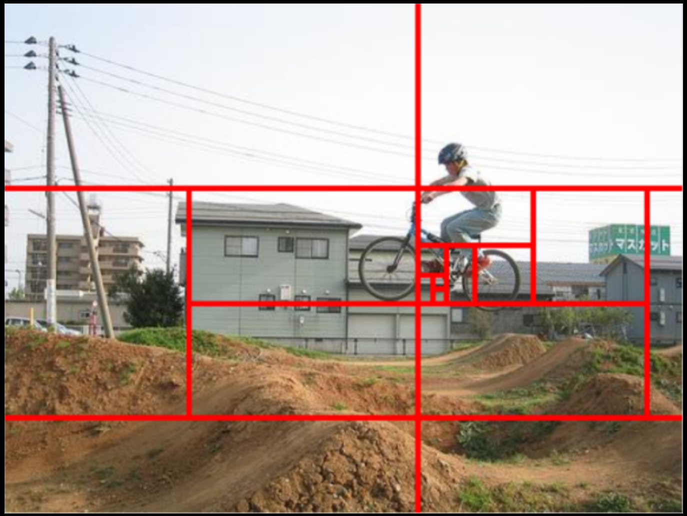
Rapid Deformable Object Detection using Branch-and-Bound



IPAM -UCLA
7 Aug 2013

Iasonas Kokkinos
Center for Visual Computing
Ecole Centrale Paris / INRIA Saclay

Sliding window classifiers

0.1

Sliding window classifiers

-0.2

Sliding window classifiers

...
1.5
...

Sliding window classifiers

0.5

Sliding window classifiers

0.3

Sliding window classifiers

0.1
-0.2
-0.1
0.1
. . .
1.5
. . .
0.5
0.4
0.3

Search/Coarse-to-fine & Object Recognition

E. Grimson. *Object Recognition by Computer.* MIT Press, 1991.

T. M. Breuel, “Fast recognition using adaptive subdivisions of transformation space,” in CVPR, 1992.

D. P. Huttenlocher, G. A. Klanderman, and W. A. Rucklidge, “Comparing images using the Hausdorff distance,” PAMI 1993

Y. Amit and D. Geman, "A computational model for visual selection," Neural Computation, 1999

Y. Amit, D. Geman and B. Jedynak, " Efficient focusing and face detection," 1999

F. Fleuret and D. Geman. Coarse-to-fine face detection. *IJCV,* 2001.

Y. Amit, D. Geman and X. Fan, A coarse-to-fine strategy for multi-class shape detection, PAMI, 2004

P. Viola and M. Jones, “Rapid object detection using a boosted cascade of simple features,” CVPR, 2001.

P. Moreels, M. Maire, and P. Perona. Recognition by probabilistic hypothesis construction. *ECCV*, 2004.

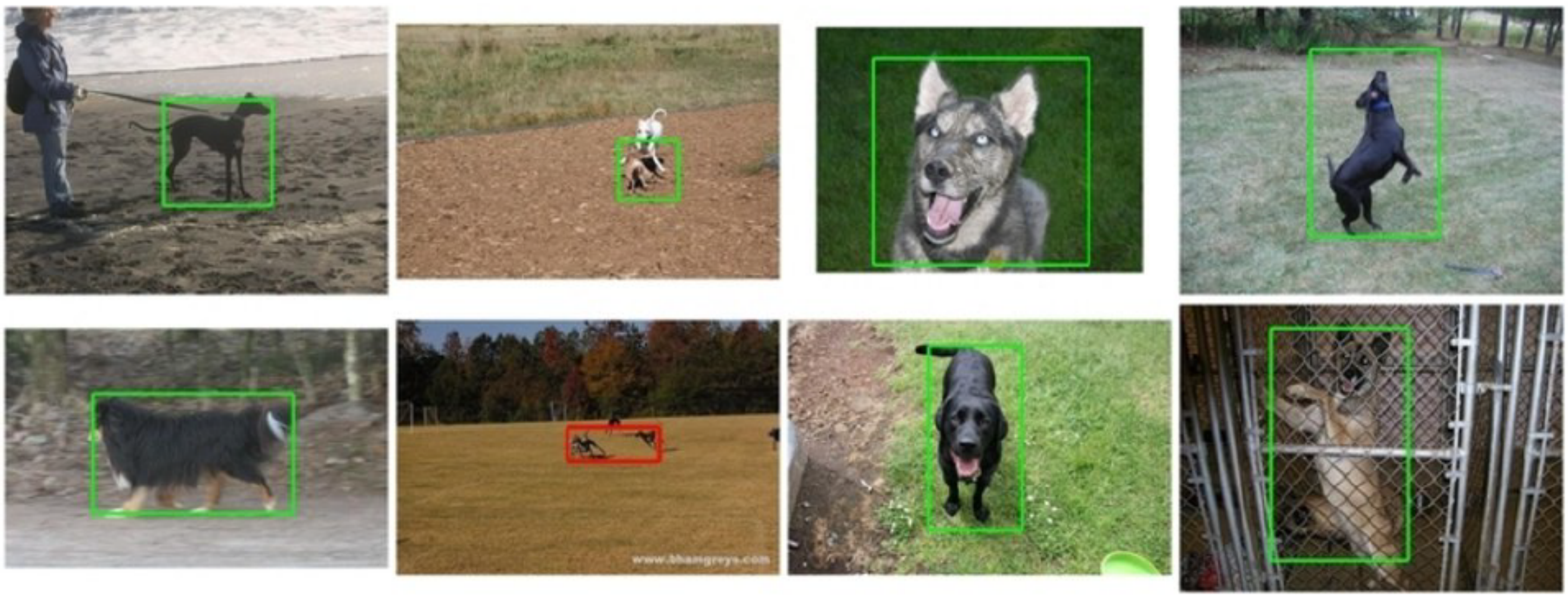
C. Lampert, M. Blaschko, and T. Hofmann. Beyond sliding windows: *CVPR,* 2008.

C. H. Lampert. An efficient divide-and-conquer cascade for nonlinear object detection. In *CVPR,* 2010.

A. Lehmann, B. Leibe, and L. V. Gool. Fast PRISM: Branch and Bound Hough Transform, IJCV, 2011.

BoW: Efficient Subwindow Search (ESS), 2008

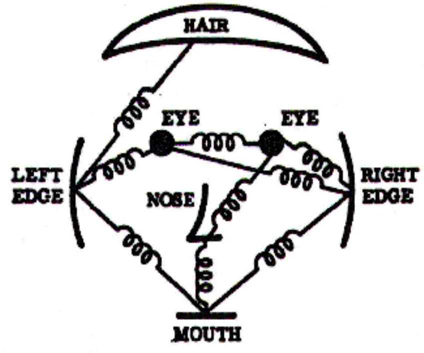
- High localization quality: first place in 5 of 20 categories.
- High speed: $\approx 40ms$ per image (excl. feature extraction)

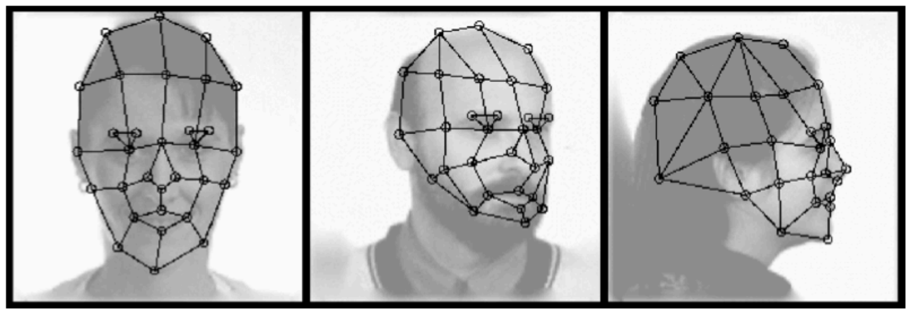


Example detections on VOC 2007 dog.

Christoph Lampert, Matthew Blaschko, Thomas Hoffman, Beyond Sliding Windows, CVPR 2008

Deformable Part Models (DPMs)





M. Fischler and R. Erschlanger. The Representation and Matching of Pictorial Structures '73.
M. Lades, et al: Distortion Invariant Object Recognition in the Dynamic Link Architecture. '93
Y. Amit, A. Kong: Graphical Templates for Model Registration. '96
A. L. Yuille, J. M. Coughlan: An A perspective on deterministic optimization for deformable templates. '00*
M. C. Burl, P. Perona: Recognition of Planar Object Classes. '96
M. C. Burl, M. Weber, P. Perona: A Probabilistic Approach to Object Recognition Using Local Photometry and Global Geometry. '98
M. Weber, M. Welling, P. Perona: Unsupervised Learning of Models for Recognition. '00
P. Felzenszwalb, and D. Huttenlocher, Pictorial Structures for Object Recognition, IJCV '05
P. Felzenszwalb, et. al., Object Detection with Discriminatively Trained DPMs,'10

Previous works on acceleration

Sparse image representations:

Y. Chen, L. Zhu, A. Yuille. Rapid inference on a novel and/or graph for object detection, NIPS 2007

I. Kokkinos and A. Yuille. HOP: Hierarchical Object Parsing, CVPR, 2009

Dense image representations:

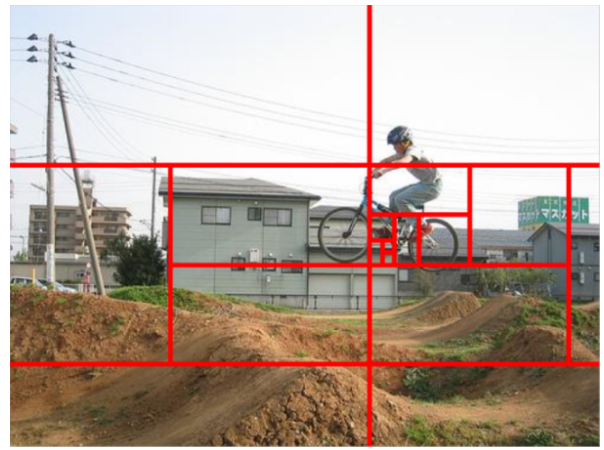
B. Sapp, A. Toshev, and B. Taskar. Cascaded models for articulated pose estimation, ECCV, 2010

M.Pedersoli, A.Vedaldi, and J.Gonzalez. A coarse-to-fine approach for object detection, CVPR 2011

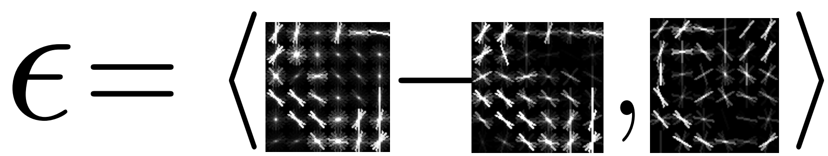
P F Felzenszwalb R B Girshick and D A McAllester Cascade object detection with DPMs CVPR 2010

