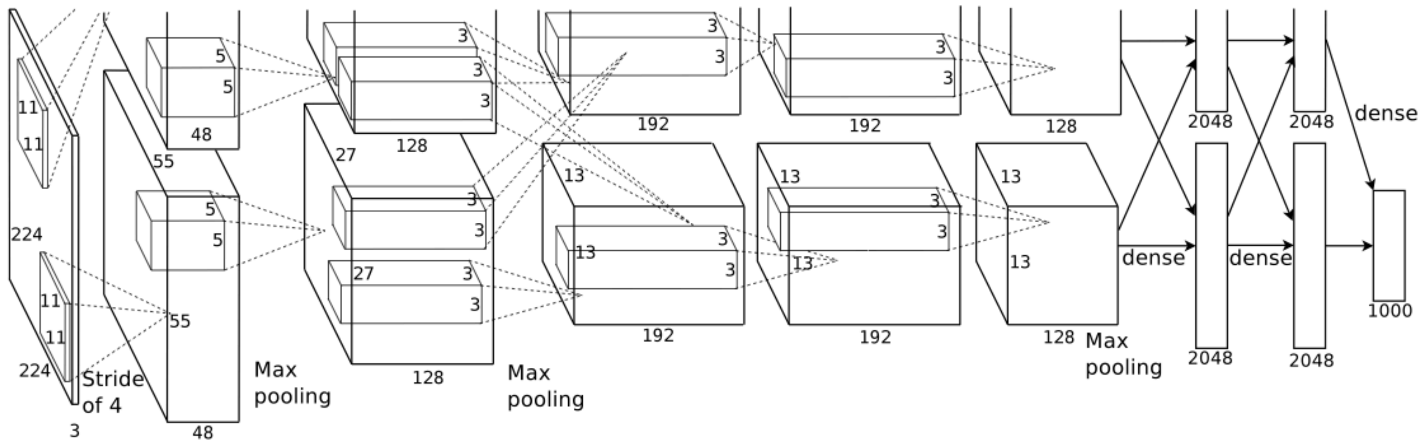
$$\begin{aligned} P(Y|X) &= \frac{P(X|Y)P(Y)}{P(X)} \\ &= \frac{P(X|Y)P(Y)}{\sum_{Y'} P(X, Y')} = \frac{P(X|Y)P(Y)}{\sum_{Y'} P(X|Y')P(Y')} \end{aligned}$$

- P(X|Y): **likelihood** of observations X, given class Y.
- P(Y): **Prior** probability of class Y
- P(Y|X): **Posterior** probability of class Y, given observations Y.

Why is this identity important?

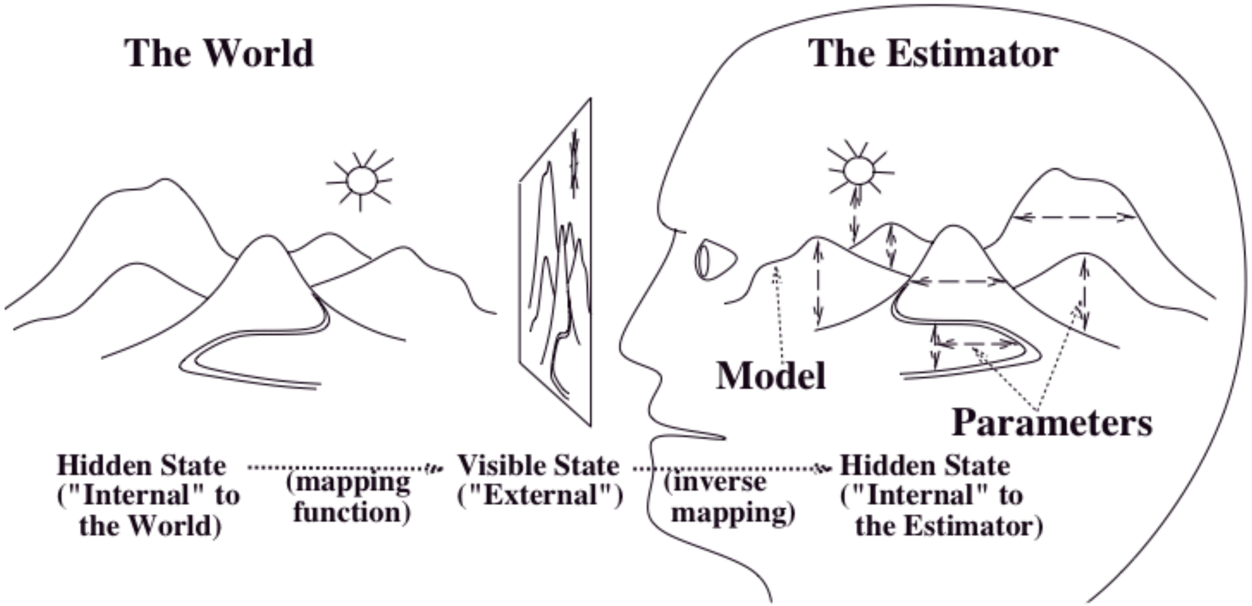
Two Main Approaches

- Discriminative



Two Main Approaches

- Generative

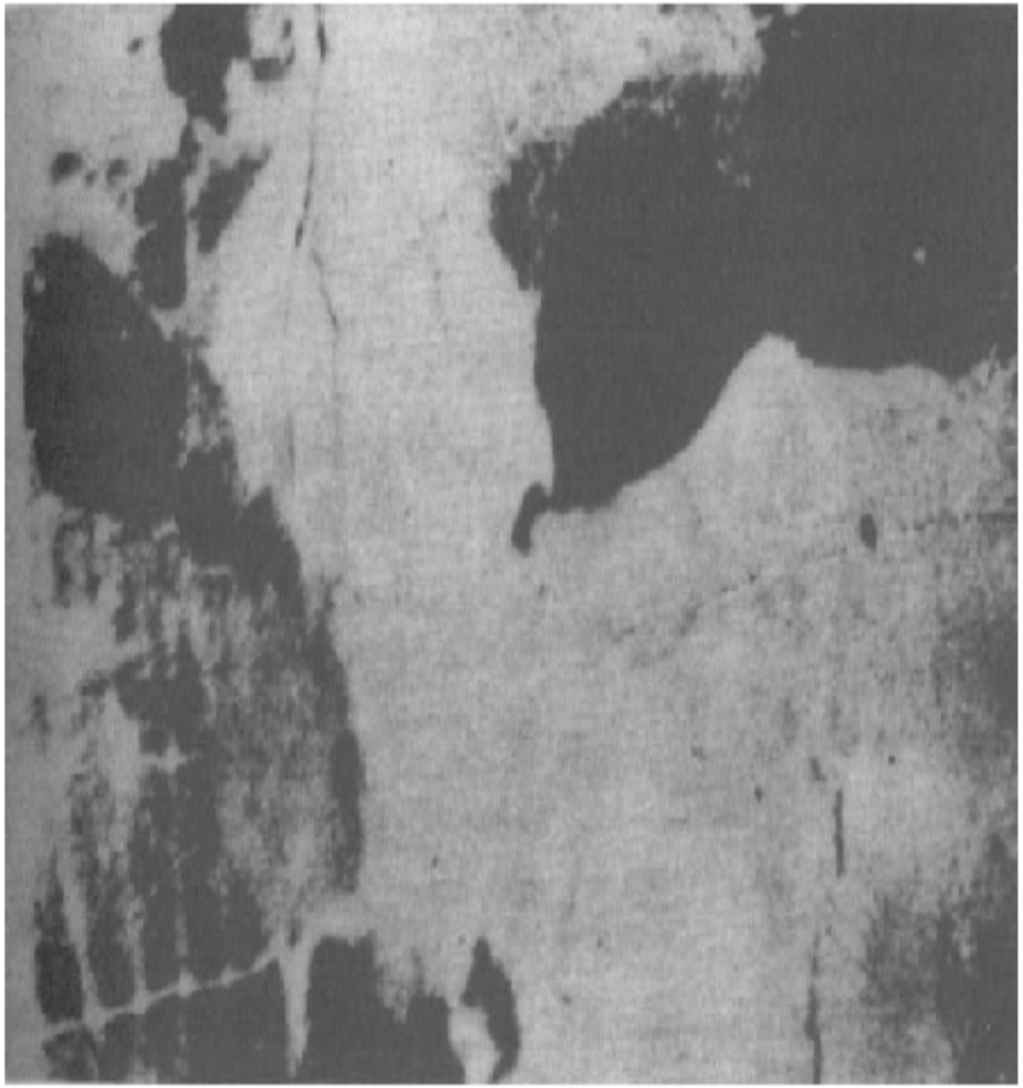


Likelihood Prior

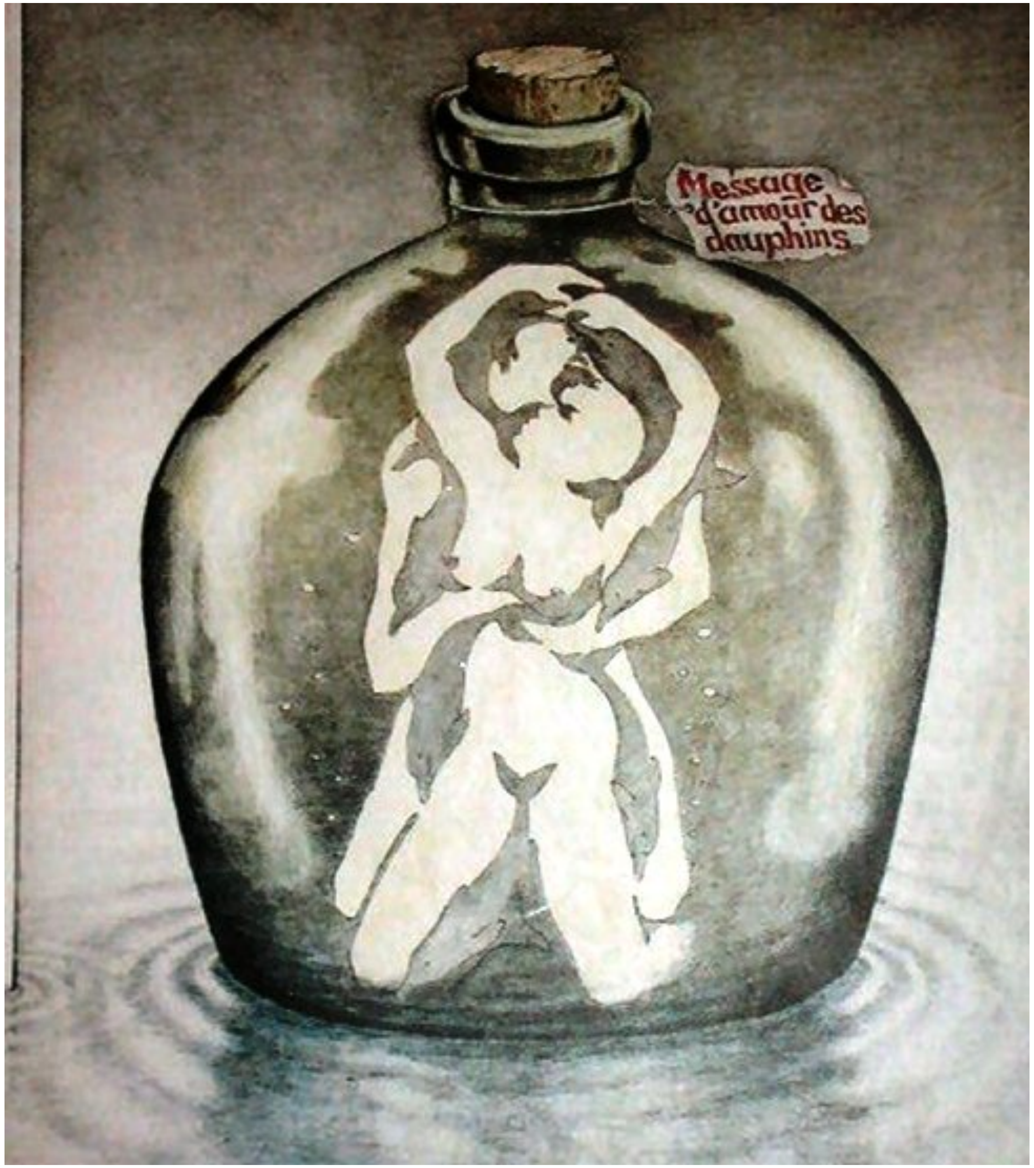
Posterior

$$P(Y|X) = \frac{P(X|Y)P(Y)}{P(X)}$$

Vision & priors



Vision & priors



Proof by eminent authority

Socrates: *..which is more correct, to say that we see with the eyes, or through the eyes?*
Theaeteus: *I should say "through," Socrates, rather than "with".*
Socrates: *Yes son, because it would be woeful if all these senses were sitting in us like Trojan horses instead of converging to some idea, soul, or whatever, by which we use such organs to perceive what is perceivable.*
Plato, *Theaetetus*, 369 BC

the perception is neither a seeing, nor a touching, nor an imagining ... rather it is an inspection.. of the mind.
Descartes *Meditations*, 1641

..it is quite possible that our empirical knowledge is a compound of that which we receive through impressions, and that which the faculty of cognition supplies from itself (sensuous impressions giving merely the occasion)
Kant, *A critique of pure reason*, 1787

to read nature is to see it, as if through a veil, in terms of an interpretation.
Cézanne, 1839-1906

…the value of a simple stimulus … for conveying information depends not only on the information conveyed by the stimulus itself but on the whole nervous constitution of the receiver of the stimulus as well.
Wiener, *Cybernetics*, 1948

... The signal never makes it to our consciousness, but gets overlaid with a clearly and precisely patterned version whose computation demands extensive use of memories, expectations, and logic.
Mumford, *Pattern Theory – a unifying perspective*, 1995

Pattern Theory

The four transformations that I propose as the basic types occurring in natural perceptual signals are:

Domain warping

What makes pattern theory hard is not that any of the above transformations are that hard to detect and decode in isolation, but rather that all of them tend to coexist, and then the decoding becomes hard.

Interruptions

Multi-scale superposition

Noise and blur

D. Mumford, *Pattern Theory – a unifying perspective*, 1995

Pattern Theory

The four transformations that I propose as the basic types occurring in natural perceptual signals are:

Domain warping

What makes pattern theory hard is not that any of the above transformations are that hard to detect and decode in isolation, but rather that all of them tend to coexist, and then the decoding becomes hard.

Interruptions

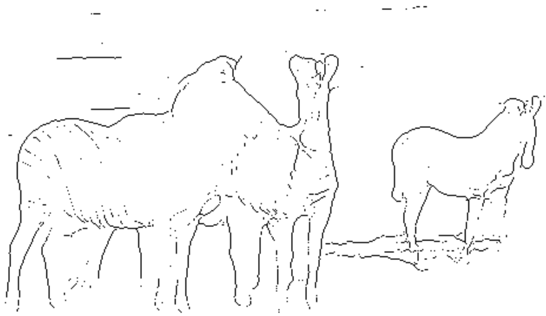
Multi-scale superposition

Noise and blur

D. Mumford, *Pattern Theory – a unifying perspective*, 1995

What is shape for computer vision?

Boundaries, contours,..



