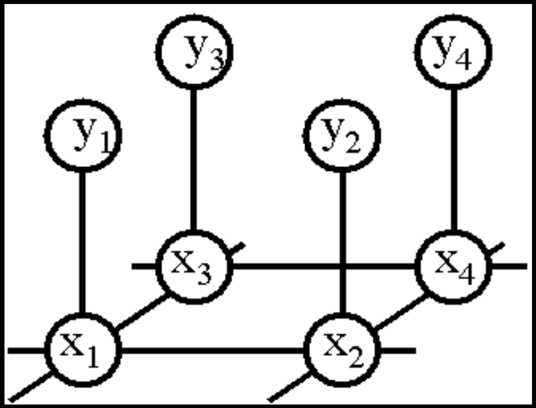
Machine Learning for Computer Vision

24 October 2013
MVA – ENS Cachan



Lecture 8: Markov Random Fields

Iasonas Kokkinos

Iasonas.kokkinos@ecp.fr

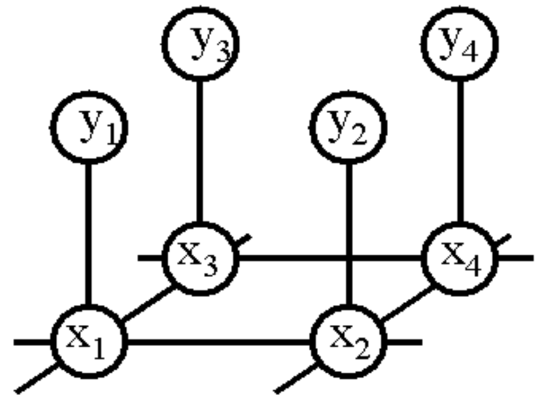
Center for Visual Computing
Ecole Centrale Paris

Galen Group
INRIA-Saclay

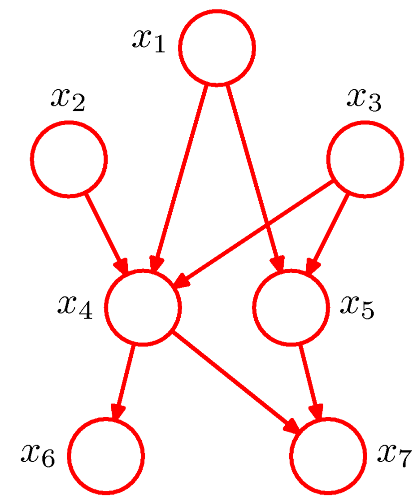
Lecture outline

Markov Random Fields

Internships & projects



Lecture 6: Bayesian Networks



Directed Acyclic Graph $G(V, E)$

Nodes - V : random variables

Edges - E : dependencies

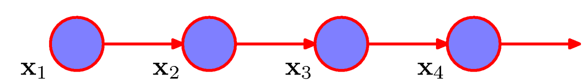
Leave from `parents', arrive at children

Distribution of each child conditioned on parent: $P(X_v | X_{\pi_v})$

$$P(X_v | \emptyset) = P(X_v)$$

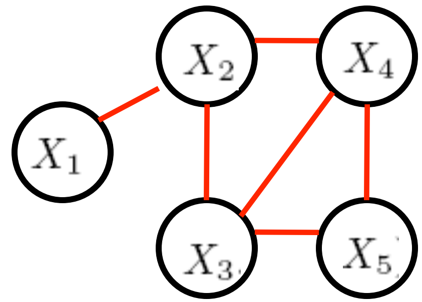
Joint probability distribution: $P(X) = \prod_v P(X_v | X_{\pi_v})$

Bayesian network for Brownian motion:



Markov Random Fields

- Undirected Graph $G(V, E)$



- Maximal graph cliques (fully-connected subsets)

$$\mathcal{C} = \{(X_1, X_2), (X_2, X_3), (X_3, X_4, X_5)\}$$

- Clique potentials $\psi_c(x_c) > 0, \forall x_c.$

$$p(x_1, \ldots, x_N) = \frac{1}{Z} \prod_{c \in \mathcal{C}} f_c\Big(\{x_j | j \in c\}\Big) = \frac{1}{Z} \prod_{c \in C} f_c(\mathbf{x}_c)$$

normalization term
(partition function)

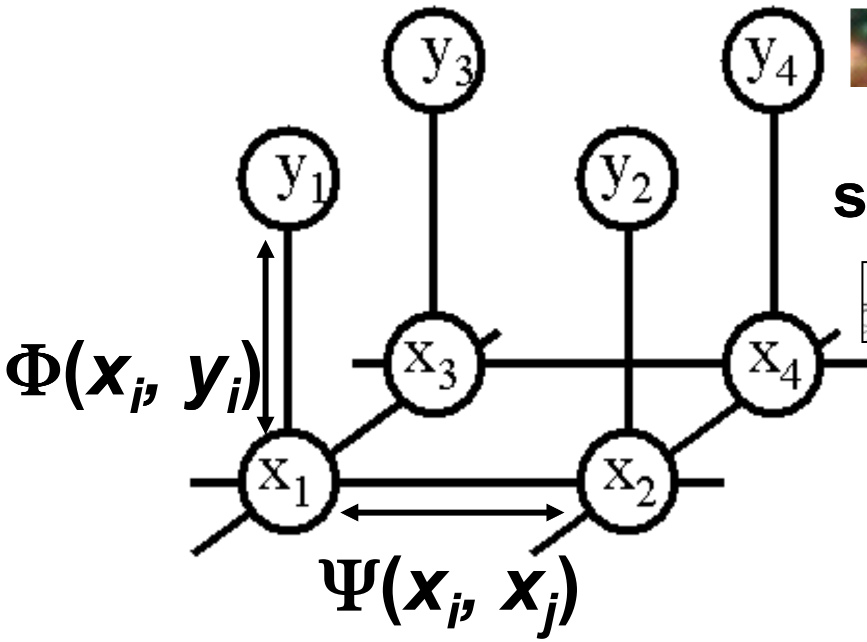
cliques of the graph

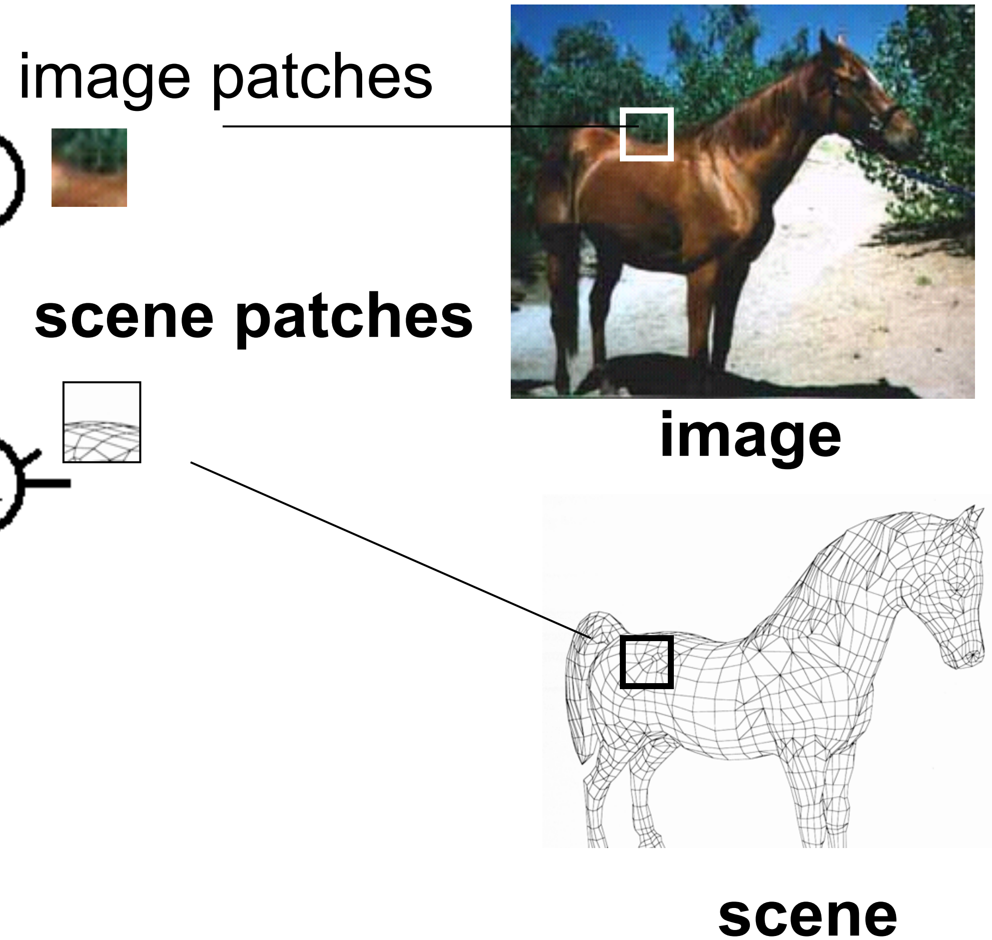
factors or potentials
(non-negative, but
unnormalized)

variables
in clique c

MRFs for Vision

$$P(X,Y) \;=\; \frac{1}{Z}\prod_i \Phi(Y_i, X_i) \prod_{(i,j)\in\mathcal{C}} \Psi(X_i, X_j)$$





$$P(X|Y) \;=\; ?$$

Network Joint Probability

$$P(X,Y) = \frac{1}{Z} \prod_i \Phi(Y_i, X_i) \prod_{i,j \in \mathcal{E}} \Psi\underbrace{(X_i, X_j)}_{\text{Neighboring scene nodes}}$$

Scene

Image

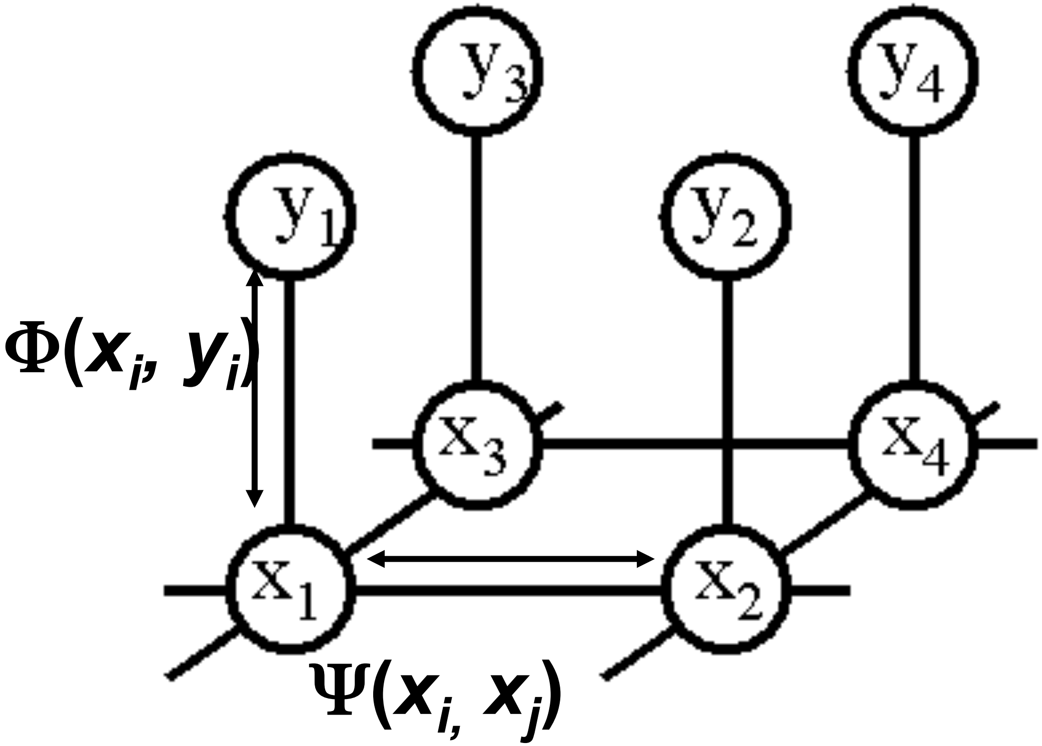
Image-scene compatibility function

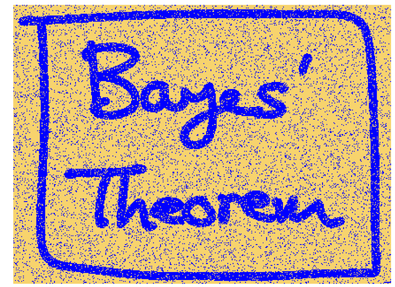
Local observations

Scene-scene compatibility function

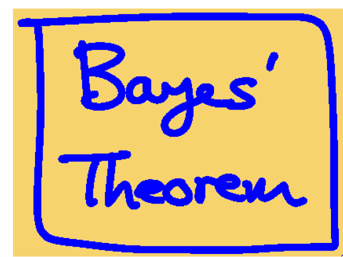
Neighboring scene nodes

MRFs for Denoising



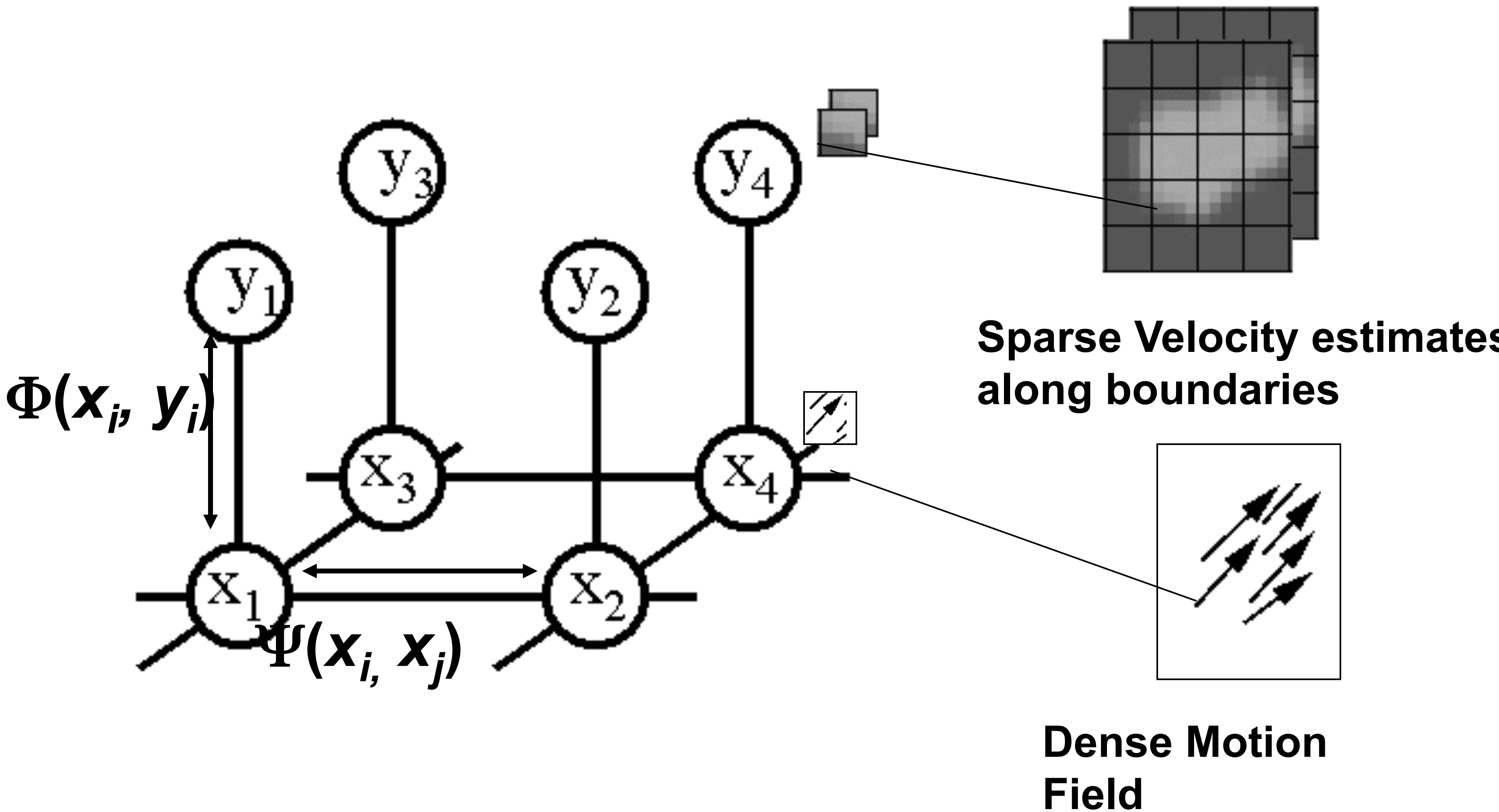


Noisy Pixel Intensities

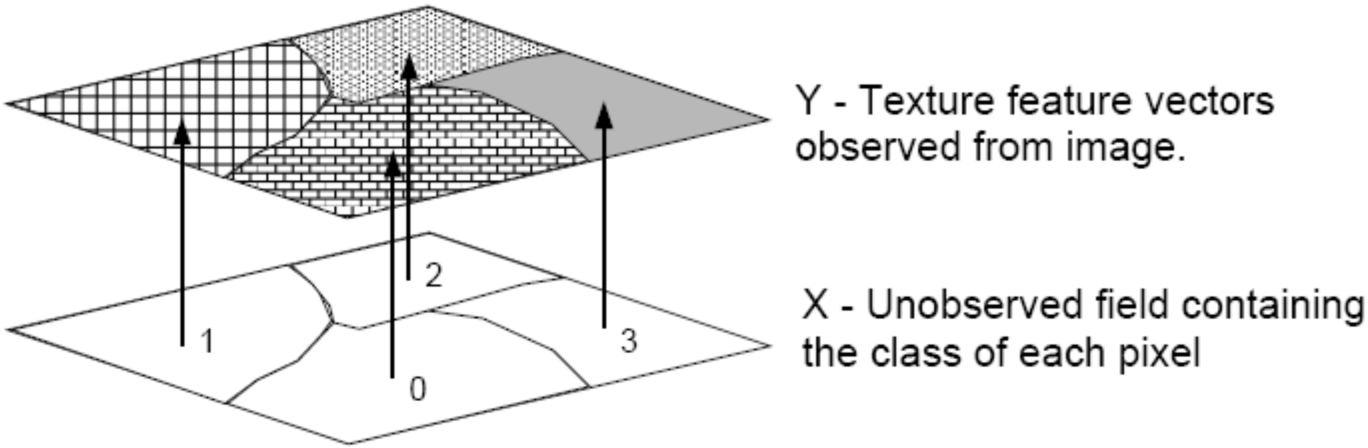


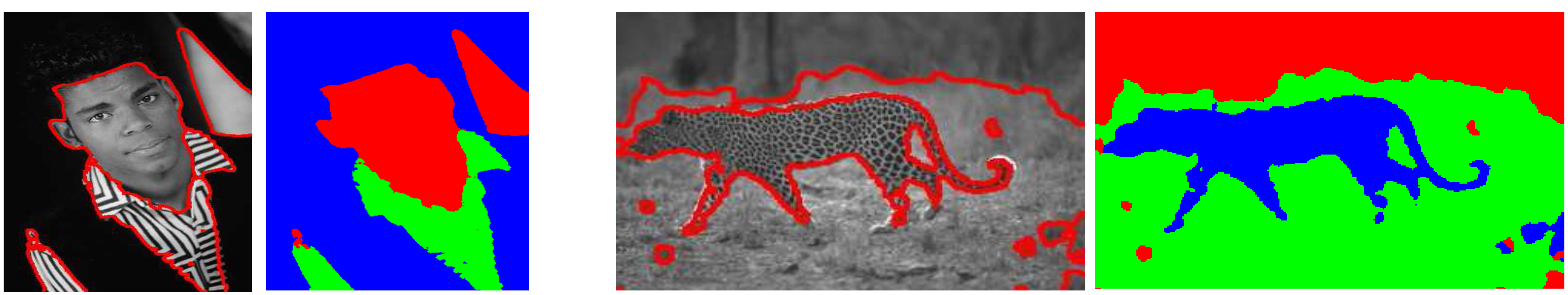
Clean Image

MRFs for Motion



MRFs for Segmentation





MRFs for Action Recognition (What, where, who)

Joint work with

Michalis Raptis
Disney Research @ CMU

Stefano Soatto
UCLA

M. Raptis, I. Kokkinos, S. Soatto, Discovering Discriminative Action Parts from Mid-Level Video Representations, CVPR 2012

Action Recognition

Analyzing Videos

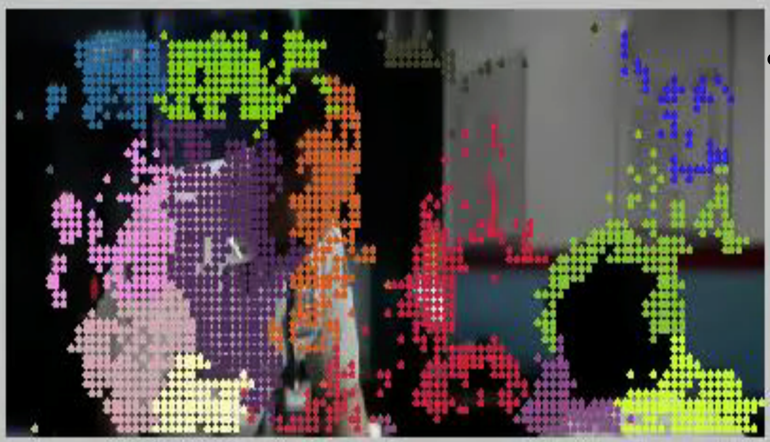
Microsoft Kinect: Dance Central

Virtual Character Animation

Surveillance

Discovery of Salient Regions

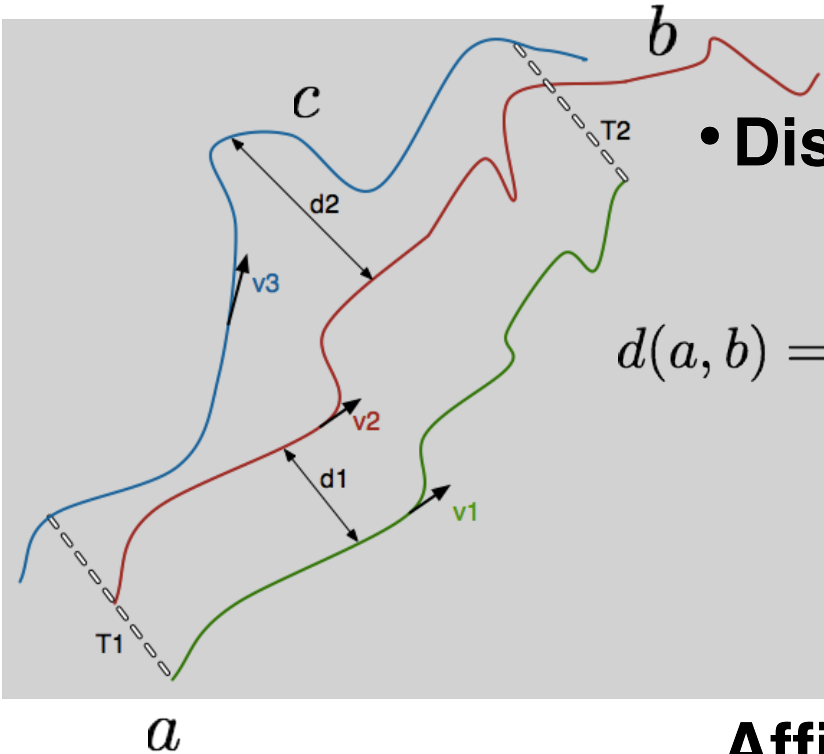
Data-Driven Segmentation of Spatio-Temporal Regions



- Object Parts Segmentation:
  - Motion as cue
  - Long temporal relations of trajectories

(Brox and Malik, ECCV' 10)

Clustering of Trajectories



- **Distance between trajectories:**

$$d(a,b) = \boxed{\max_{t\in[\tau_1,\tau_2]} d_{\text{spatial}}[t]} \cdot \boxed{\frac{1}{\tau_2-\tau_1}\sum_{t=\tau_1}^{\tau_2} d_{\text{velocity}}[t]}$$

Spatial Distance

Average Velocity Difference

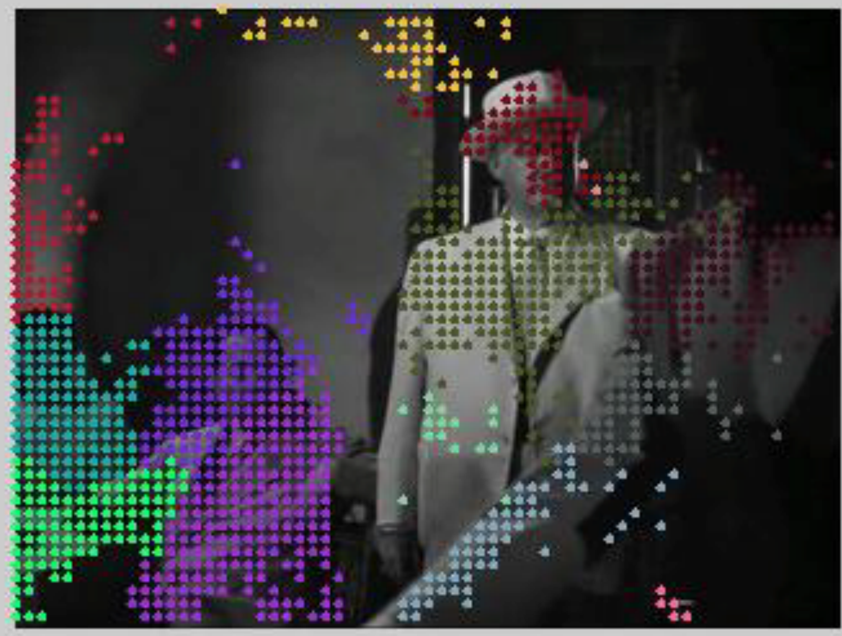
Affinity Matrix: $w(a,b) = \exp(-d(a,b))$

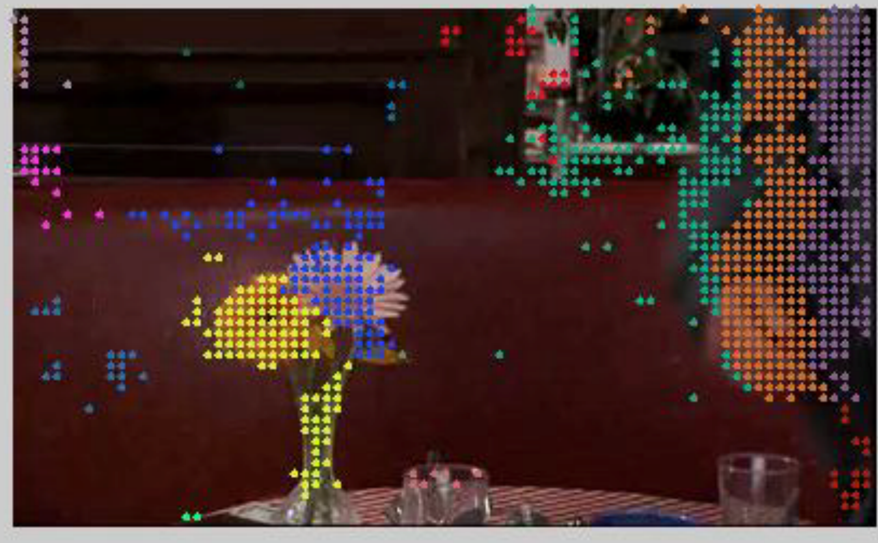
Greedy agglomerative clustering

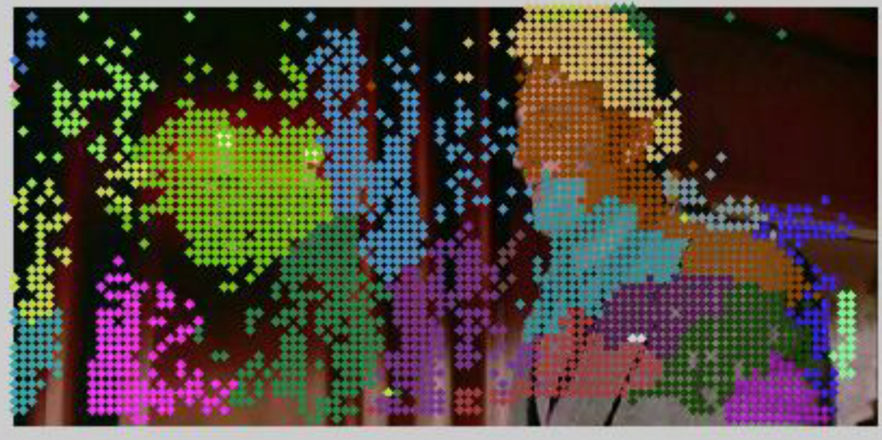
↓

Trajectory Groups: $\{G_k\},\ k = 1,\ldots,N$

(Brox and Malik, ECCV' 10)