Talk outline

Part 1: Branch and Bound for DPMs



Part 2: Bounds for part scores

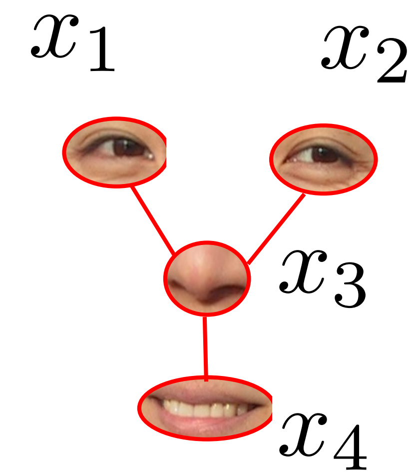


I. K., Rapid DPM detection using Dual Tree Branch-and-Bound, NIPS, 2011/INRIA TR 2012

I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

http://vision.mas.ecp.fr/Personnel/iasonas/dpms.html

Deformable Part Models (DPMs)



$$E(x) = \sum_i m_i(x_i) + \sum_{i,j \in E} \phi_{i,j}(x_i, x_j)$$

Local appearance **Pairwise compatibility**

Local appearance

$$m_i(x_i) = \langle w_i, H(x_i) \rangle$$

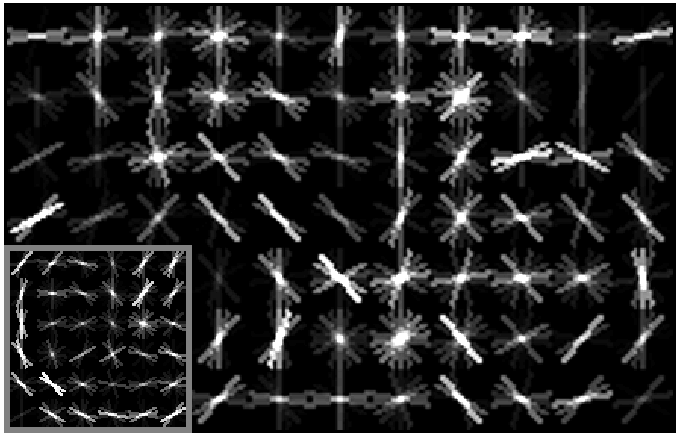
Pairwise compatibility

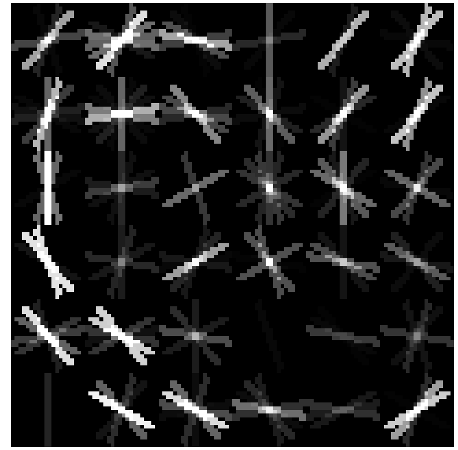
$$\phi_{i,j}(x_i, x_j) = -(h_i - h_j - \hat{h}_{i,j})^2 \eta - (v_i - v_j - \hat{v}_{i,j})^2 \nu$$

P. Felzenszwalb, and D. Huttenlocher, Pictorial Structures for Object Recognition, IJCV 2005
P. Felzenszwalb, et. Al., Object Detection with Discriminatively Trained Part-Based Models, 2010

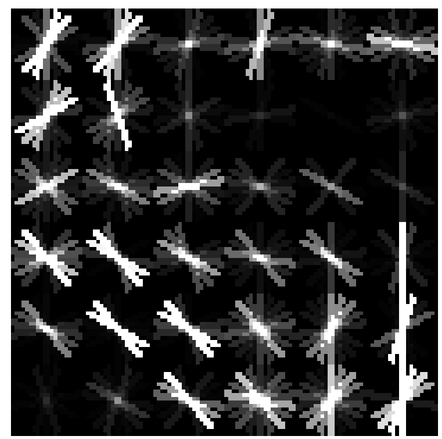
Part score computation

$$E(x) = \sum_i \boxed{m_i(x_i)} + \sum_{i,j \in E} \phi_{i,j}(x_i, x_j)$$





$\mathbf{w}[y]$

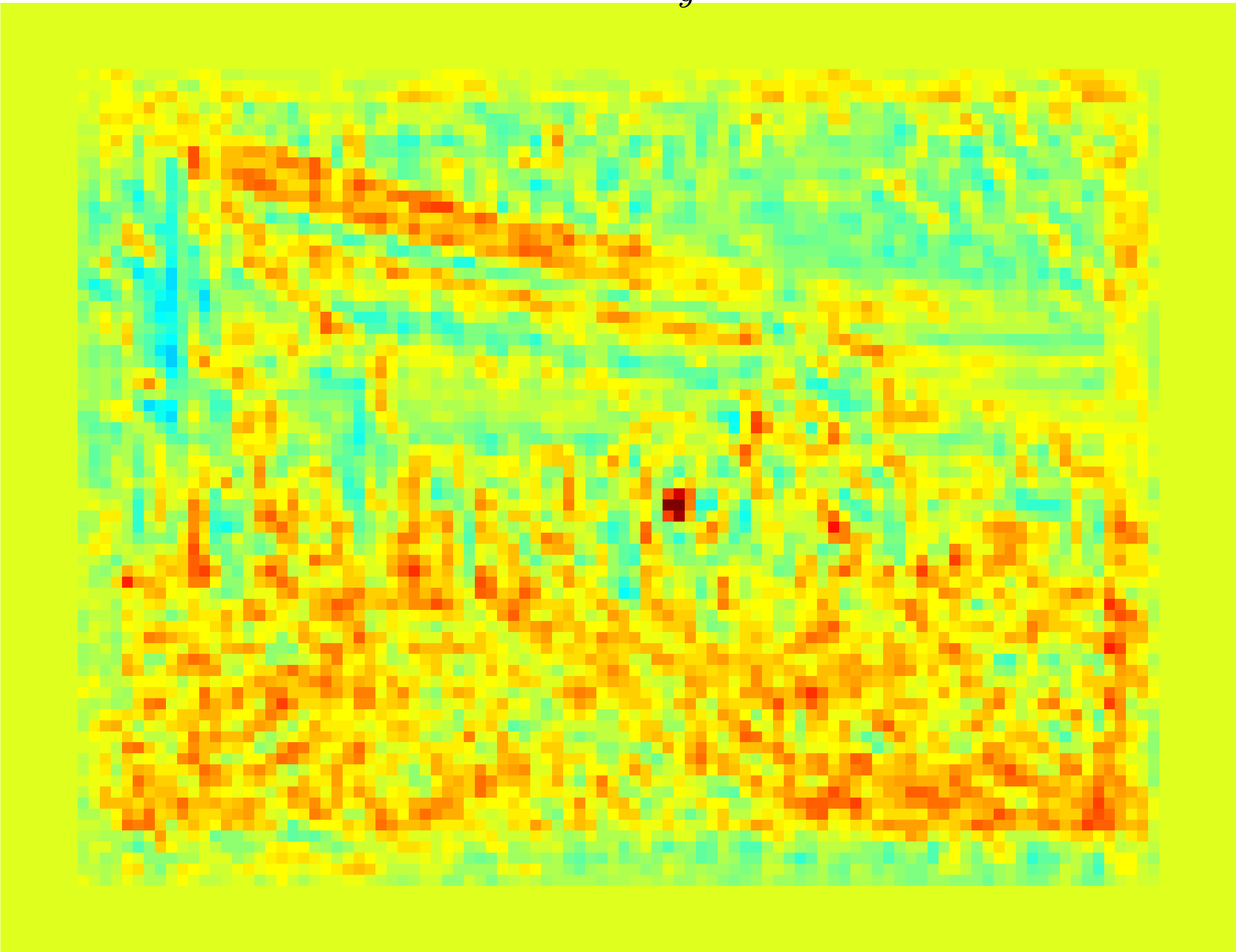


$\mathbf{h}[x+y]$

$$s[x] = \sum_y \langle \mathbf{h}[x+y], \mathbf{w}[y] \rangle$$

Part score

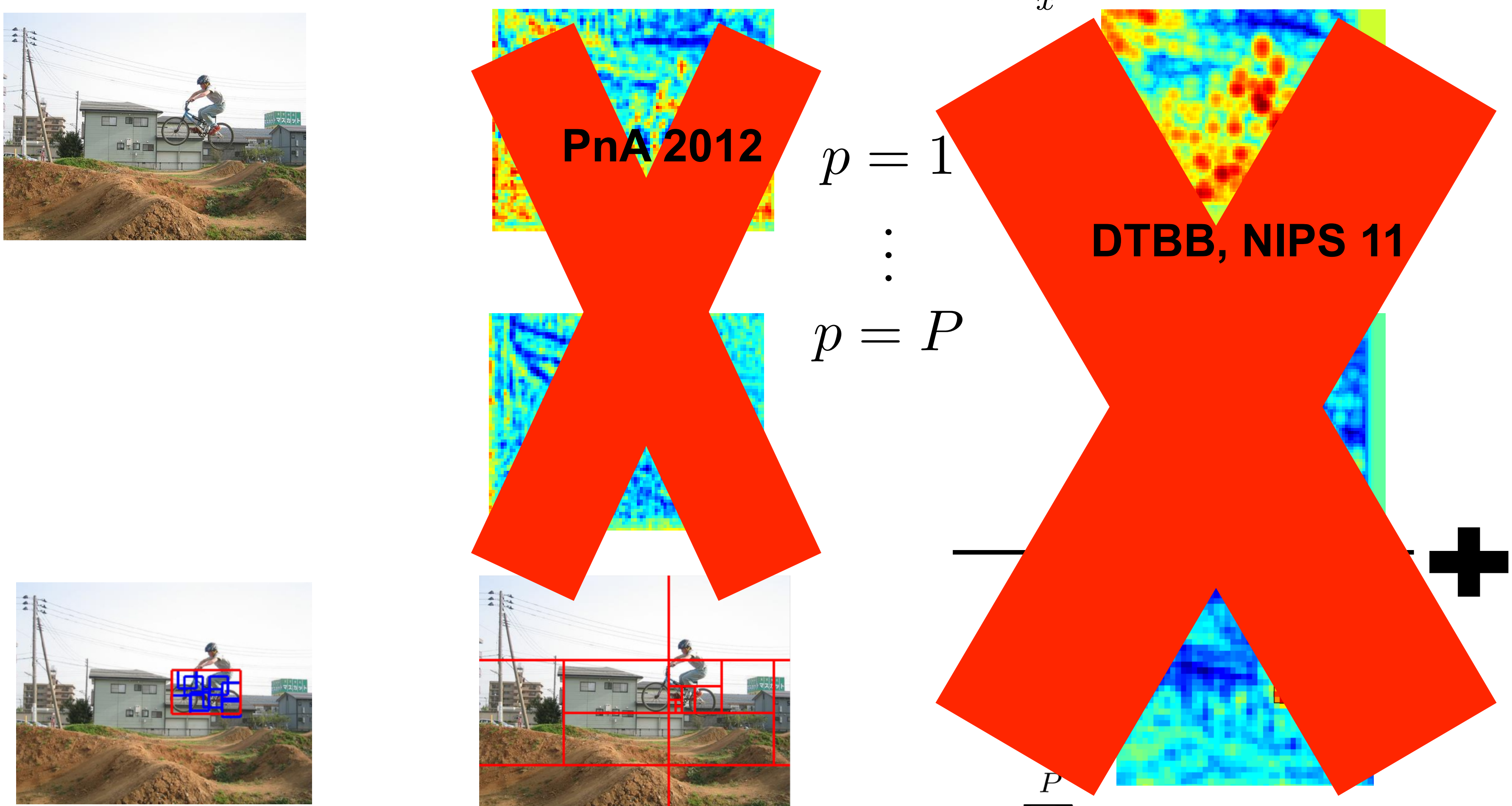
$$\mathbf{h}[x] \qquad s[x] = \sum_y \langle \mathbf{h}[x+y], \mathbf{w}[y] \rangle$$



Object detection with DPMs

$U_p(x') = \langle \mathbf{w}_p, \mathbf{H}(x') \rangle$

$\max_{x'}[U_p(x') + B_p(x, x')]$



$$S(x) = \sum_{p=1}^{P} \max_{x'}[U_p(x') + B_p(x, x')]$$

I. K., Rapid DPM detection using Dual Tree Branch-and-Bound, NIPS, 2011/INRIA TR 2012
I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Object detection complexity

Score for object hypothesis: $$S(x) = \sum_{p=1}^{P} \max_{x'} [U_p(x') + B_p(x, x')]$$