Active models

Point clouds

Surfaces

Deformable part models

Take home message: many things

Breakdown of variability

Application:

Low-level ('generic', image-)

High-level ('object-specific', object-)

Representation:

O-D: Points

1-D: Contours

2-D: Deformations

3-D: Surface

Many representation-specific bifurcations

Core problem: correspondence

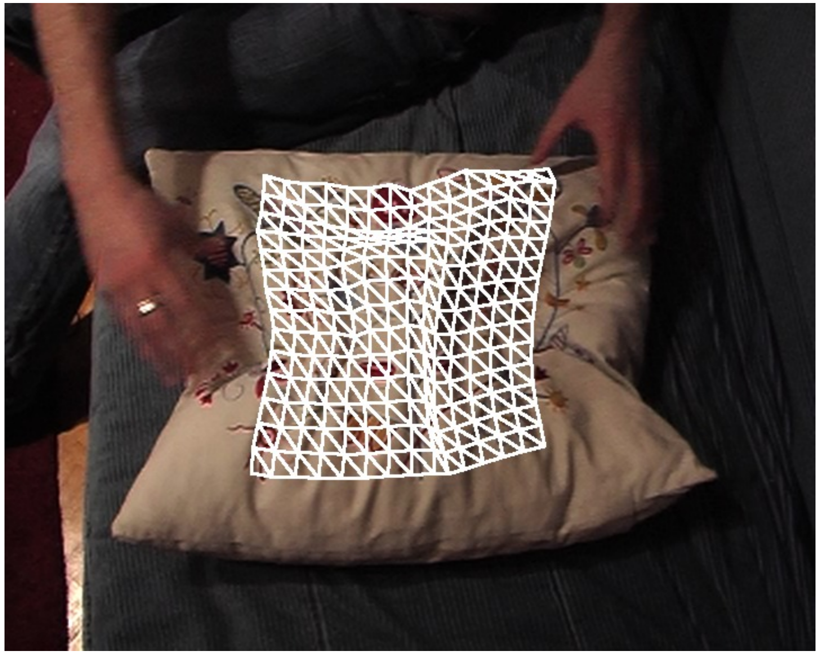
'what is here comes from here'

Simple case: global rigid transformation (4-9 DoF)

Core problem: correspondence

'what is here comes from here'

Complications: non-rigid transformation



Core problem: correspondence

‘what is here comes from here’

Complications: occlusions

Core problem: correspondence

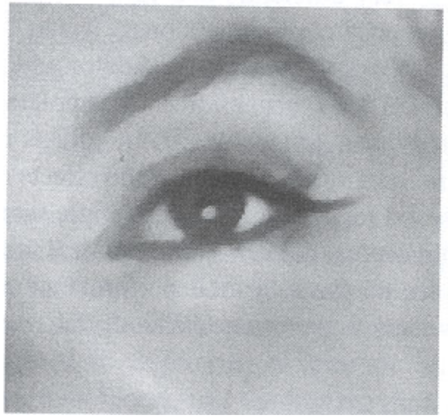
‘what is here comes from here’

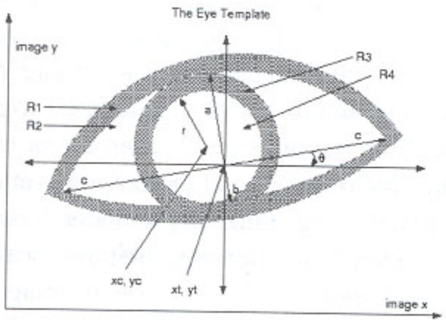
Complications: background changes

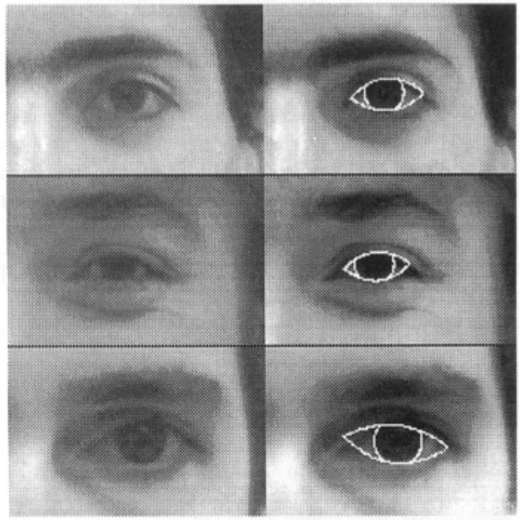
Core problem: correspondence

'what is here comes from there'

Complications: intra-category variation







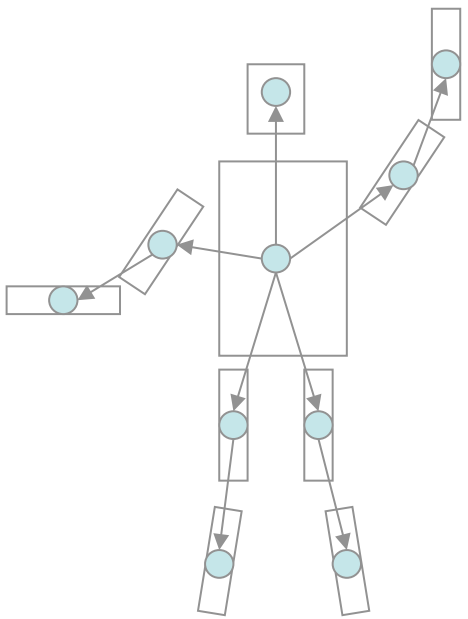
'One shape fits all'

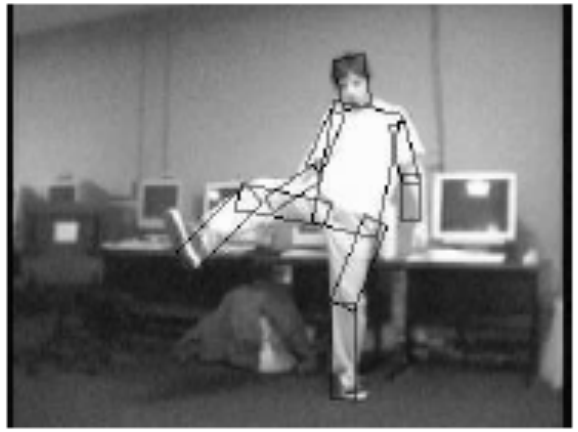
A.L. Yuille, D.S. Cohen and P.W. Hallinan. Feature extraction from faces using deformable templates. CVPR 1989.

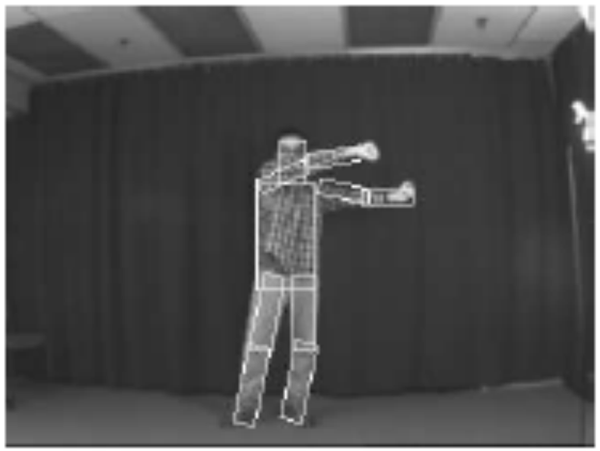
Core problem: correspondence

‘what is here comes from there’

Complications: computation







Is there a shape? Where?

Core problem: correspondence

‘what is here comes from there’

Complications: 3D, pose variability

T. J. Cashman, A. W. Fitzgibbon: What Shape Are Dolphins? Building 3D Morphable Models from 2D Images, 2013
http://research.microsoft.com/en-us/um/people/awf/dolphins/

Lecture outline

Introduction

Motion estimation

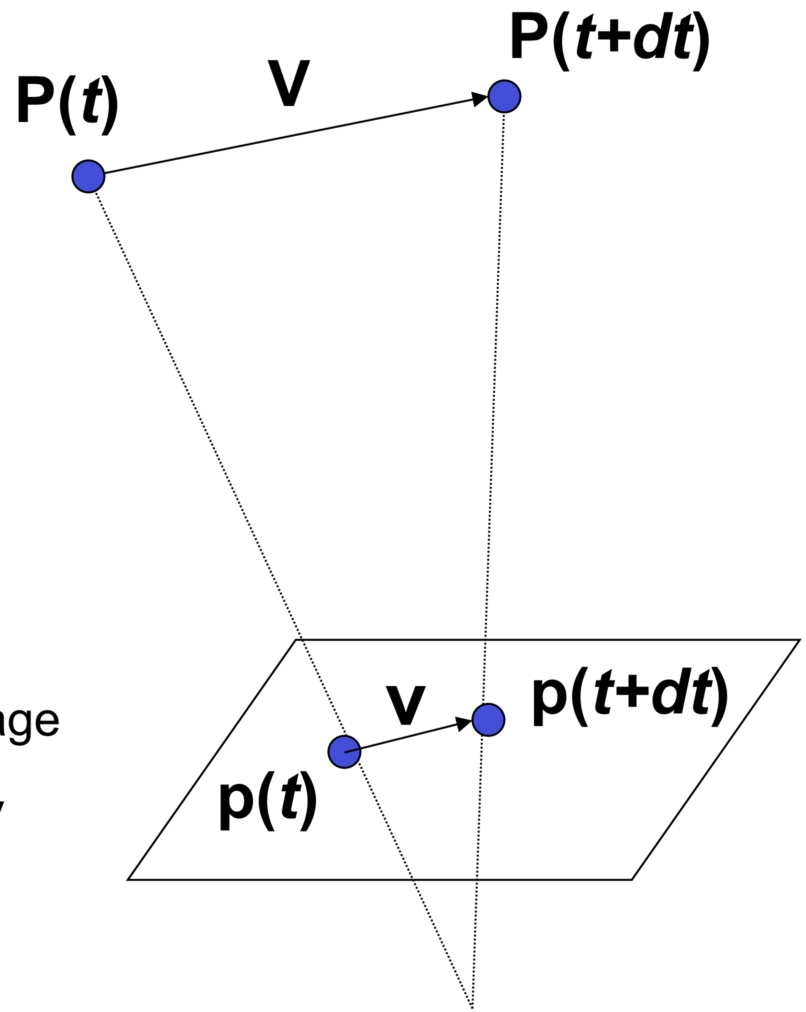
Estimating motion (aka optical flow/registration/warping/ deformation/morphing/…)

T. Brox, J. Malik, Large displacement optical flow: descriptor matching in variational motion estimation, PAMI 2011

Motion Field

$P(t)$ moving 3D point

$V(t) = P'(t)$ 3D velocity of P



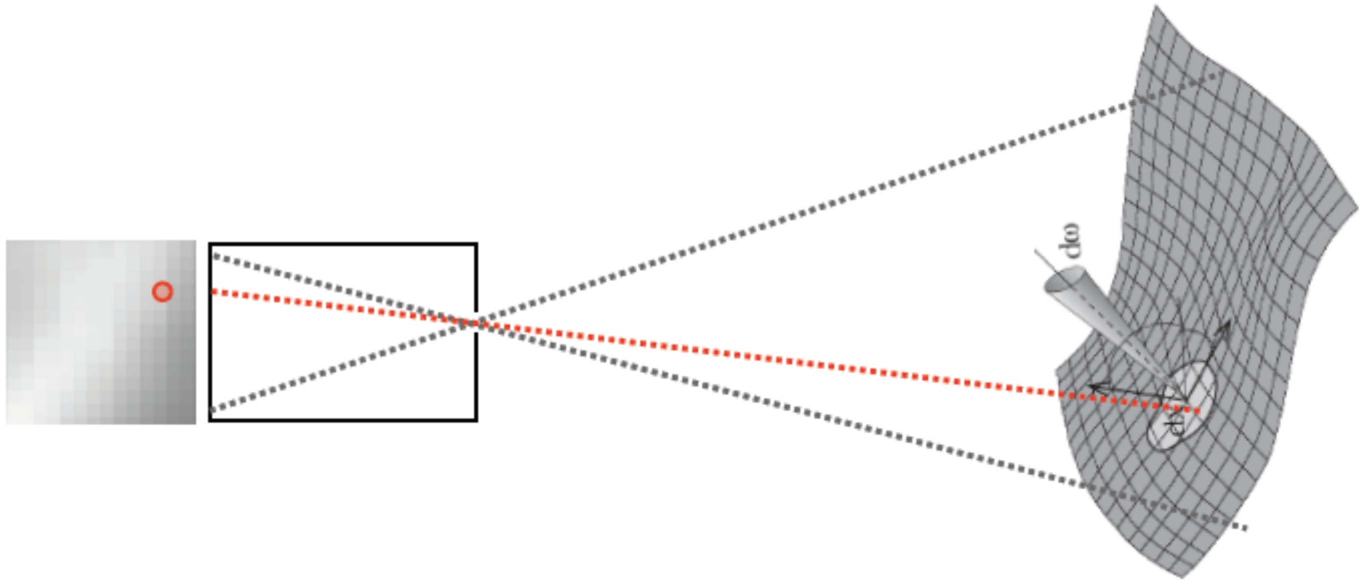
$p(t) = (x(t), y(t))$ projection on image

$\mathbf{v}(t) = (u(t), v(t))$ apparent velocity

Task: estimate
$$u(t) = x'(t)$$
$$v(t) = y'(t)$$

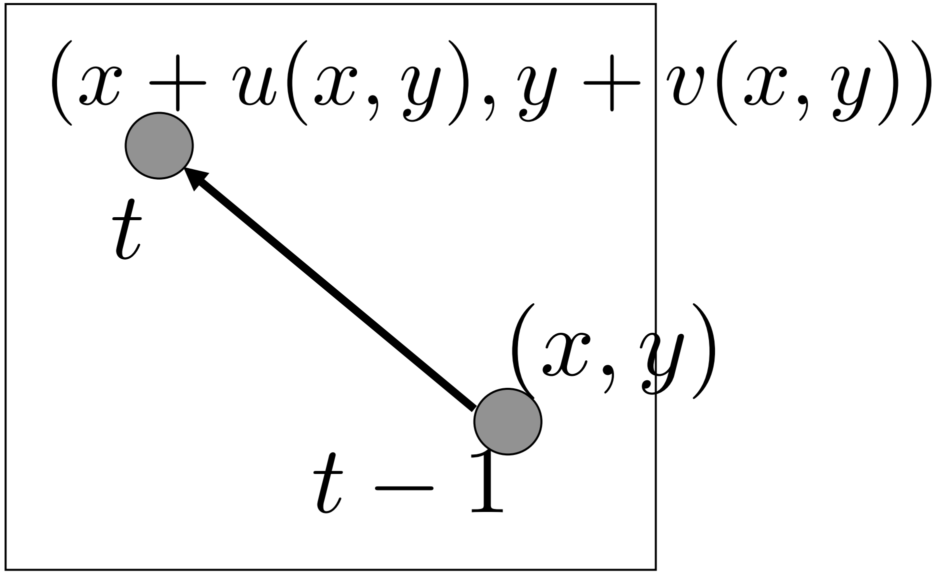
Brightness constancy constraint

- Image is projection of 3D environment
- Each pixel: projection of a surface patch
- Pixel intensity influenced by 3D surface, incident light, camera …
- Assumption: intensity of surface patch remains constant



Brightness constancy constraint

Optical flow estimation:



Constraint:

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

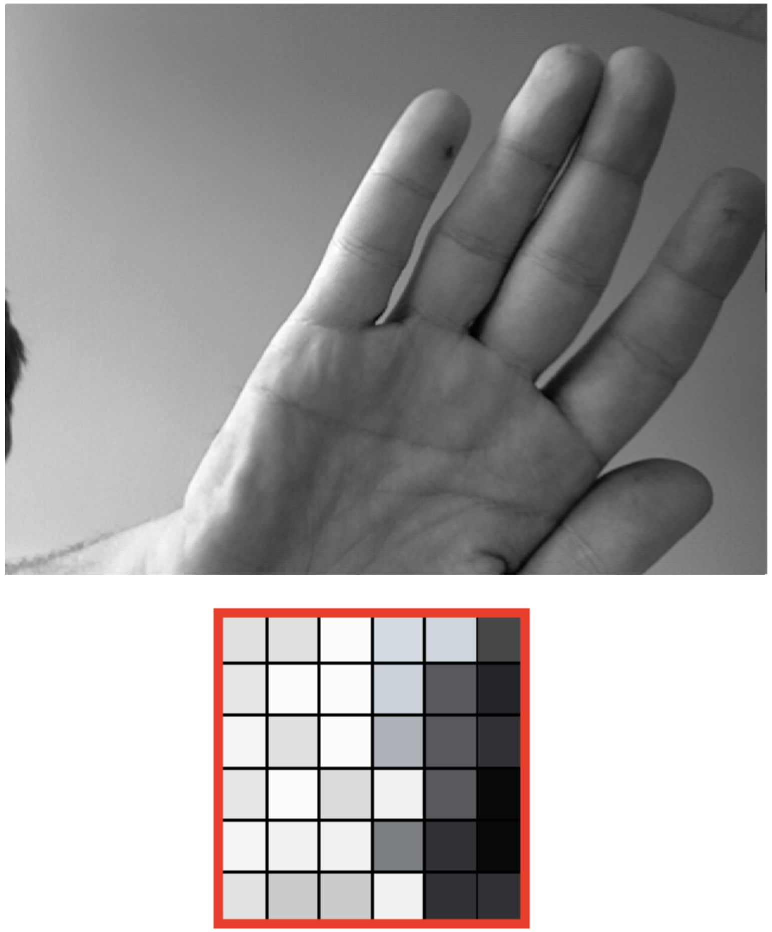
Taylor expansion of RHS:

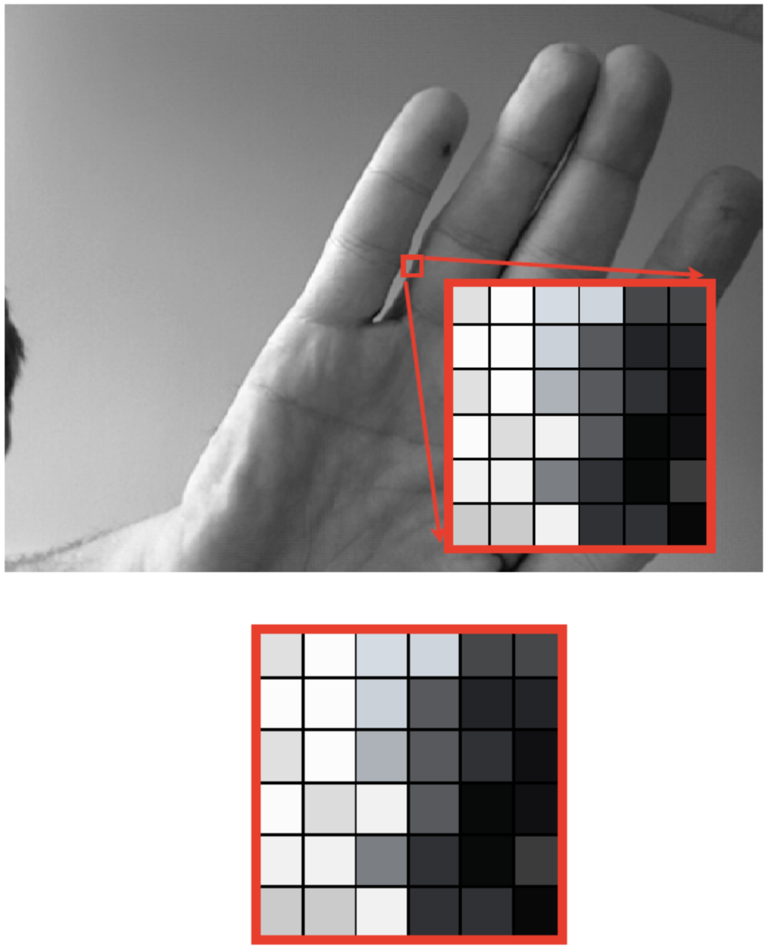
$$I(x, y, t-1) \simeq I(x, y, t) + \frac{\partial I}{\partial x} u(x,y) + \frac{\partial I}{\partial y} v(x,y)$$