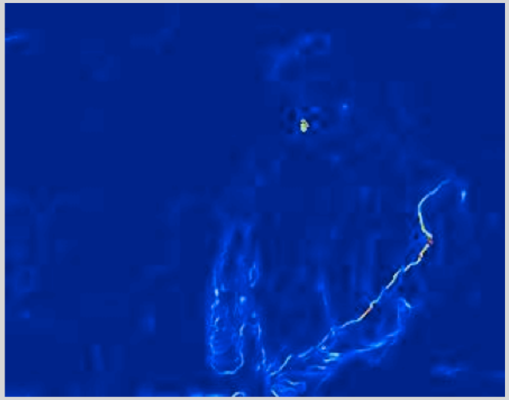


Trajectory Group Descriptor

- Densely Sampled :
  - Histogram of Gradients
  - Histogram of Optical Flow
  - Histogram of Oriented Edges of Motion Boundaries

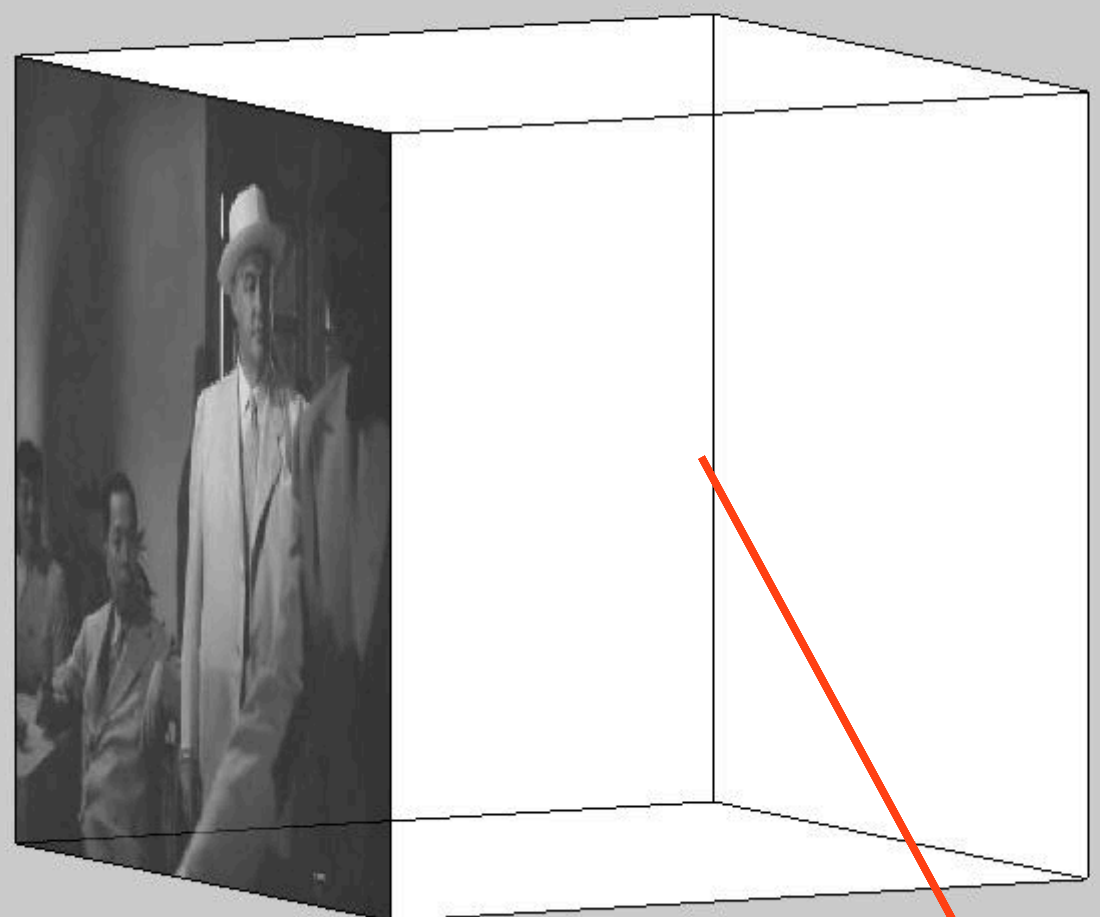
Optical Flow



Motion Boundaries

- Quantization step of all descriptors
- Bag of Features :
  - Accumulate the labels of descriptors in the neighborhood of a trajectory of the group

Trajectory Group Descriptor



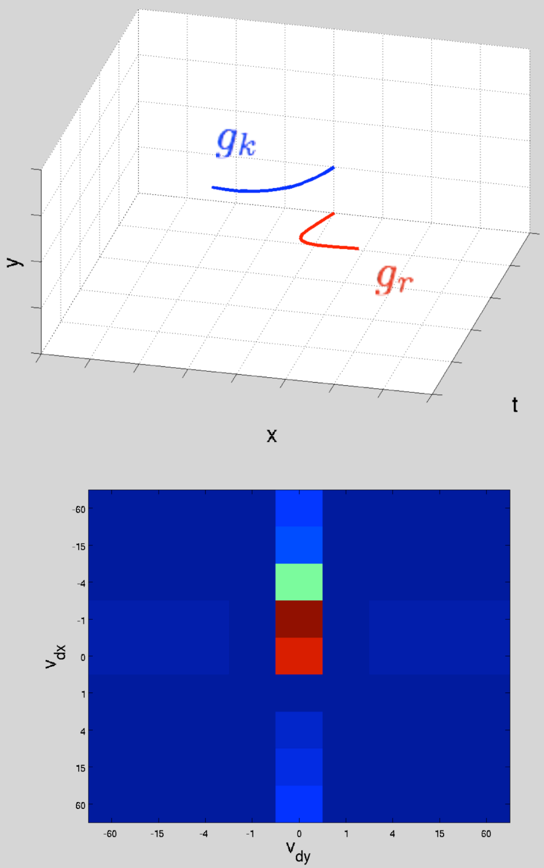
- Histogram of dense descriptors: h_k
- 3D “average” trajectory: g_k

$$G_k = \{h_k, g_k\}$$

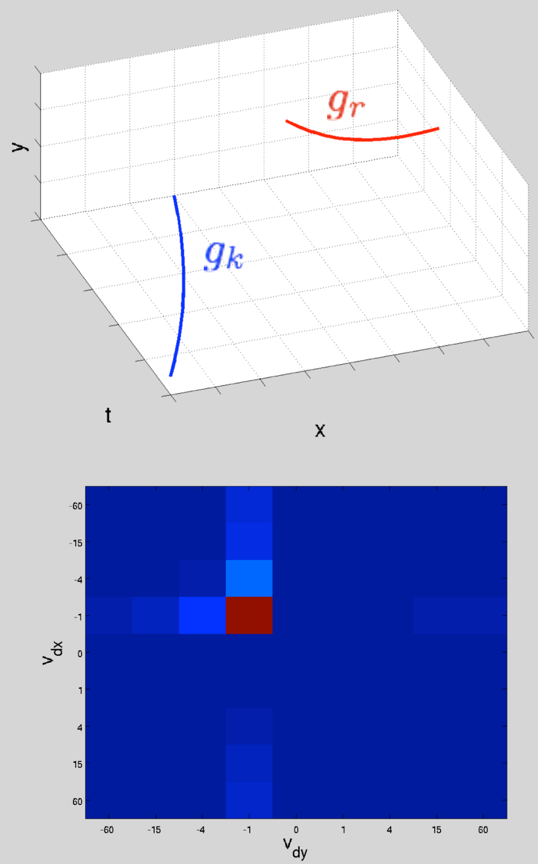
Pairwise Relationships

- **Characterize the change of the relative position of two "average" trajectories**

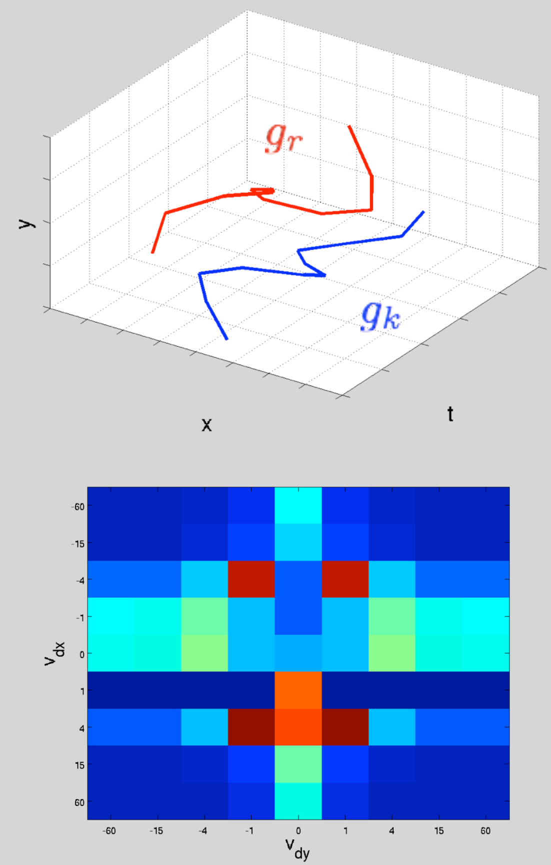
Converging in x dimension



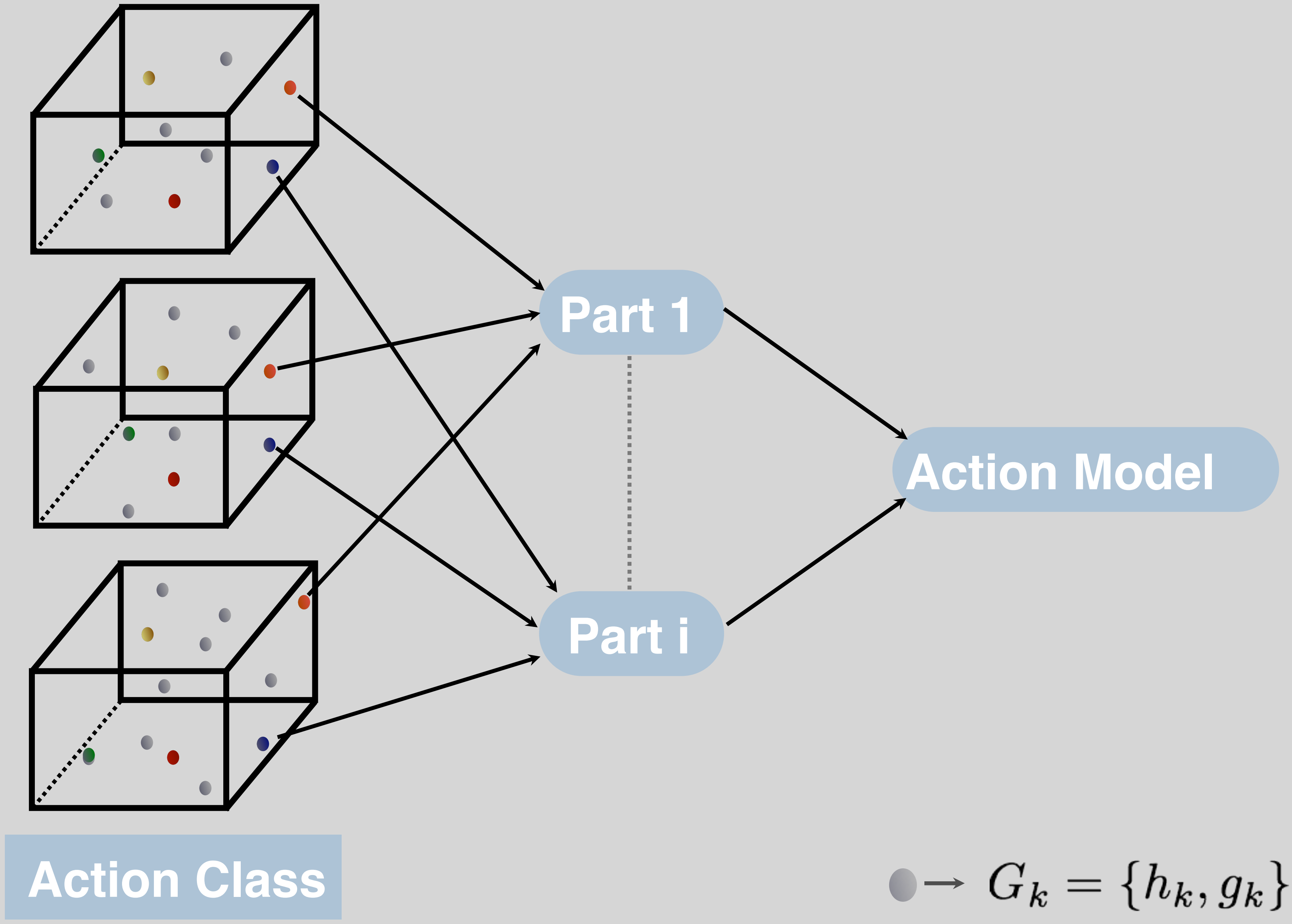
Slowing Converging in both dimensions



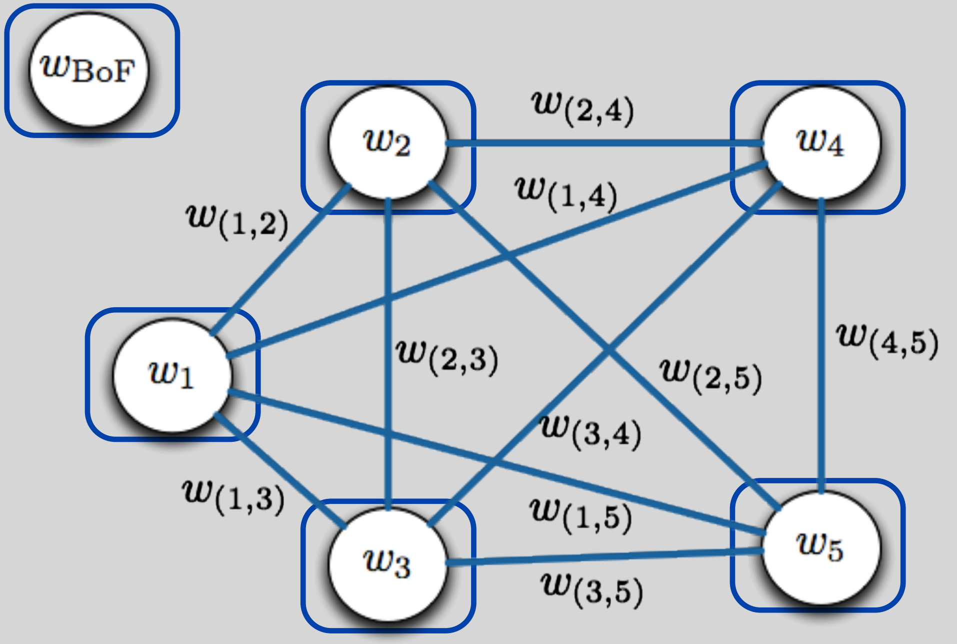
Converging and Diverging



Mid-level Representation



Action Model



- BoF Term
- Unary Terms
- Pairwise Terms

$$\text{score}(P,\sigma;\mathbf{w}) = \boxed{\langle w_0, h_{\text{BoF}}\rangle} + \boxed{\sum_{i=1}^{5}\langle w_i, h_{p_i}\rangle} + \boxed{\sum_{i=1}^{5}\sum_{j=i+1}^{5}\langle w_{i,j}, \psi(g_{p_i}, g_{p_j}, \sigma)\rangle} = \langle \mathbf{w}, \Phi(x, P, \sigma)\rangle$$

Latent Variables: $P = \{p_1, p_2, p_3, p_4, p_5\}$, σ

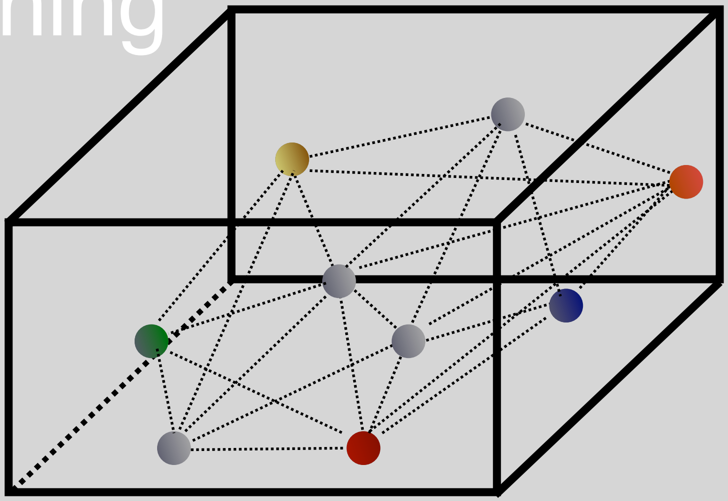
Cluster Association: $p_i = k \Rightarrow \text{ Part } i \rightarrow G_k = \{h_k, g_k\}$

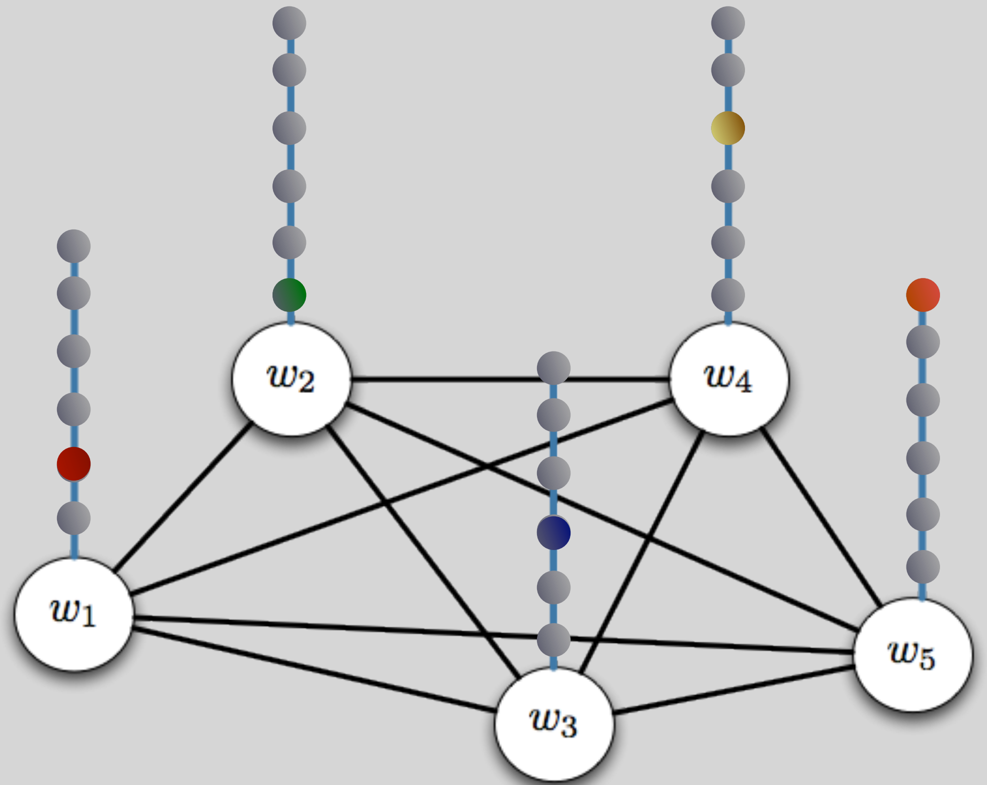
Set of Scales: $\sigma \in \{\sigma_1, \sigma_2, \ldots, \sigma_N\}$

Subgraph Matching

assifying a video:

$$P^* = \arg\max_P \text{score}(P; \mathbf{w}^*, \sigma_j)$$





Subgraph matching as an MRF labeling problem

Tree-reweighted message passing, Kolmogorov , PAMI 06)

Learning

Find $\mathbf{w}$ that leads to the maximum margin classification

Latent SVM (Felzenszwalb et al. 2008)

ernate between estimating P and $\mathbf{w}$

a. Given $\mathbf{w}$:

$$P_i^* = \arg\max_{P_i} \langle \mathbf{w}, \Phi(x_i, P_i, \sigma_j) \rangle$$

b. Given P_i^*:

$$\min_{\mathbf{w},\xi} \ \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_i \xi_i$$

$$\text{s.t.} \ \ y_i(\langle \mathbf{w}, \Phi(\mathbf{x}_i, P_i^*)\rangle + b) \geq 1 - \xi_i, \ \forall i$$

$$\xi_i \geq 0, \quad \forall i$$

Ranking SVM

Non-convex optimization -- CCCP: (Yuille and Rangarajan 2003)

Learning

Find $\mathbf{w}$ that leads to the maximum margin classification

Latent SVM (Felzenszwalb et al. 2008)

ernate between estimating and P $\mathbf{w}$

a. Given $\mathbf{w}$:

$$P_i^* = \arg\max_{P_i} \langle \mathbf{w}, \Phi(x_i, P_i, \sigma_j) \rangle$$

b. Given P_i^*:

$$\min_{\mathbf{w},\xi} \quad \frac{1}{2}\|\mathbf{w}\|^2 + C\sum_{i,j}\xi_{ij}$$

$$\text{s.t.} \quad \langle \mathbf{w}, \Phi(\mathbf{x}_i, P_i^*)\rangle - \langle \mathbf{w}, \Phi(\mathbf{x}_j, P_j^*)\rangle \geq \Delta_{0/1}(y_i, y_j) - \xi_{ij}, \ \forall i, j, y_j \in \mathcal{Y}\backslash\{y_i\}$$

$$\xi_{ij} \geq 0 \ \forall i, j$$

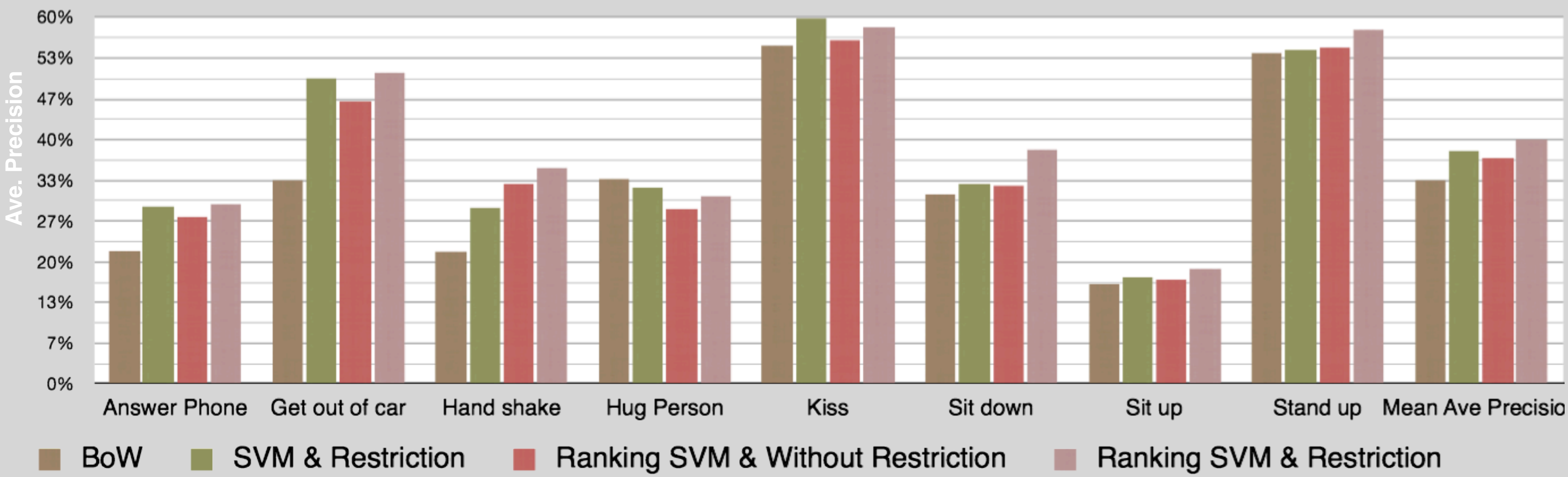
Ranking SVM

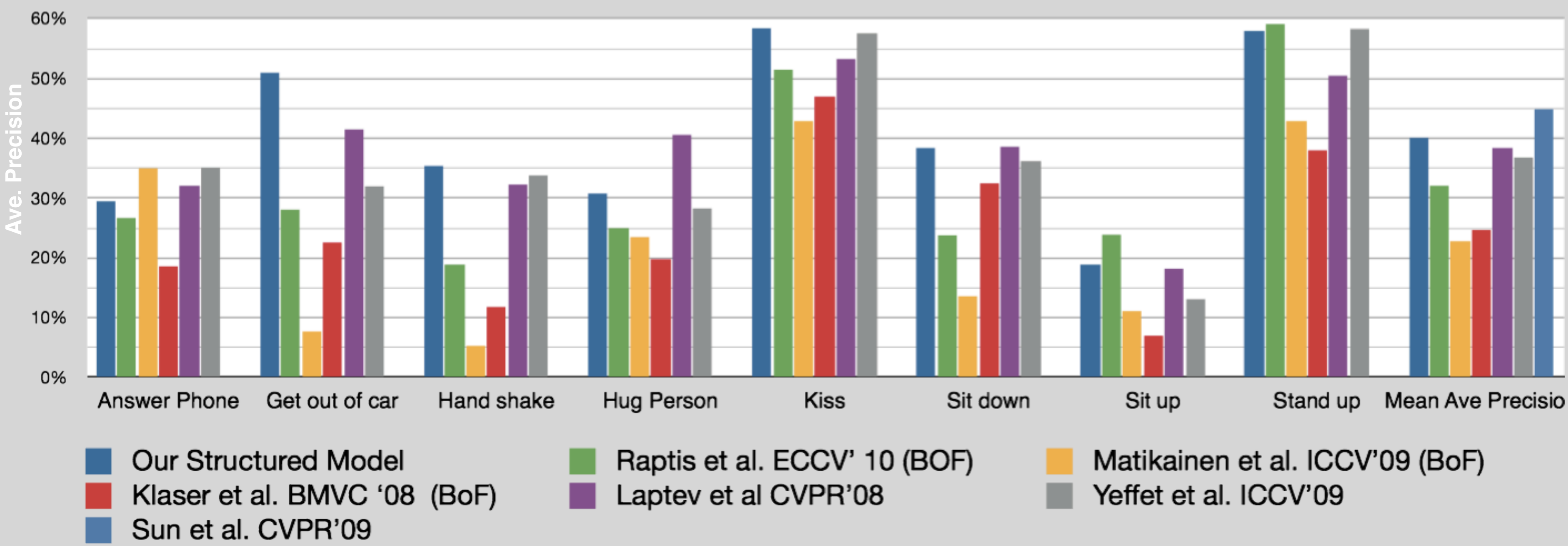
Non-convex optimization -- CCCP: (Yuille and Rangarajan 2003)

Initialization

- Weights of pairwise terms set to zero
- Unary terms weights:
    - Set equal to the center of a cluster produced by K-means on the descriptors of the positive training videos
- Weak Annotations with Bounding Boxes
    - Restrict the selections of parts to groups intersecting the bounding boxes