Threshold-based detection: recover all locations above threshold

$$M^\theta = \{x : S(x) \geq \theta\}$$

1-best detection: recover top-scoring candidate

$$M^* = \{\arg\max_x S(x)\}$$

Naïve implementation (Dynamic Programming) : $O(PN^2)$ $\quad N = |\{x\}|$

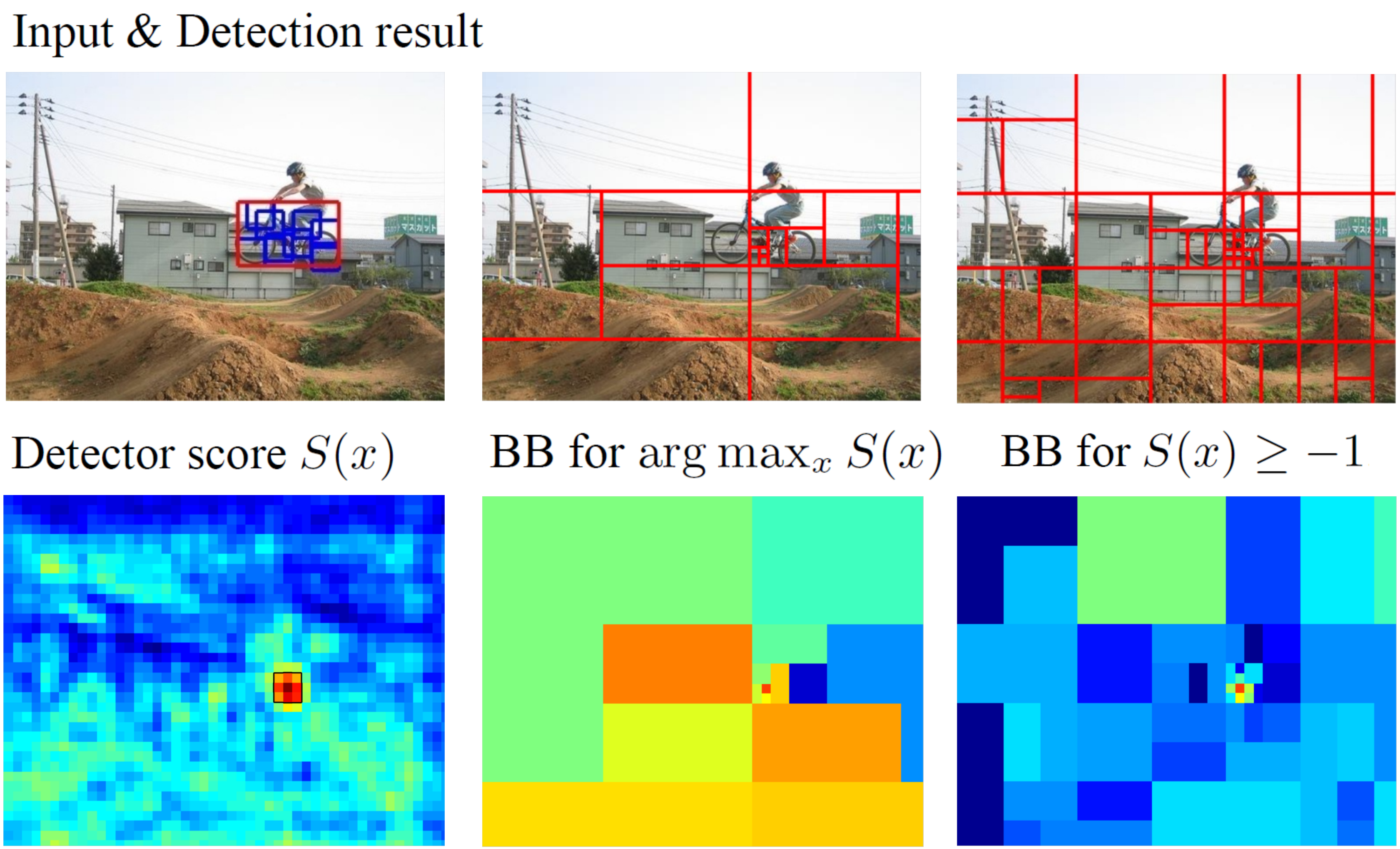
Generalized Distance Transforms (GDT): $O(PN)$

Our work (best-case): $\simeq |M| P \log_2 N$

Key idea: avoid the exact evaluation of S(x)

I. K., Rapid DPM detection using Dual Tree Branch-and-Bound, NIPS, 2011/INRIA TR 2012

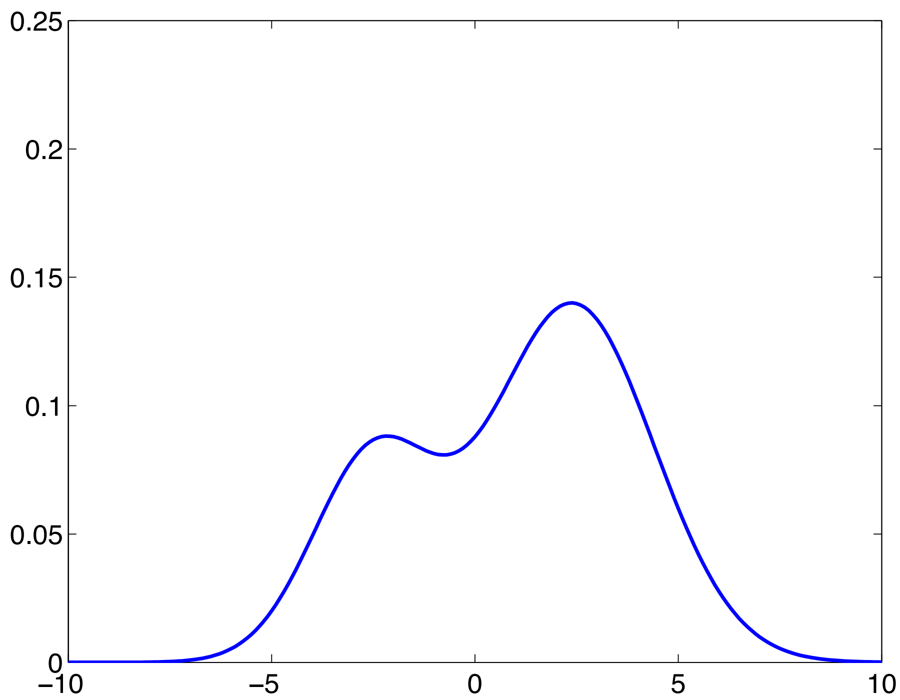
Branch-and-Bound for Deformable Part Models



I. K., Rapid DPM detection using Dual Tree Branch-and-Bound, NIPS, 2011/INRIA TR 2012

Branch and bound technique

- Task: $$\max_{x \in X} f(x)$$
- Bounding function: $$\overline{f}(X) \geq \max_{x \in X} f(x), \quad \overline{f}(\{x\}) = f(x)$$



Branch-and-bound algorithm

Branch-and-Bound

$M^* = BB(X_0, \overline{S})$

INITIALIZE: $\mathcal{Q} = \{(X_0, \overline{S}(X_0)\}$

while 1 **do**

  $X = \text{Pop}[\mathcal{Q}]$

  if $\text{Singleton}[X]$ **then**

    RETURN X {First singleton: best X}

  end if

  $[X_1, X_2] = \text{Branch}[X]$

  $\text{Push}[\mathcal{Q}, (X_1, \overline{S}(X_1))]$, $\text{Push}[\mathcal{Q}, (X_2, \overline{S}(X_2))]$

end while

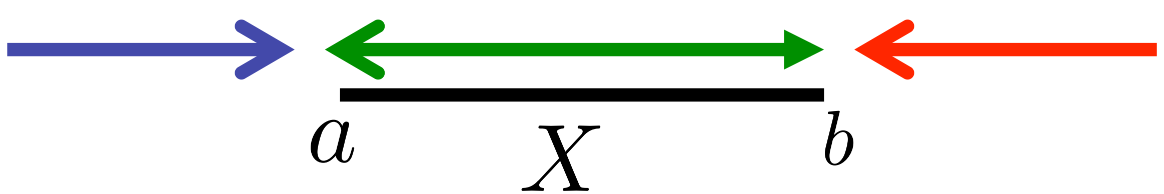
Bounding a mixture-of-gaussians

- Property: $$\max_{x \in X} h(x) + g(x) \leq \max_{x \in X} h(x) + \max_{x \in X} g(x)$$
- Function: $$f(x) = \pi_1 N(x; \mu_1, \sigma_1) + \pi_2 N(x; \mu_2, \sigma_2)$$

$$\max_{x \in X} f(x) \leq \max_{x \in X} [\pi_1 N(x; \mu_1, \sigma_1)] + \max_{x \in X} [\pi_2 N(x; \mu_2, \sigma_2)]$$

$$= \pi_1 N(d(X, \mu_1, \sigma_1); 0, 1) + \pi_2 N(d(X, \mu_2, \sigma_2); 0, 1)$$

$$d(X, \mu, \sigma) = \begin{cases} 0 & \boxed{a \leq \mu \leq b} \\ \frac{1}{\sigma^2}(\mu - a)^2 & \boxed{\mu \leq a} \\ \frac{1}{\sigma^2}(\mu - b)^2 & \boxed{b < \mu} \end{cases}$$



Bounding the DPM cost function

Goal: $$\overline{S}(X) \geq S(X) = \max_{x \in X} S(x) \quad \forall X, \quad \overline{S}(\{x\}) = S(x)$$

where $$S(x) = \sum_{p=1}^{P} \underbrace{\max_{x'}[U_p(x') + B_p(x, x')]}_{m_p(x)}$$

$$\max_{x \in X} \sum_{p=1}^{P} m_p(x) \leq \sum_{p=1}^{P} \max_{x \in X} m_p(x)$$

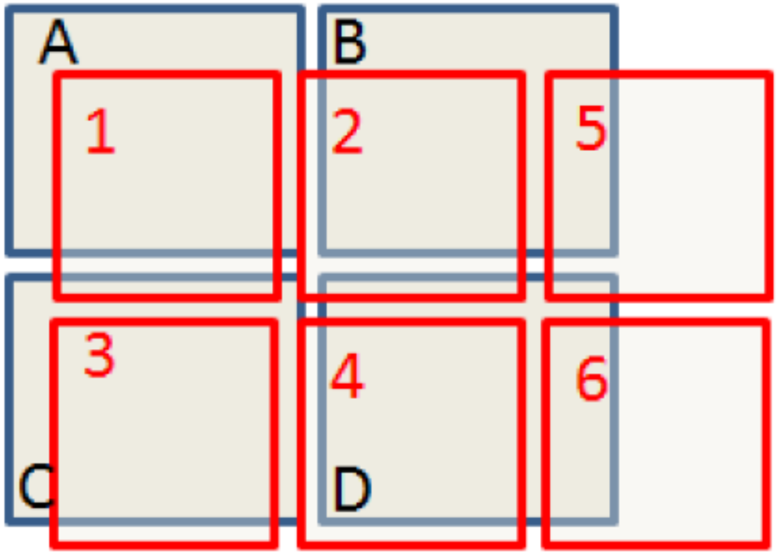
$$\max_{x \in X} S(x) \leq \sum_{p=1}^{P} \overline{m}_p(X)$$

Focus on constructing $\overline{m}_p(X)$

Breaking up the message bound

$$m(X) \doteq \max_{x \in X} \max_{x' \in X'} U(x') + B(x', x)$$

Partition domains: $X = \cup_{d \in D} X_d, \quad X' = \cup_{s \in S} X_s$



Define: $\mu_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} U(x') + B(x', x)$

$$m(X) = \max_d \max_s \mu_d^s$$

Bounding the source-to-domain contributions

$$\mu_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} U(x') + B(x', x)$$

Upper bound:

$$\mu_d^s \leq \max_{x' \in X_s} U(x') + \max_{x \in X_d} \max_{x' \in X_s} B(x', x)$$