Brightness Constancy Constraint:

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

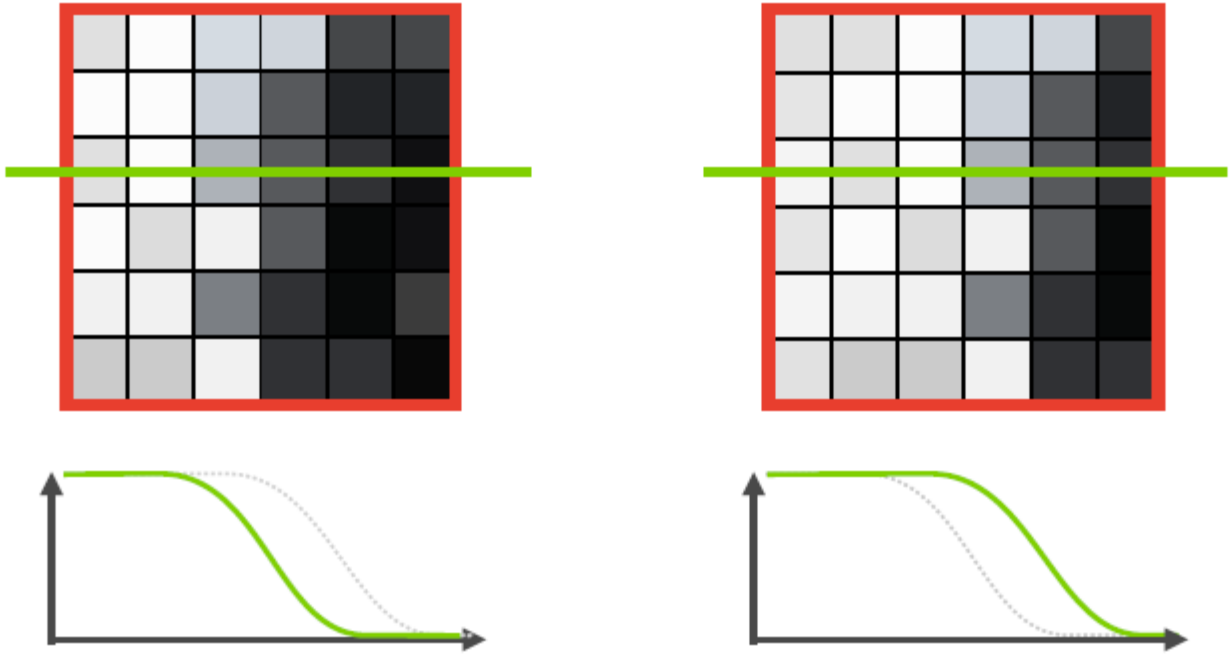
Optical flow estimation - 2D to 1D





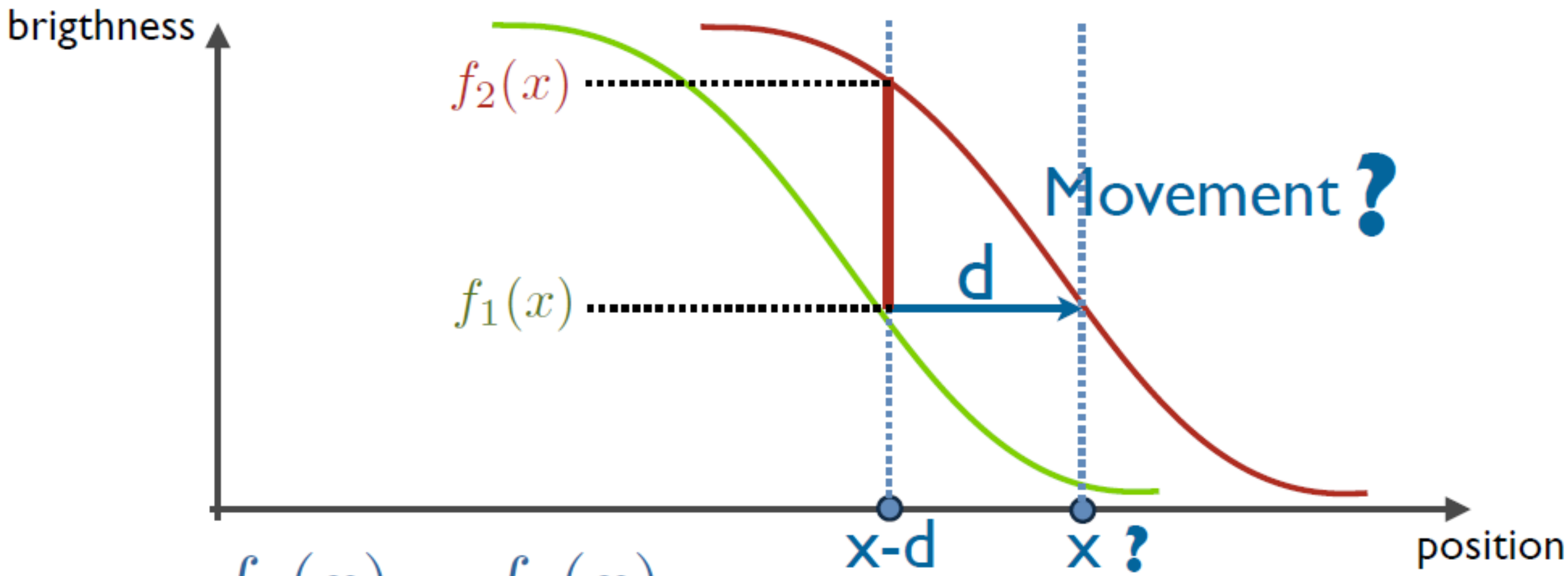
Optical flow estimation - 1D case

- How can we estimate the displacement?



Optical Flow Estimation - 1D case

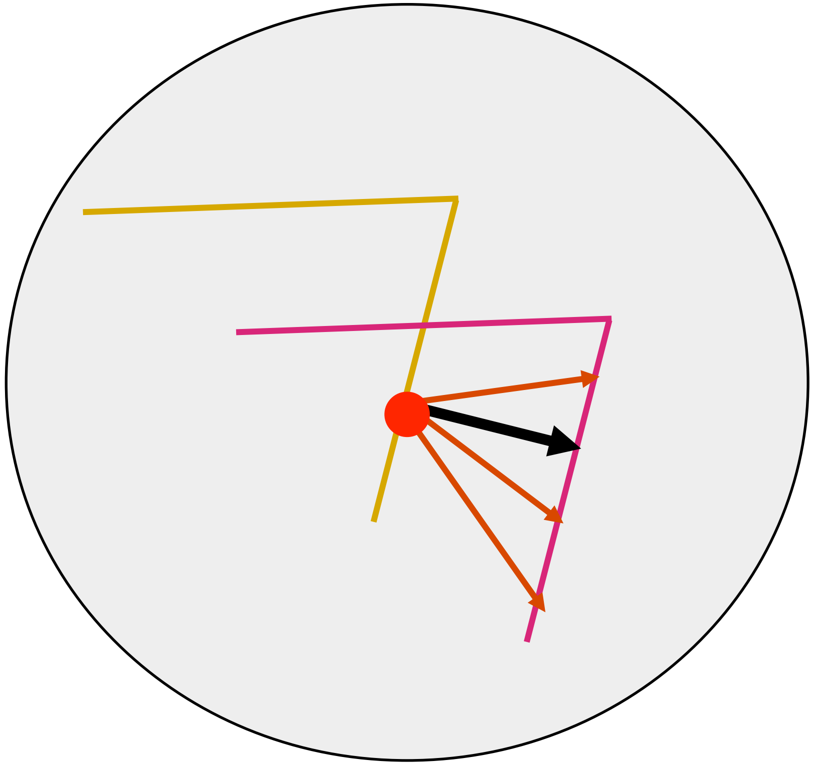
- Known: Gradient, Difference of intensities at x-d
- Unknown: d



$$d = \frac{f_1(x) - f_2(x)}{f_1'(x)}$$

How about 2D?

Brightness constraint: not enough for 2D!