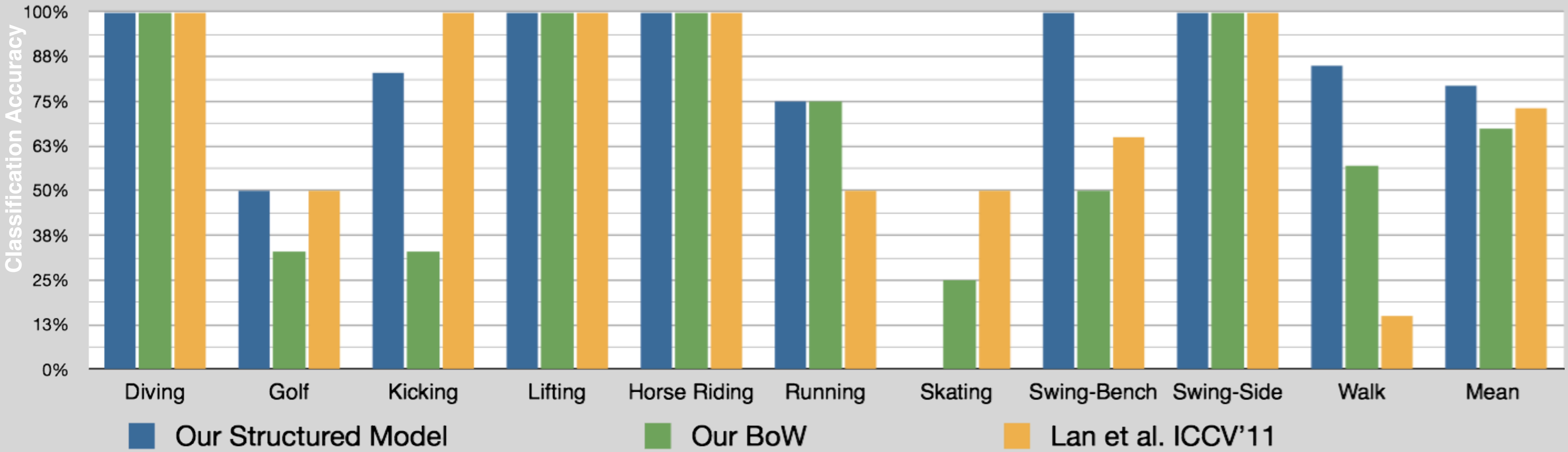
HOHA Dataset

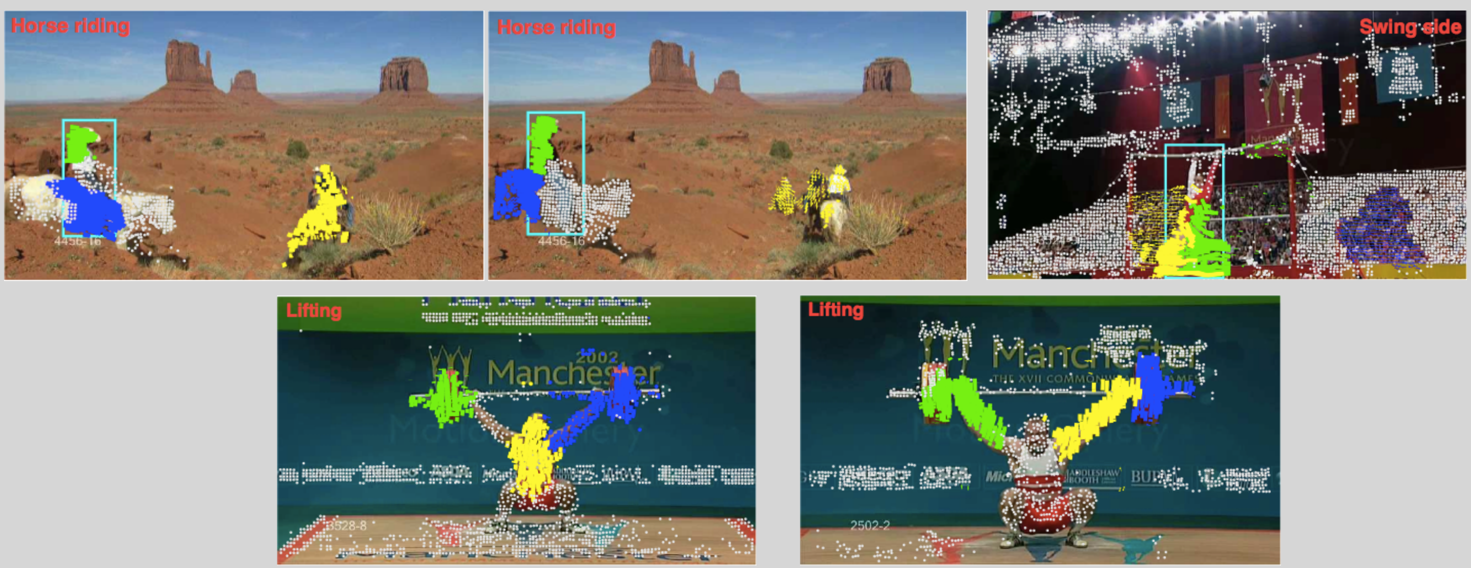




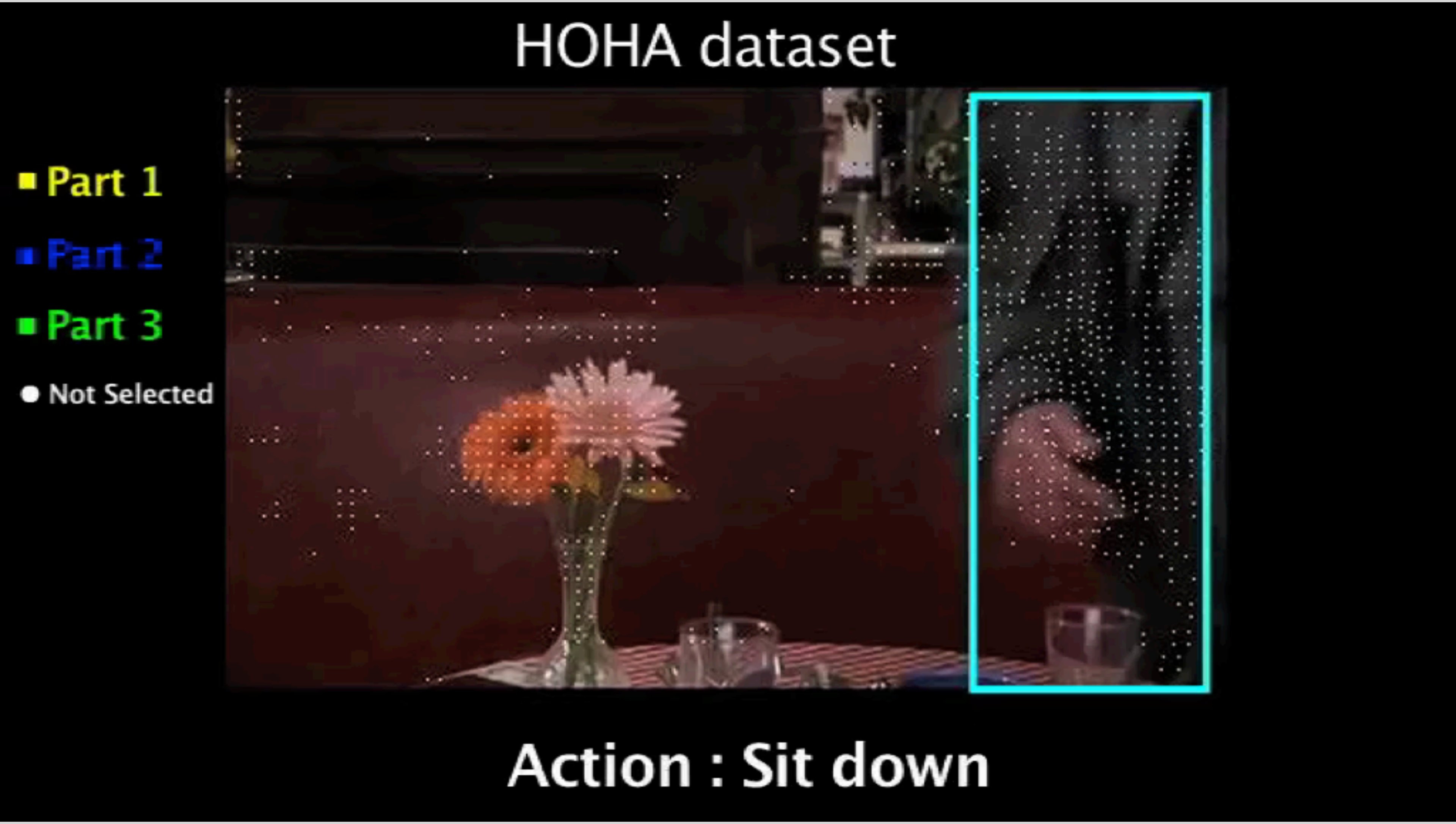
UCF Sports Dataset

Per-class Classification Accuracy



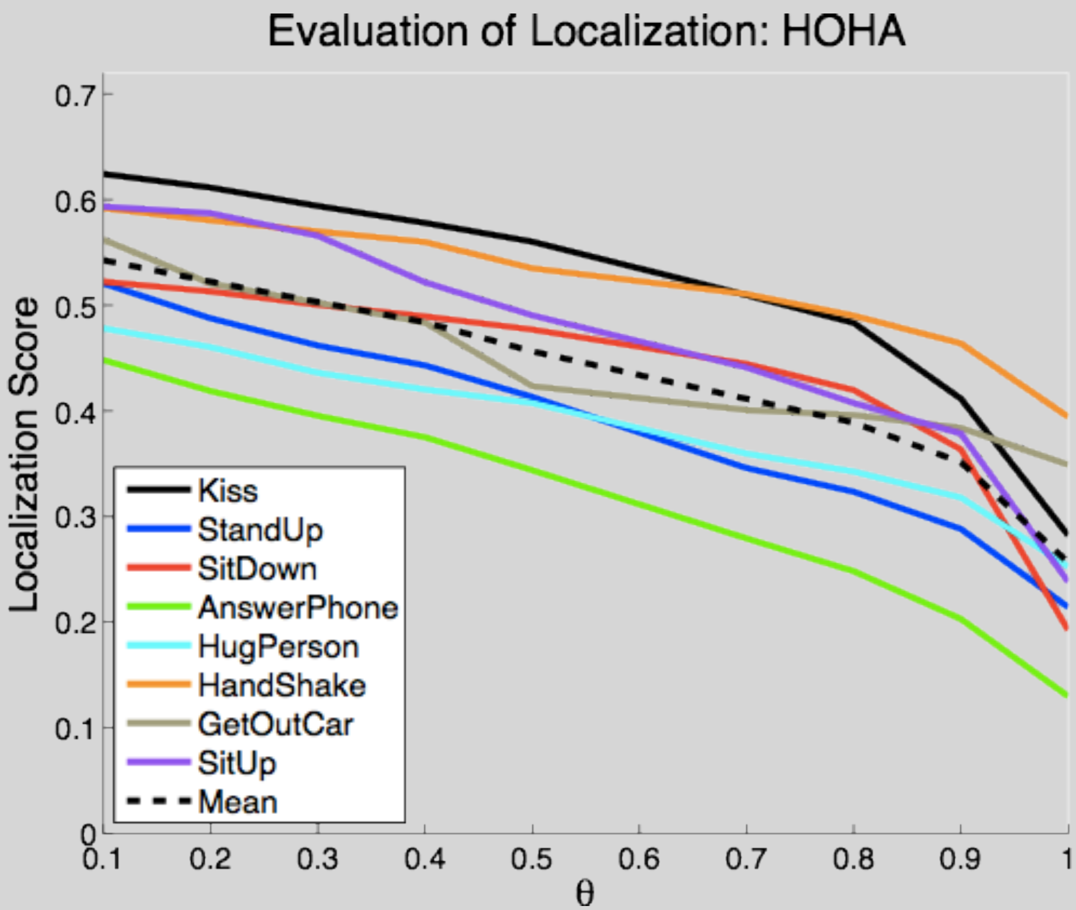


Test Examples



Colored Groups of Trajectories are associated with the model
White colored trajectories are not selected by the model

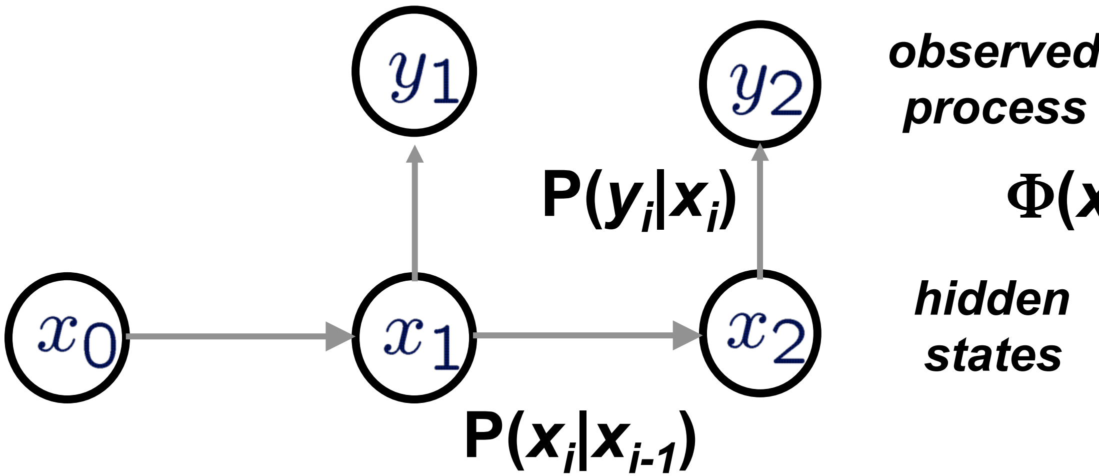
Localization Performance

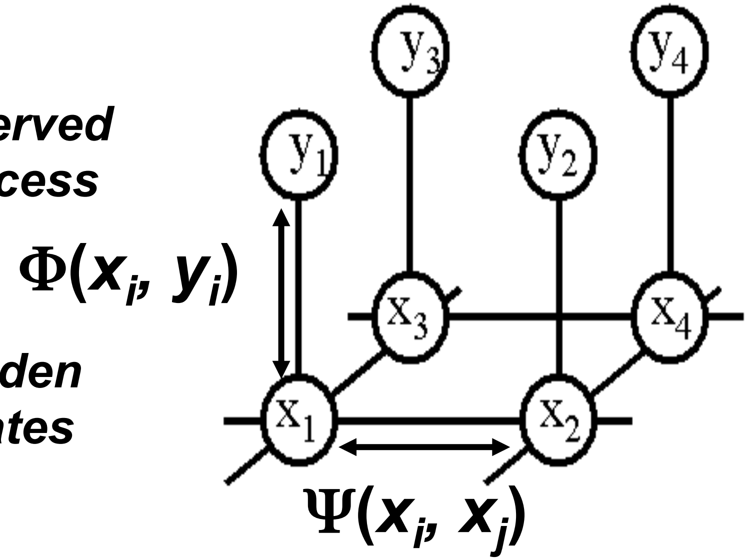


θ: Percentage of trajectories inside the bounding box

$$\text{Localization score} = \frac{1}{|V| \cdot T} \sum_{i=1}^{|V|} \sum_{t=1}^{T} [\frac{|D_{i,t} \cap L_t|}{|D_{i,t}|} \geq \theta]$$

HMMs vs MRFs





Inference: **Max/Sum Product (Polynomial)**

NP-hard

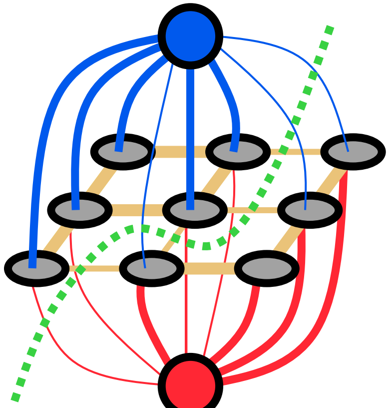
Lecture outline

Mutliple Instance Learning

Markov Random Fields

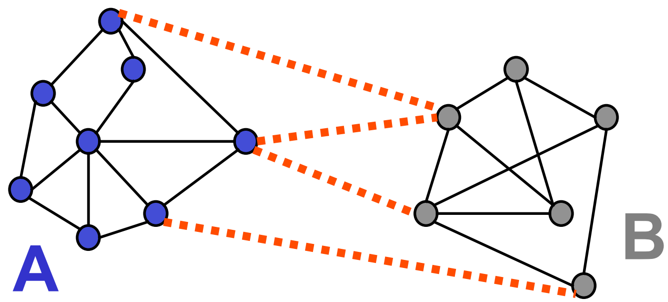
Graph Cuts

Dual Decomposition



Internships & projects

Graph Cut

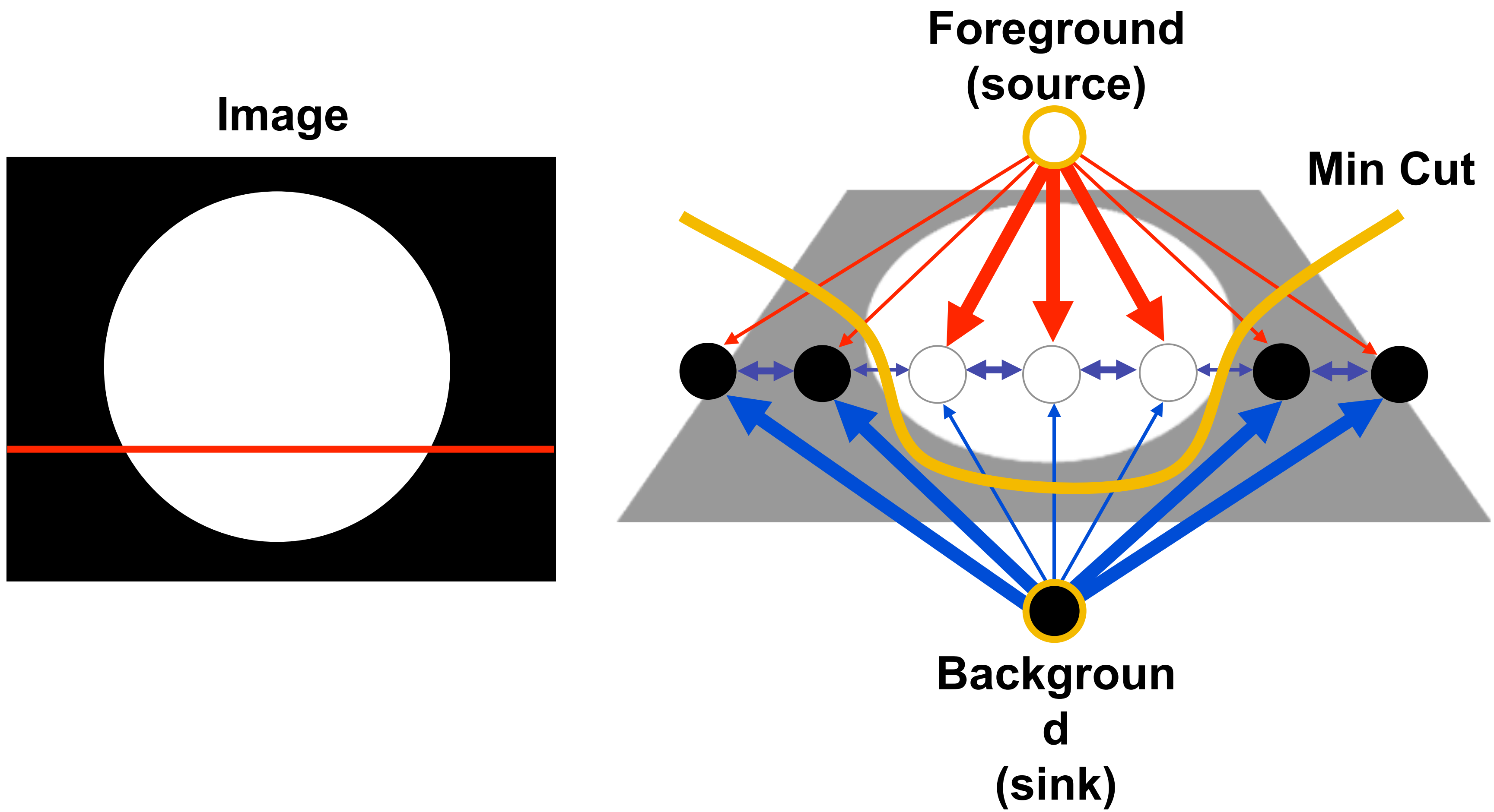


- Set of edges whose removal makes a graph disconnected
- Cost of a cut
  - Sum of weights of cut edges:

$$cut(A,B) = \sum_{p \in A, q \in B} w_{p,q}$$

- A graph cut gives us a segmentation
  - What is a “good” graph cut and how do we find one?

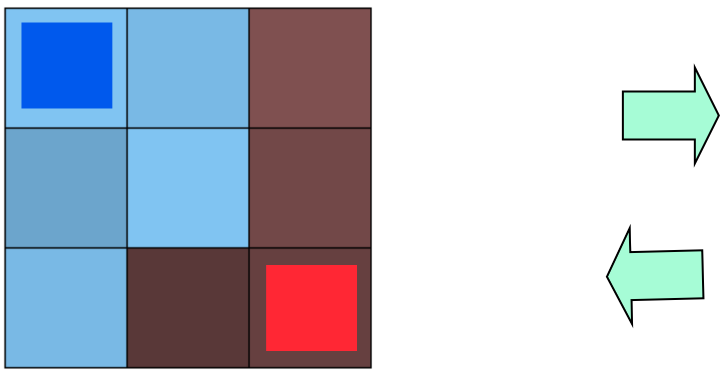
Graph Cuts for Image Segmentation

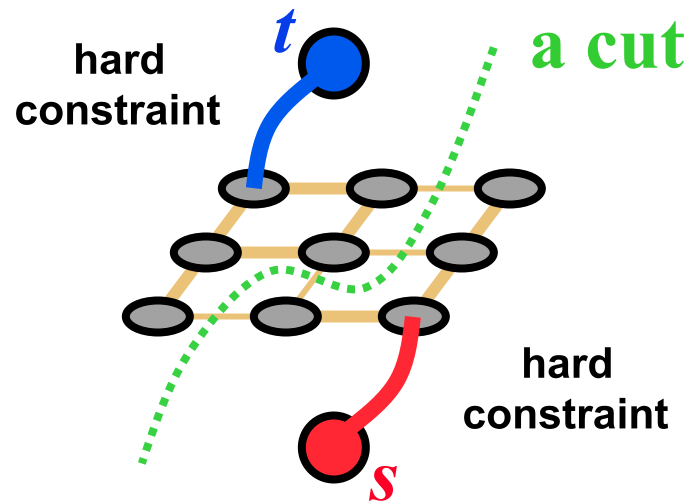


- Segmentation by s-t mincut
  - *Cut:* separating source and sink
  - *Min Cut:* Global minimal energy in polynomial time (1MPixel/sec)

Graph Cuts for Optimal Boundary Detection

- Idea: convert MRF into source-sink graph





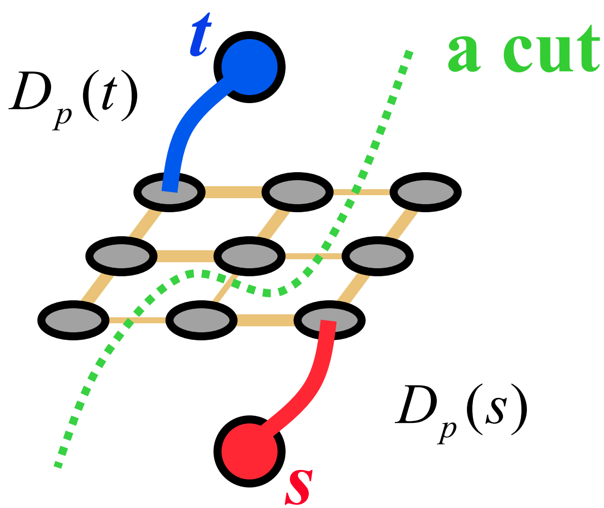
Minimum cost cut can be computed in polynomial time

(max-flow/min-cut algorithms)

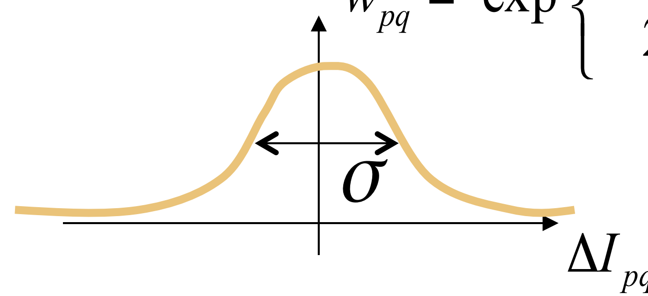
Simple Example of Energy

Regional term **Boundary term**

$$E(L) = \underbrace{\sum_{p} D_p(L_p)}_{\text{t-links}} + \underbrace{\sum_{pq \in N} w_{pq} \cdot \delta(L_p \neq L_q)}_{\text{n-links}}$$



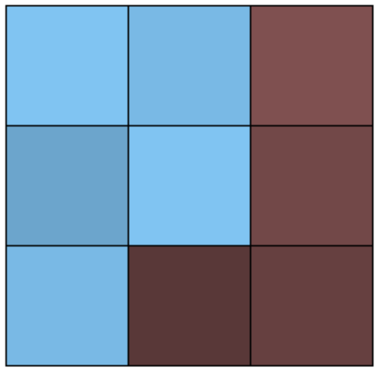
$$w_{pq} = \exp\left\{-\frac{\Delta I_{pq}}{2\sigma^2}\right\}$$

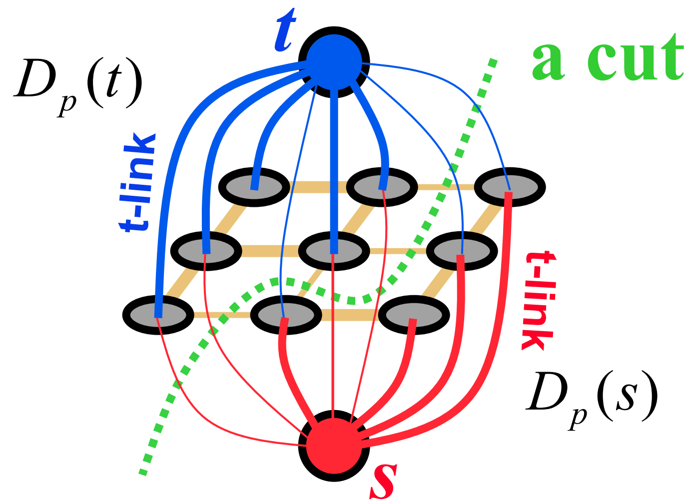


$$L_p \in \{s, t\}$$

(binary object segmentation)

Adding Regional Properties





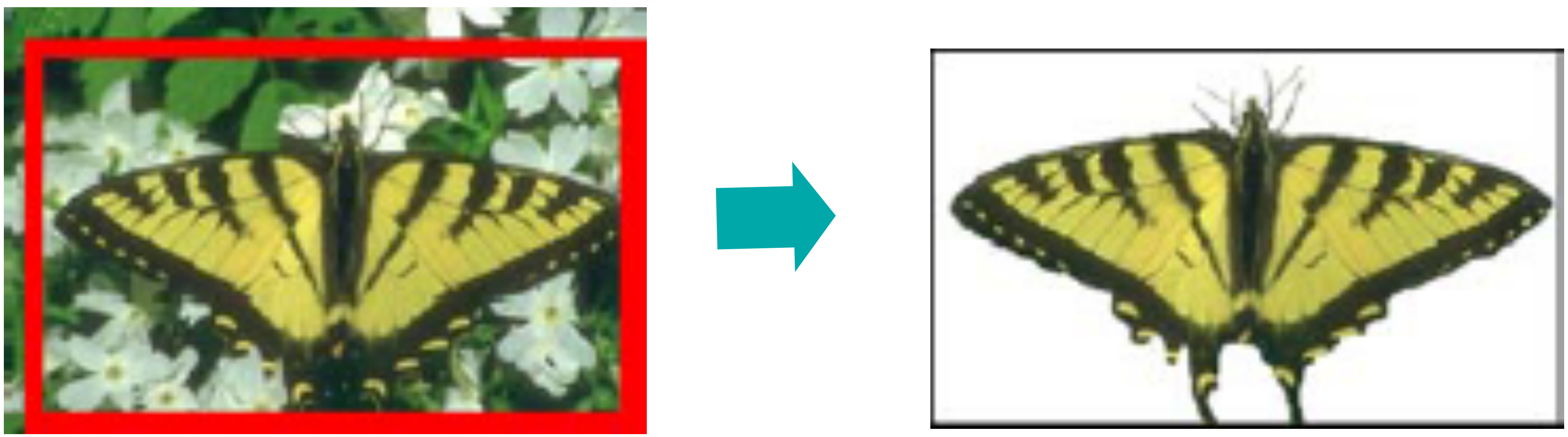
"expected" intensities of
object and background
I^s and I^t
can be re-estimated

$$D_p(s) \propto \exp\left(-\| I_p - I^s \|^2 / 2\sigma^2\right)$$
$$D_p(t) \propto \exp\left(-\| I_p - I^t \|^2 / 2\sigma^2\right)$$

EM-style optimization

Iterative learning of regional color-models

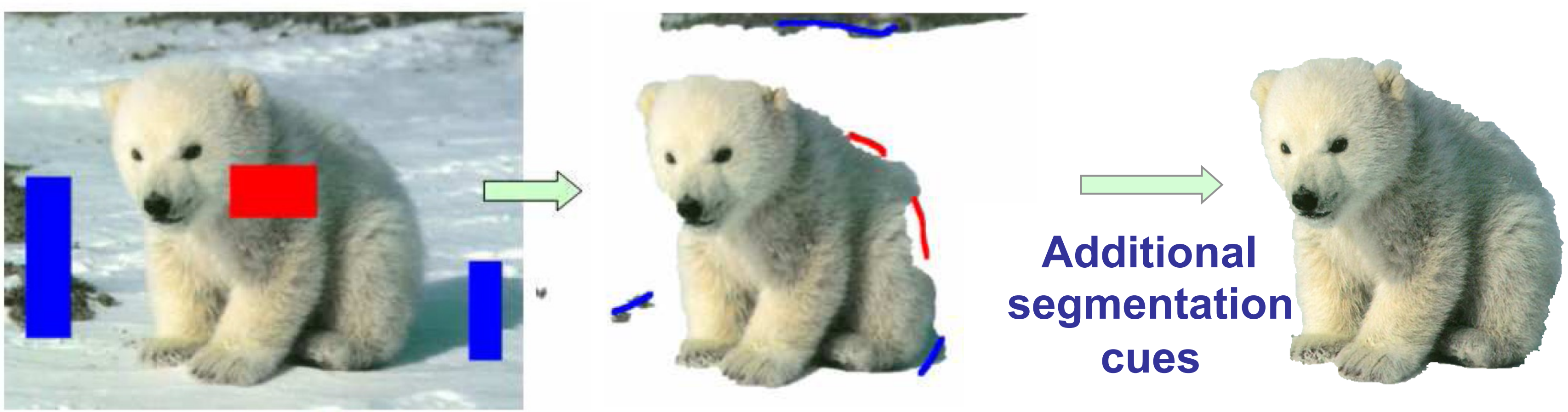
- GMMRF cuts (Blake et al., ECCV04)
  Region competition with Gaussian mixture distributions
- Grab-cut (Rother et al., SIGGRAPH 04)



parametric regional model – Gaussian Mixture (GM)
designed to guarantee convergence

GraphCut Applications: “GrabCut”

- Procedure
  - User marks foreground and background regions with a brush.
  - This is used to create an initial segmentation
    which can then be corrected by additional brush strokes.