Intuitively: take best source point, put it at best possible source location

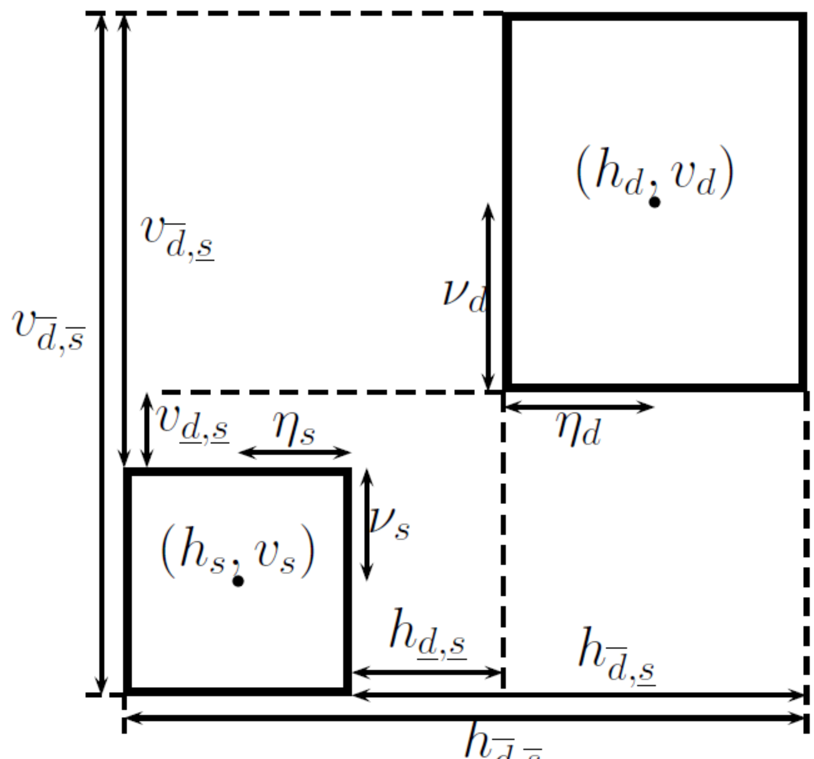
Quantities involved: $\overline{u}^s \doteq \max_{x' \in X_s} U(x'), \quad \overline{g}_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} B(x', x)$

Bounding the binary term

$$\mathcal{G}_{\overline{d},\overline{s}} \doteq \max_{x \in X_d} \max_{x' \in X_s} \mathcal{G}_{x,x'} \qquad \mathcal{G}_{x,x'} = -H(h-h')^2 - V(v-v')^2$$

$$= -H h^2_{\underline{d},\underline{s}} - V v^2_{\underline{d},\underline{s}}$$

$$h_{\underline{d},\underline{s}} \doteq \min_{h \in X^h_d} \min_{h' \in X^h_s} |h-h'| \qquad v_{\underline{d},\underline{s}} \doteq \min_{v \in X^v_d} \min_{v' \in X^c_s} |v-v'|$$



Done so far

$$m(X) \doteq \max_{x \in X} \max_{x' \in X'} U(x') + B(x', x)$$

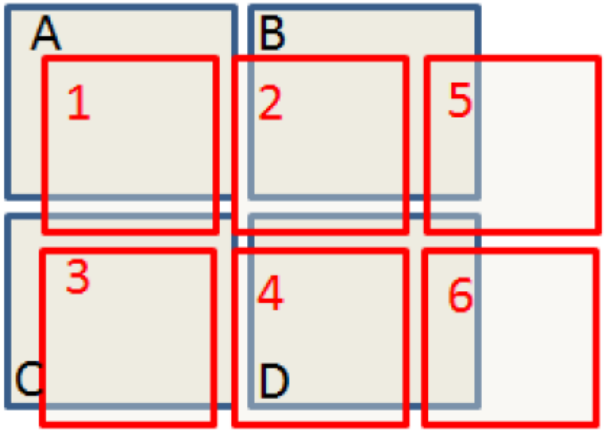
$$X = \cup_{d \in D} X_d, \quad X' = \cup_{s \in S} X_s$$

$$\mu_d^s \doteq \max_{x \in X_d} \max_{x' \in X_s} U(x') + B(x', x)$$

$$m(X) = \max_d \max_s \mu_d^s$$

$$\overline{m}(X) = \max_d \max_s \overline{\mu}_d^s$$

$$X = \cup_{d \in D} X_d, \quad X' = \cup_{s \in S} X_s$$



Large intervals: loose bounds

Small intervals: tight bounds, but maximization over multiple terms

$$|S| = \frac{|X|}{|X_s|}$$

Need to prune

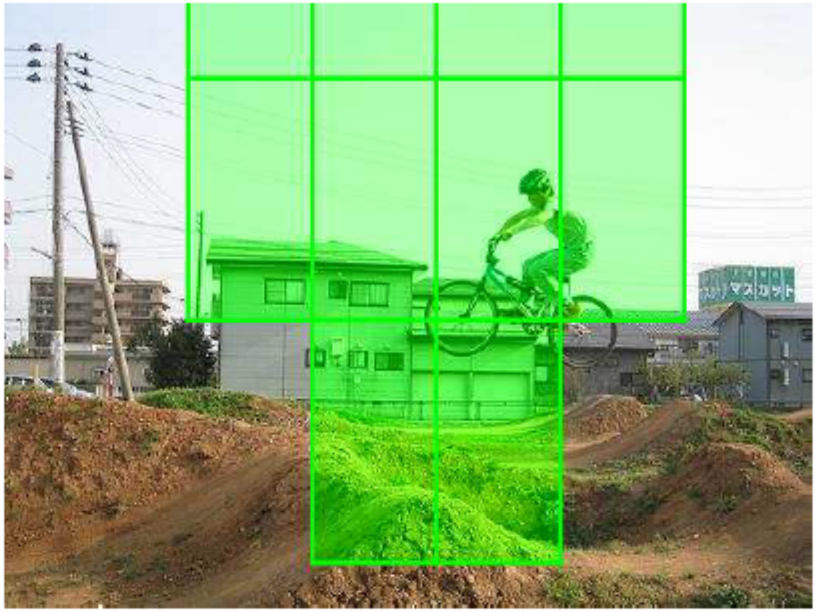
A.G. Gray and A.W. Moore. Nonparametric density estimation: Toward computational tractability. ICDM 2003