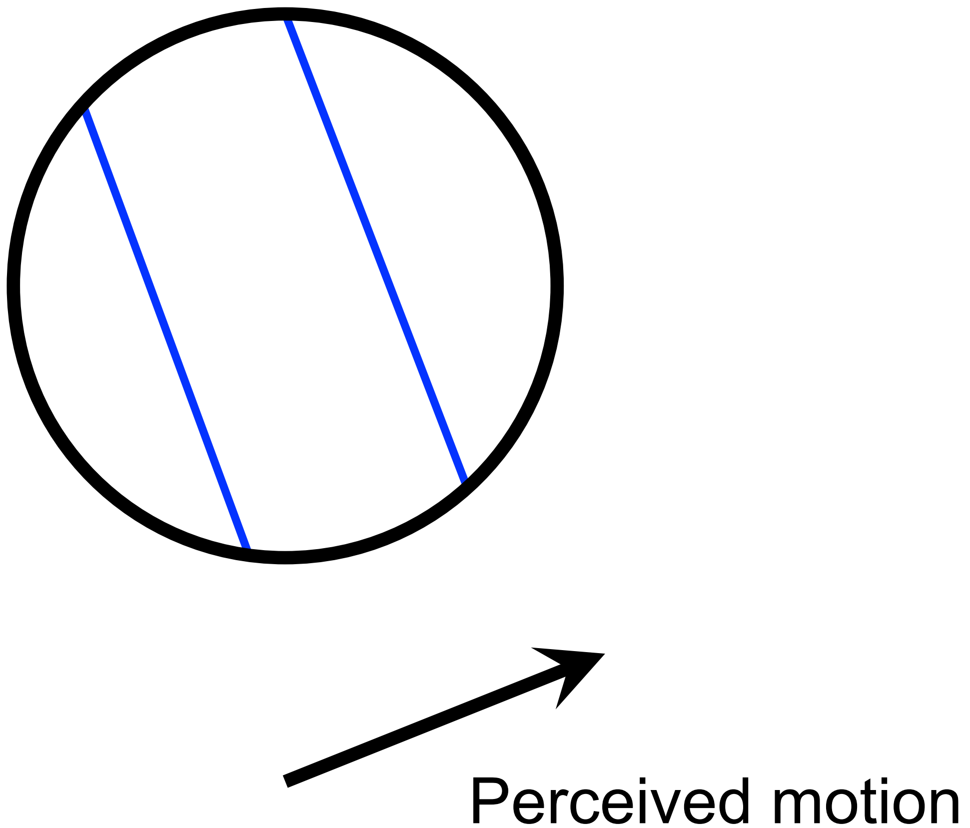
$$I_x u + I_y v + I_t = 0$$

$$(u', v') = (u, v) + c(-I_y, I_x)$$

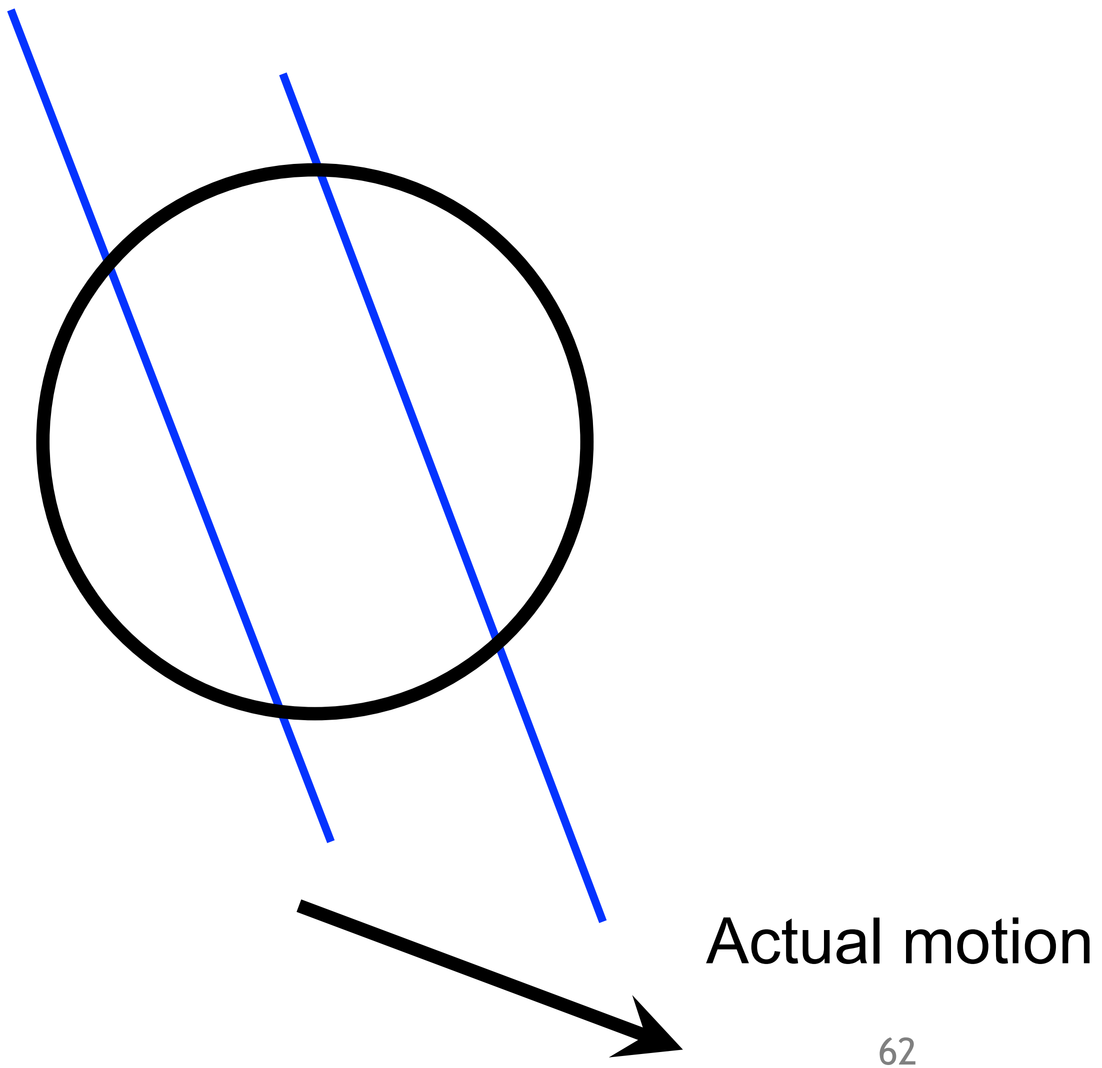
unknown flow parallel to the edge

$$\nabla I \cdot (u, v) + I_t = 0$$

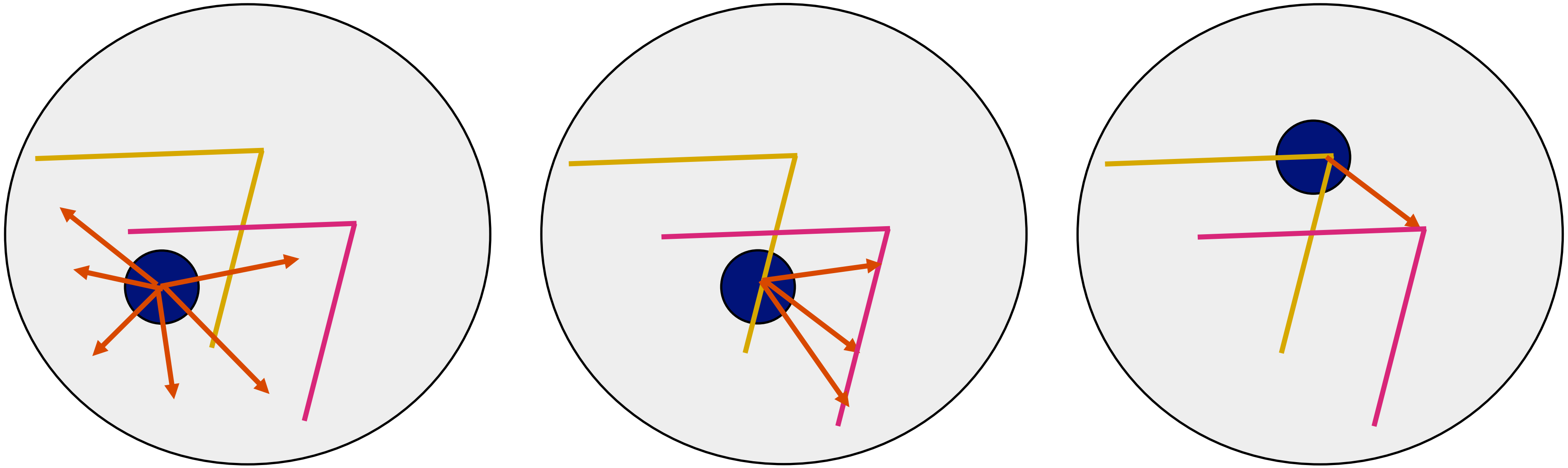
The Aperture Problem



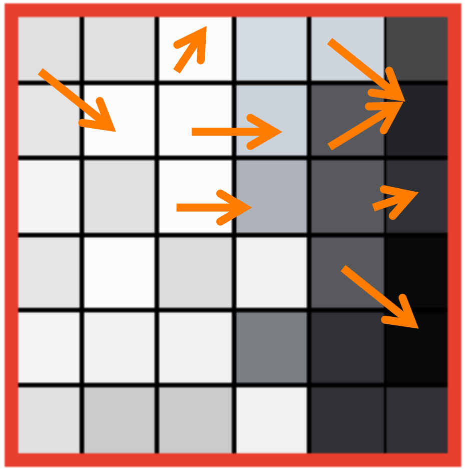
The Aperture Problem



Optical flow uncertaintly



Overcoming the aperture effect

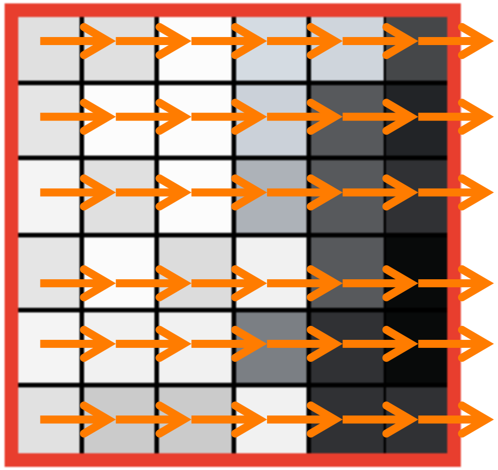


Brightness constancy constraint

$$I_x(p_i)u_i + I_x(p_i)v_i + I_t(p_i) = 0$$

Overcoming the aperture effect

united we move



Brightness constancy constraint

$$I_x(p_i)u + I_x(p_i)v + I_t(p_i) = 0$$

Lucas-Kanade

$$I_x(p_i)u + I_x(p_i)v + I_t(p_i) = 0$$

$$\begin{bmatrix} I_{x,1} & I_{y,1} \\ \vdots & \\ I_{x,25} & I_{y,25} \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = \begin{bmatrix} -I_{t,1} \\ \vdots \\ -I_{t,25} \end{bmatrix}$$

Rewrite: $\mathbf{A}\mathbf{u} = \mathbf{b}$ 25 equations, 2 unknows

Residuals: $\epsilon = \mathbf{A}\mathbf{u} - \mathbf{b}$

Cost: $\epsilon^T\epsilon = \mathbf{b}^T\mathbf{b} - 2\mathbf{u}^T\mathbf{A}^T\mathbf{b} + \mathbf{u}^T\mathbf{A}^T\mathbf{A}\mathbf{u}$

Minimization: $\mathbf{A}^T\mathbf{A}\mathbf{u} = \mathbf{A}^T\mathbf{b}$

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. IJCAI, 1981.

Lucas-Kanade, continued

$$\mathbf{A}^T\mathbf{A}\mathbf{u} = \mathbf{A}^T\mathbf{b}$$

$$\mathbf{u} = \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{A}^T\mathbf{b}$$

Is it invertible?

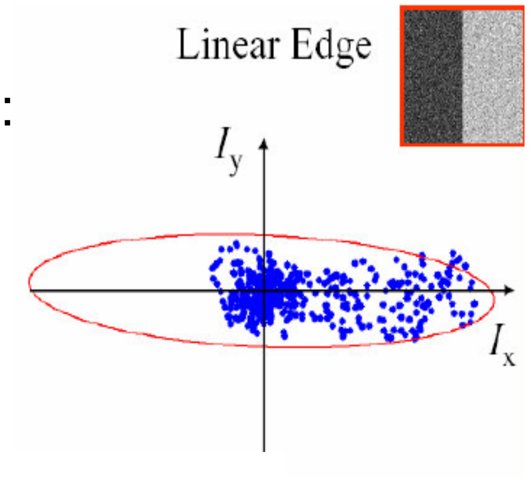
$$\mathbf{A} = \left[\begin{array}{cc} I_{x,1} & I_{y,1} \\ & \vdots \\ I_{x,25} & I_{x,25} \end{array}\right]$$

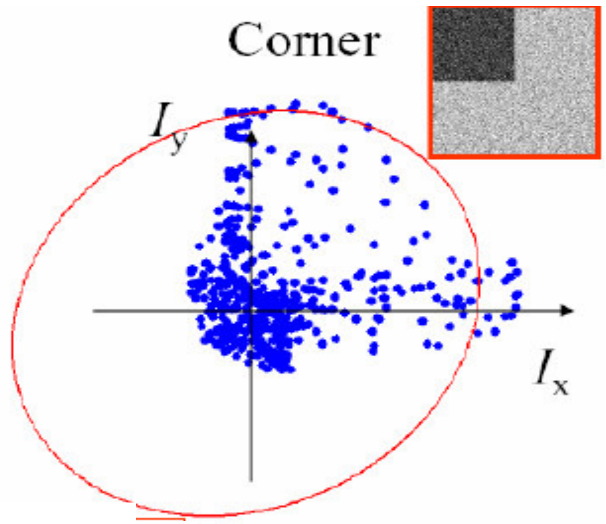
$$\mathbf{A}^T\mathbf{A} = \left[\begin{array}{cc} \sum_i I_{x,i}^2 & \sum_i I_{x,i}I_{y,i} \\ \sum_i I_{x,i}I_{y,i} & \sum_i I_{y,i}^2 \end{array}\right]$$

B. Lucas and T. Kanade. An iterative image registration technique with an application to stereo vision. IJCAI, 1981.

Second Moment Matrix

Distribution of gradients:

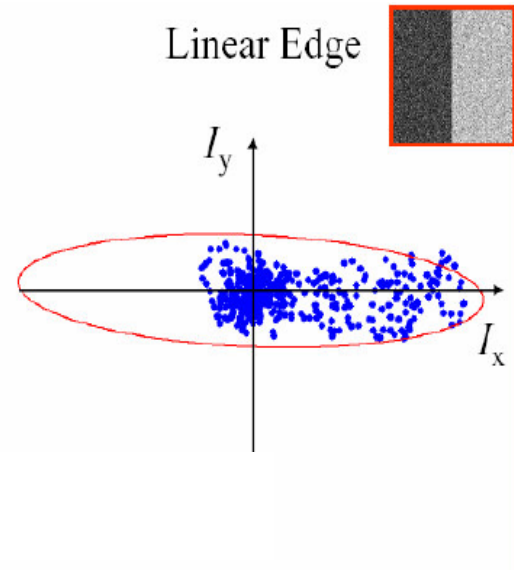


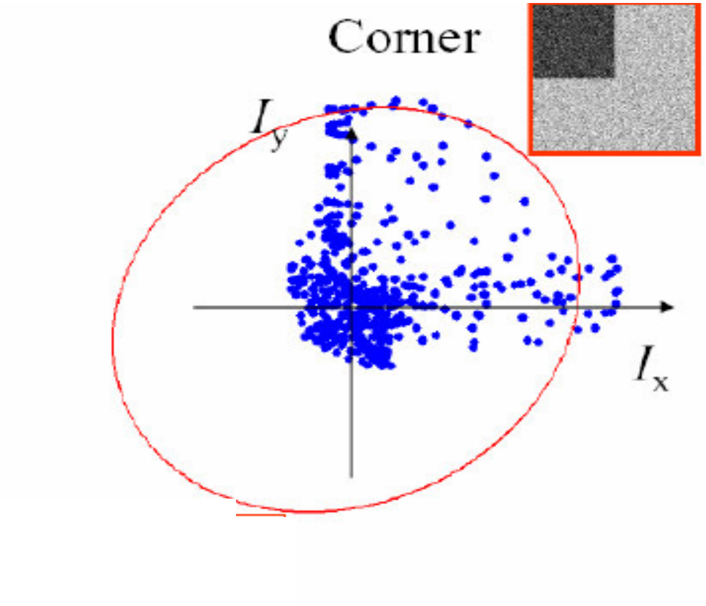


$$J = \begin{bmatrix} \sum_{x',y'} I_x^2 & \sum_{x',y'} I_x I_y \\ \sum_{x',y'} I_x I_y & \sum_{x',y'} I_y^2 \end{bmatrix}$$

Second Moment Matrix

Distribution of gradients:

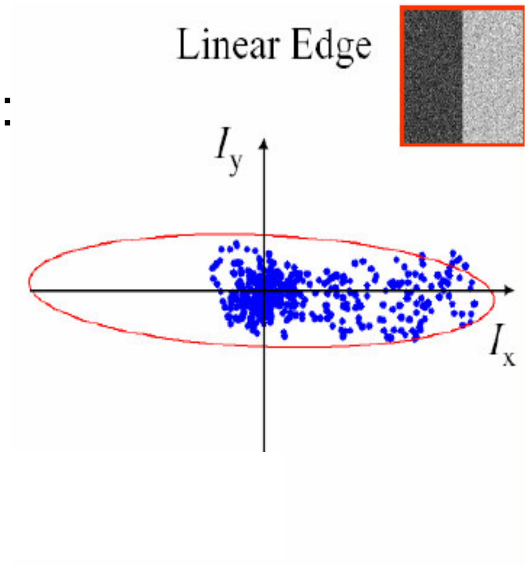


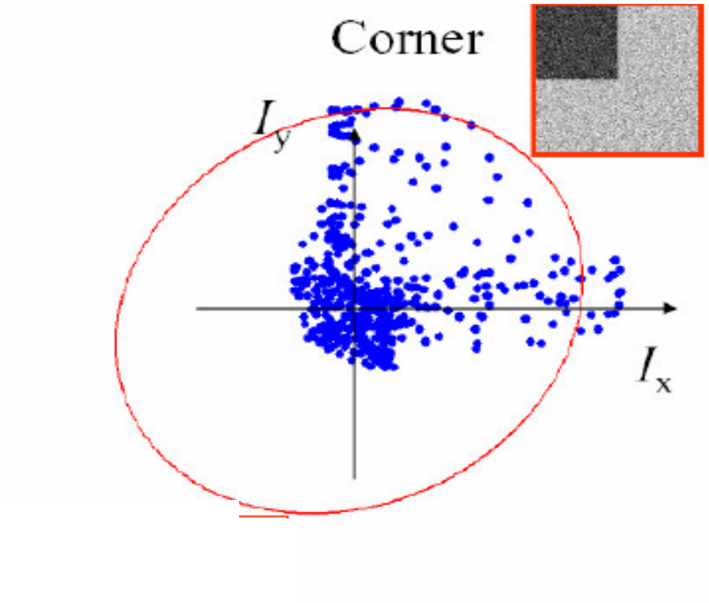


$$J = \sum_{x',y'} (I_x, I_y)^T (I_x, I_y)$$

Second Moment Matrix

Distribution of gradients:

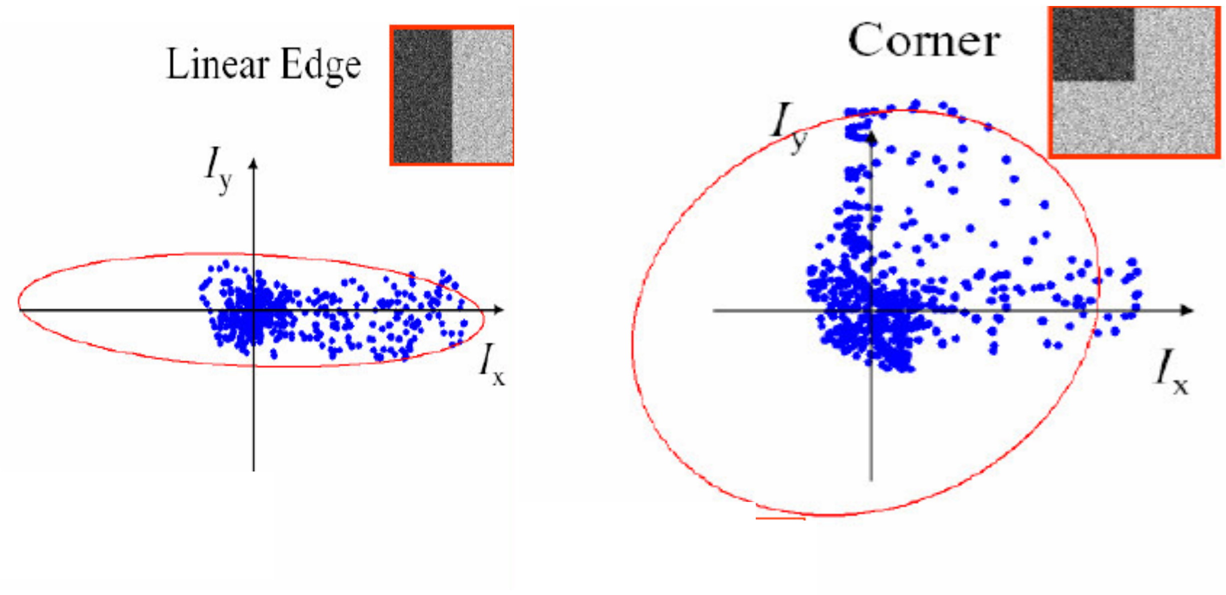




$$J = \sum_{x',y'} (\nabla G_\sigma * u)^T (\nabla G_\sigma * u)$$

Second Moment Matrix

Distribution of gradients:

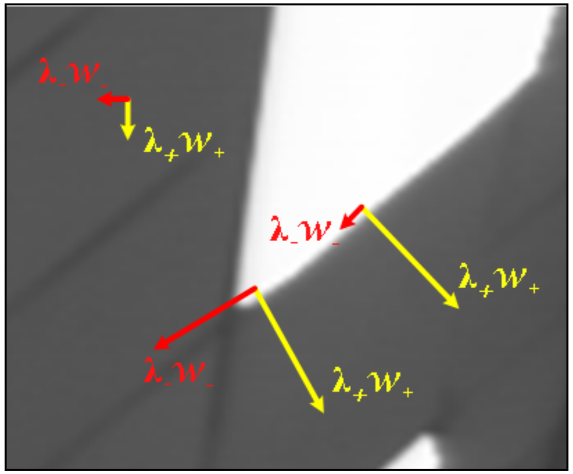


$$J = G_\rho * \left[(\nabla G_\sigma * u)^T (\nabla G_\sigma * u) \right]$$

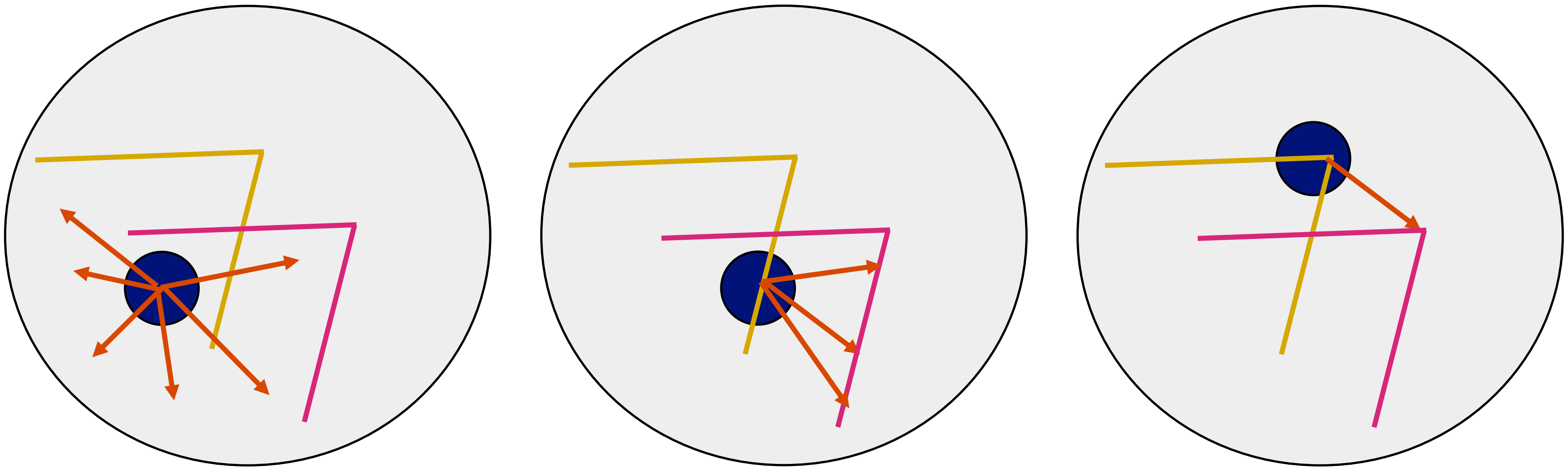
Second Moment Matrix

$$J = G_\rho * \left[(\nabla G_\sigma * u)^T (\nabla G_\sigma * u) \right]$$

- Eigenvectors w_+, w_- : directions of maximal and minimal variation of u
- Eigenvalues: amounts of minimal and maximal variation u