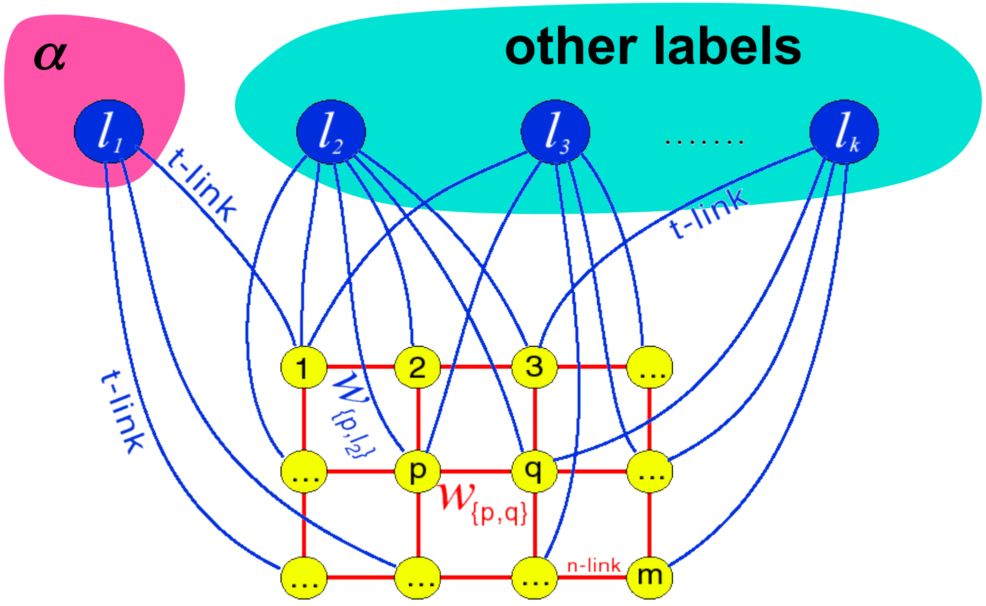


User segmentation cues

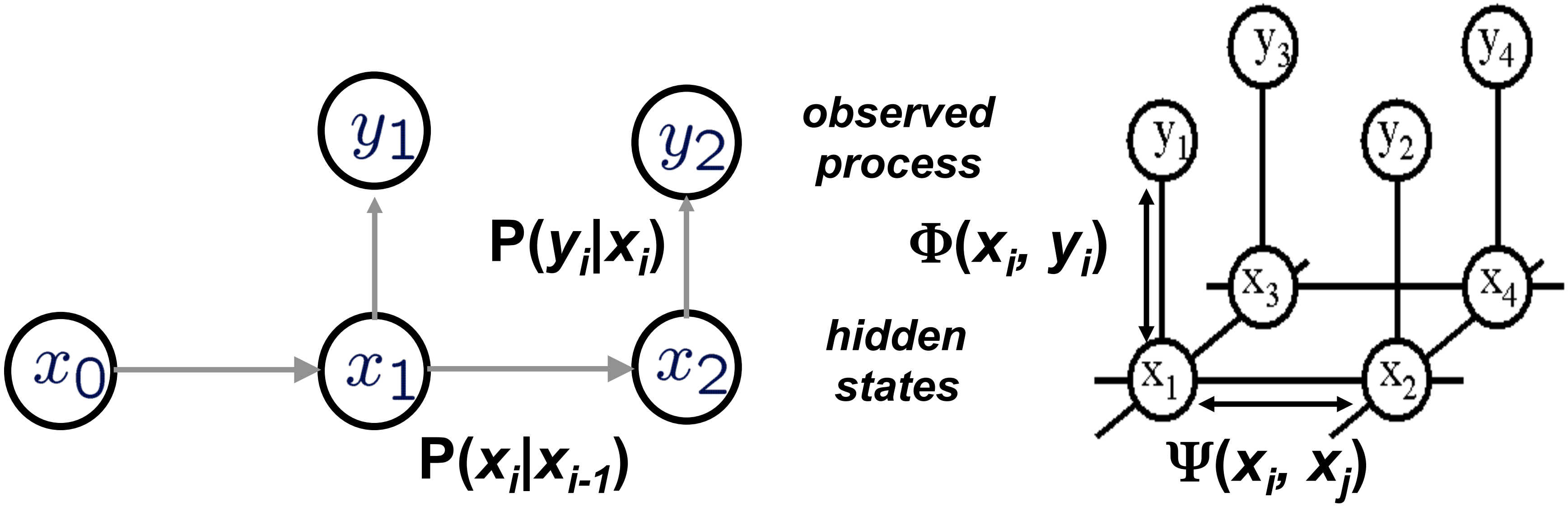
GrabCut: Example Results

More labels: α-Expansion

- Basic idea: k vs rest
  - Break multi-way cut computation into a sequence of binary s-t cuts.



HMMs vs MRFs



Inference: Max/Sum Product (Polynomial) **NP-hard**

alpha-expansion
Dual decomposition
Gibbs sampling
Loopy Belief Propagation
TRW-s
...

Parameter Estimation: EM

Lecture outline

Mutliple Instance Learning

Markov Random Fields

- Graph Cuts
- Dual Decomposition

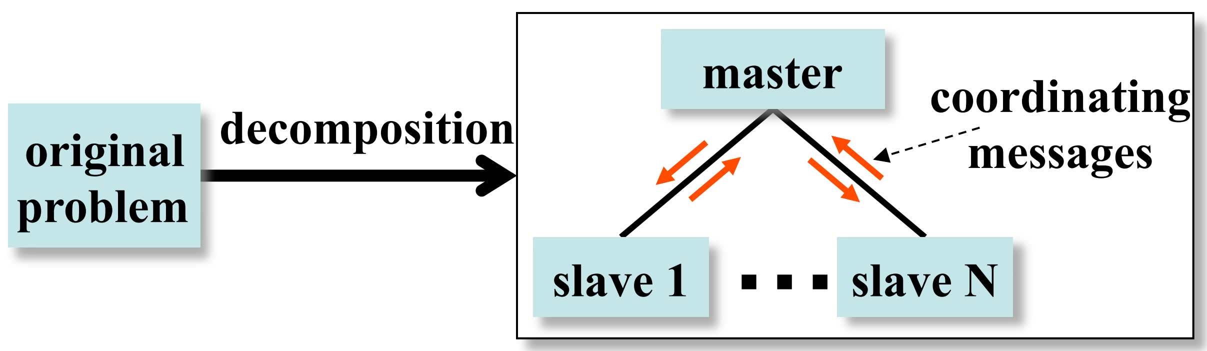
Internships & projects

Slide Credits: N. Komodakis

MRF optimization via dual-decomposition

Decompose hard optimization problem into easy subproblems (slaves)

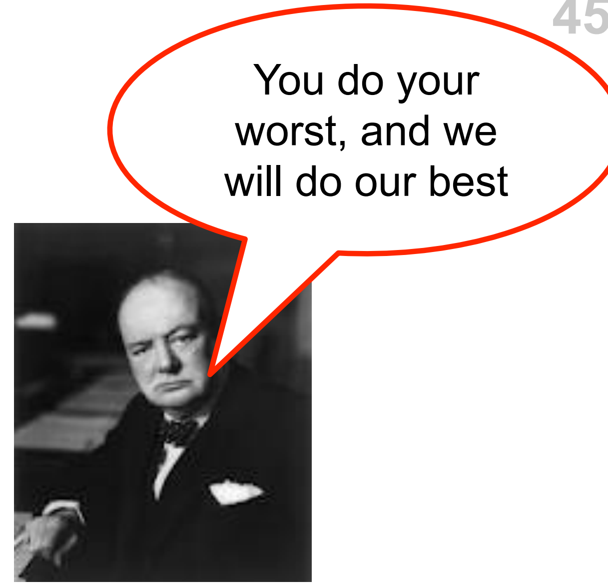
Master: extract a solution by combining subproblem solutions



- Stronger theoretical properties than state-of-the-art
- New insights into existing message-passing techniques

N. Komodakis and N. Paragios, ICCV 2007

Lagrangian reminder (SVM class)



- Constrained optimization problem:

$$\begin{aligned} \min_{w} \quad & f(w) \\ s.t. \quad & h_i(w) = 0, \quad i = 1 \ldots l \\ & g_i(w) \leq 0, \quad i = 1 \ldots m \end{aligned}$$

- Equivalent: $\min_{w} \quad f_{uc}(w) = f(w) + \sum_{i=1}^{l} I_0(h_i(w)) + \sum_{i=1}^{m} I_+(g_i(w))$

$$I_0(x) = \begin{cases} 0, & x = 0 \\ \infty, & x \neq 0 \end{cases}, \qquad I_+(x) = \begin{cases} 0, & x \leq 0 \\ \infty, & x > 0 \end{cases}$$

- \`Soften' constraints: $L(w, \lambda, \mu) = f(w) + \sum_{i=1}^{l} \lambda_i h_i(w) + \sum_{i=1}^{m} \mu_i g_i(w), \quad \mu_i > 0 \forall i$

- Solve: $f(w^*) = \min_{w} \max_{\lambda, \mu : \mu_i > 0} L(w, \lambda, \mu)$

Dual Decomposition technique

- Optimization problem: $$\begin{array}{ll} \min_x & \sum_i f^i(x) \\ \text{s.t.} & x \in \mathcal{C} \end{array}$$

Easy to optimize summands, hard to optimize sum

Equivalent problem:

$$\begin{array}{ll} \min_{\{x^i\},x} & \sum_i f^i(x^i) \\ \text{s.t.} & x^i \in \mathcal{C}, \boxed{x^i = x} \end{array}$$

Lagrange multipliers:

$$g(\{\lambda^i\}) = \min_{\{x^i \in \mathcal{C}\},x} \sum_i f^i(x^i) + \sum_i \lambda^i \cdot (x^i - x)$$

Dual:

$$\max_{\{\lambda^i\} \in \Lambda} \; g(\{\lambda^i\}) = \sum_i g^i(\lambda^i)$$

$$g^i(\lambda^i) = \min_{x^i \in \mathcal{C}} f^i(x^i) + \lambda^i \cdot x^i$$

$$\{\lambda^i\} \in \Lambda = \{\{\lambda^i\} \mid \sum_i \lambda^i = 0\}$$

or else $g(\{\lambda^i\}) = -\infty$

Dual Decomposition

- i-th slave problem $g^i(\lambda^i) = \min_{x^i \in \mathcal{C}} f^i(x^i) + \lambda^i \cdot x^i$

  $\bar{x}^i(\lambda^i) \equiv$ minimizer of i-th slave problem for given λ^i

- Master problem: $\max_{\{\lambda^i\} \in \Lambda} \; g(\{\lambda^i\}) = \sum_i g^i(\lambda^i)$

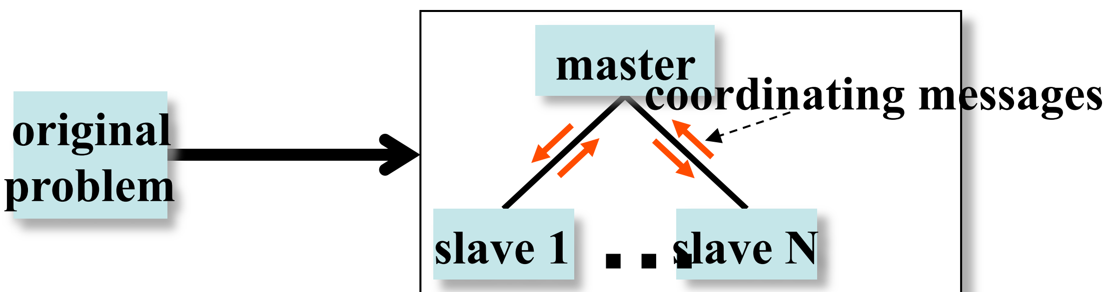
  Projected subgradient:

  $\lambda^i \leftarrow \left[\lambda^i + \alpha_t \nabla g^i(\lambda^i)\right]_\Lambda$

  $\nabla \equiv$ subgradient w.r.t. λ^i

  $[\cdot]_\Lambda \equiv$ projection on feasible set Λ

  $\nabla g^i(\lambda^i) = \bar{x}^i(\lambda^i)$



Slaves solve their subproblems, and return $\bar{x}^i(\lambda^i)$

Master updates each λ^i by setting $\lambda^i \leftarrow \left[\lambda^i + \alpha_t \bar{x}^i(\lambda^i)\right]_\Lambda$

Optimizing MRFs via dual decomposition

Cast MRF optimization as a linear integer program:

$$\min \left[\sum_{p \in G} \sum_{a \in L} V_p(a) x_{p,a} + \sum_{pq \in E} \sum_{a,b \in L} V_{pq}(a,b) x_{pq,ab} \right]$$

$$\text{s.t.} \sum\nolimits_{a \in L} x_{p,a} = 1$$

(only one label assigned per vertex)

$$\left.\begin{aligned} \sum\nolimits_{a \in L} x_{pq,ab} &= x_{q,b} \\ \sum\nolimits_{b \in L} x_{pq,ab} &= x_{p,a} \end{aligned}\right\}$$

enforce consistency between variables $x_{p,a}$, $x_{q,b}$ and variable $x_{pq,ab}$

$$x_{p,a},\ x_{pq,ab} \in \{0,1\}$$

Optimizing MRFs via dual decomposition

Cast MRF optimization as a linear integer program:

$$\begin{aligned} \min_{\mathbf{x}} \quad & E(\theta, \mathrm{x}) = \theta \cdot \mathrm{x} = \sum_{p \in \mathcal{V}} \theta_p \cdot \mathbf{x}_p + \sum_{pq \in \mathcal{E}} \theta_{pq} \cdot \mathbf{x}_{pq} \\ \text{s.t.} \quad & \mathrm{x} \in \mathcal{C} \end{aligned}$$

$\theta = \{\{\theta_p\}, \{\theta_{pq}\}\}$ is the vector of MRF-parameters consisting of all unary $\theta_p = \{\theta_p(\cdot)\}$ and pairwise $\theta_{pq} = \{\theta_{pq}(\cdot, \cdot)\}$ vectorized potentials.

$\mathrm{x} = \{\{\mathbf{x}_p\}, \{\mathbf{x}_{pq}\}\}$ is the vector of binary MRF-variables consisting of unary subvectors $\mathbf{x}_p = \{x_p(\cdot)\}$ and pairwise subvectors $\mathbf{x}_{pq} = \{x_{pq}(\cdot, \cdot)\}$

Constraints $\mathcal{C}$ enforce consistency between variables $\{\mathbf{x}_p\}$ and $\{\mathbf{x}_{pq}\}$

Slave problems for MRFs

- Tree-structured MRFs (solvable with max-product).
  - To each tree T from a set of trees $\mathcal{T}$, we can associate a slave MRF with parameters θ^T
  - These parameters must initially satisfy:
    $\sum_{T \in \mathcal{T}(p)} \theta_p^T = \theta_p, \qquad \sum_{T \in \mathcal{T}(pq)} \theta_{pq}^T = \theta_{pq},$
    (Here $\mathcal{T}(p), \mathcal{T}(pq)$ denote all trees in $\mathcal{T}$ containing respectively p and pq)

Theoretical properties

- Guaranteed convergence
- Provably optimizes LP-relaxation (unlike existing tree-reweighted message passing algorithms)
  - In fact, distance to optimum is guaranteed to decrease per iteration

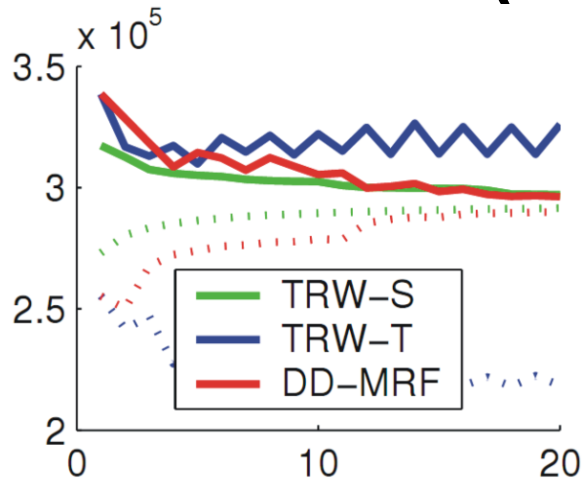
Experimental results

- Resulting algorithm is called DD-MRF
- It has been applied to:
  - stereo matching
  - optical flow
  - binary segmentation
  - synthetic problems
- Lower bounds produced by the master certify that solutions are almost optimal

More: N. Komodakis & P. Kumar course, 2nd semester of MVA

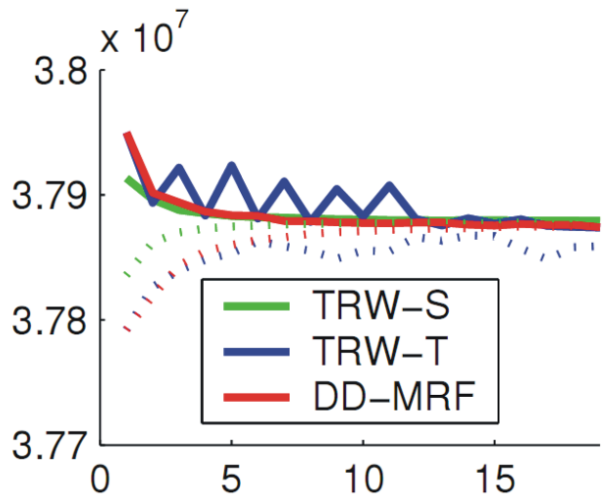
Experimental results (2/4)

estimated disparity for
Tsukuba stereo pair



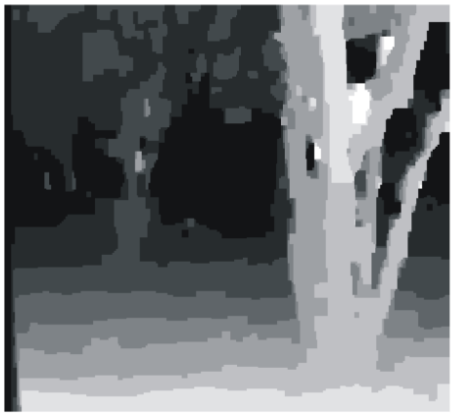
lower bounds (dual costs) and
MRF energies (primal costs)

estimated disparity for
Map stereo pair

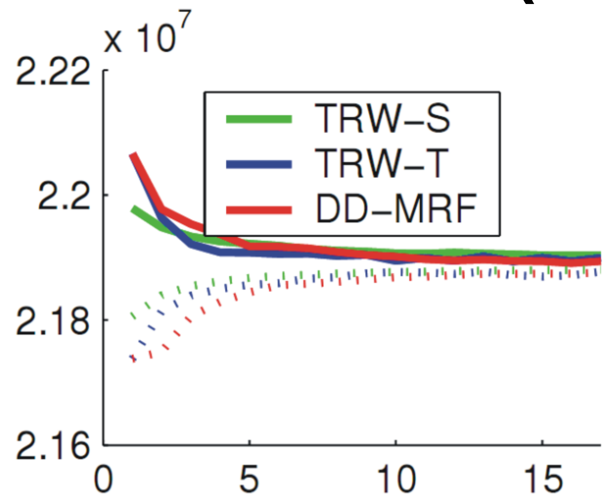


lower bounds (dual costs) and
MRF energies (primal costs)

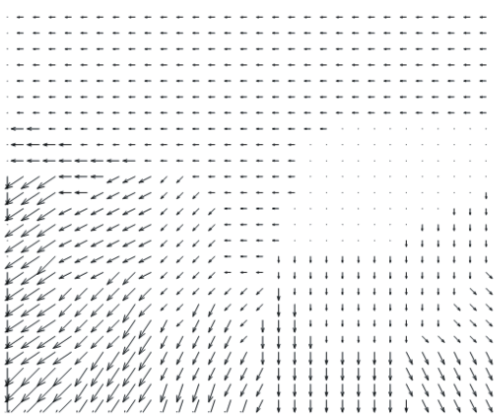
Experimental results (3/4)



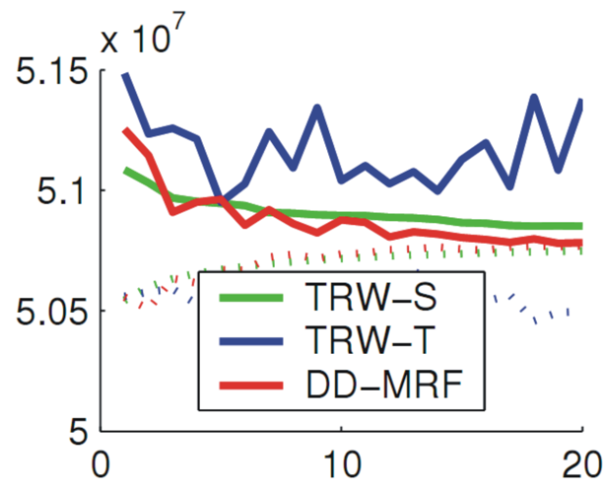
estimated disparity for SRI stereo pair



lower bounds (dual costs) and MRF energies (primal costs)

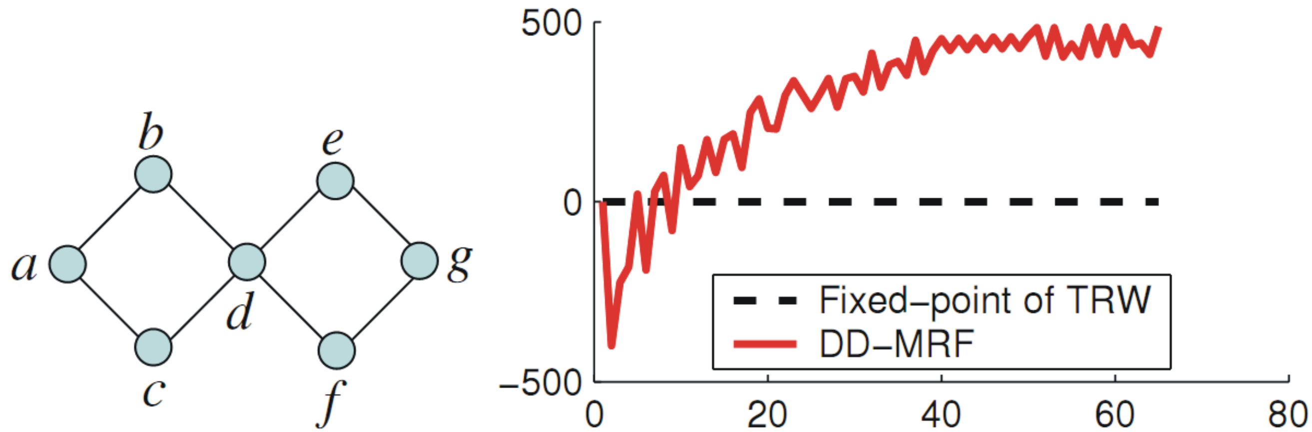


estimated optical flow for Yosemite sequence

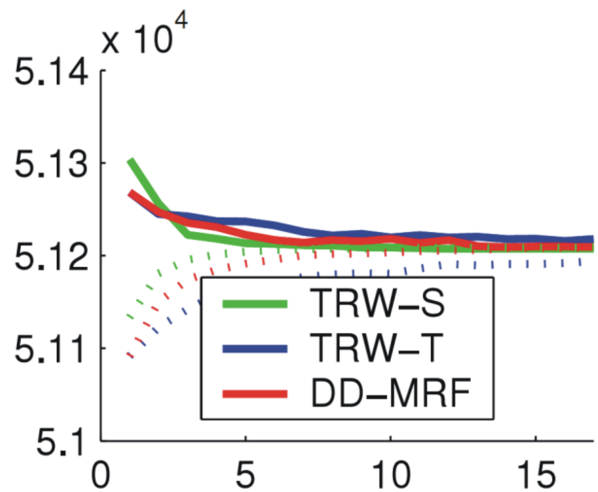


lower bounds (dual costs) and MRF energies (primal costs)

Experimental results (4/4)



a simple synthetic example illustrating that TRW methods are not able to maximize the dual lower bound, whereas DD-MRF can.