Supporter pruning demonstration

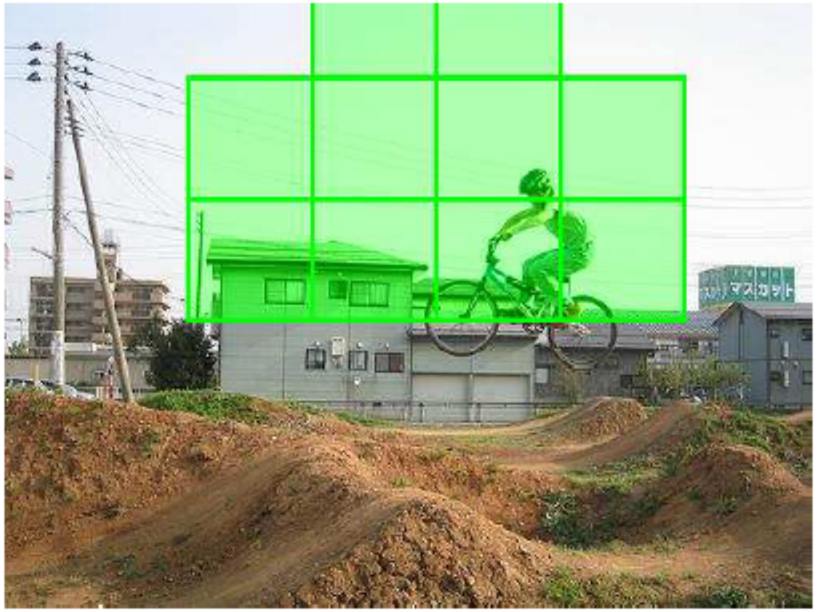
Supporter pruning demonstration

Supporter pruning demonstration

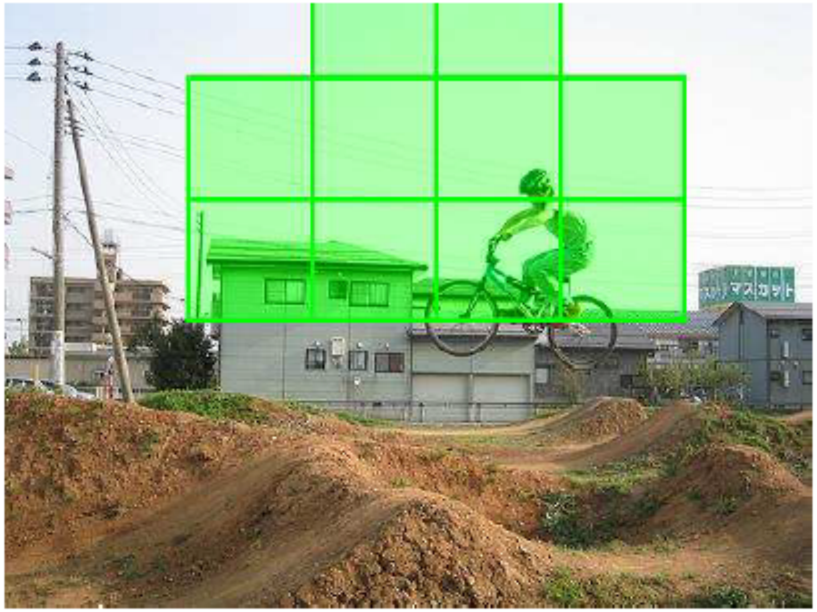
Supporter pruning demonstration



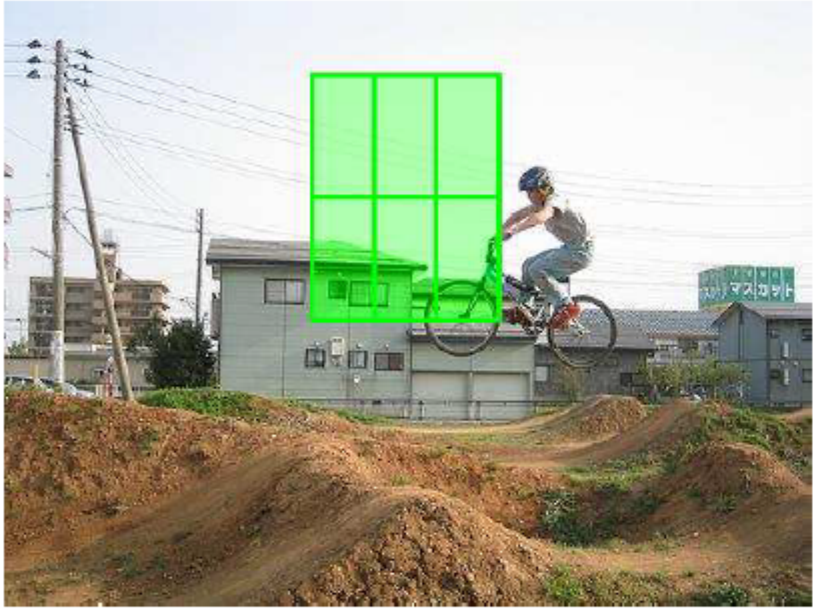
Supporter pruning demonstration

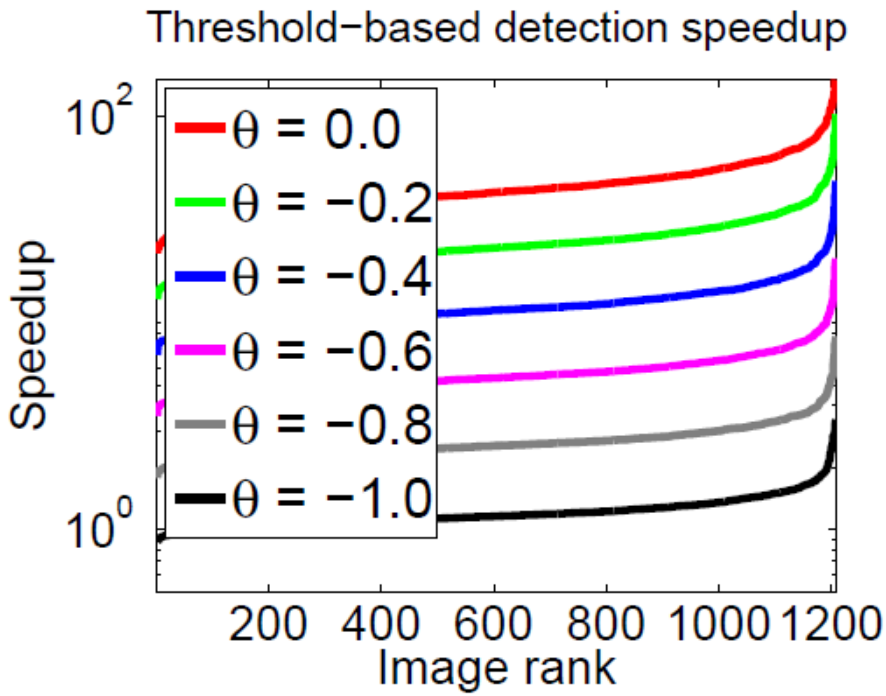
Supporter pruning demonstration



Supporter pruning demonstration



Results on Pascal VOC



1200 images, 20 Categories per image

Results on Pascal VOC

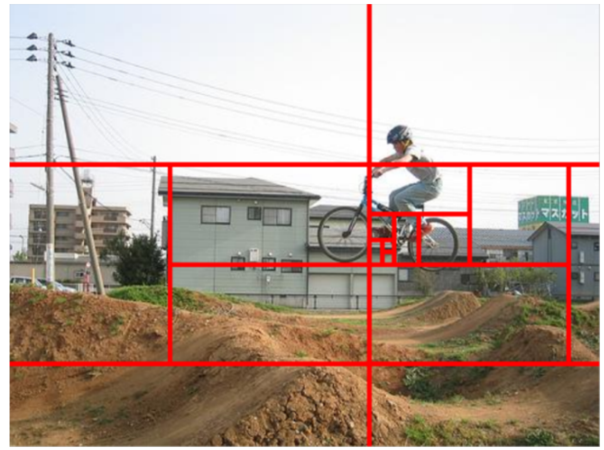
1200 images, 20 Categories per image

	Our algorithm	[4]
Unary terms	13.20 ± 1.49	159.41 ± 15.82
KD-trees	1.72 ± 0.21	0.00 ± 0.00
Detection, $\theta = 0.0$	0.25 ± 0.07	10.74 ± 1.02
Detection, $\theta = -.2$	0.47 ± 0.12	10.74 ± 1.02
Detection, $\theta = -.4$	0.93 ± 0.22	10.74 ± 1.02
Detection, $\theta = -.6$	1.95 ± 0.42	10.74 ± 1.02
Detection, $\theta = -.8$	4.17 ± 0.84	10.74 ± 1.02
Detection, $\theta = -1$	9.14 ± 1.79	10.74 ± 1.02
Detection, 1-best	0.41 ± 0.08	10.74 ± 1.02
Detection, 5-best	0.47 ± 0.09	10.74 ± 1.02
Detection, 10-best	0.48 ± 0.10	10.74 ± 1.02

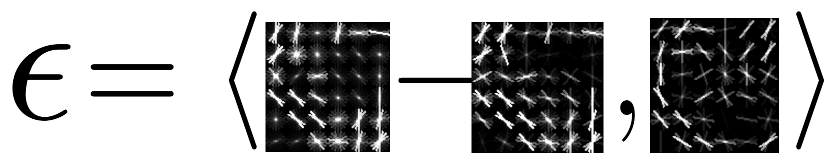
[4] P.F. Felzenszwalb, R. Girshick, D. A. McAllester, and D. Ramanan, Object Detection with Discriminatively Trained Part-Based Models, 2010

Talk outline

Part 1: B&B for DPMs



Part 2: Bounds for part scores

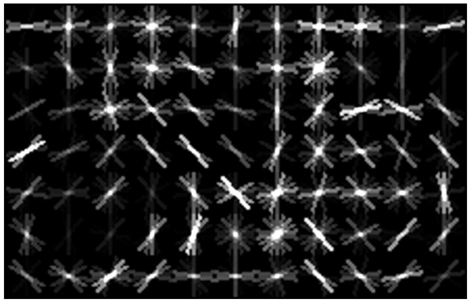


I. K., Rapid DPM detection using Dual Tree Branch-and-Bound, NIPS, 2011/INRIA TR 2012

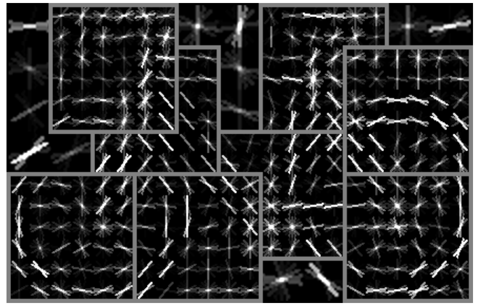
I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

http://vision.mas.ecp.fr/Personnel/iasonas/dpms.html

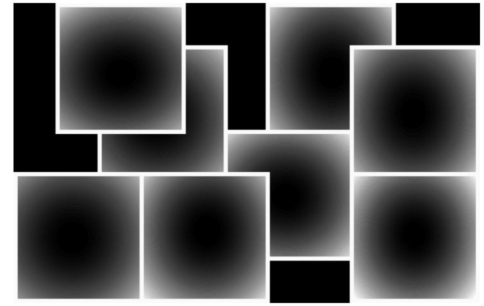
Bounding the part scores



w_1



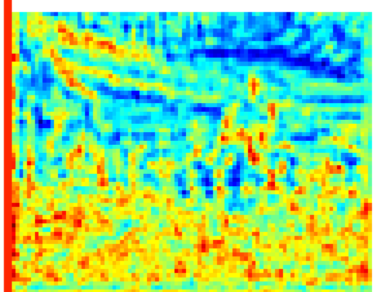
$w_2 \dots w_P$



$B_2 \dots B_P$

$$U_p(x) = \langle w_p, H(x) \rangle$$

$$\mu_p(x) = \max_{x'}[U_p(x') + B_p(x, x')]$$

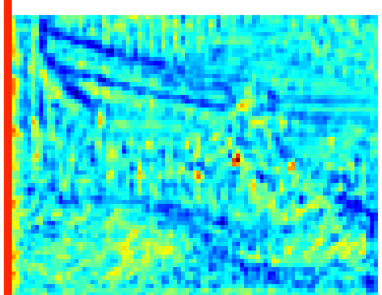


$\vdots$

GDT

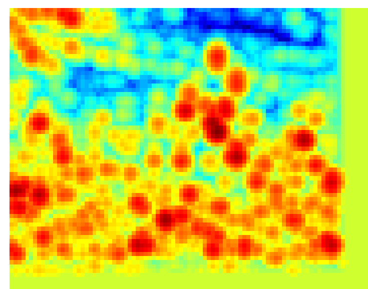
$p = 3$

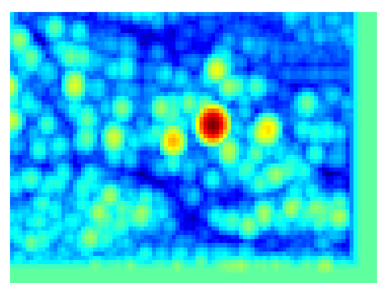
$\vdots$



$p = 5$

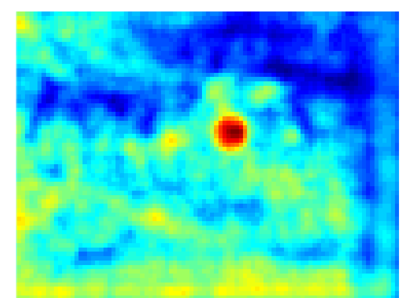
$\vdots$





DTBB, NIPS 2011

$$S(x) = \sum_{p=1}^{P} \mu_p(x)$$



Accelerating detection with DPMs

Efficient detection with DPMs

P F Felzenszwalb R B Girshick and D A McAllester Cascade object detection with DPMs CVPR 2010

B. Sapp, A. Toshev, and B. Taskar. Cascaded models for articulated pose estimation, ECCV, 2010

M. Pedersoli, A.Vedaldi, and J.Gonzalez. A coarse-to-fine approach for object detection, CVPR 2011

I. Kokkinos, Rapid DPM Detection using Dual-Tree Branch-and-Bound, NIPS 2011

Efficient part score computation/approximation

H. Pirsiavash and D. Ramanan, Steerable Part Models, CVPR 2012

A. Vedaldi and A. Zisserman, Sparse Kernel Maps and Faster Product Quantization Learning, CVPR 2012

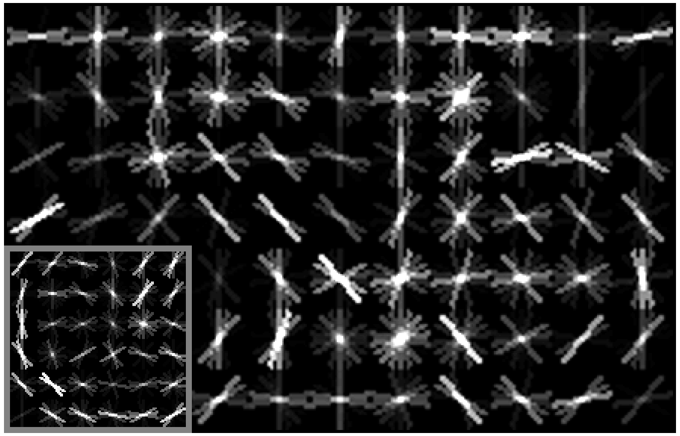
H.O. Song, S. Zickler, T. Althoff, R. Girschick, M. Fritz, C. Geyer, P. Felzenszwalb, T. Darrell, Sparselet Models for Efficient Multiclass Object Detection, ECCV 2012

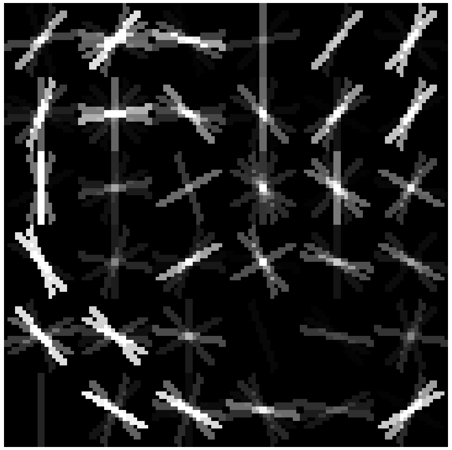
C. Dubout and F. Fleuret. Exact Acceleration of Linear Object Detectors, ECCV 2012

Dean et al. Fast Accurate Detection of 100.000 object categories, CVPR 2013

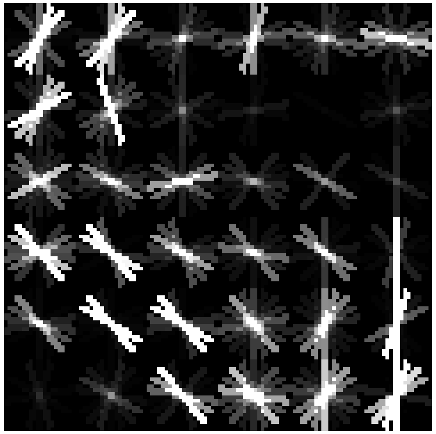
I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Part score computation