Local structure and motion estimation



$\lambda_1 \simeq 0, \lambda_2 \simeq 0$ λ_1 large, $\lambda_2 \simeq 0$ $\lambda_1, \lambda_2 :$ large

Lucas-Kanade and the aperture effect

Brightness constancy + neighboring pixels have same (u,v)

$$I_x(p_i)u + I_x(p_i)v + I_t(p_i) = 0$$

5x5 window: 25 equations, 2 unknowns

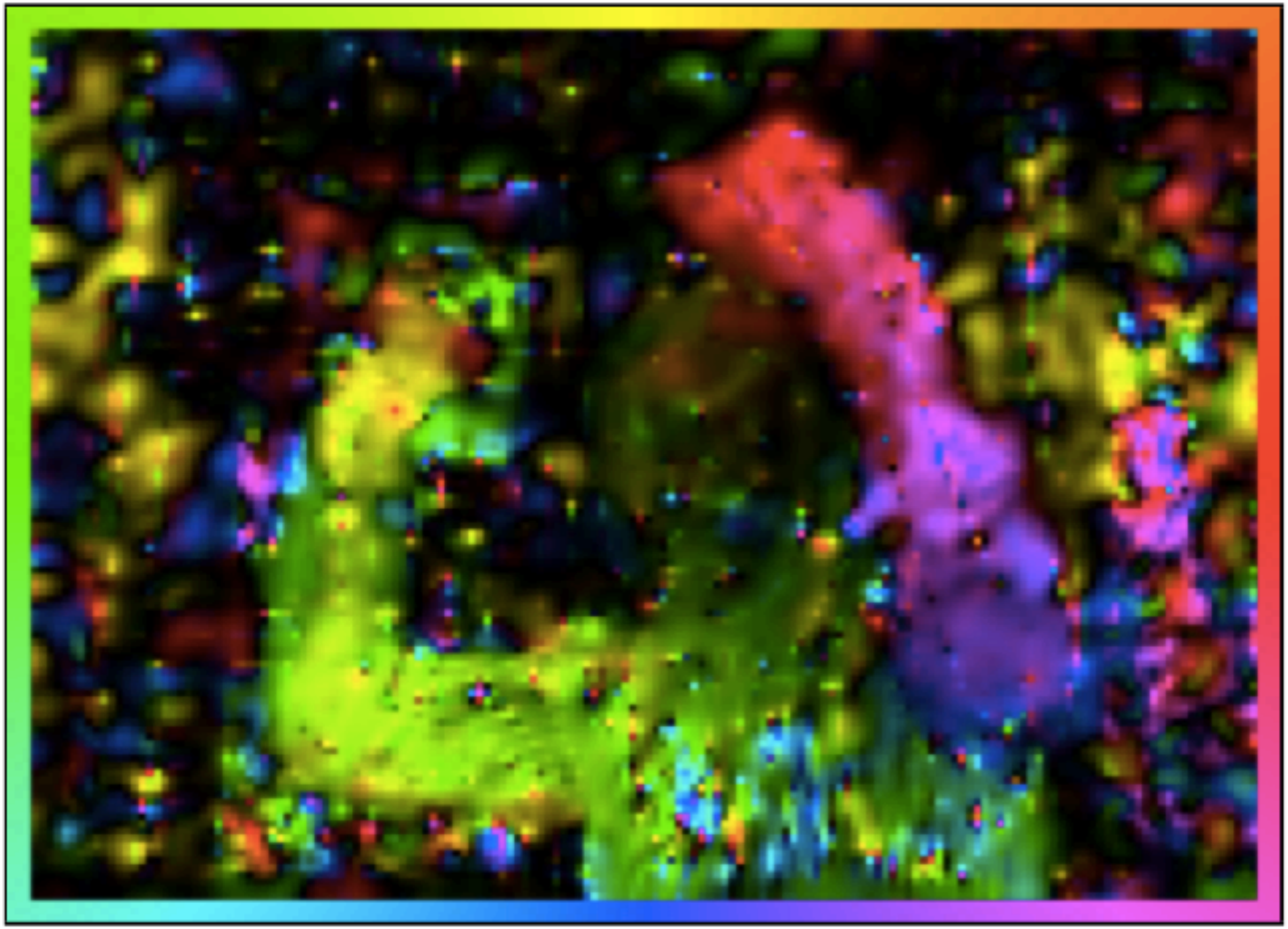
$$\begin{bmatrix} I_x(\mathrm{p}_1) & I_y(\mathrm{p}_1) \\ I_x(\mathrm{p}_2) & I_y(\mathrm{p}_2) \\ \vdots & \vdots \\ I_x(\mathrm{p}_{25}) & I_y(\mathrm{p}_{25}) \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} I_t(\mathrm{p}_1) \\ I_t(\mathrm{p}_2) \\ \vdots \\ I_t(\mathrm{p}_{25})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

Solve the system around each pixel separately

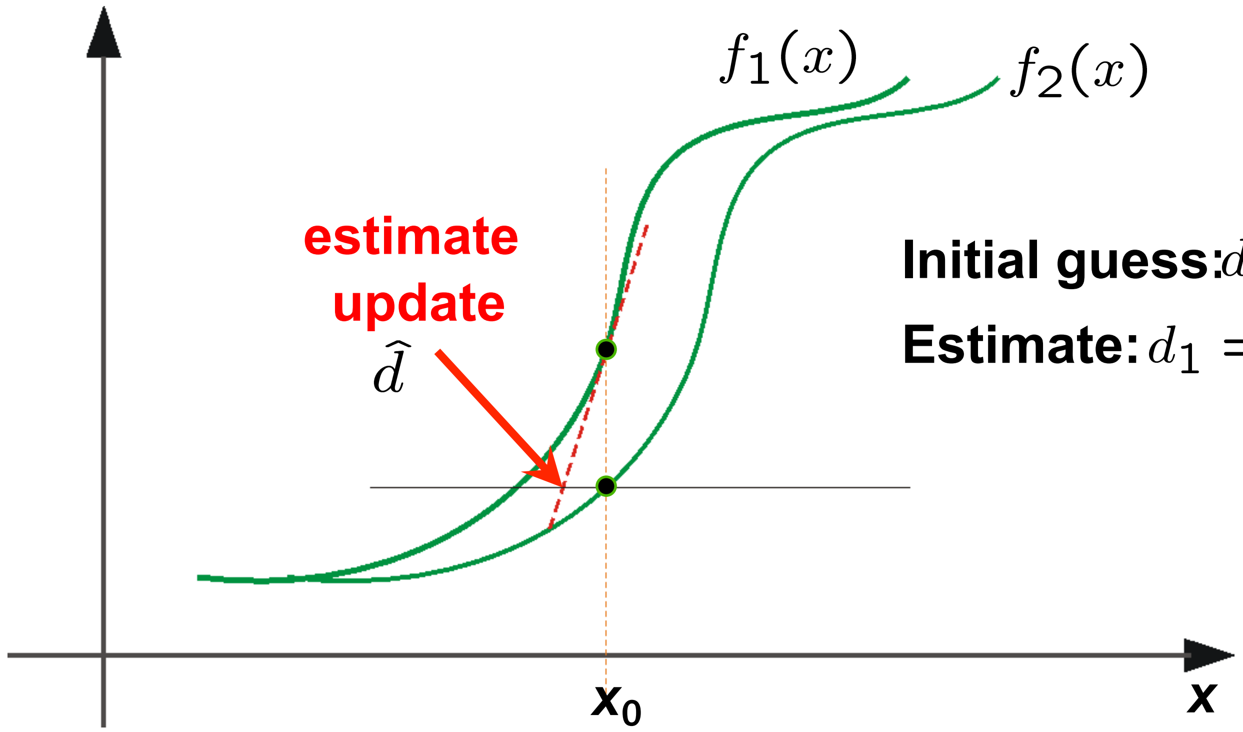
Dense Lukas Kanade

Dense Lukas Kanade

Dense Lukas Kanade



Optical Flow: Iterative Estimation

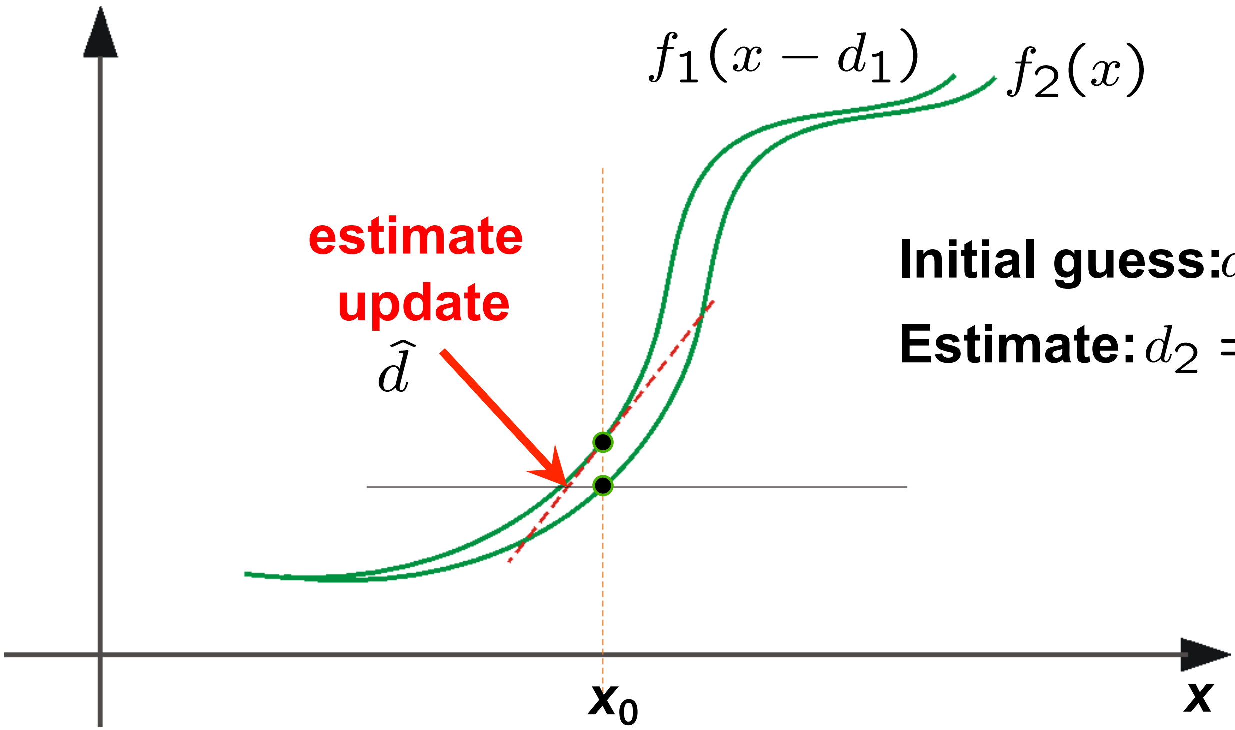


Initial guess: $d_0 = 0$

Estimate: $d_1 = d_0 + \hat{d}$

(using *d* for *displacement* here instead of *u*)