lower bounds (dual costs) and MRF energies (primal costs) for binary segmenation

2nd part outline

Internship proposals

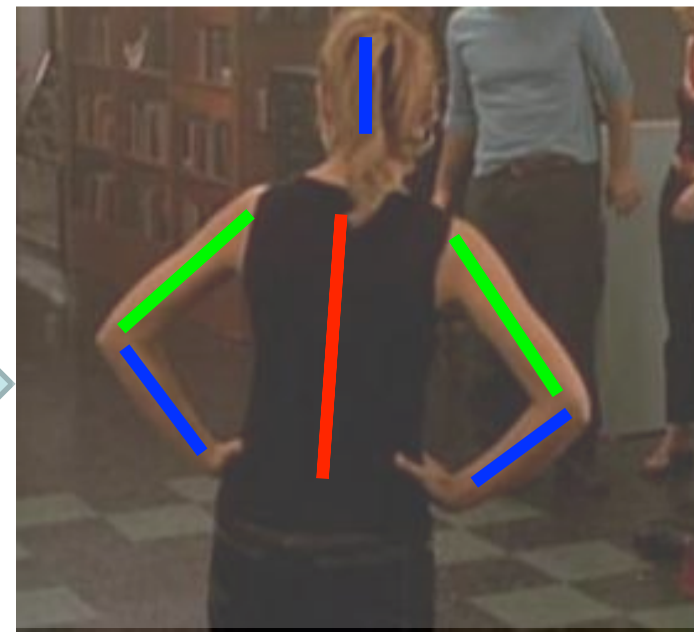
Human Pose Estimation

- **Goal**: Image -> Stick Figure
  2D locations of anatomical parts from a single image

input

efficient inference

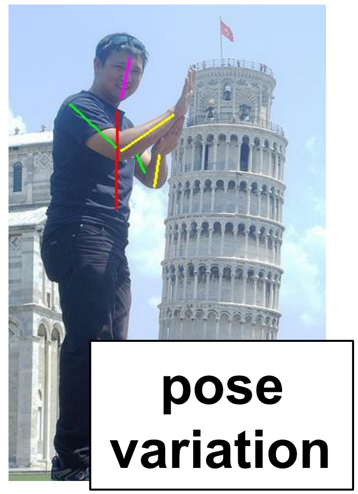
output



Slide Credits: B. Sapp, D. Weiss, B. Taskar

Challenges


lighting variation


pose variation


intrinsic scale variations

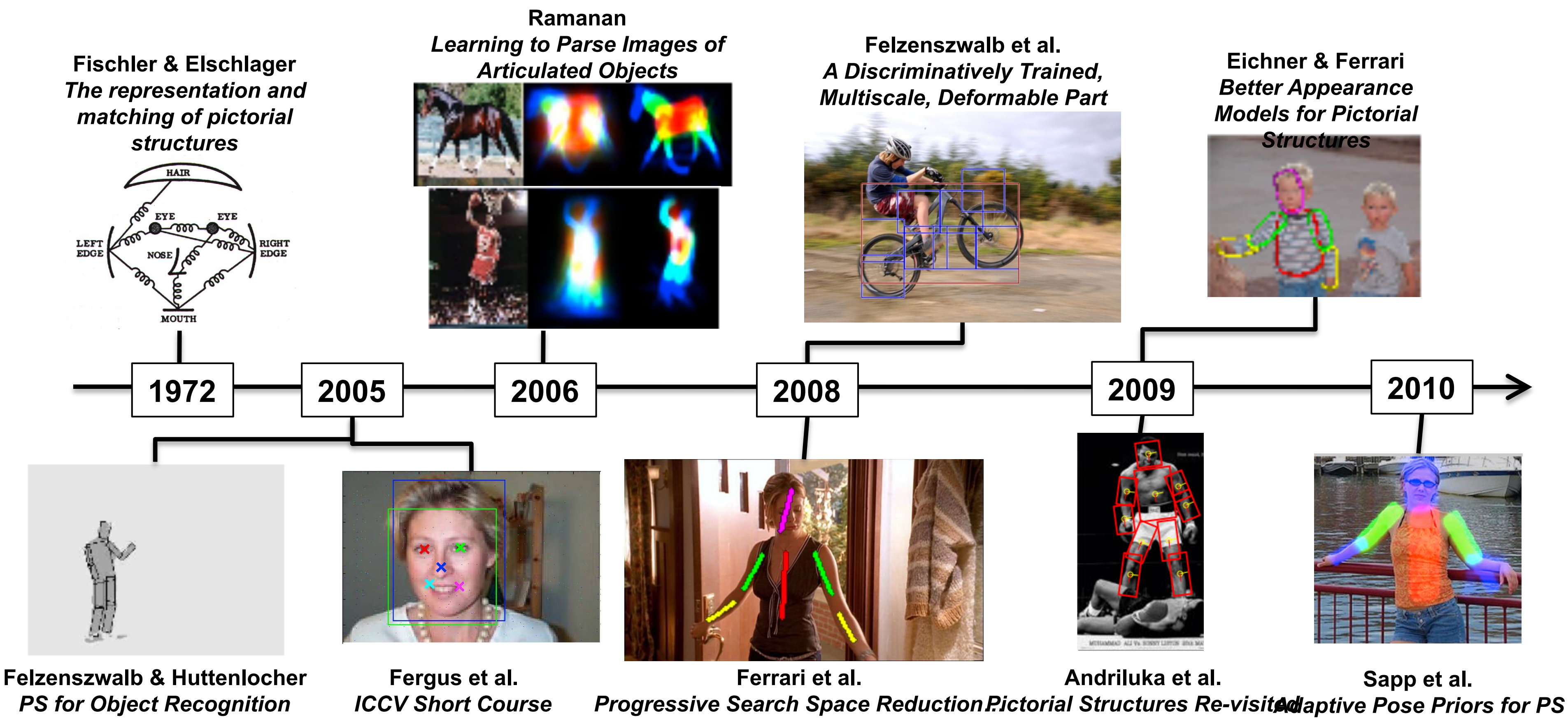

del.icio.us
background clutter


foreshortening

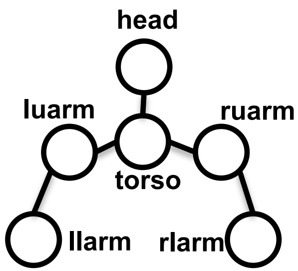
Articulated Pose and Pictorial Structures

A popular choice for (articulated) parts-based models

A non-exhaustive timeline

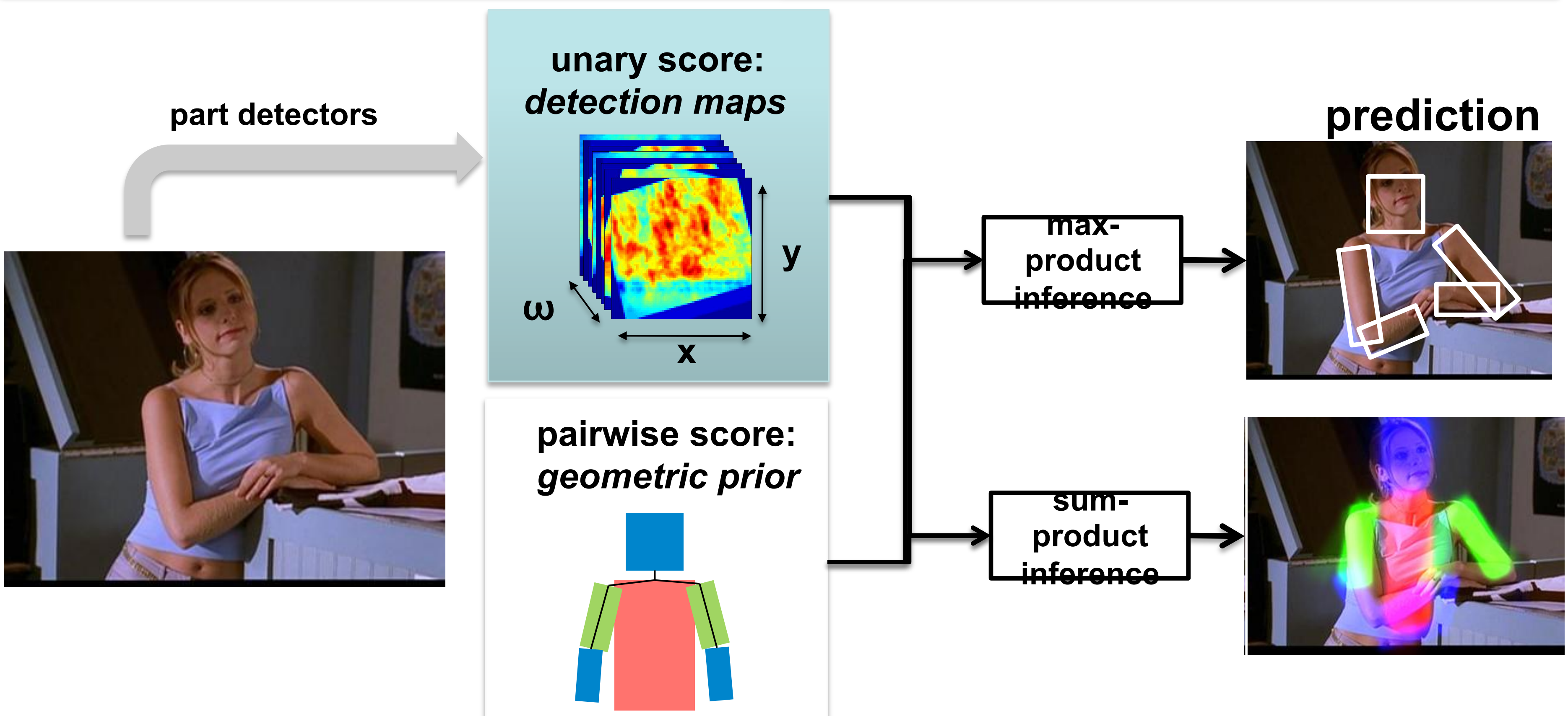


Pictorial Structures for Pose Estimation

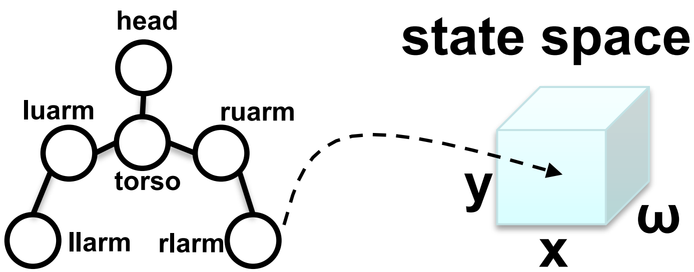


l_i : **location for part *i***

$$p(l_1, \cdots, l_M | Im) \propto \exp \left[\boxed{\sum_i s_i(l_i, Im)} + \boxed{\sum_{ij} s_{ij}(l_i, l_j)} \right]$$



The Complexity of PS



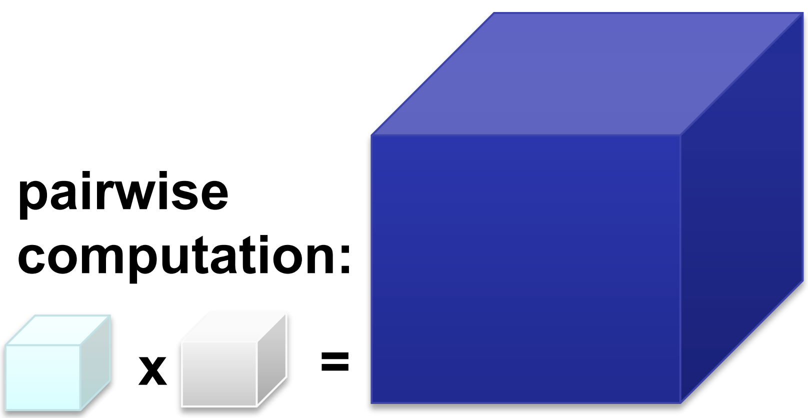
typical state space size:

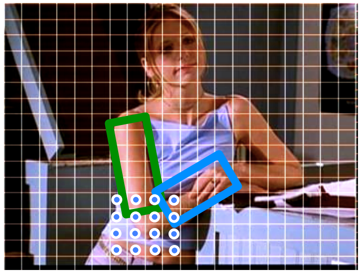
$$n = x \times y \times \omega = 80 \times 80 \times 24$$

n > 150,000 states

- Standard inference in a tree graphical model is $O(\#parts \cdot n^2)$
- Typical # of valid combinations for two neighboring parts: (80 x 80 x 24)•(80/5 x 80/5 x 24) ≈ **1 billion state-pairs!**

pairwise computation:



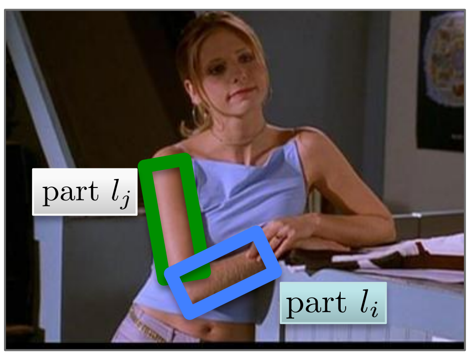


The Complexity of PS

If $s_{ij}(l_i, l_j, Im) = s_{ij}(l_i - l_j)$, **efficient inference tricks** can be used:

[Felzenszwalb & Huttenlocher, 2005]

- Max-prod w/ unimodal cost: Distance transform for $O(\#parts \cdot n)$
- Sum-prod w/ linear filter cost: Convolution for $O(\#parts \cdot n \log n)$



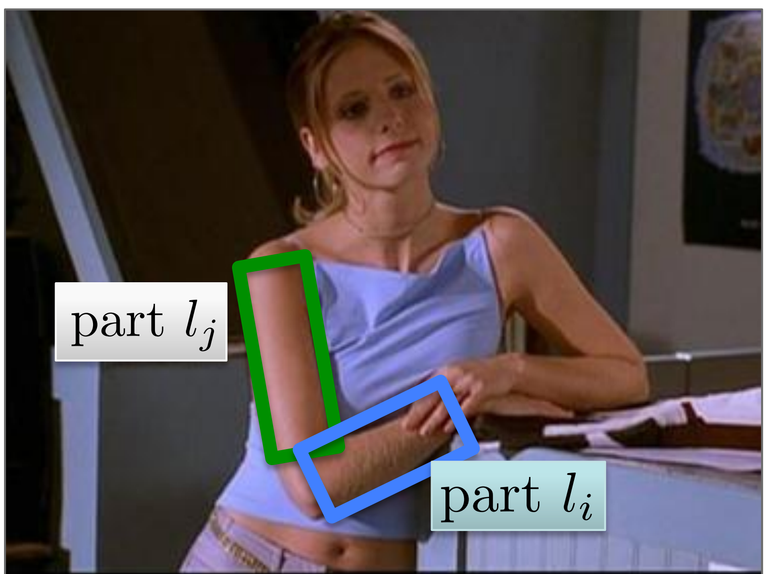
score for part-state pair:

Internship Proposal: Dual Tree Branch and Bound for pose estimation

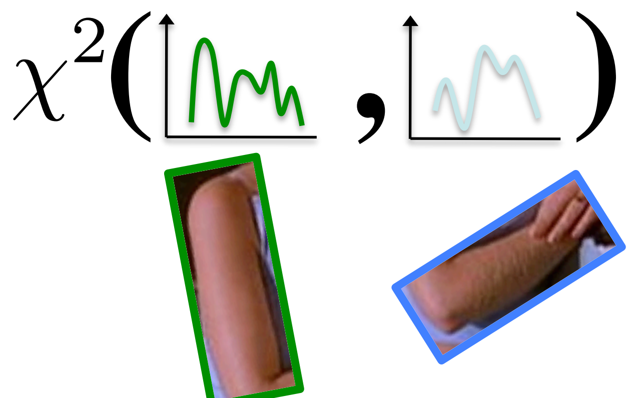
Challenge: Integrating richer pairwise terms

$$p(l_1, \cdots, l_M | Im) \propto \exp\left[\sum_i s_i(l_i, Im) + \sum_{ij} s_{ij}(l_i, l_j, \boxed{Im})\right]$$

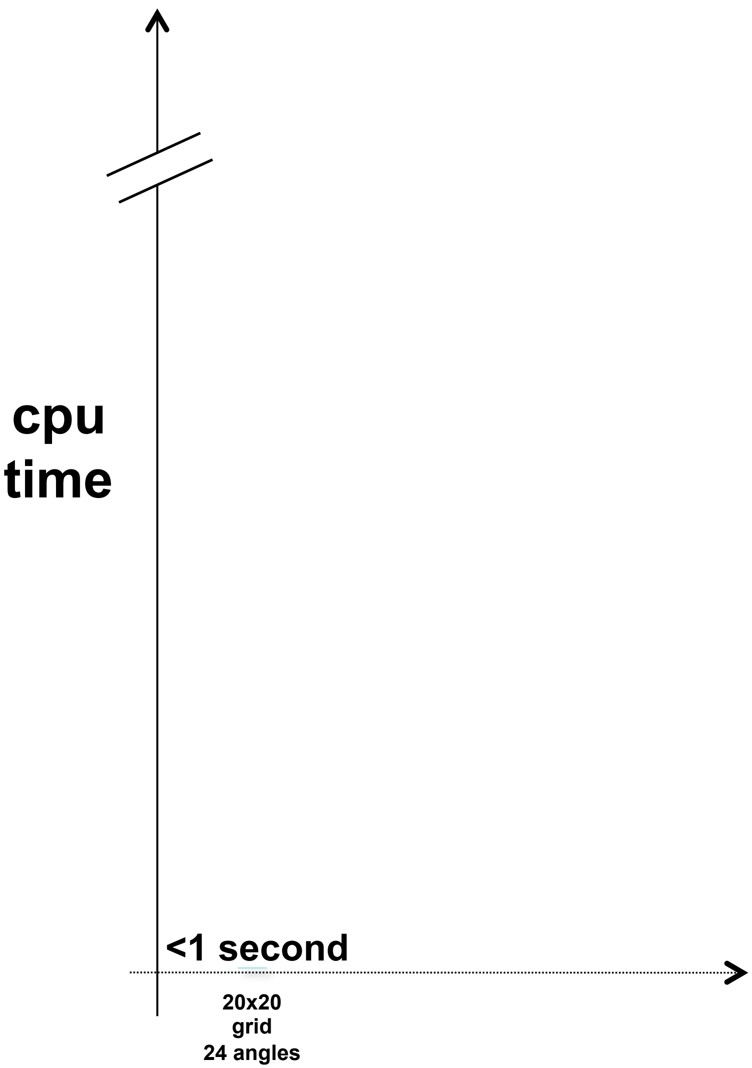
incorporate image evidence



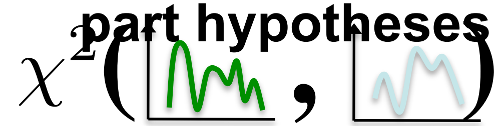
e.g., distance in color distribution:

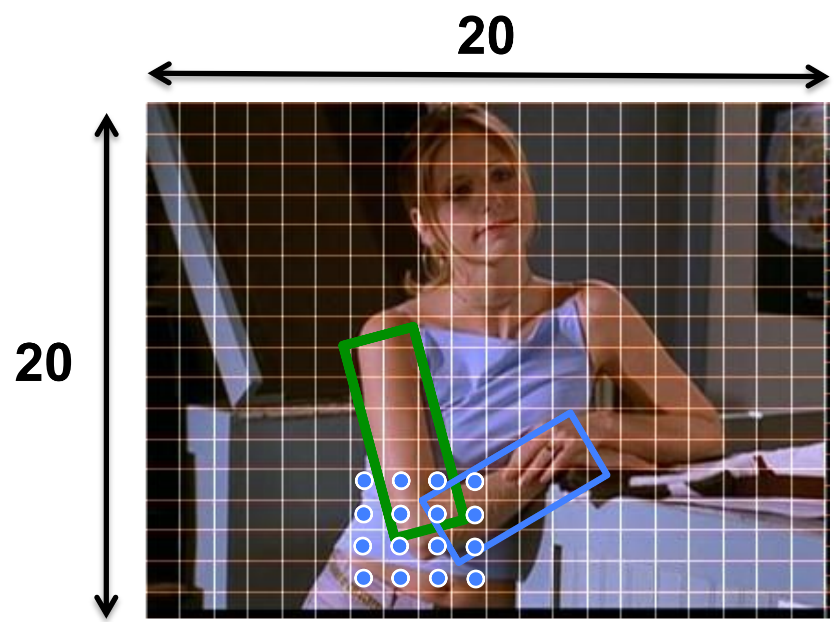


Computation example



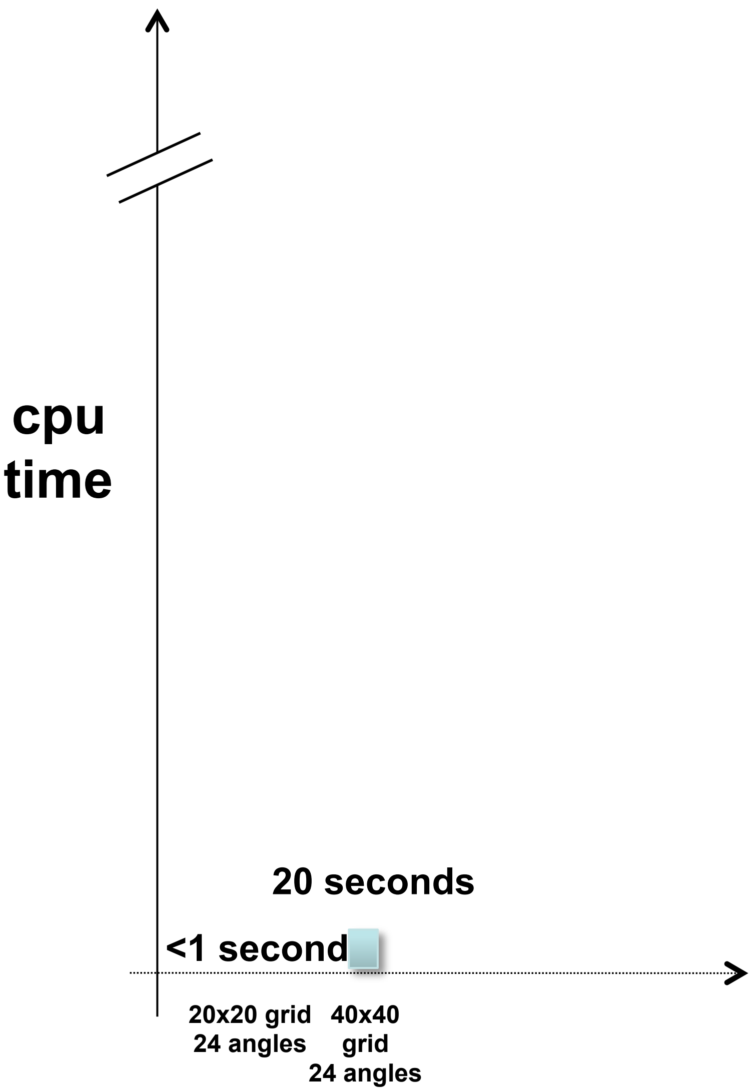
color histogram χ^2 distance computation between all pairs of part hypotheses