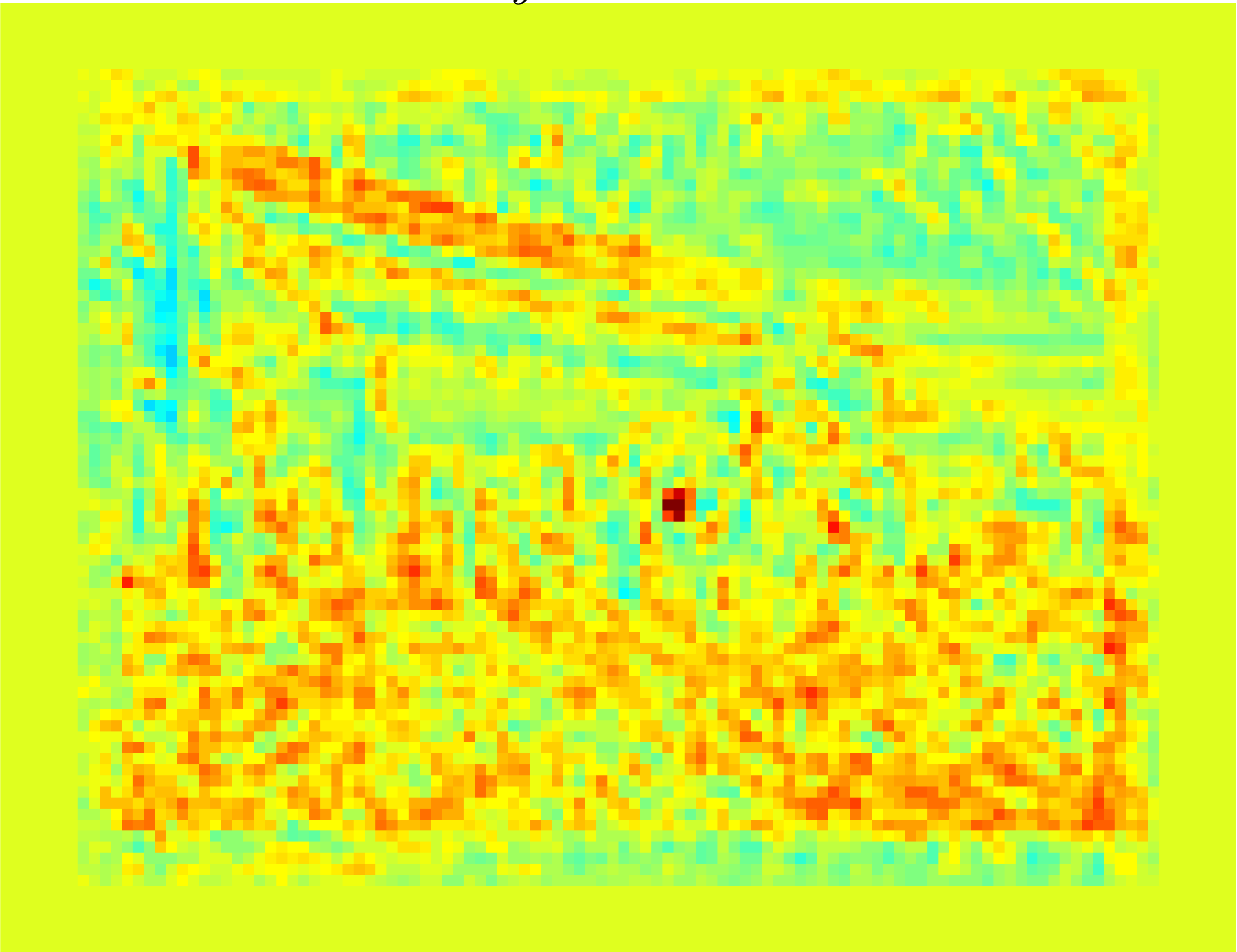
$$\mathbf{w}[y]$$



$$\mathbf{h}[x+y]$$

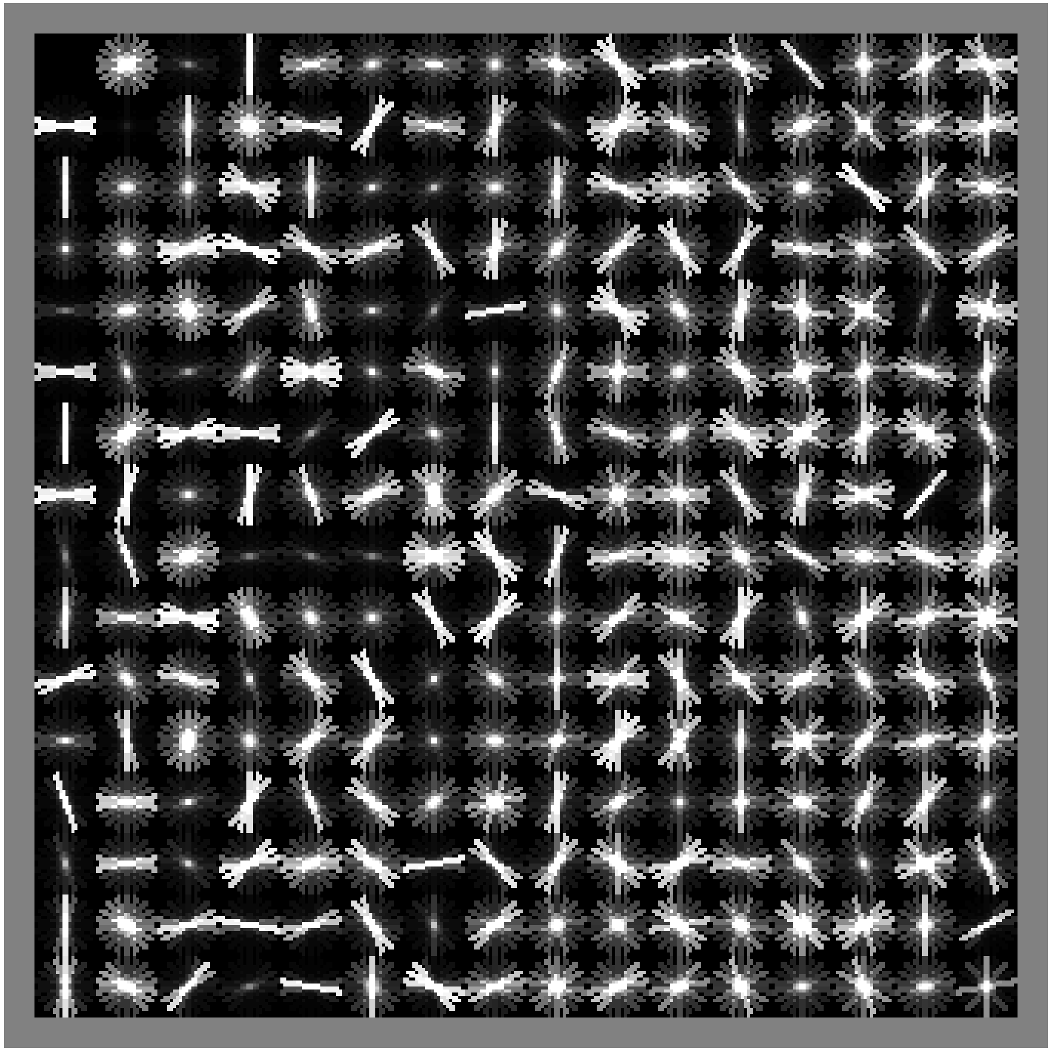
$$s[x] = \sum_y \langle \mathbf{h}[x+y], \mathbf{w}[y] \rangle$$

Part scores $s[x] = \sum_y \langle \mathbf{h}[x+y], \mathbf{w}[y] \rangle$



HOG cell quantization: visual ‘letters’

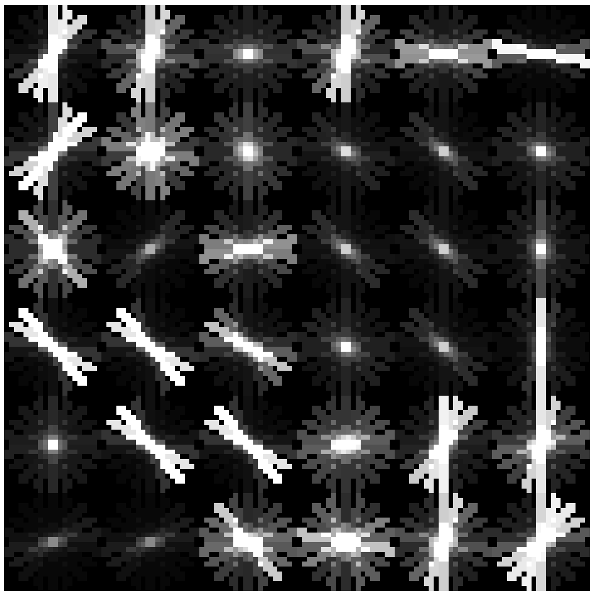
$$\mathcal{C} = \{C_1, \ldots, C_{256}\}$$



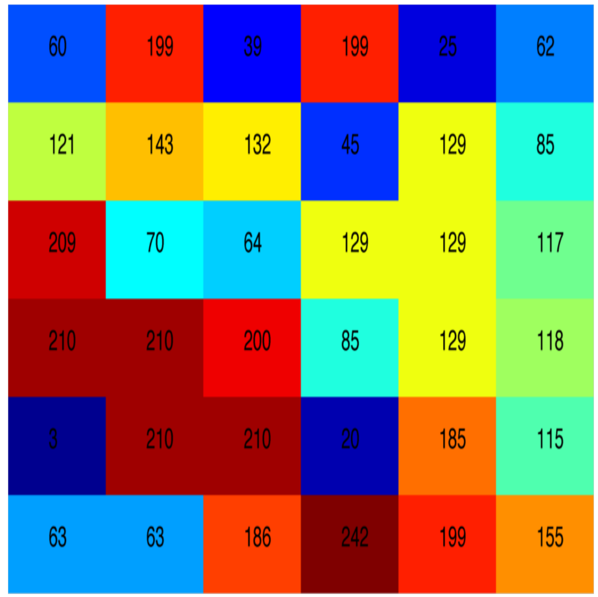
I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

HOG feature quantization

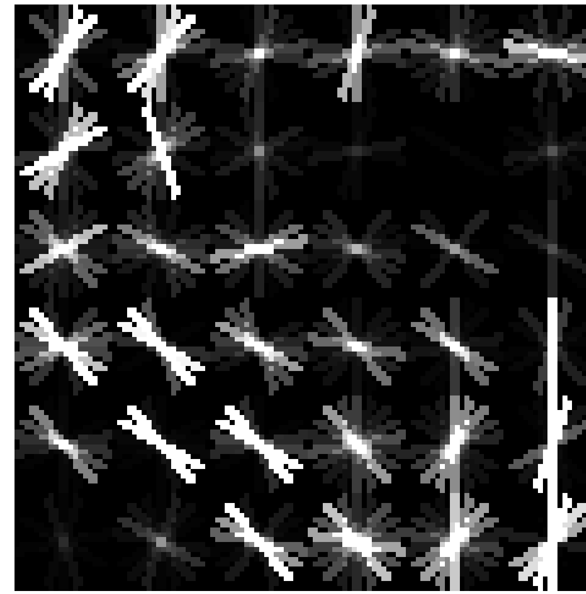
HOG detail



Codebook indices



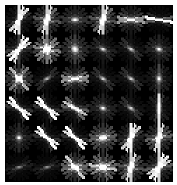
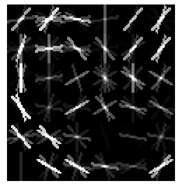
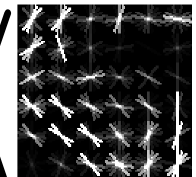
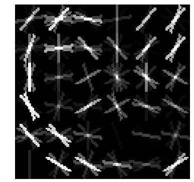
Quantized HOG



$$\mathbf{h}[x] \qquad i[x] = \arg\min_k d(\mathbf{h}[x], C_k) \qquad \hat{\mathbf{h}}[x] = C_{i[x]}$$

I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Efficient inner product approximation

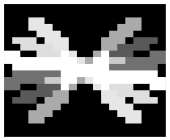
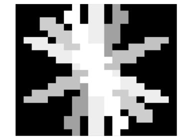
$\langle$  ,  $\rangle \simeq \langle$  ,  $\rangle$

$$s[x] \simeq \hat{s}[x]$$

$$\sum_y \langle \mathbf{h}[x+y], \mathbf{w}[y] \rangle \simeq \sum_y \langle \hat{\mathbf{h}}[x+y], \mathbf{w}[y] \rangle$$

I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Efficient inner product approximation