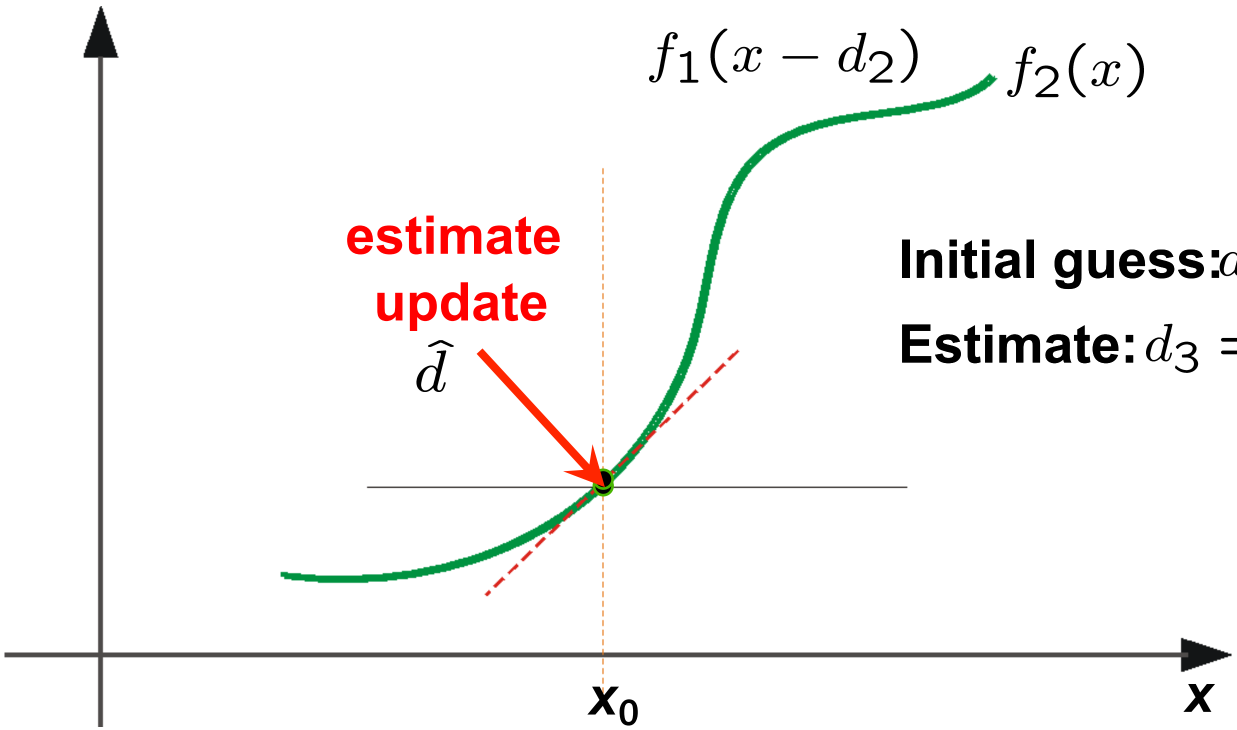
Optical Flow: Iterative Estimation



Initial guess: d_1

Estimate: $d_2 = d_1 + \hat{d}$

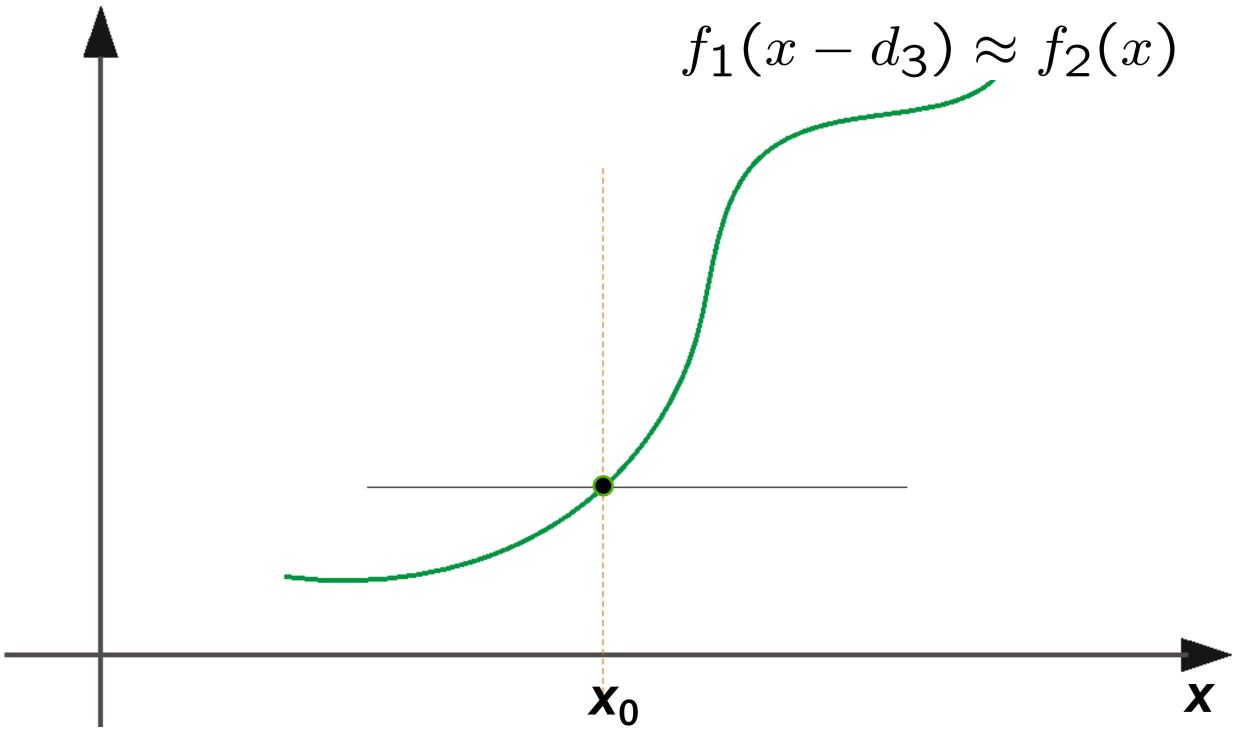
Optical Flow: Iterative Estimation



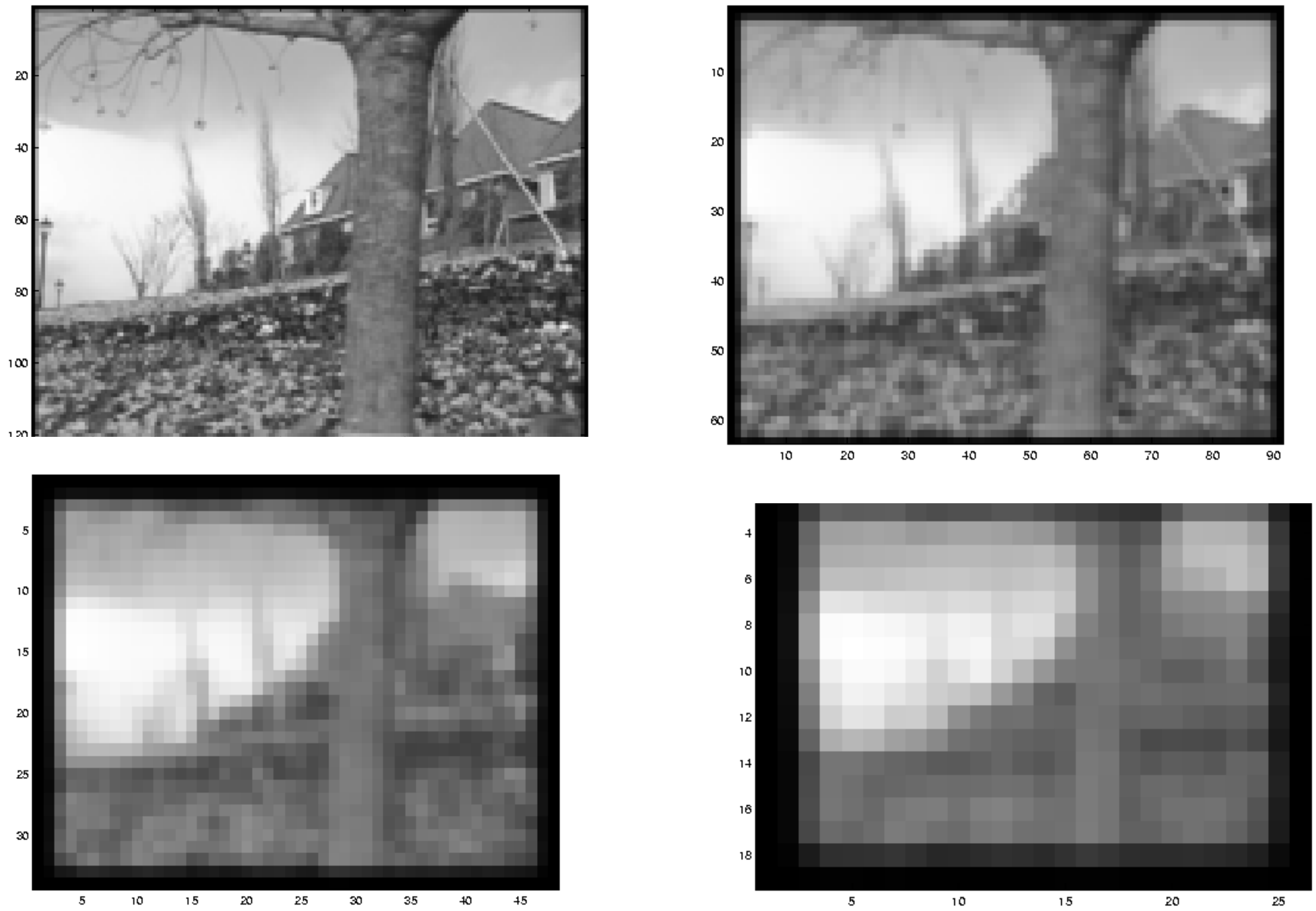
Initial guess: d_2

Estimate: $d_3 = d_2 + \hat{d}$

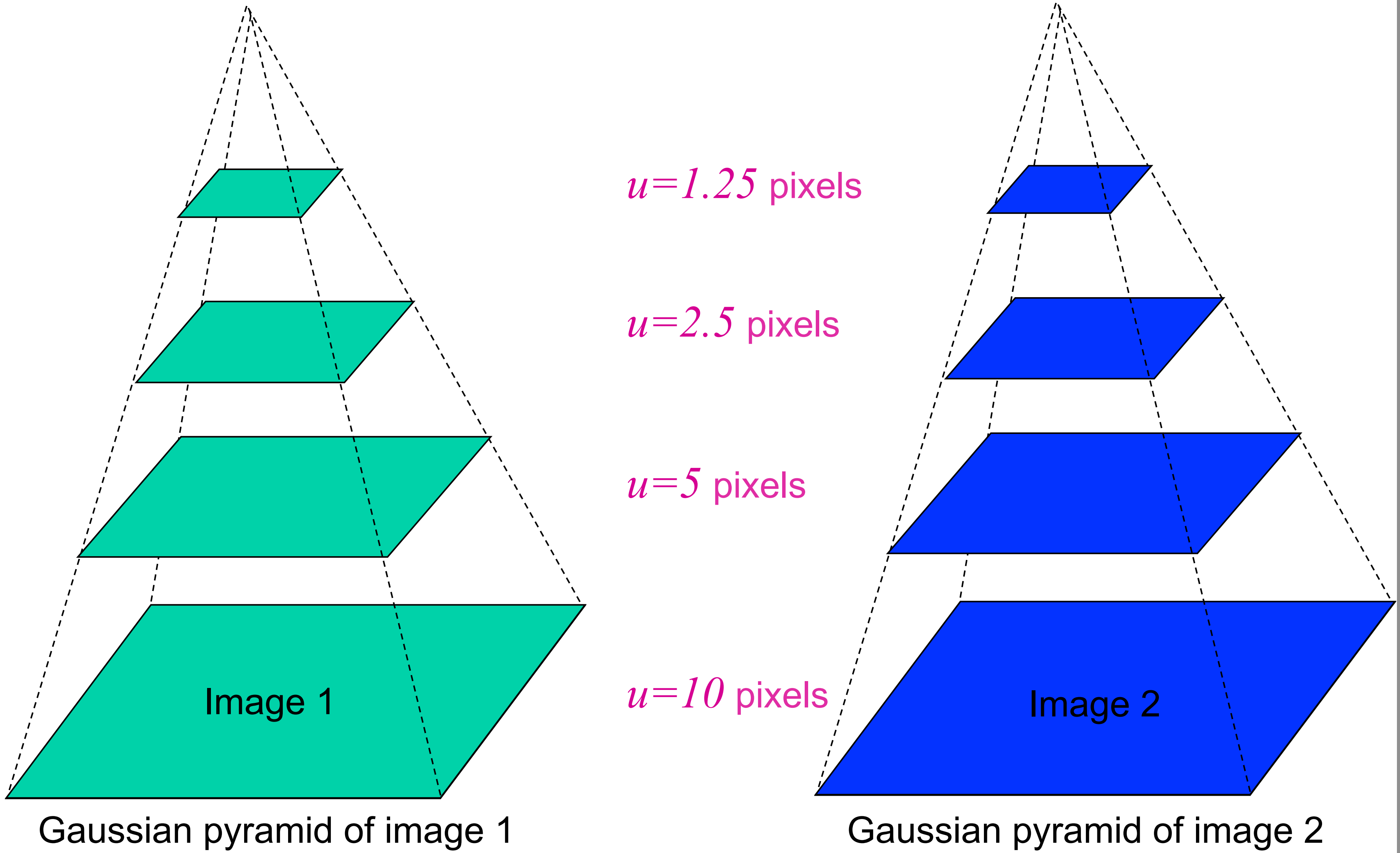
Optical Flow: Iterative Estimation



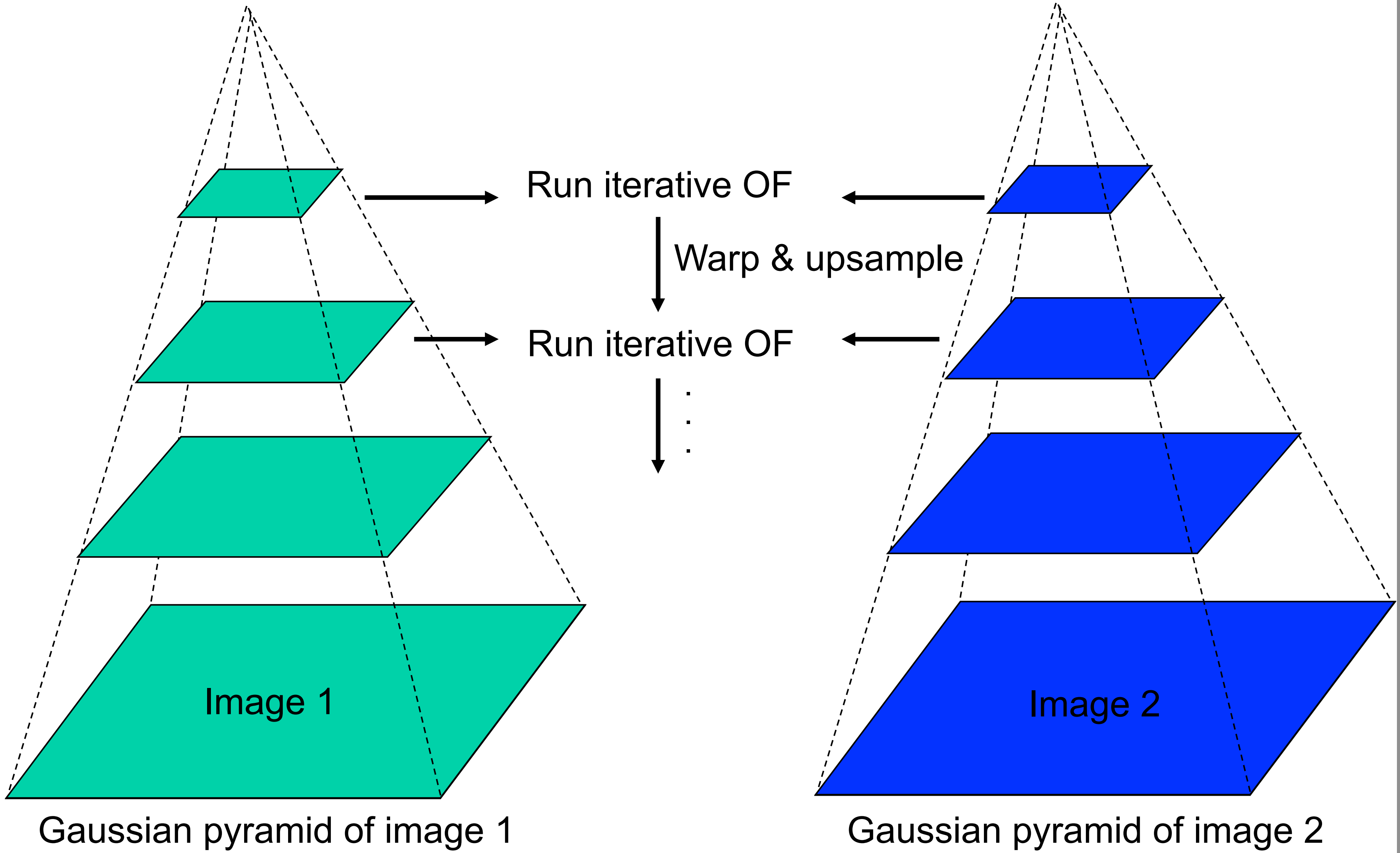
Large Displacements: Reduce Resolution!



Coarse-to-fine Optical Flow Estimation



Coarse-to-fine Optical Flow Estimation



Horn-Schunk and the aperture effect

Brightness constancy + flow smoothness:

$$J(u,v) = \frac{1}{2}\iint_\Omega \left[I_t + \nabla I \cdot (u,v)\right]^2 + \lambda\left[(\nabla u)^2 + (\nabla v)^2\right] \mathrm{d}x\mathrm{d}y$$

Minimize with respect to $u(x,y), v(x,y)$

Euler-Lagrange derivative:

$$J_u = \frac{\partial J}{\partial u} - \frac{\partial}{\partial x}\frac{\partial J}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial J}{\partial u_y}$$

Minimum conditi $J_u = 0, \quad J_v = 0$

B.K.P. Horn and B.G. Schunck, "Determining optical flow." Artificial Intelligence, 1981.

Horn-Schunk and the aperture effect

Brightness constancy + flow smoothness:

$$J(u,v) = \frac{1}{2}\iint_{\Omega} \left[I_t + \nabla I \cdot (u,v)\right]^2 + \lambda \left[(\nabla u)^2 + (\nabla v)^2\right] \mathrm{d}x\mathrm{d}y$$

Minimize with respect to $u(x,y), v(x,y)$

Euler-Lagrange derivative:

$$\begin{aligned} J_u &= \frac{\partial J}{\partial u} - \frac{\partial}{\partial x}\frac{\partial J}{\partial u_x} - \frac{\partial}{\partial y}\frac{\partial J}{\partial u_y} \\ &= I_x \left(I_t + (\nabla I) \cdot [u,v]\right) - \lambda \left(u_{xx} + u_{yy}\right) \\ &= I_x I_t + I_x^2 u + I_x I_y v - \lambda \left(\nabla^2 u\right) \\ J_v &= I_y I_t + I_x I_y v + I_y^2 v - \lambda \left(\nabla^2 v\right) \end{aligned}$$

Minimum conditi $J_u = 0, \quad J_v = 0$

B.K.P. Horn and B.G. Schunck, "Determining optical flow." Artificial Intelligence, 1981.

Horn-Schunk results

Horn-Schunk results

Horn-Schunk results

Lucas-Kanade revisited

- General expression for LK criterion:

$$
\begin{aligned}
E_{LK}(u,v)\big|_{x,y} &= \sum_{\mathcal{N}(x,y)} (f_x u + f_y v + f_t)^2 \\
&= K_\rho * (f_x u + f_y v + f_t)^2\big|_{x,y}
\end{aligned}
$$

- LK flow

$$
\begin{bmatrix} K_\rho * f_x^2 & K_\rho * (f_x f_y) \\ K_\rho * (f_x f_y) & K_\rho * (f_y)^2 \end{bmatrix} \begin{bmatrix} u \\ v \end{bmatrix} = - \begin{bmatrix} K_\rho * (f_x f_t) \\ K_\rho * (f_y f_t) \end{bmatrix}
$$

- How does this compare with Horn-Schunk criterion?