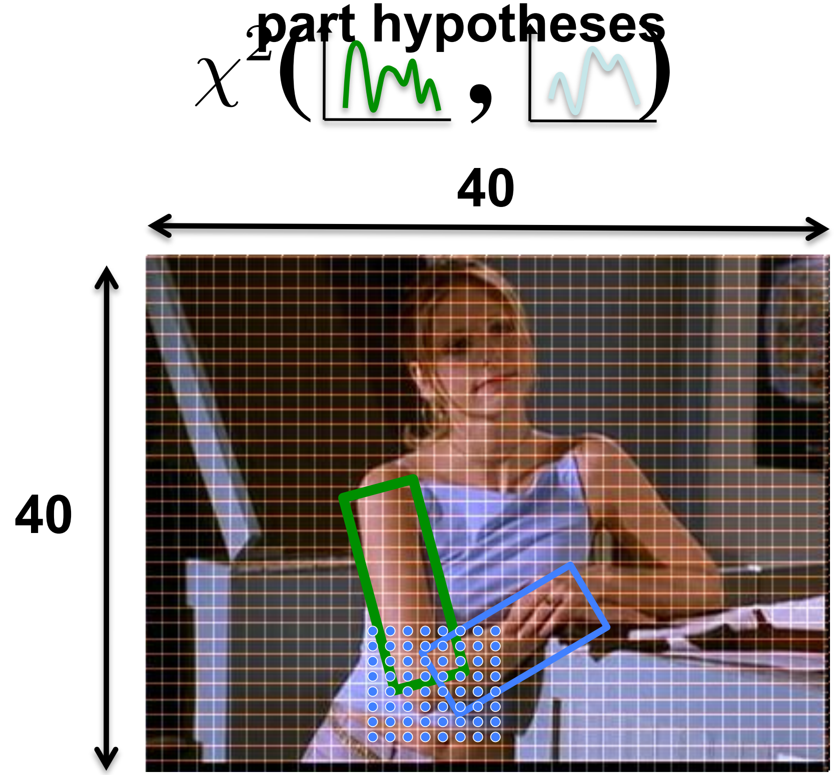


3.7 million comparisons

Computation example

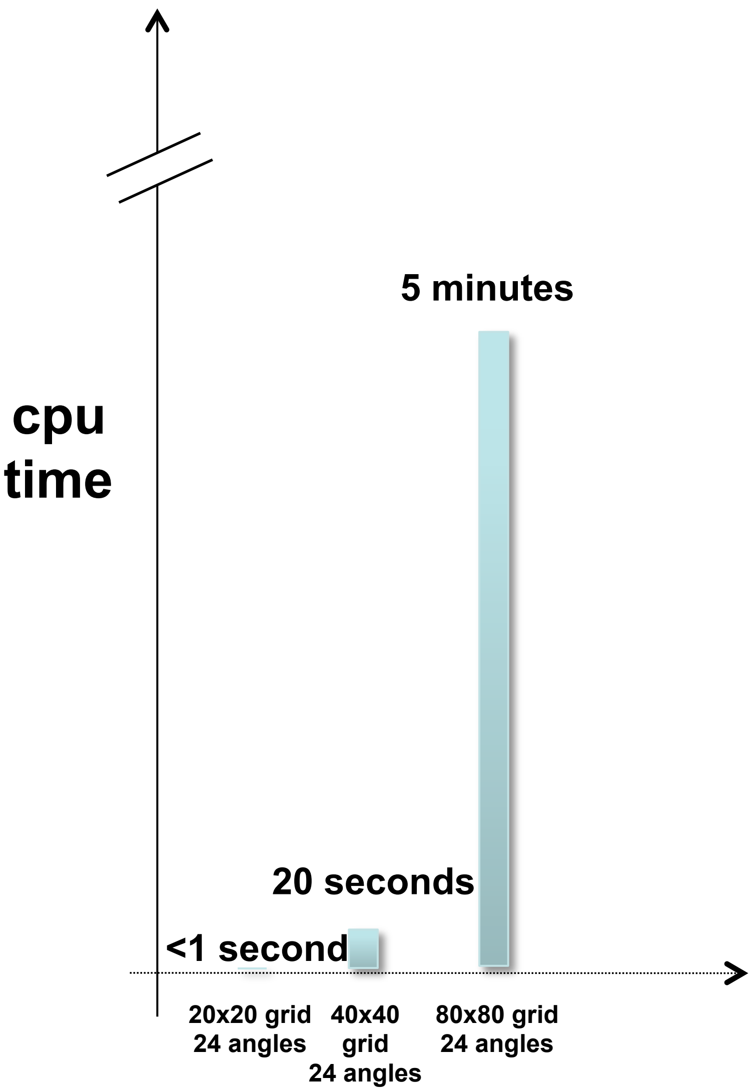


color histogram χ^2 distance computation between all pairs of part hypotheses

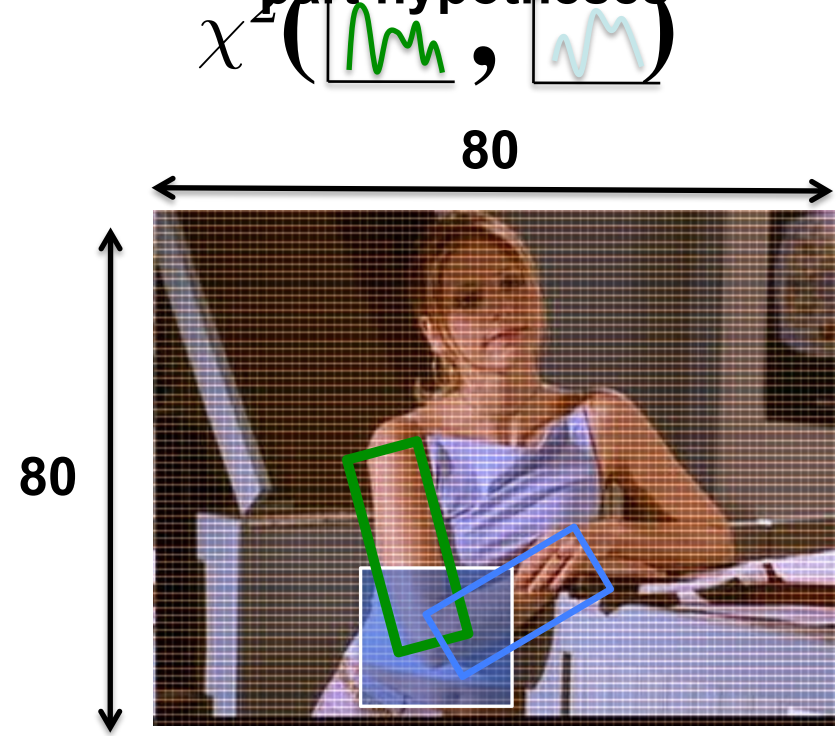


59 million comparisons

Computation example

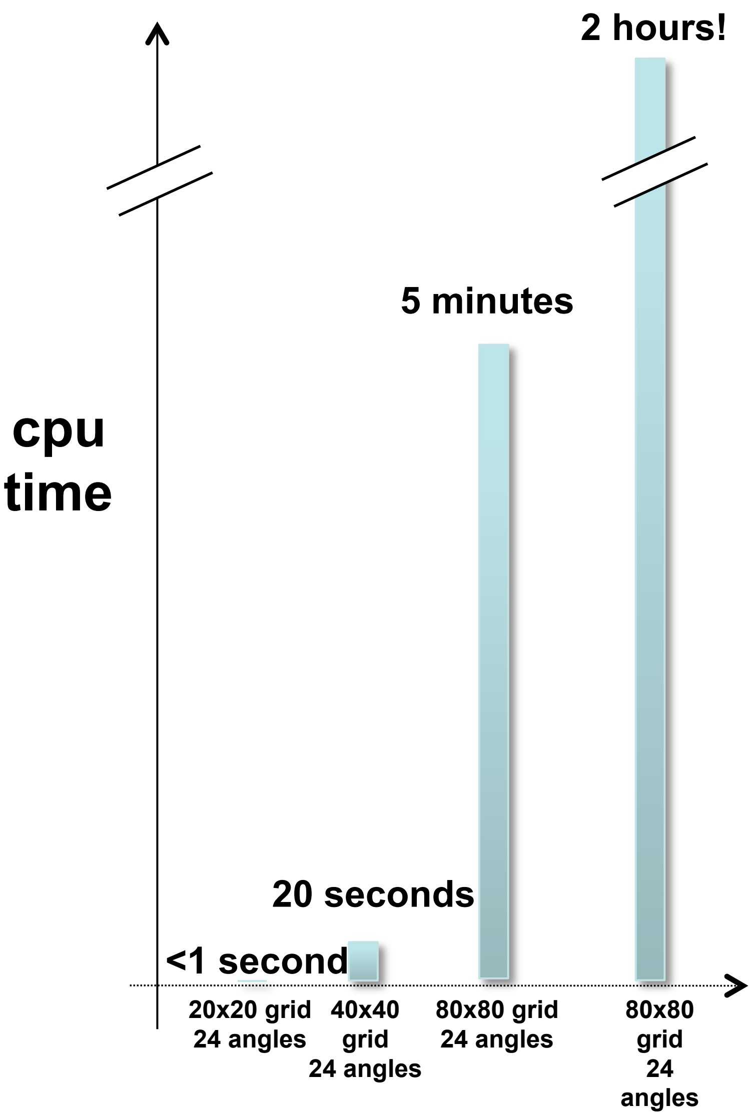


color histogram χ^2 distance computation between all pairs of part hypotheses

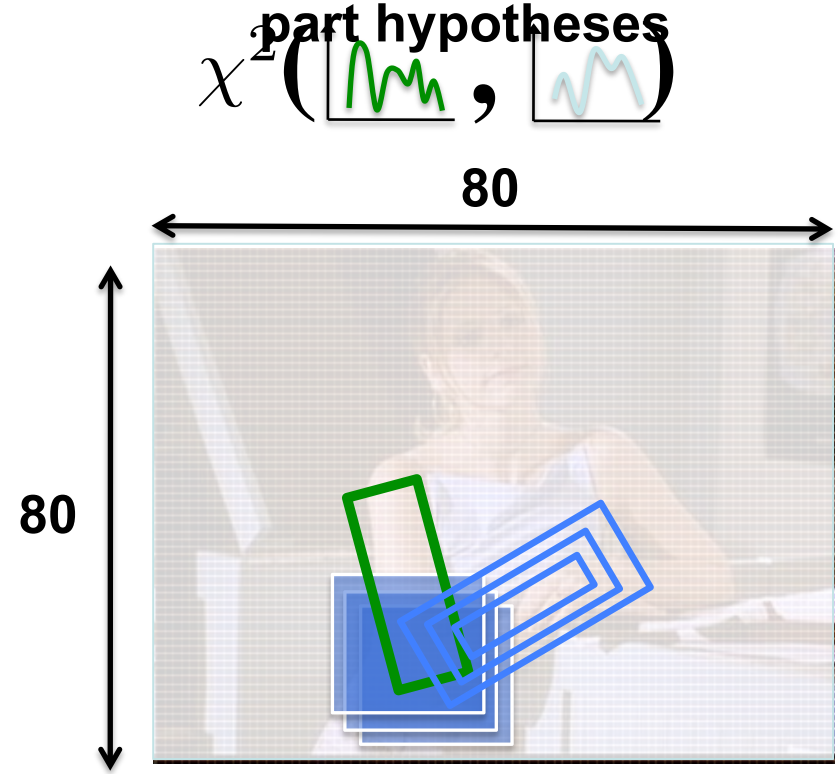


1 billion comparisons

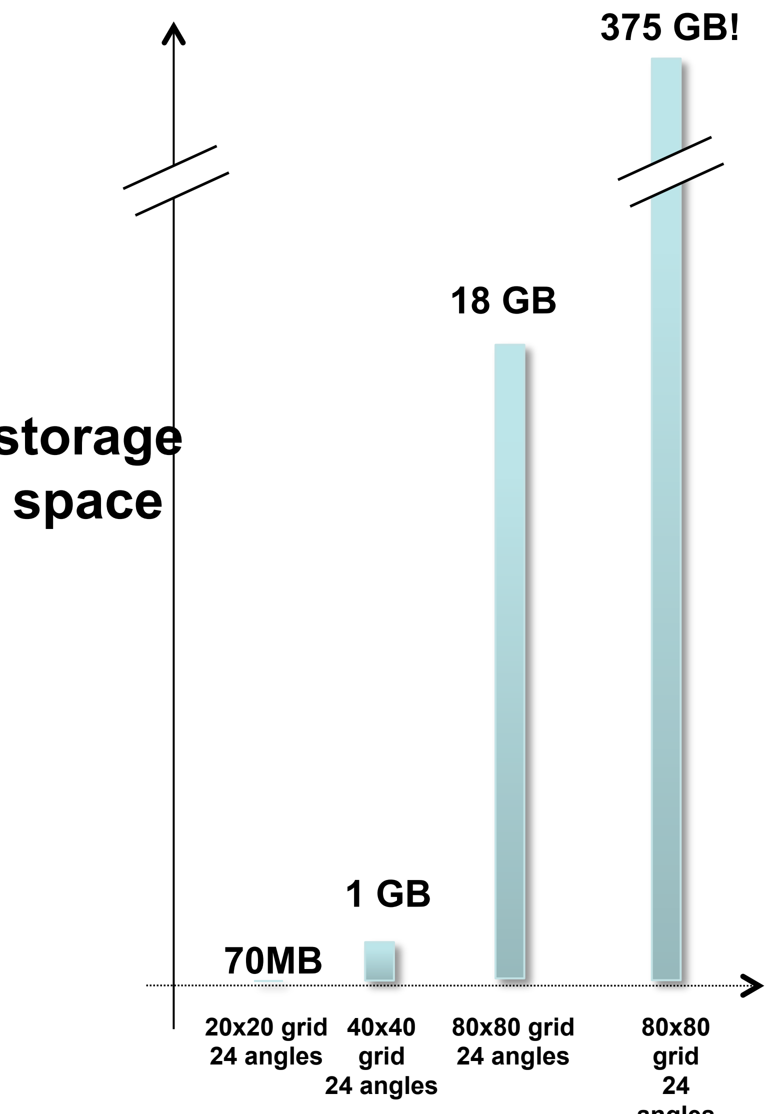
Computation example



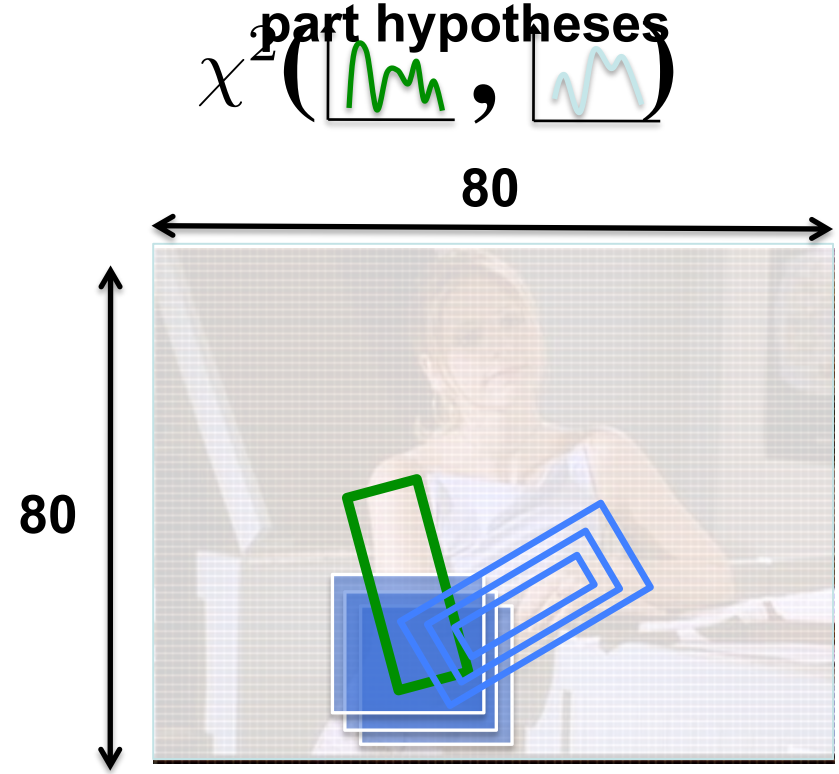
color histogram χ^2 distance computation between all pairs of part hypotheses



Computation example



color histogram χ^2 distance computation between all pairs of part hypotheses



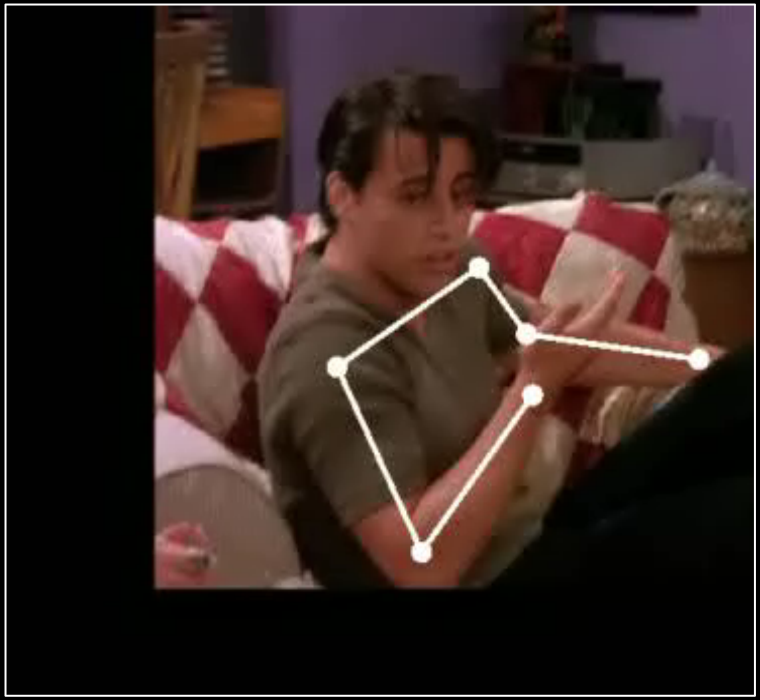
Exhaustive Inference

Branch–and –bound!

Parsing Human Motion

Input

Desired Output



Parsing Human Motion with Stretchable Models,Sapp, Weiss, Taskar CVPR 2011

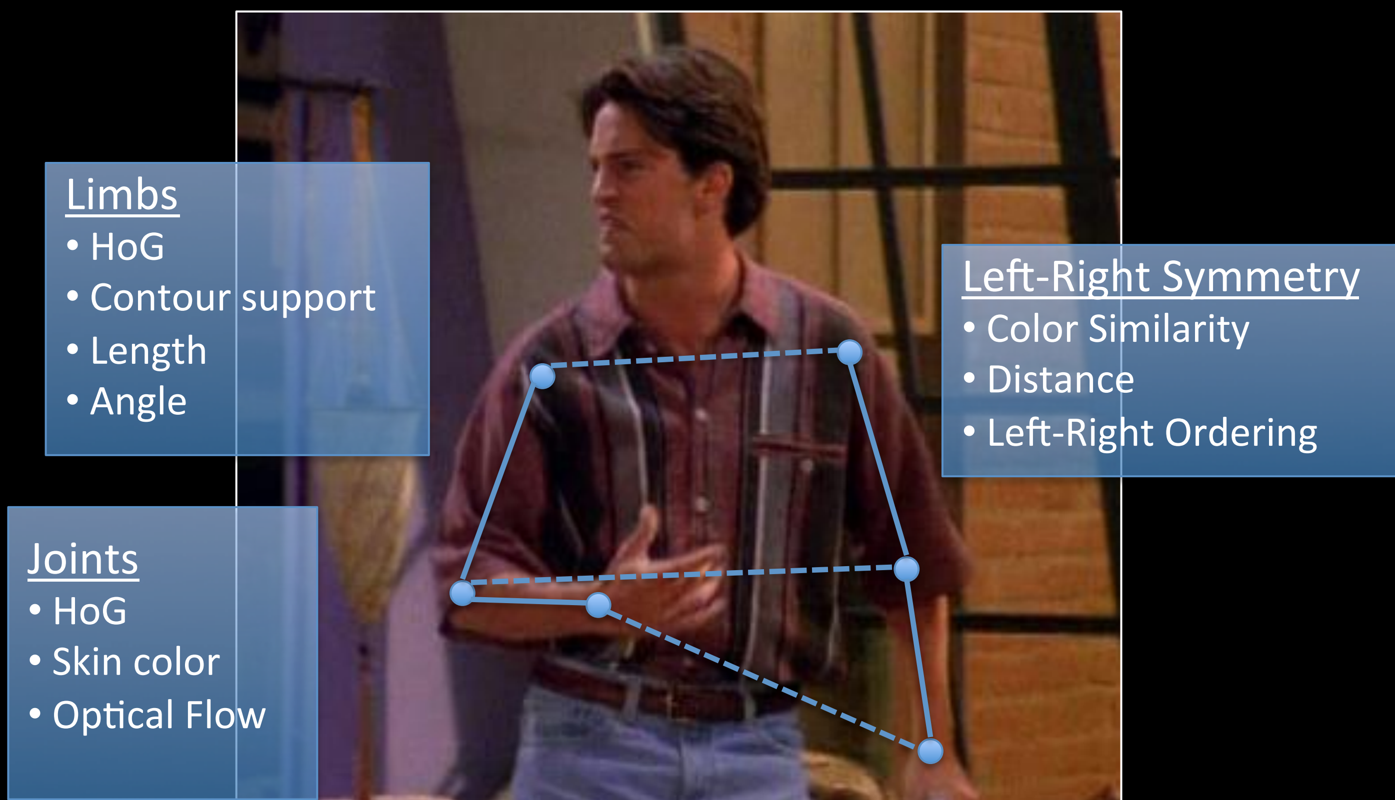
What to Model

- Detecting joints in isolation is hard

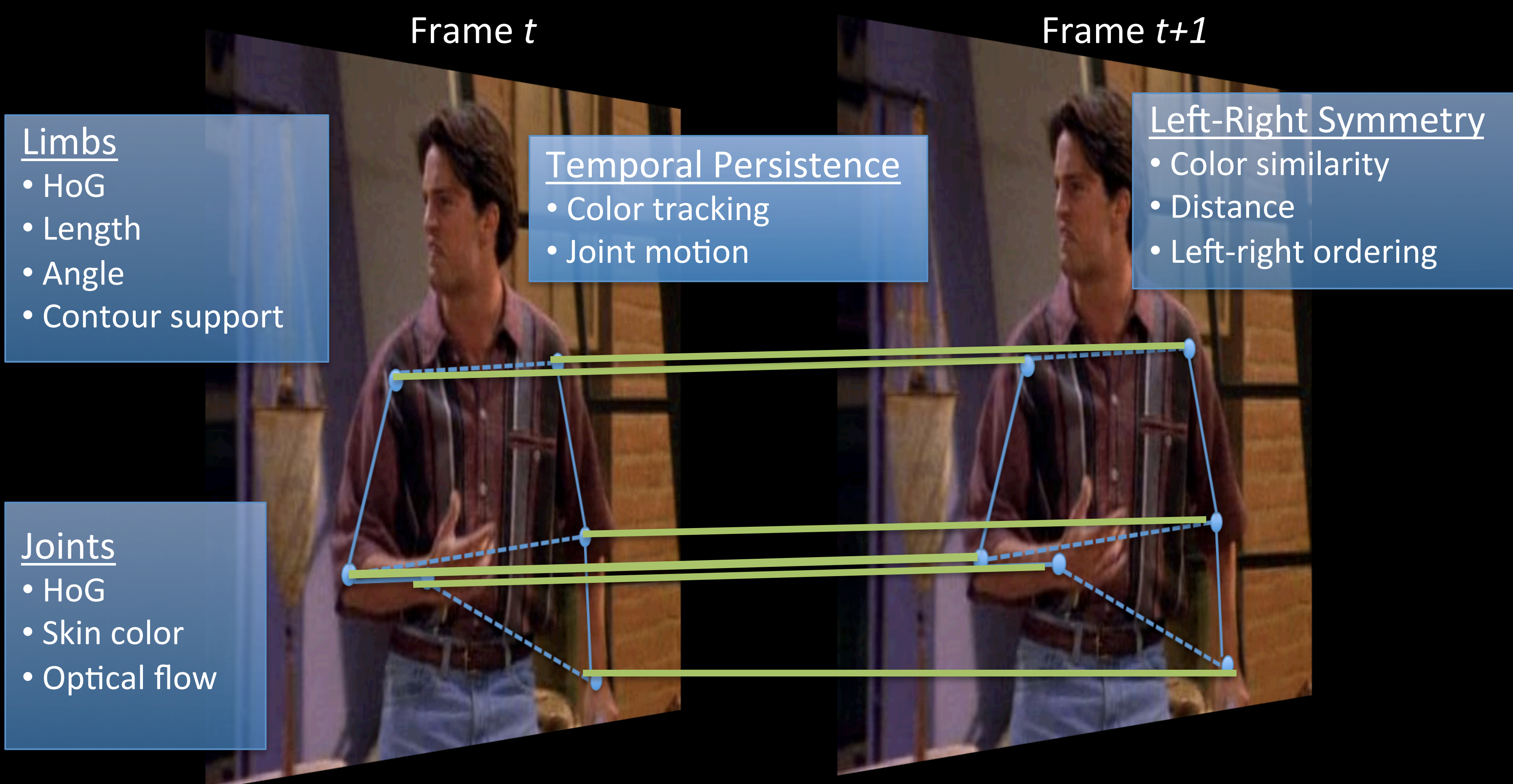
Where are the elbows?

What to Model

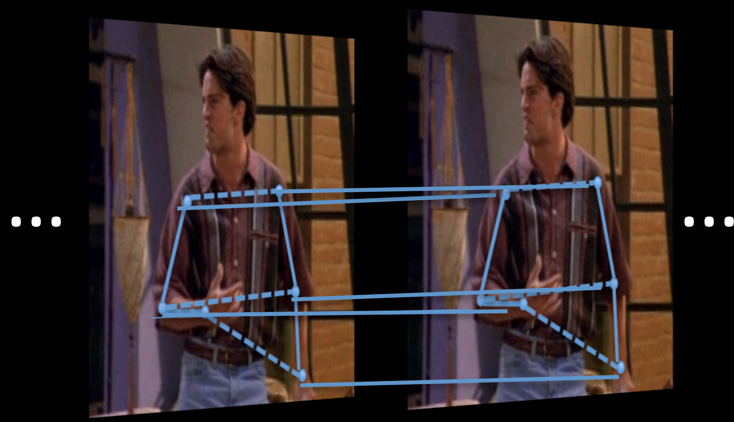
- Detecting joints in isolation is hard
- Need to describe relationships *between* joints



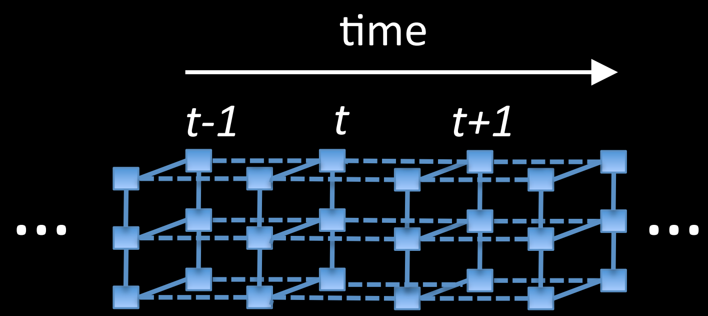
What to Model



Full Model



=

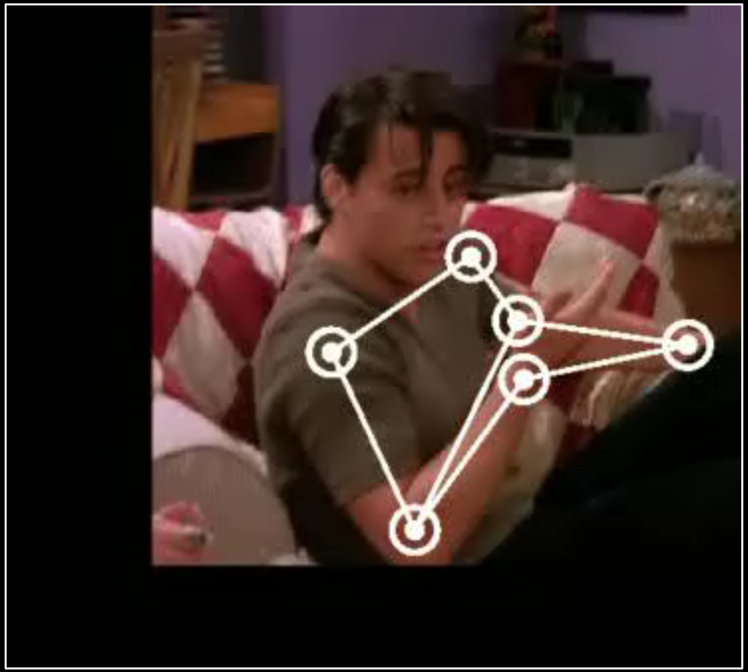


<u>Joints</u>
- HoG
- Skin color
- Optical flow

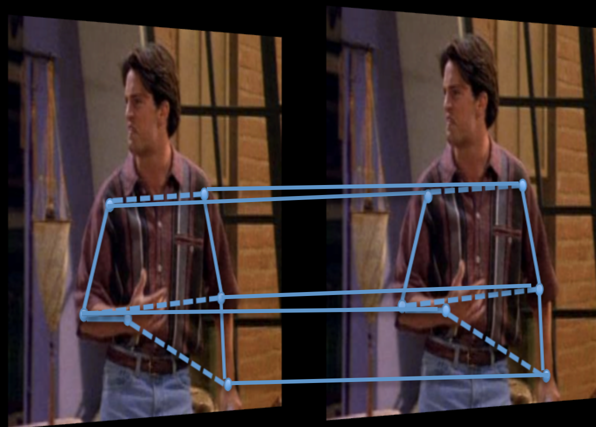
<u>Limbs</u>
- HoG
- Length
- Angle
- Contour support

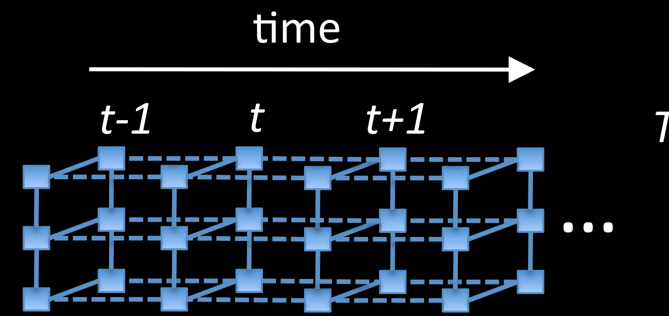
<u>Symmetry</u>
- Color similarity
- Distance
- Ordering

<u>Time Persistence</u>
- Color tracking
- Joint motion



Full Model: an MRF

…  … = …



…

$$G = (V, E)$$

N joints, T frames

$$\mathbf{y} \equiv \{y_1, \ldots, y_{NT}\} \equiv \text{joint locations in every frame}$$

$$x \equiv \text{video data}$$

$$\boxed{p(\mathbf{y}|x;G) = \frac{1}{Z(x)} \prod_{i \in V} \phi(x, y_i) \prod_{(i,j) \in E} \phi(x, y_i, y_j)}$$

MAP placement:

$$\mathbf{y}^\star = \arg\max_{\mathbf{y}} p(\mathbf{y}|x;G)$$

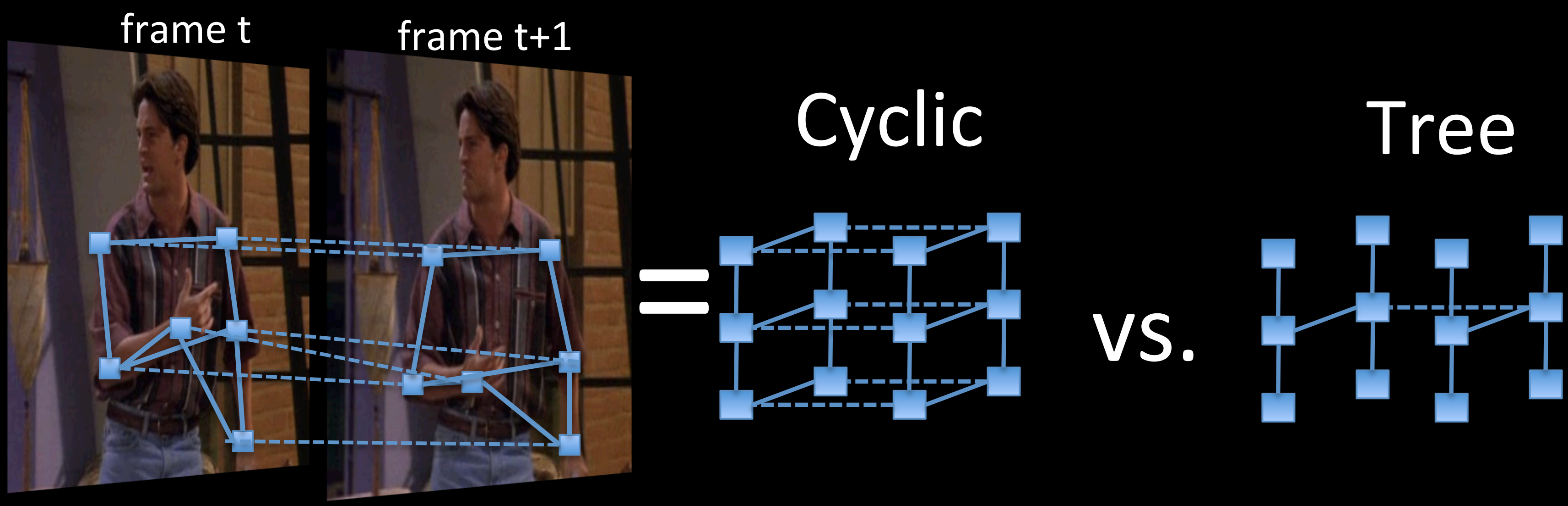
INTRACTABLE

How do we do it?