$\Big\langle$  ,  $\Big\rangle \simeq \Big\langle$  ,  $\Big\rangle$

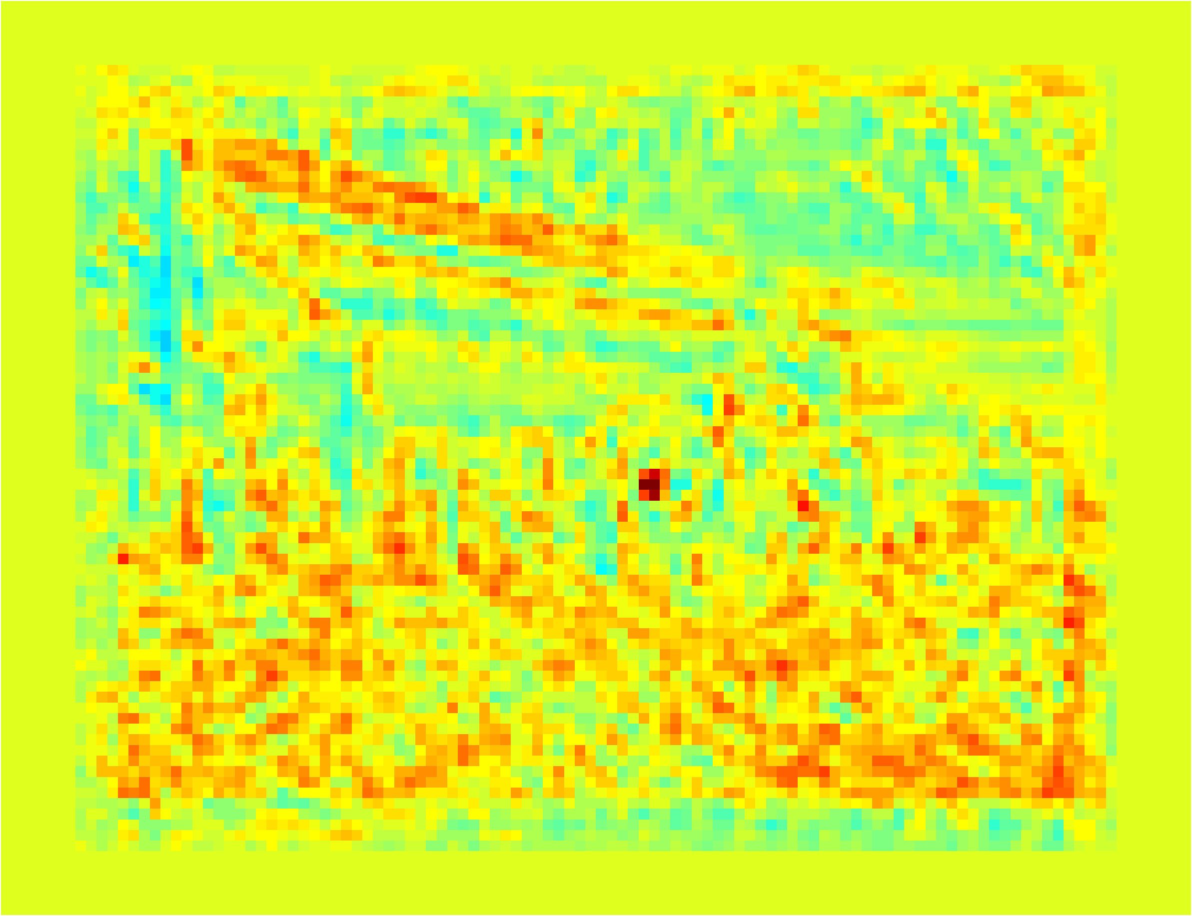
$$\begin{aligned}\langle \mathbf{h}[x+y], \mathbf{w}[y] \rangle &\simeq \langle \hat{\mathbf{h}}[x+y], \mathbf{w}[y] \rangle \\ &= \langle C_{I[x+y]}, \mathbf{w}[y] \rangle \\ &= \mathit{\Pi}[I[x+y], y]\end{aligned}$$

$$\mathit{\Pi}[k, y] = \langle C_k, \mathbf{w}_y \rangle$$

$$s[x] \simeq \hat{s}[x] = \sum_y \mathit{\Pi}[I[x+y], y]$$

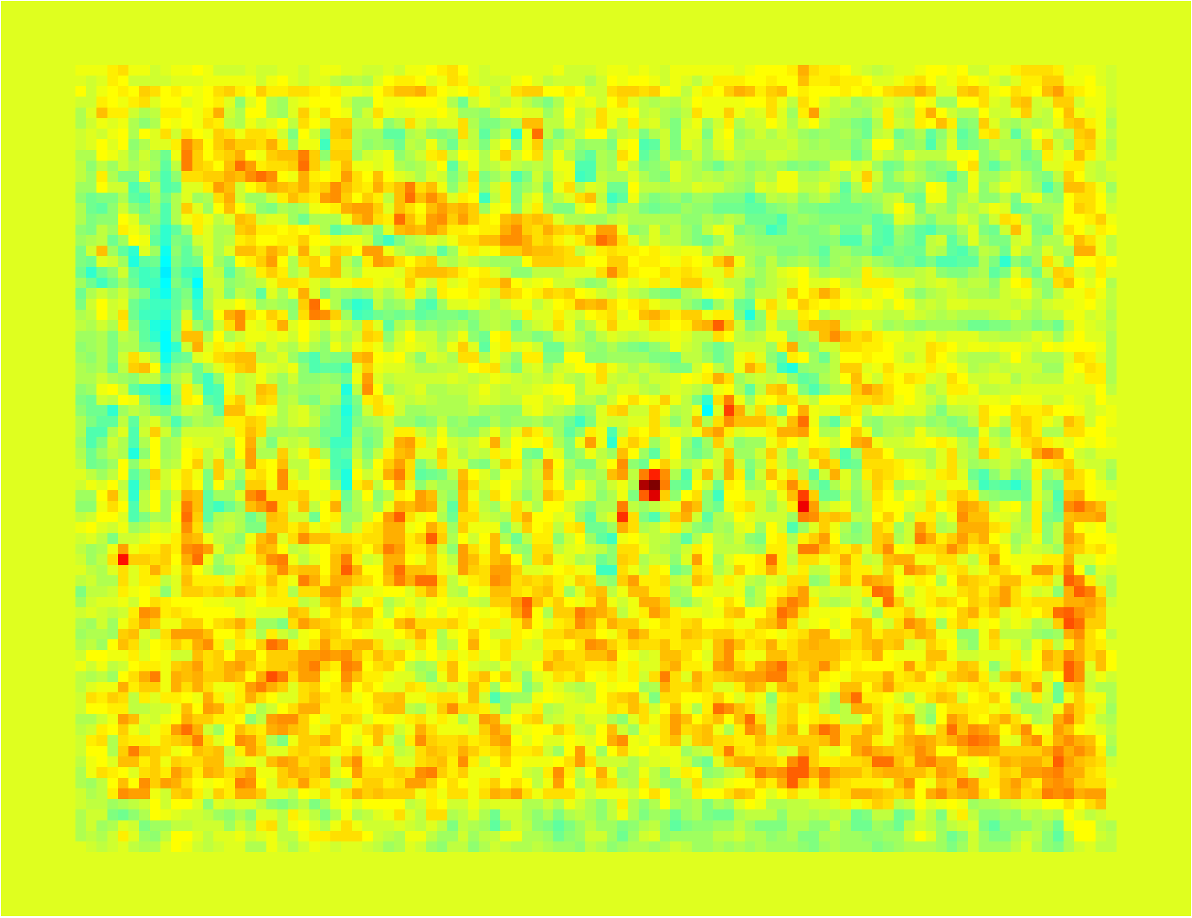
I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Lookup-based estimate demonstration: $s[x]$



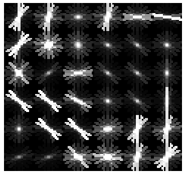
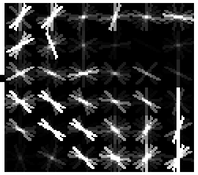
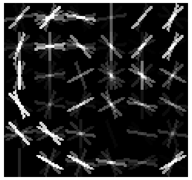
I. K., *Bounding Part Scores for Rapid Object Detection with DPMs*, PnA, 2012

Lookup-based estimate demonstration: $\hat{s}[x]$



I. K., *Bounding Part Scores for Rapid Object Detection with DPMs*, PnA, 2012

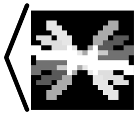
Part-level approximation error

$\epsilon \doteq s - \hat{s} = \langle$  $-$  ,  $\rangle$

$$= \sum_y \underbrace{\langle \mathbf{h}[y] - \hat{\mathbf{h}}[y], \mathbf{w}[y] \rangle}_{e[y]}$$

I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Cell-level approximation error

$$e[y] = \langle \mathbf{h}[y] - \hat{\mathbf{h}}[y], \mathbf{w}[y] \rangle = \langle$$  $-$  $,$  $\rangle$

$$= \langle \mathbf{e}[y], \mathbf{w}[y] \rangle$$

$$= \sum_{f=1}^{32} \mathbf{e}_y[f] \mathbf{w}_y[f]$$

Chebyshev inequality

For any zero-mean random variable, and any value of α:

$$P(|X| > \alpha) \leq \frac{E\{X^2\}}{\alpha^2}$$

Equivalently, with probability of error smaller than p_e:

$$X \in \left[-\sqrt{\frac{E\{X^2\}}{p_e}}, \sqrt{\frac{E\{X^2\}}{p_e}}\right]$$

I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Chebyshev-based bounds

Lookup-based approximation:

$$s[x] \simeq \hat{s}[x] = \sum_y \Pi[I[x+y], y]$$

With probability of error at most p_e:

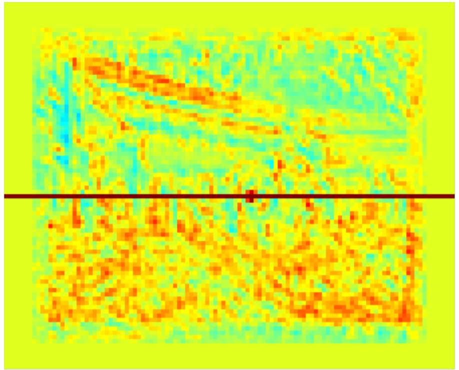
$$\underline{s}[x] \leq s[x] \leq \overline{s}[x]$$

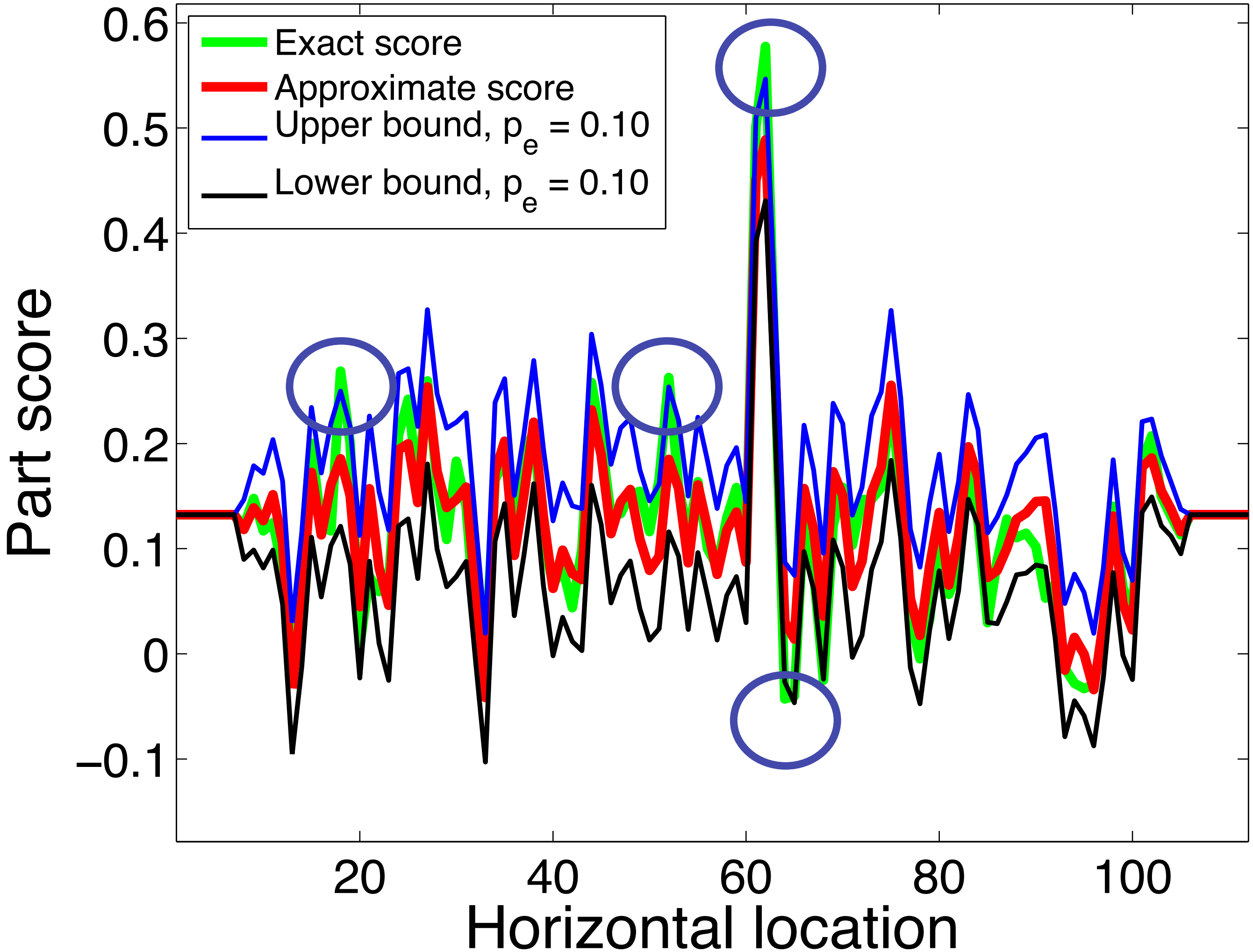
$$\underline{s}[x] = \hat{s}[x] - \sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[x+y]\|^2}{p_e F}}$$

$$\overline{s}[x] = \hat{s}[x] + \sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[x+y]\|^2}{p_e F}}$$