Lucas-Kanade meets Horn-Schunk

Introduce: $w \doteq \begin{bmatrix} u \\ v \\ 1 \end{bmatrix}$ $\nabla_3 f \doteq \begin{bmatrix} f_x \\ f_y \\ f_t \end{bmatrix}$

$$|\nabla w|^2 \doteq |\nabla u|^2 + |\nabla v|^2 \quad J_\rho \doteq G_\rho * \left[\nabla_3 f \nabla_3 f^T\right]$$

Lucas-Kanade

$$E_{LK}(w) = \int_\Omega w^T J_\rho w \ \mathrm{d}x\mathrm{d}y$$

Horn-Schunk

$$E_{HS}(w) = \int_\Omega w^T J_0 w + a|\nabla w| \mathrm{d}x\mathrm{d}y$$

Bruhn-Weickert-Schnoerr

$$E_{BWS}(w) = \int_\Omega w^T J_\rho w + a|\nabla w| \mathrm{d}x\mathrm{d}y$$

A. Bruhn, J. Weickert, C. Schnörr: 'Lucas/Kanade meets Horn/Schunck: Combining local and global optic flow methods.' IJCV 2005

Horn-Schunk

Bruhn-Weickert-Schnoerr

Bruhn-Weickert-Schnoerr, continued

$$E_{BWS}(w) = \int_{\Omega} w^T J_\rho w + a|\nabla w| \mathrm{d}x\mathrm{d}y$$

Spatio-temporal regularization

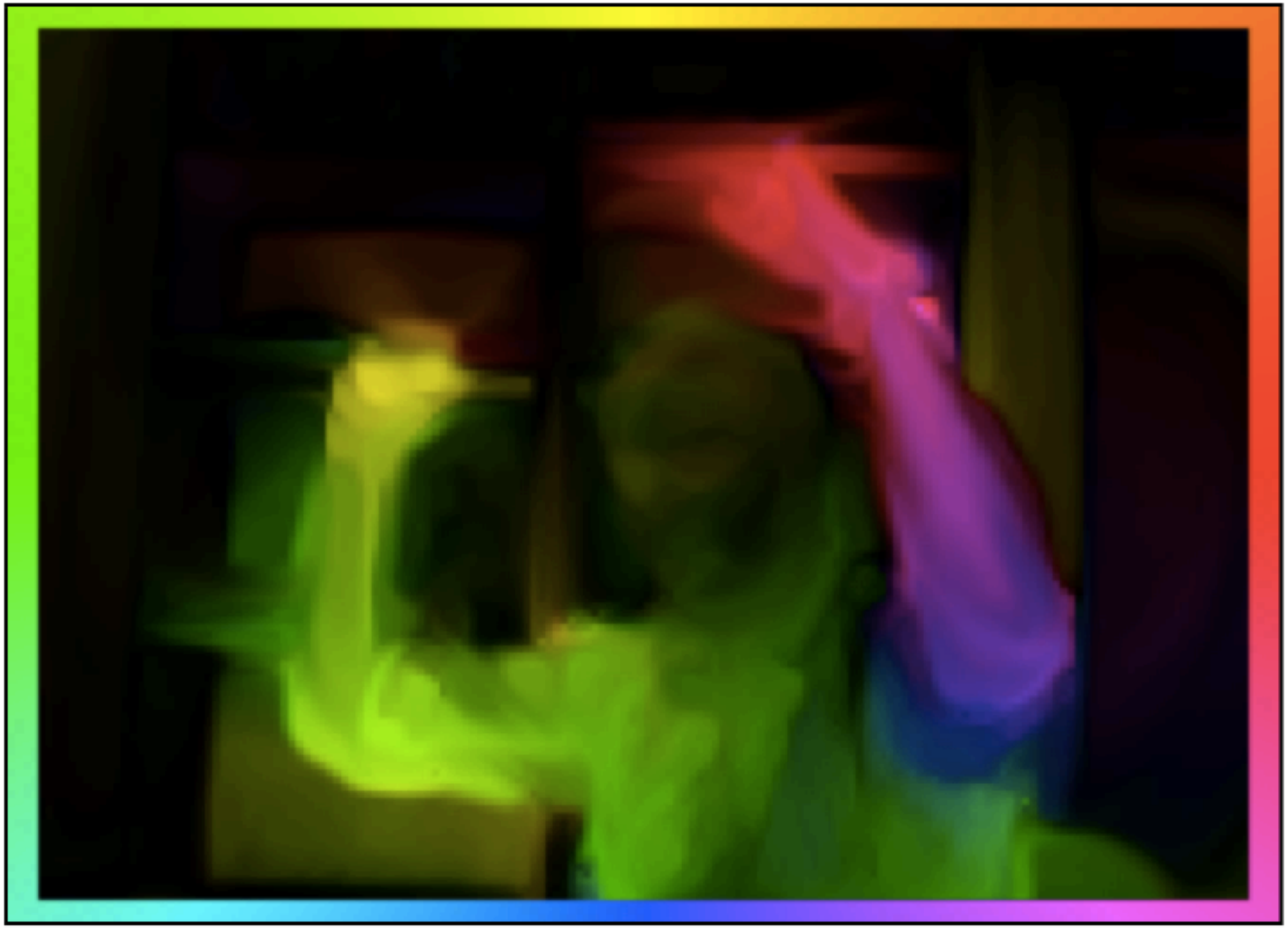
$$E'_{BWS}(w) = \int_{\Omega \times [0,T]} w^T J_\rho w + a|\nabla_3 w| \mathrm{d}x\mathrm{d}y$$

Robust norms:

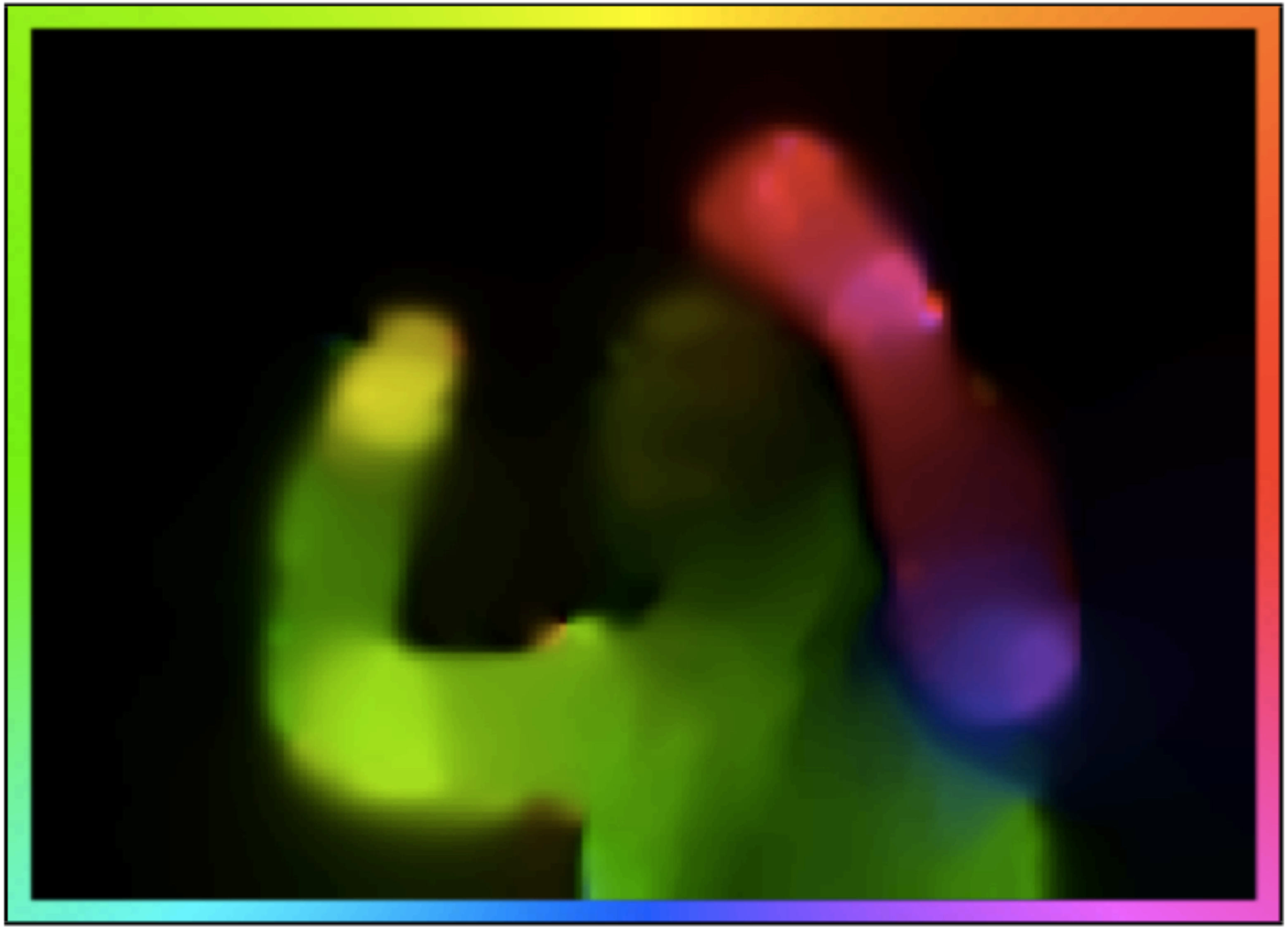
$$E''_{BWS}(w) = \int_{\Omega \times [0,T]} \psi_1 \left(w^T J_\rho w\right) + a\psi_2 \left(|\nabla_3 w|\right) \mathrm{d}x\mathrm{d}y$$

Bruhn-Weickert-Schnoerr

Bruhn-Weickert-Schnoerr - anisotropic



Bruhn-Weickert-Schnoerr – flow regularization



Numerical solutions

- Euler-Lagrange:

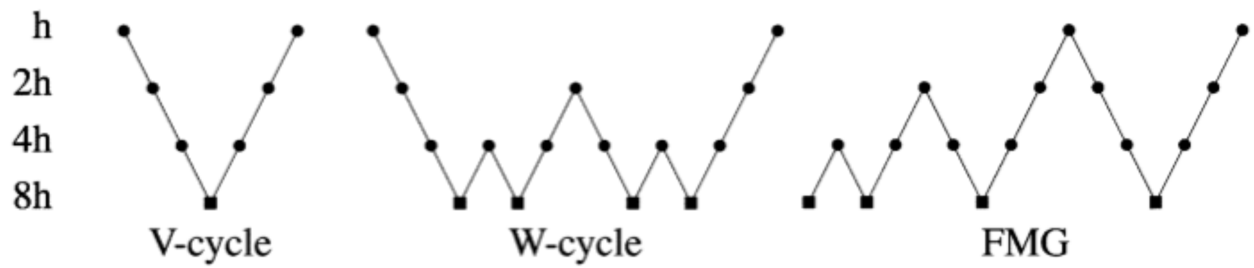
$$I_x^2 u + I_x I_y v = \lambda \nabla^2 u - I_x I_t$$
$$I_x I_y u + I_y^2 v = \lambda \nabla^2 v - I_y I_t$$

- Numerical approximation to Laplacian

$$\nabla^2 u \simeq 4\hat{u}(x,y) - u(x,y)$$

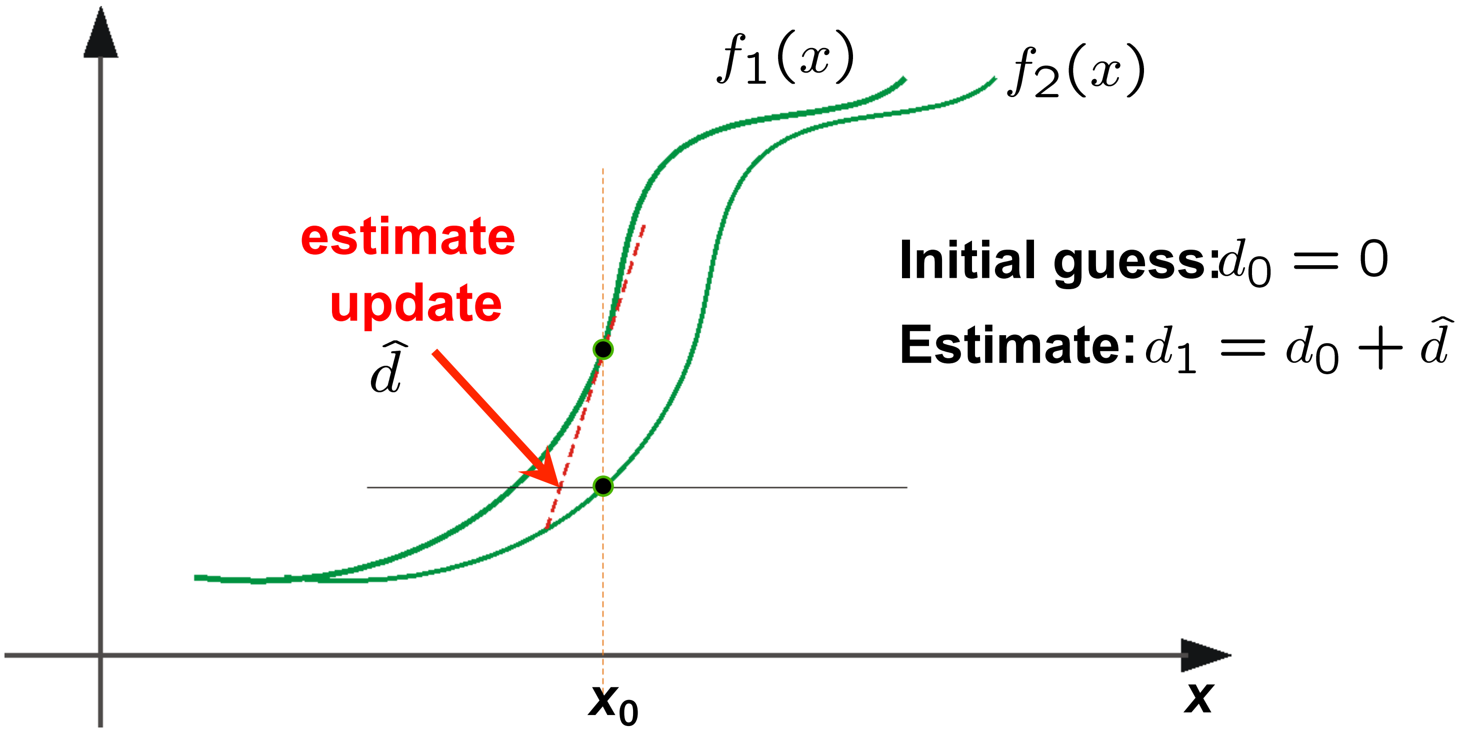
$$\hat{u}(x,y) = \frac{1}{4}\left(u(x-1,y) + u(x+1,y) + u(x,y-1) + u(x,y+1)\right)$$

- Sparse linear system in u,v
  - Gauss-Seidel, Successive Over Relaxation (SOR)
  - Multigrid
  - 12 fps, 2006



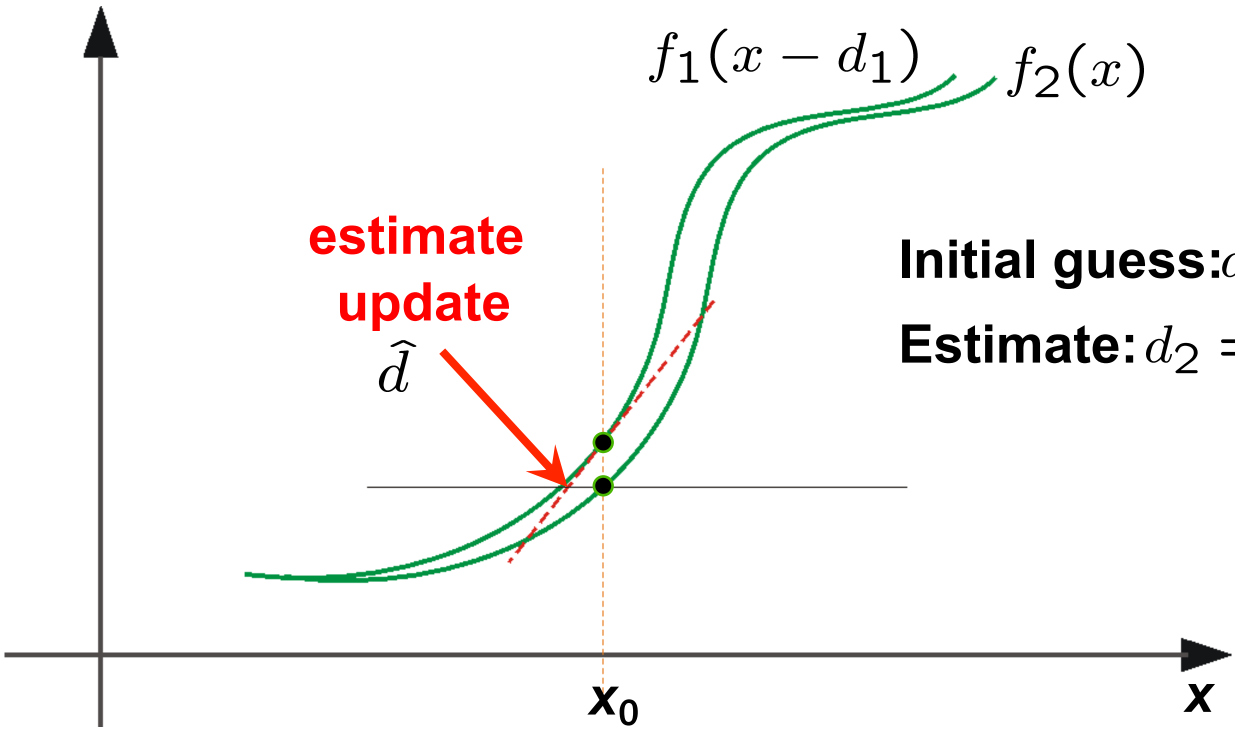
A. Brandt, "Multi-level adaptive solutions to boundary-value problems," Math. Comput., 1977.
D. Terzopoulos, Image Analysis Using Multigrid Relaxation Methods, PAMI, 1986
A. Kenigsberg, R. Kimmel, and I. Yavneh, "A Multigrid Approach for Fast Geodesic Active Contours," 2001
G. Papandreou and P. Maragos, "Multigrid Geometric Active Contour Models", TIP 2007
A. Bruhn, J. Weickert, T. Kohlberger, C. Schnörr: A multigrid platform for real-time motion computation with discontinuity-preserving variational methods. IJCV 2006

Optical Flow: Iterative Estimation



(using *d* for *displacement* here instead of *u*)