Tree Decomposition + Agreement Algorithms

+ efficient
+ not greedy

Sidestepping Intractability



Inference *exponential* in number of joints

Inference *linear* in number of joints

Lose some of our edges!

DTBB: Inference sub*linear* in number of joints

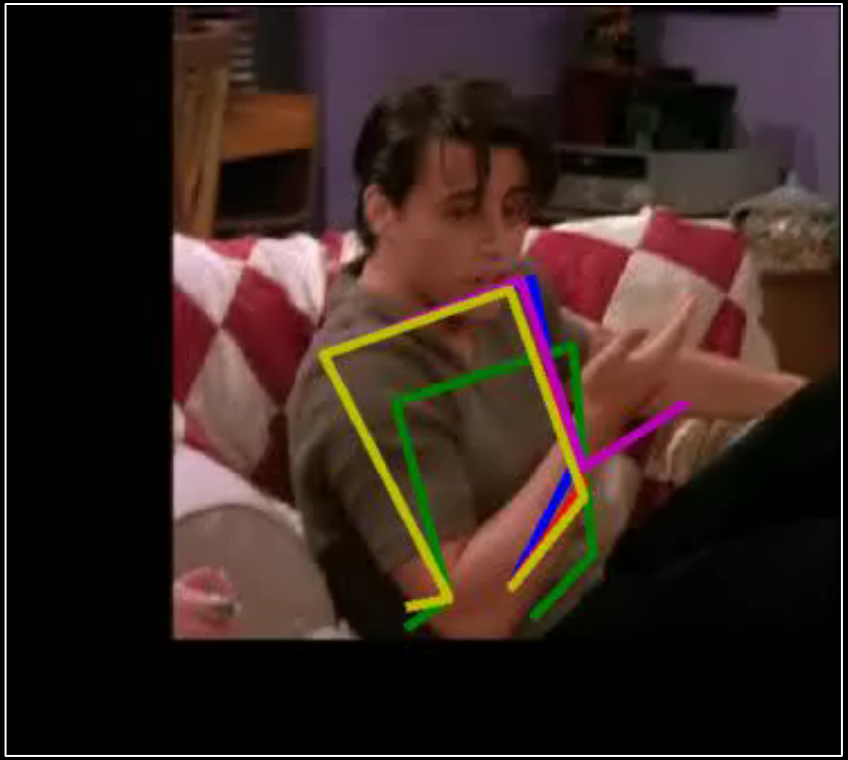
Model (dis)agreement

$$\operatorname*{argmax}_{\mathbf{y}} p(\mathbf{y}|x; G_m) \neq \operatorname*{argmax}_{\mathbf{y}} p(\mathbf{y}|x; G_{m'})$$

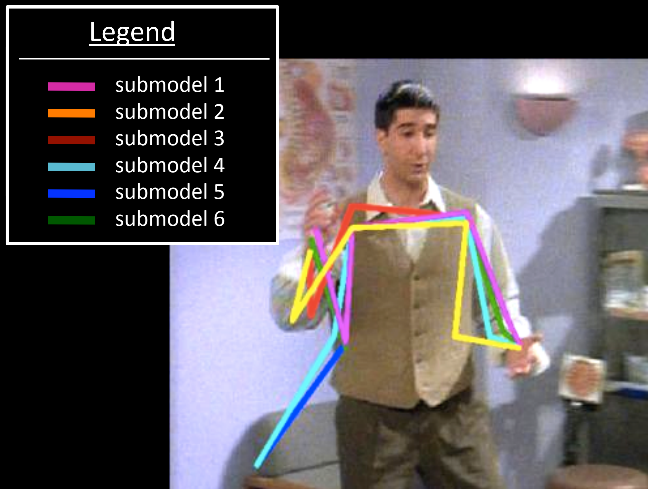
different tree models

Legend

- submodel 1
- submodel 2
- submodel 3
- submodel 4
- submodel 5
- submodel 6



The Value of Agreement



force agreement

No Agreement

Single Variable Agreement

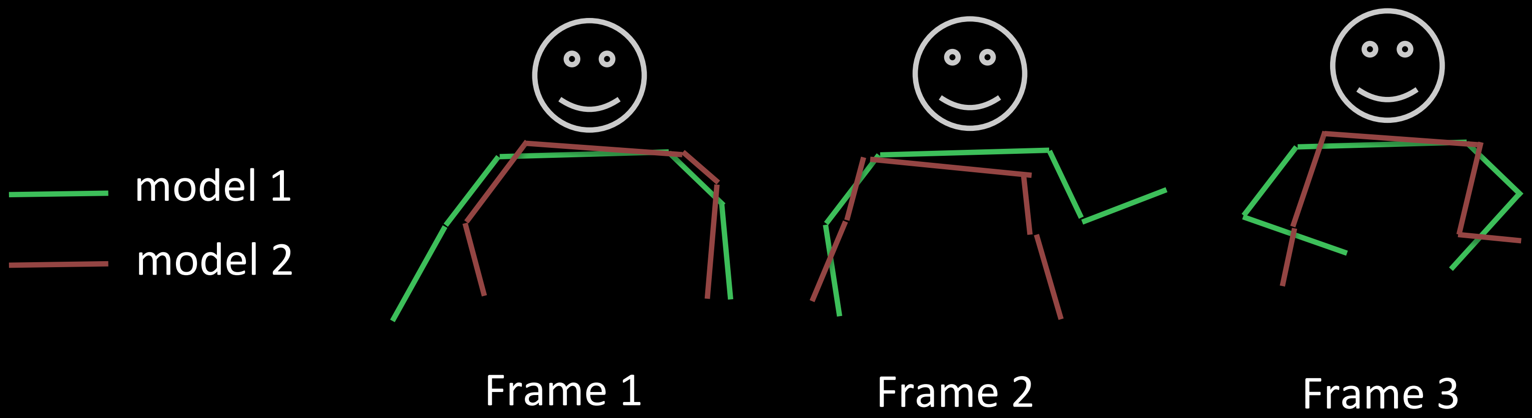
Single Frame Agreement

Full Agreement (Dual Decomposition)

Degree of Agreement / Computational Cost

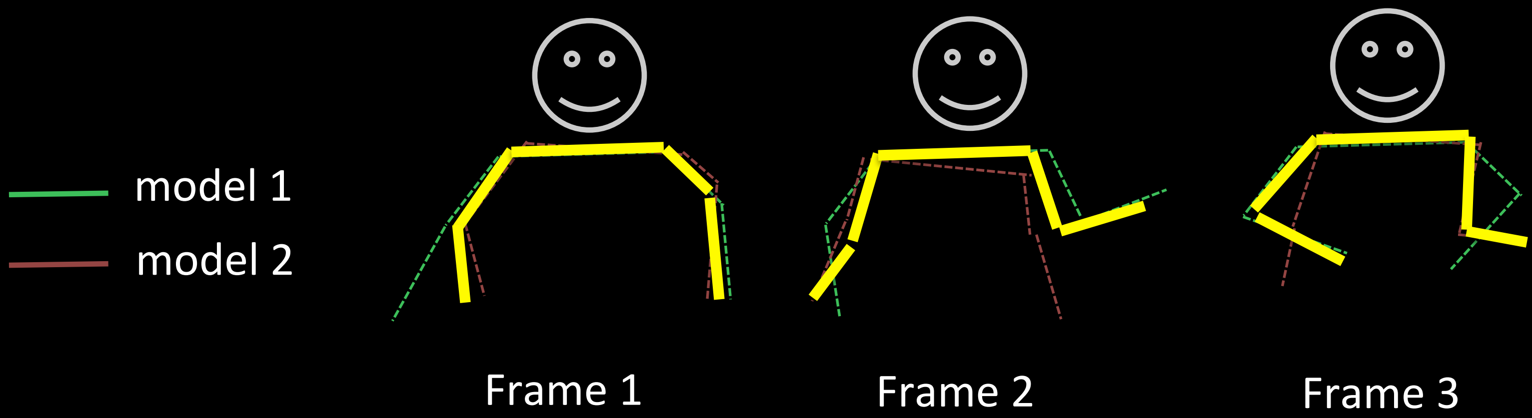
Does more agreement = better performance?

Agreement Methods



Toy example: 2 models in disagreement, 3 frames.

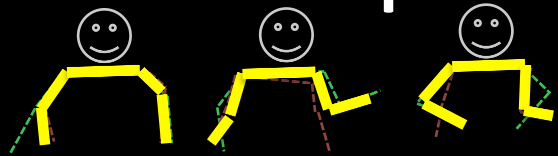
Agreement: Dual Decomposition



Force all models to agree on
every joint in *every* frame.

[Bertsekas, 1999], [Komodakis et al., 2007]

Dual Decomposition



Subgradient descent on dual to reach agreement:

```
while (!converged) {
    1. run modified inference in all M submodels
    2. adjust dual variables
}
```

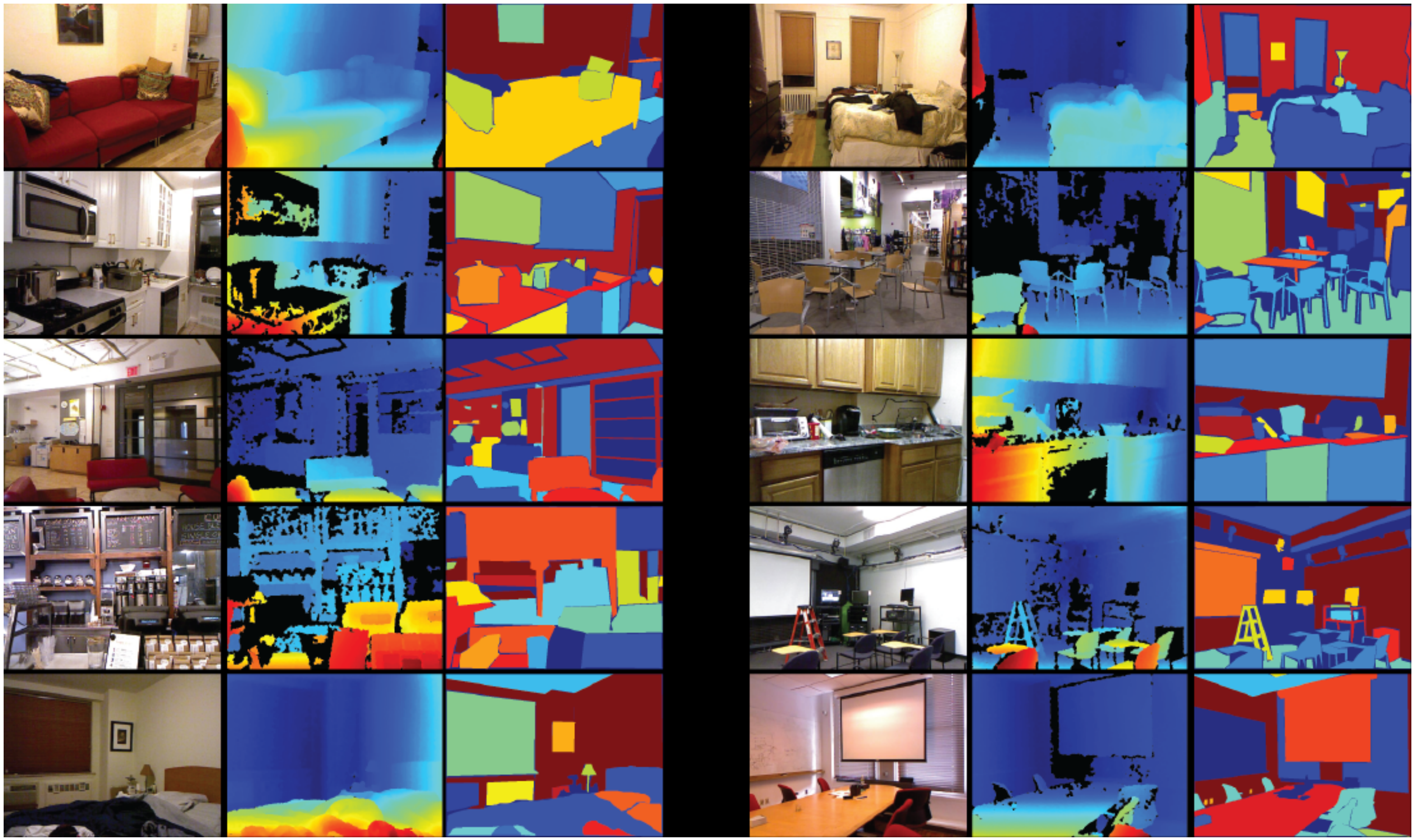
Cost: $O((\#\ \text{iters}) \cdot M \cdot N \cdot T \cdot |\mathcal{Y}_i|^2)$

Except, it may never converge (in which case, we round).

And, 100 to 500 iterations (typical) → 100-500x slower!

Vision for the coming decade

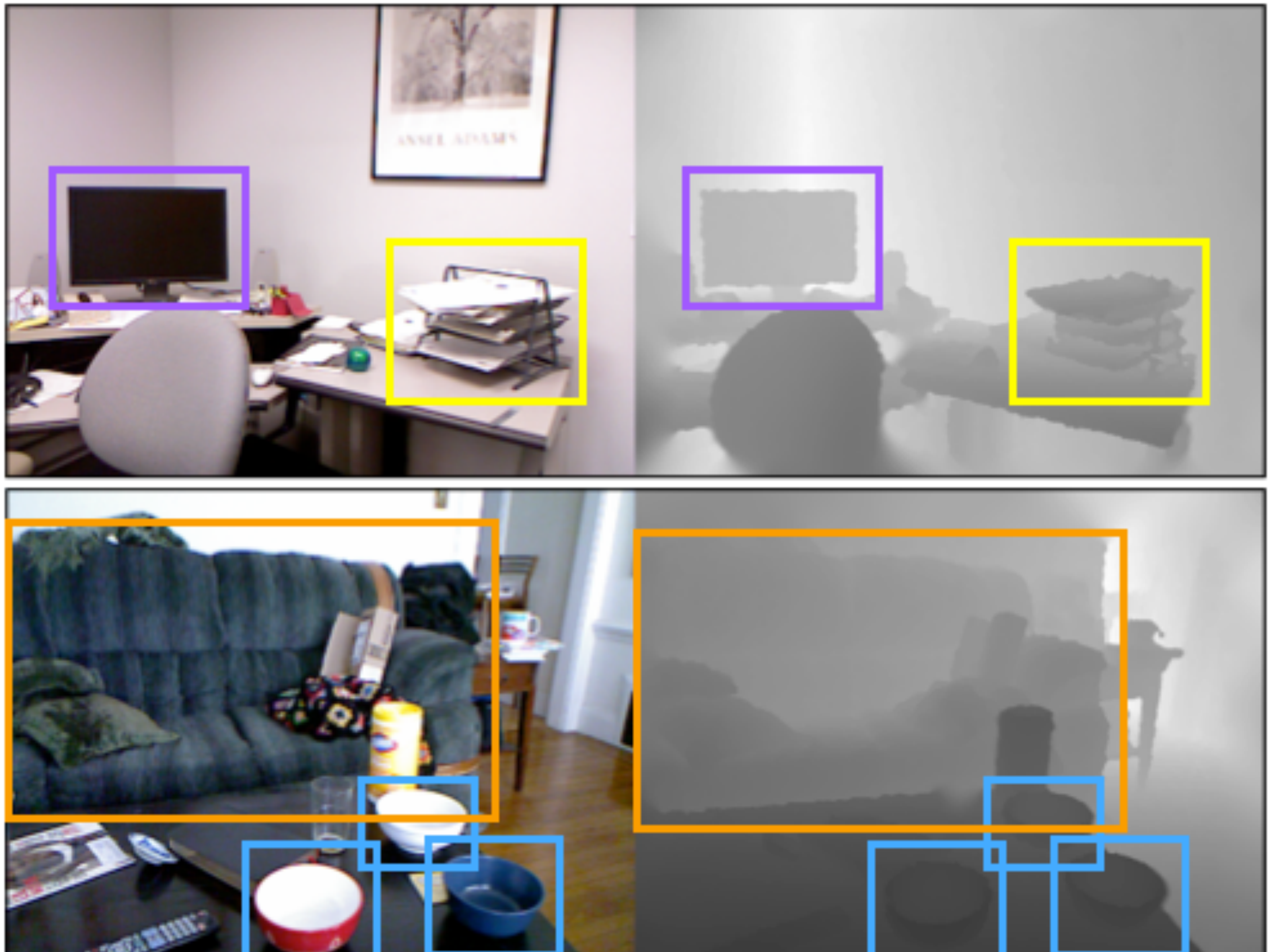
Datasets (NYU)



Indoor Scene Segmentation using a Structured Light Sensor
Nathan Silberman and Rob Fergus, 3DRR Workshop, ICCV 2011

.

Datasets (Berkeley)



A. Janoch, S. Karayev, Y. Jia, J.T. Barron, M. Fritz, K. Saenko, T. Darrell. A Category-Level 3-D Object Dataset: Putting the Kinect to Work. ICCV Workshop on Consumer Depth Cameras in Computer Vision 2011. .

Datasets (Washington/Intel)

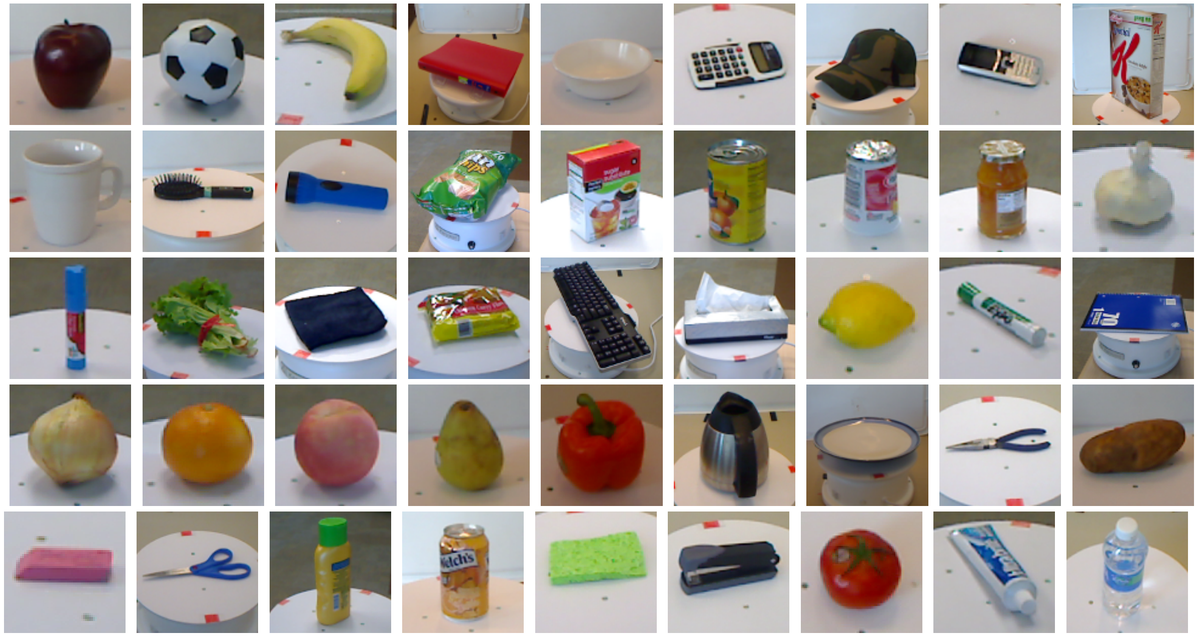


Fig. 3. Objects from the RGB-D Object Dataset. Each object shown here belongs to a different category.

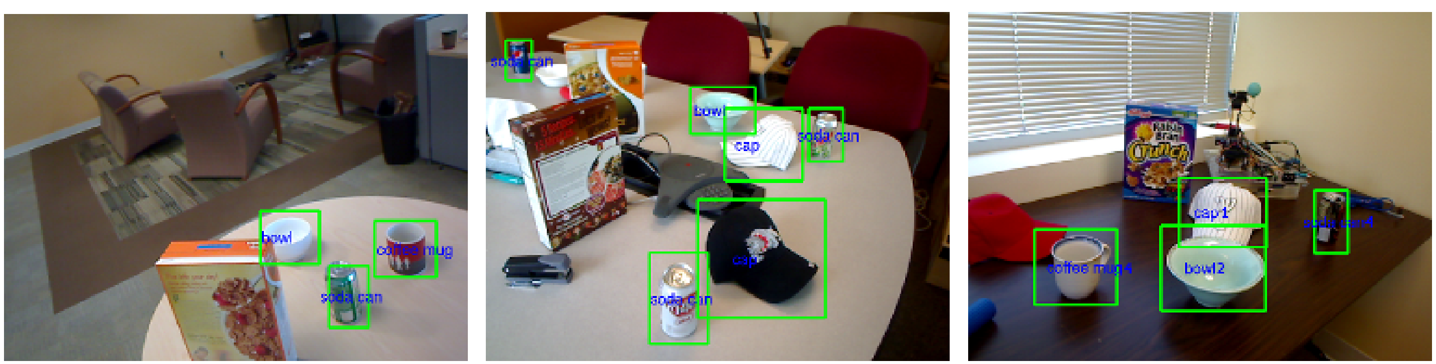
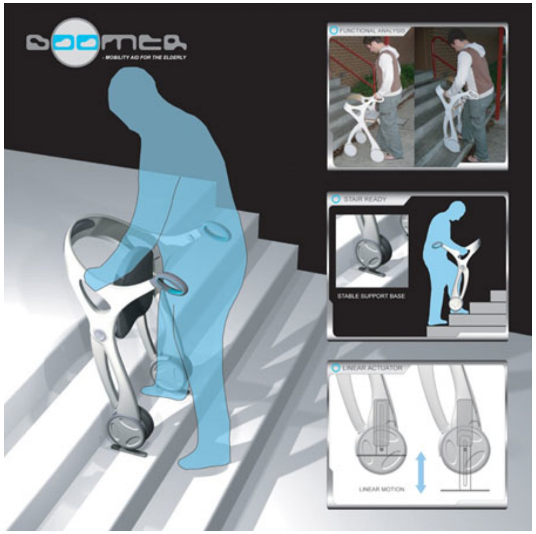


Fig. 11. Three detection results in multi-object scenes. From left to right, the first two images show multi-category detection results, while the last image shows multi-instance detection results.

Depth-based 3D pose estimation and action recognition



Project MOBOT: Robotic assistants for the elderly

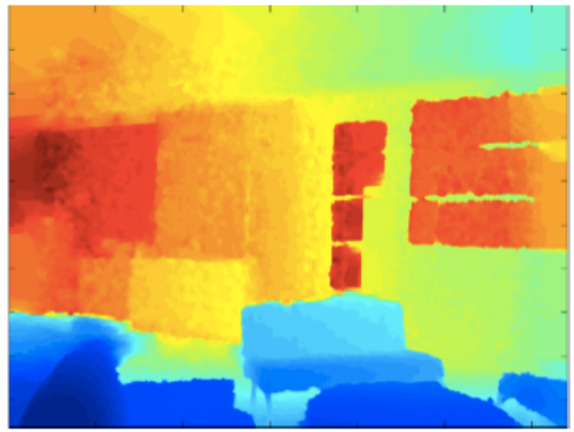
Partners: TU Munich (DE), ICCS (GR), U Heidelberg (DE)

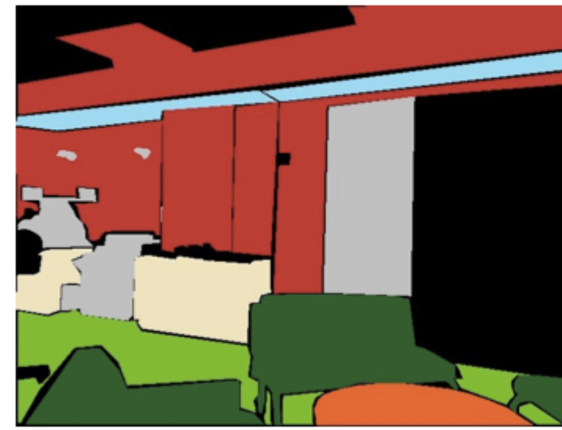
Rapid pose estimation using depth data (DTBB, Depth features, pose estimation)
Action recognition from multiple cues (depth, motion, intensity, detectors)

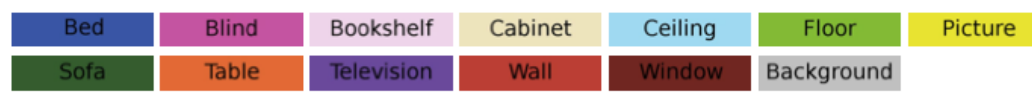
Fully funded for 3 years, starting as soon as February

Depth-based 3D object recognition and segmentation

RGB input | **Depth input** | **Desired symbolic output**





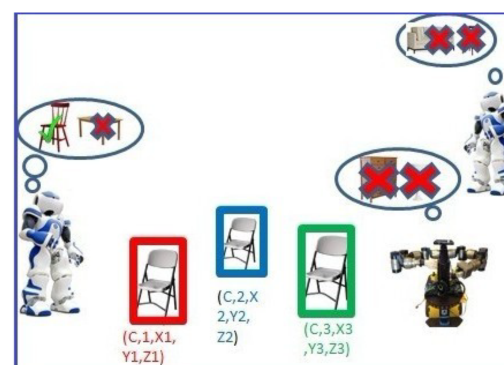


Project RECONFIG: Reconfigurable, heterogeneous multi-agent systems

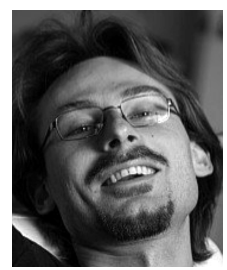
Partners: KTH (SE), U Aalto (FI) , NTUA (GR)

Multi-view object recognition using depth data
Segmentation, registration, tracking of object surfaces

Fully funded for 3 years, starting as soon as February



Internship proposal: Surface-based 3D recognition



Alex Bronstein
Tel-Aviv University

Michael Bronstein
USI Lugano

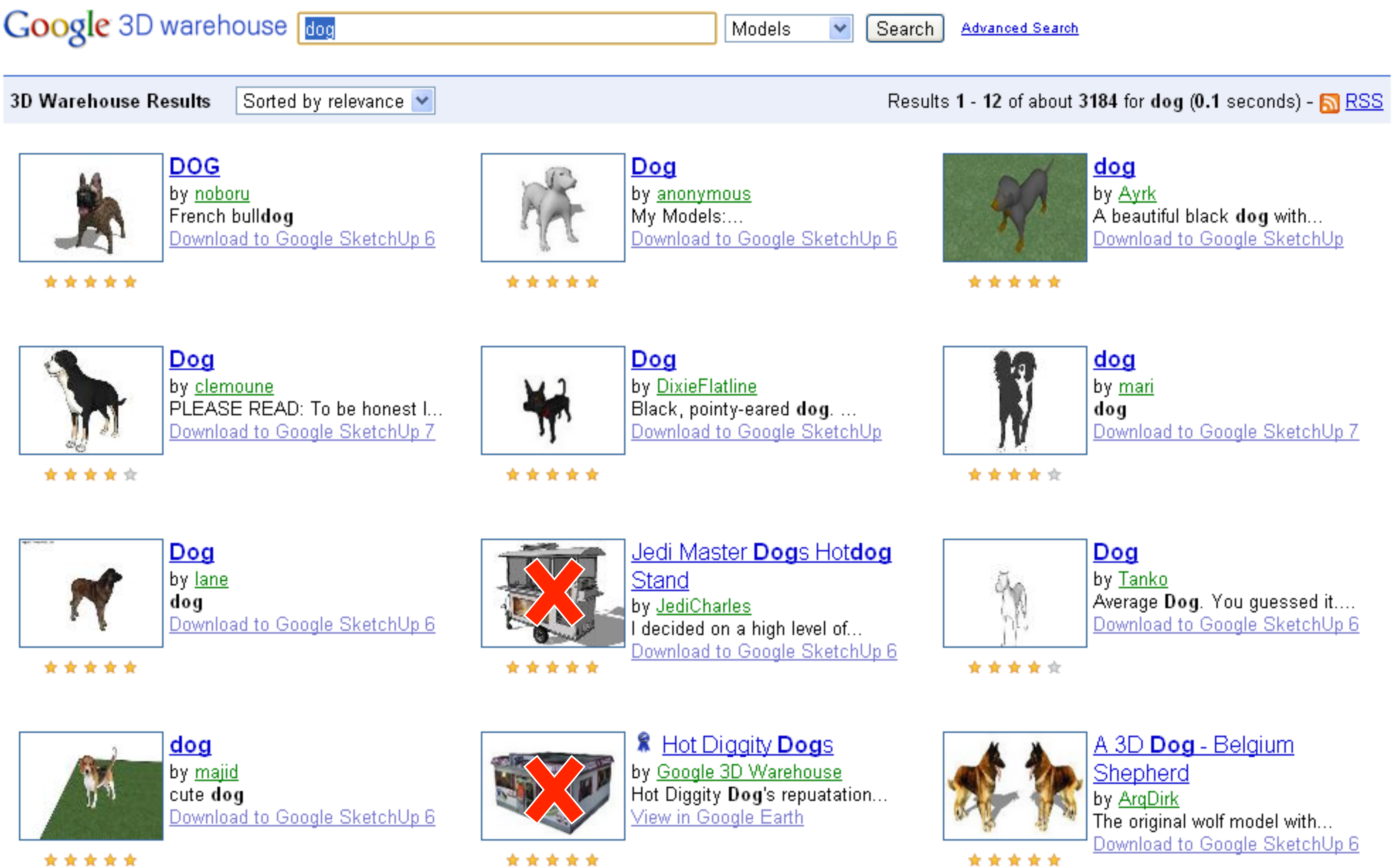
The next challenge

Text

Visual data

Geometric data

Shape retrieval today



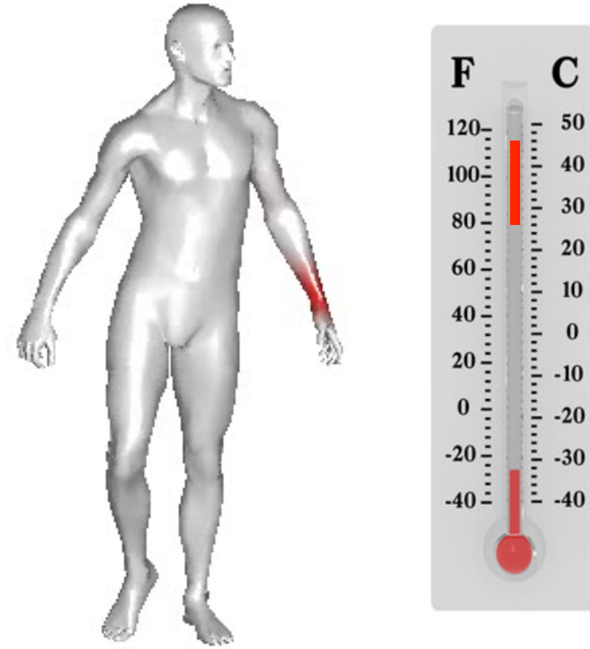
Diffusion geometry

Heat equation $$\left(\Delta_S + \frac{\partial}{\partial t}\right) u(t,x) = 0,$$

where

Δ_S - positive semidefinite Laplace-Beltrami operator

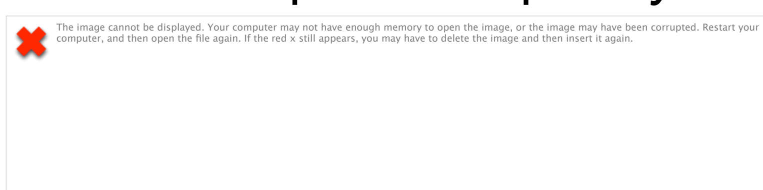
u - heat distribution



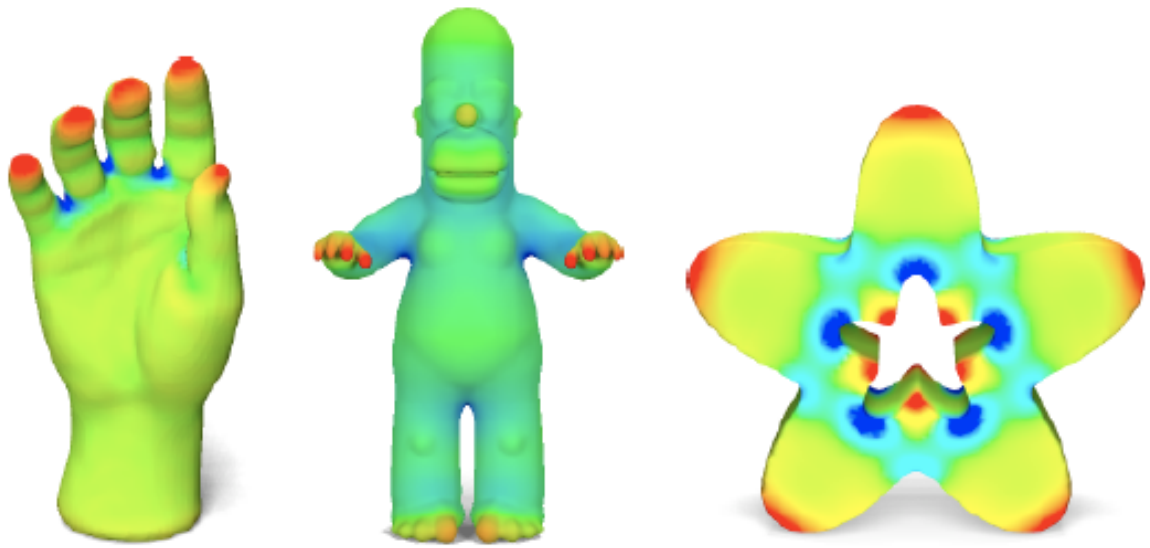
Fundamental solution (**heat kernel**, $k_t(x,y)$) – heat equation solution for initial conditions $u(x,0) = \delta(x-y)$

Amount of heat transferred from point *x* to point *y* in time *t*

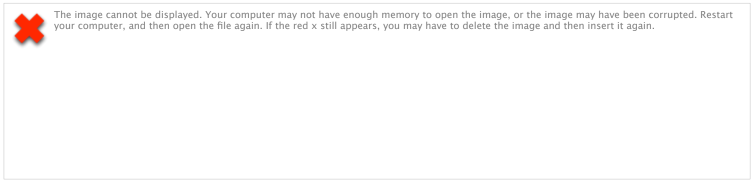
Spectral expression



Heat kernel interpretation



Geometric interpretation: "multiscale Gaussian curvature"

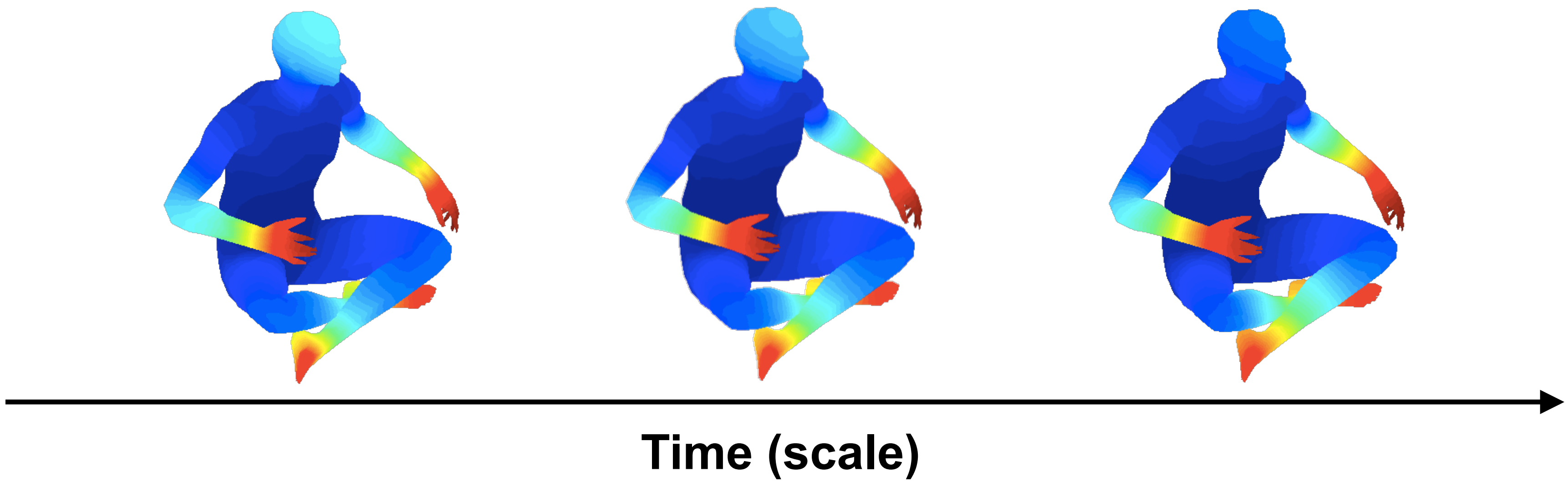


Probabilistic interpretation: the probability of a random walk to remain at point *x* after time *t*.

Sun, Ovsjanikov, Guibas, 2009

Heat kernel signature

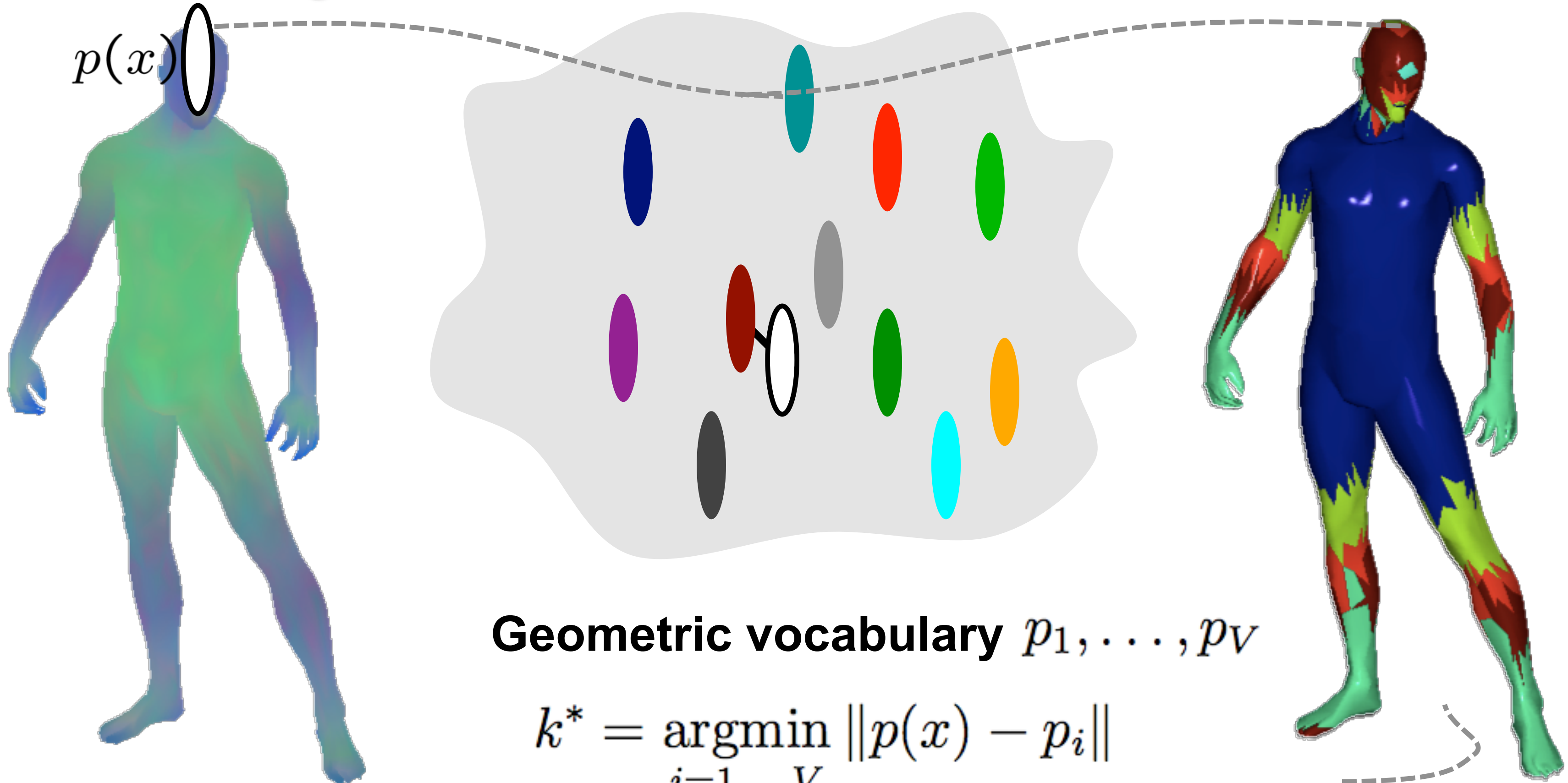
Multiscale descriptor $p(x) = (k_{t_1}(x,x), \ldots, k_{t_n}(x,x))$



- Intrinsic, hence deformation-invariant
- Provably informative
- Efficiently computable on different shape representations
- Multiscale

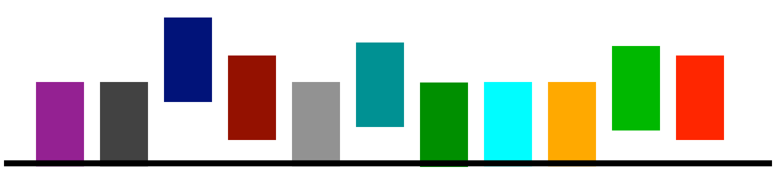
Sun, Ovsjanikov, Guibas, 2009

Shape Google



Geometric vocabulary $p_1, \ldots, p_V$

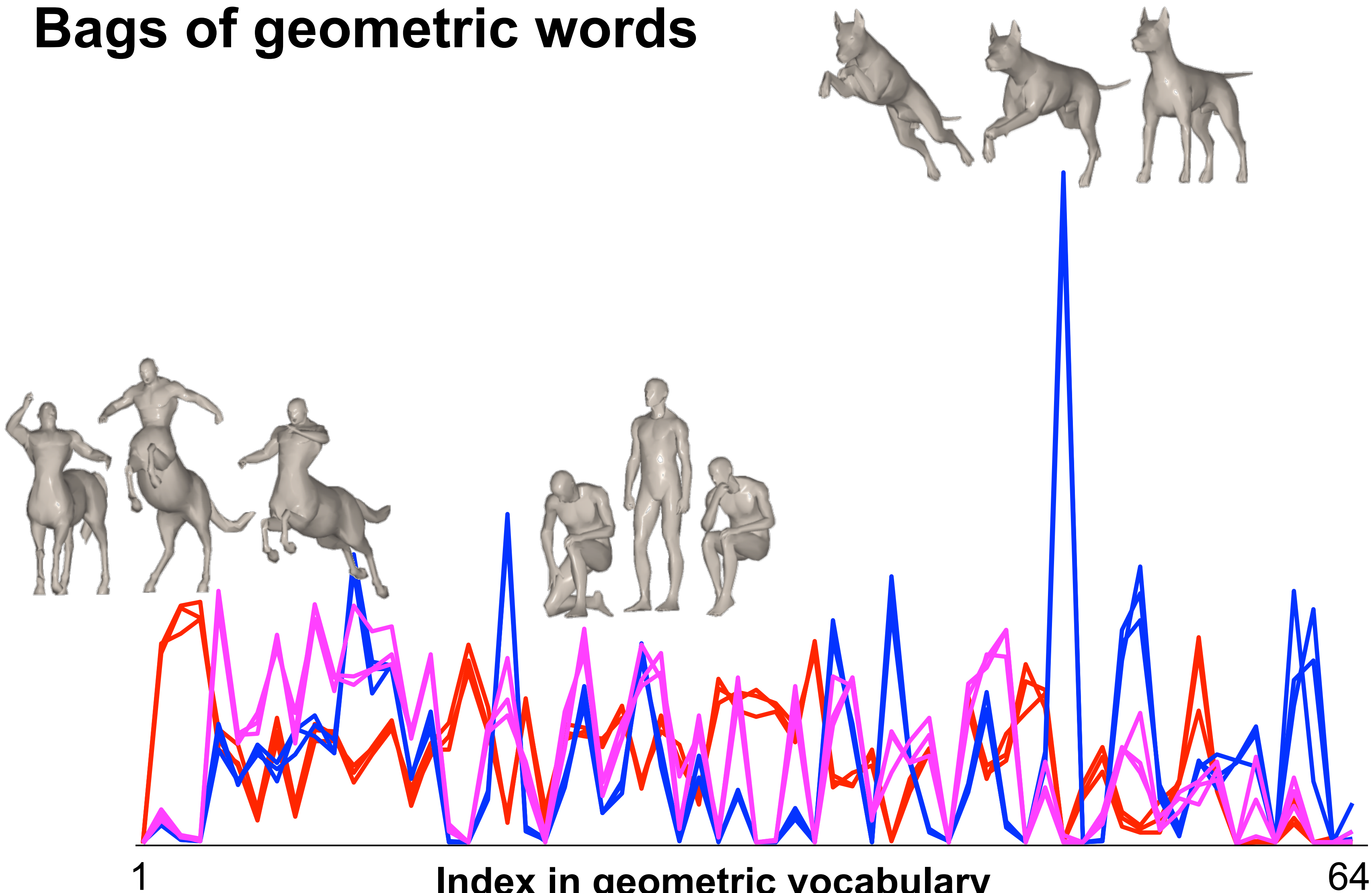
$$k^* = \underset{i=1,\ldots,V}{\operatorname{argmin}} \|p(x) - p_i\|$$



Bag of geometric words

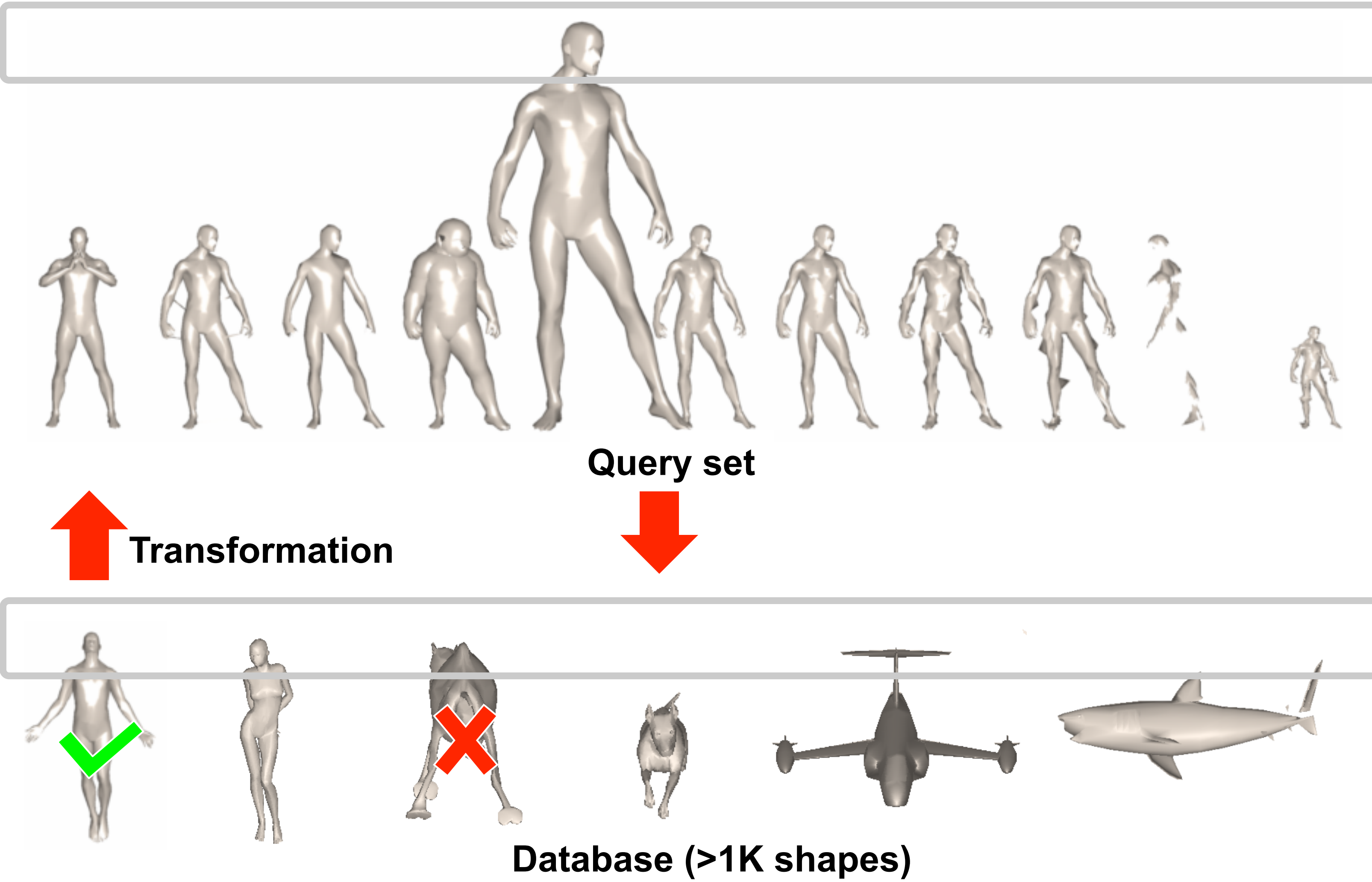
Ovsjanikov, BB, Guibas, 2009
BB. Ovsjanikov, Guibas 2010

Bags of geometric words



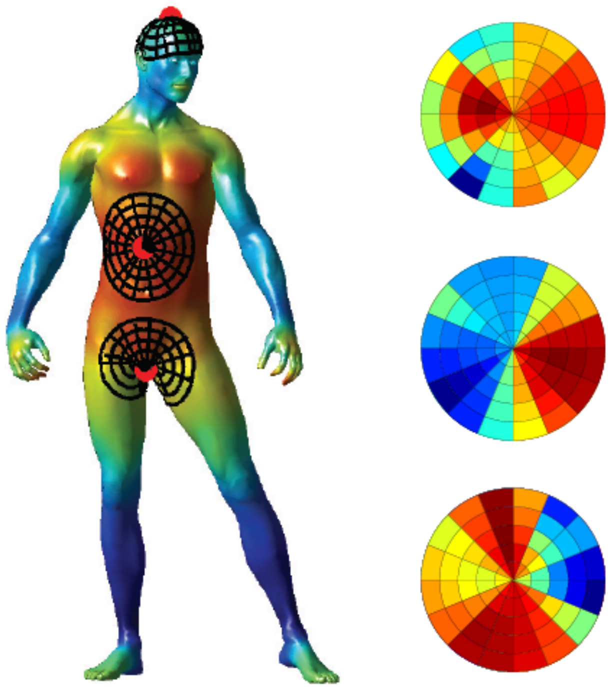
Ovsjanikov, BB, Guibas, 2009
BB. Ovsjanikov, Guibas 2010

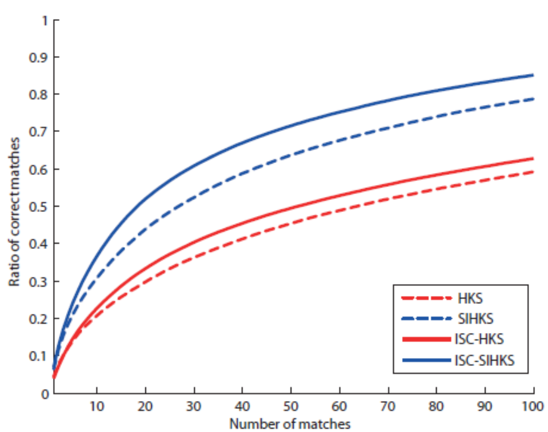
SHREC 2010: Robust shape retrieval benchmark



B *et al.* 2010

Intrinsic Shape Contexts



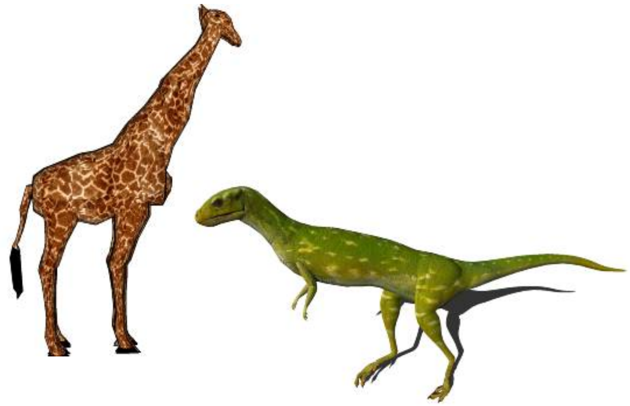
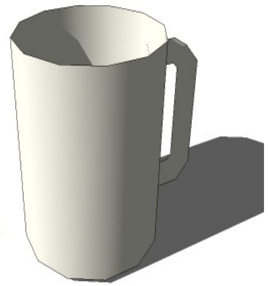


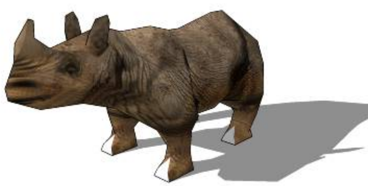
Pointwise ISC

Kokkinos, Bronstein, Bronstein, CVPR 2012


How Kinect Views Your Living Room

http://sketchup.google.com/3dwarehouse/

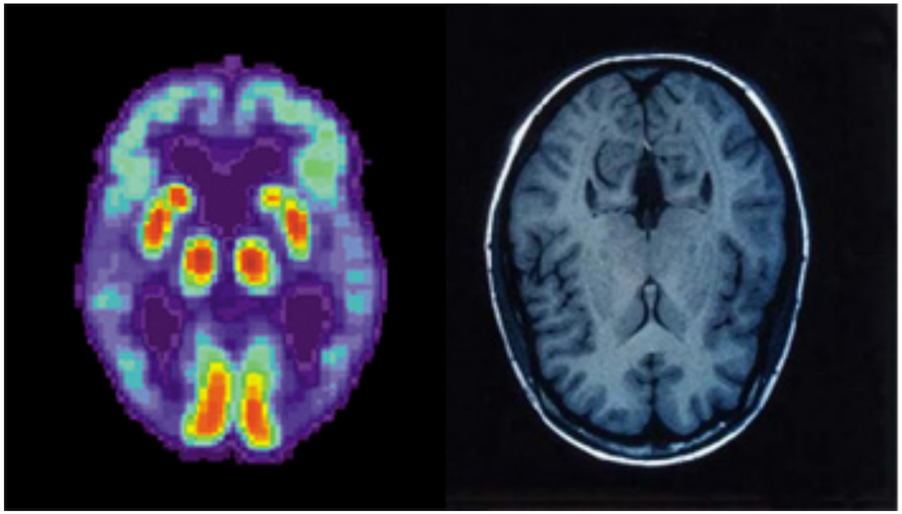






How to connect these?

Comparing apples with oranges: metric learning



G. Shakhnarovich. Learning task-specific similarity. PhD thesis, MIT, 2005.

M. M. Bronstein, A. M. Bronstein, F. Michel, N. Paragios, "Data fusion through cross-modality metric learning using similarity-sensitive hashing", CVPR, 2010.

E. Xing, A. Ng, M. Jordan, and S. Russell. Distance metric learning with application to clustering with side-information. NIPS, 2003.