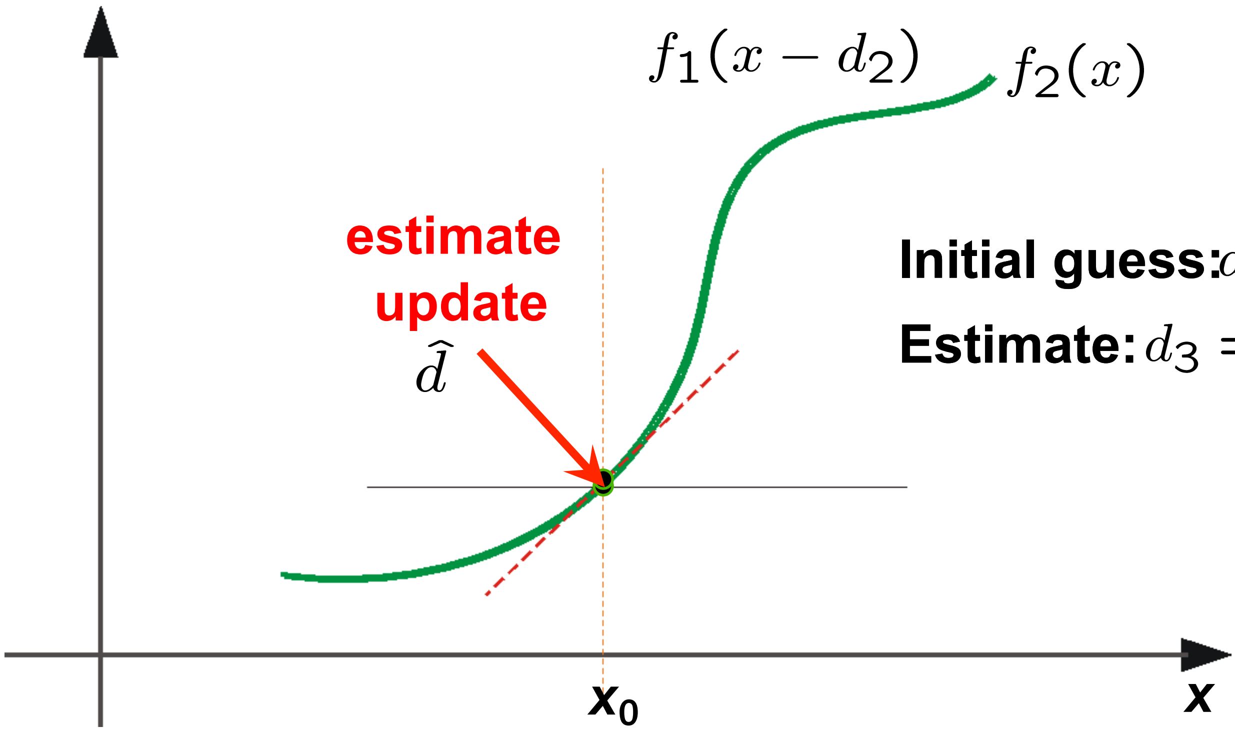
Optical Flow: Iterative Estimation



Initial guess: d_1

Estimate: $d_2 = d_1 + \hat{d}$

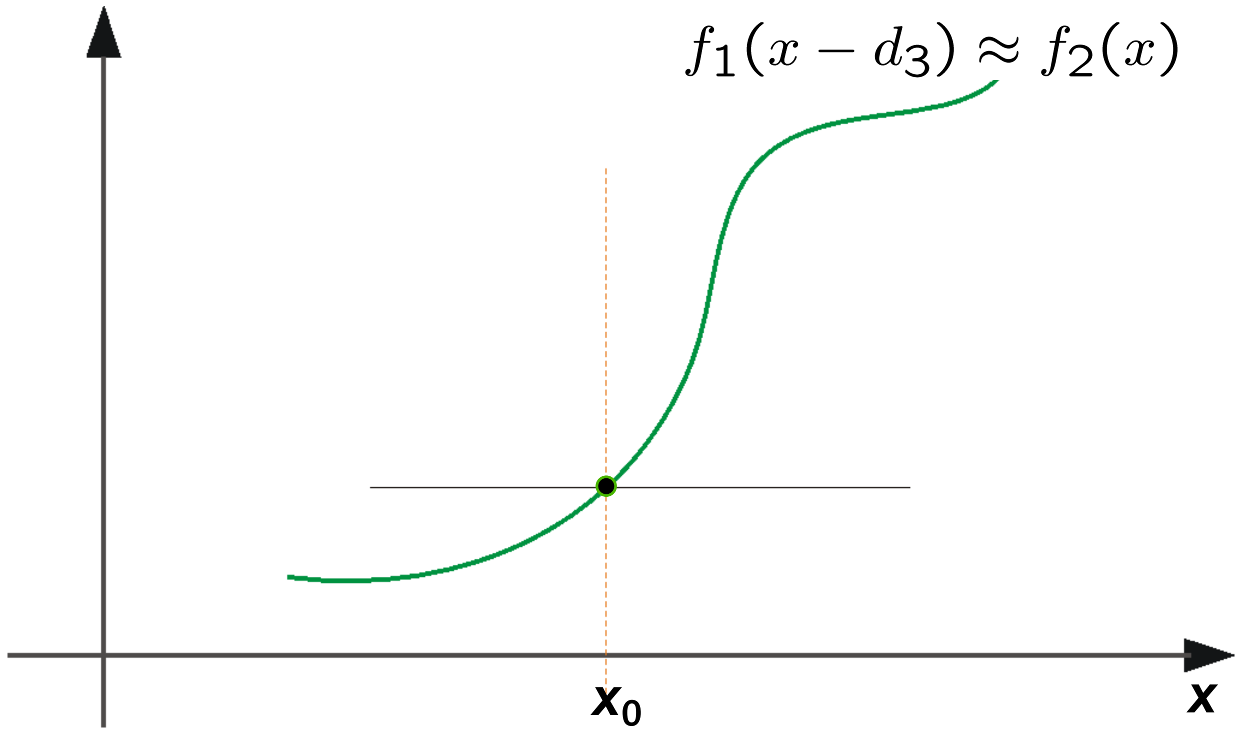
Optical Flow: Iterative Estimation



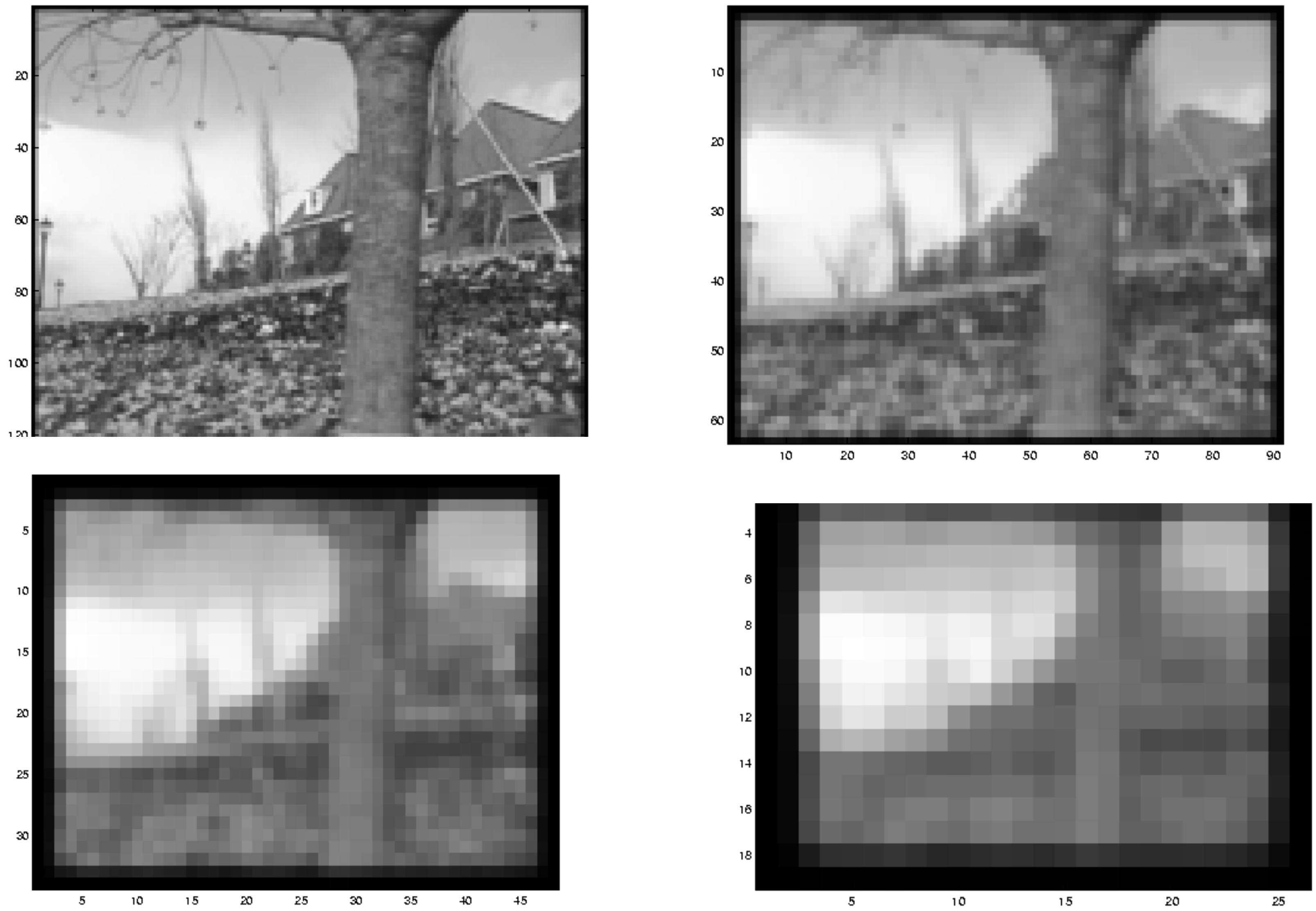
Initial guess: d_2

Estimate: $d_3 = d_2 + \hat{d}$

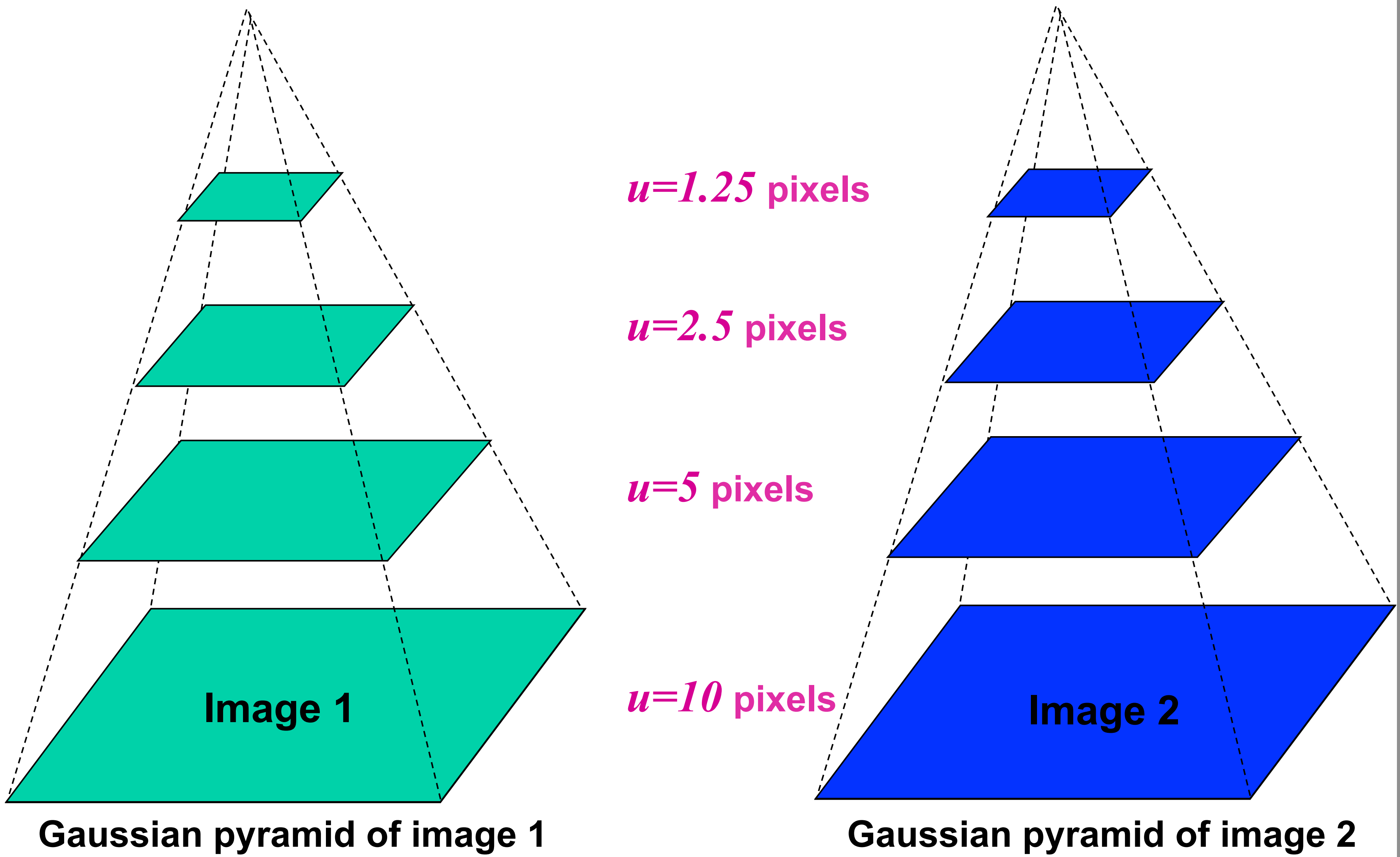
Optical Flow: Iterative Estimation



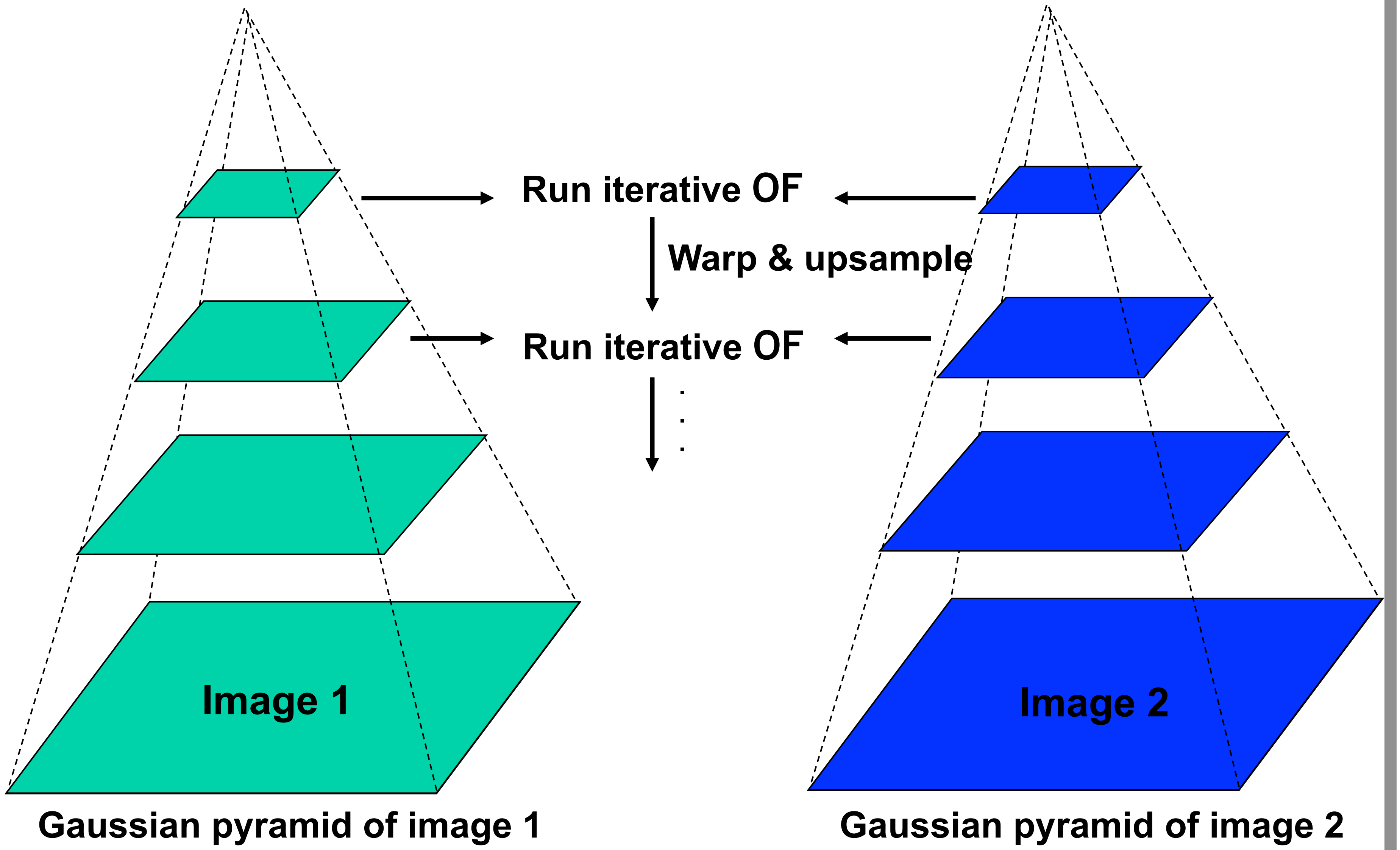
Large Displacements: Reduce Resolution!



Coarse-to-fine Optical Flow Estimation



Coarse-to-fine Optical Flow Estimation



Beyond the brightness constancy constraint

SIFT flow: dense correspondence across different scenes, C Liu, J Yuen, A Torralba, J. Sivic, W. Freeman, ECCV 2008

T. Brox, J. Malik, Large displacement optical flow: descriptor matching in variational motion estimation, PAMI 2011