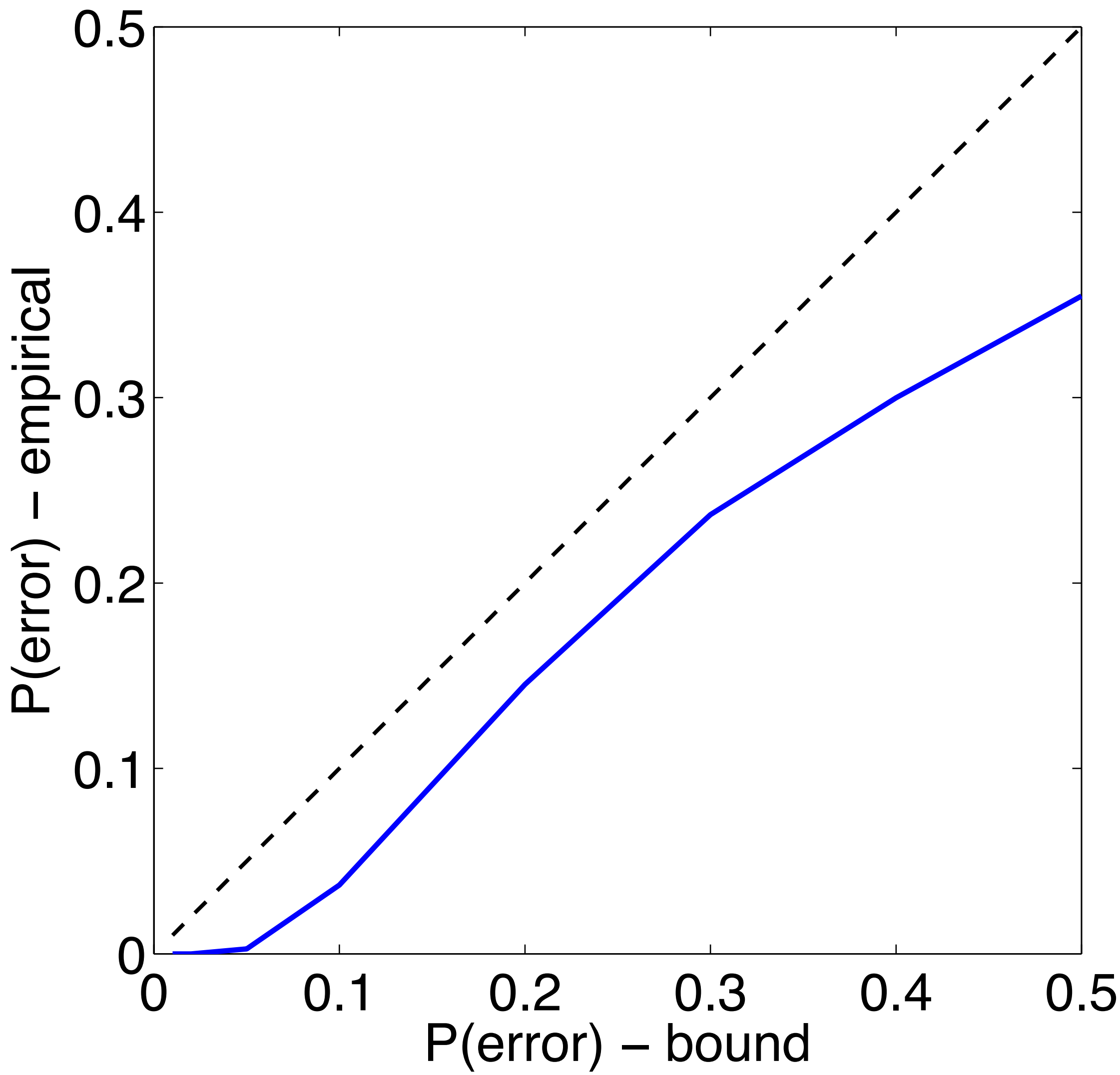
I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Bound demonstration for varying confidence

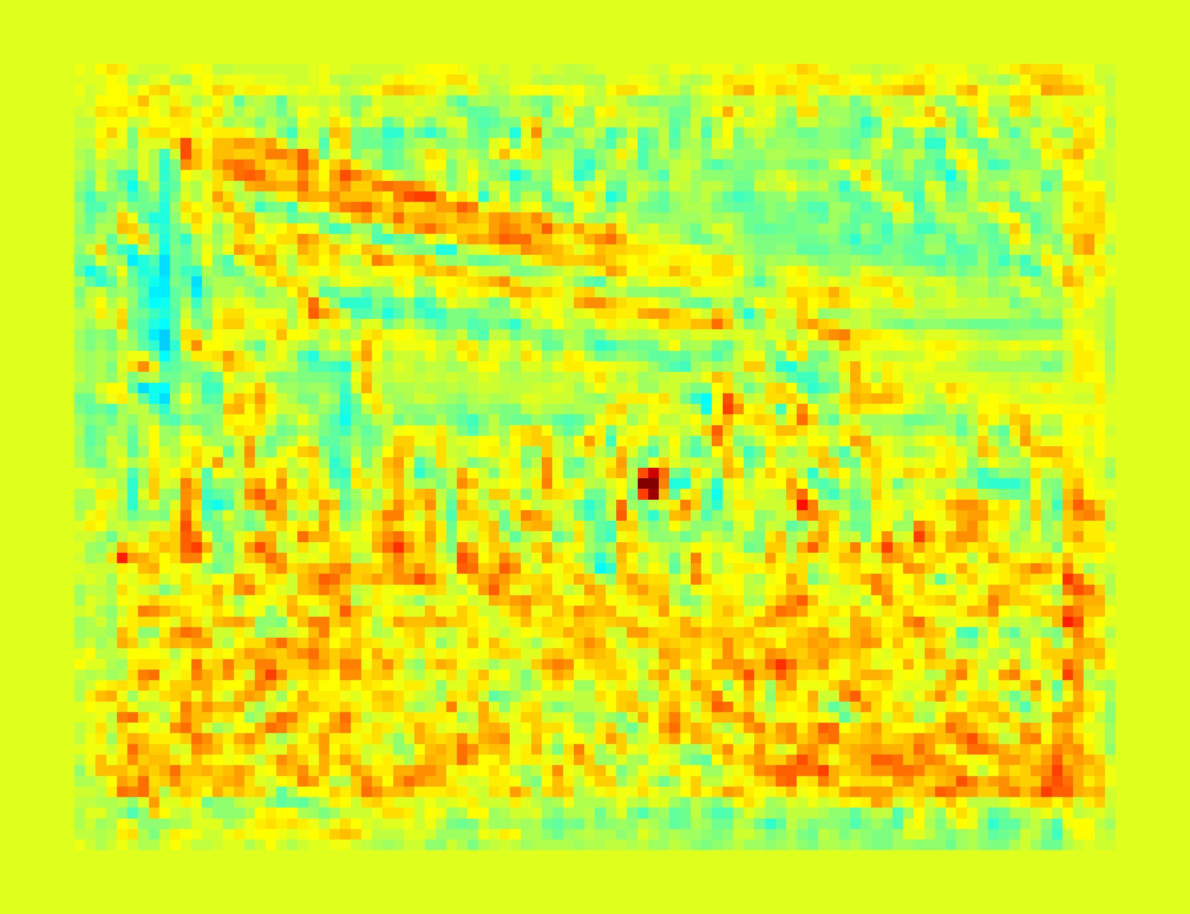




Bound tightness

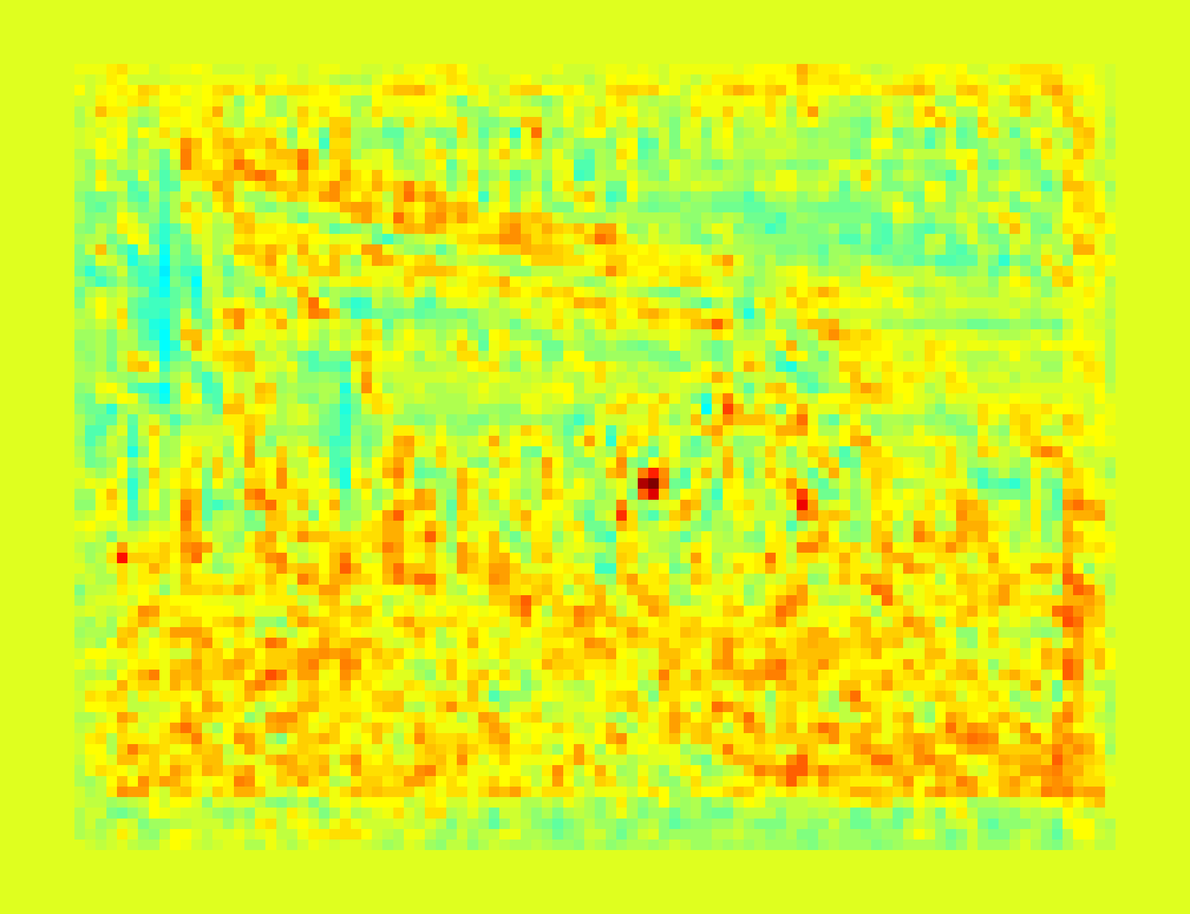


Bound demonstration: $s[x]$



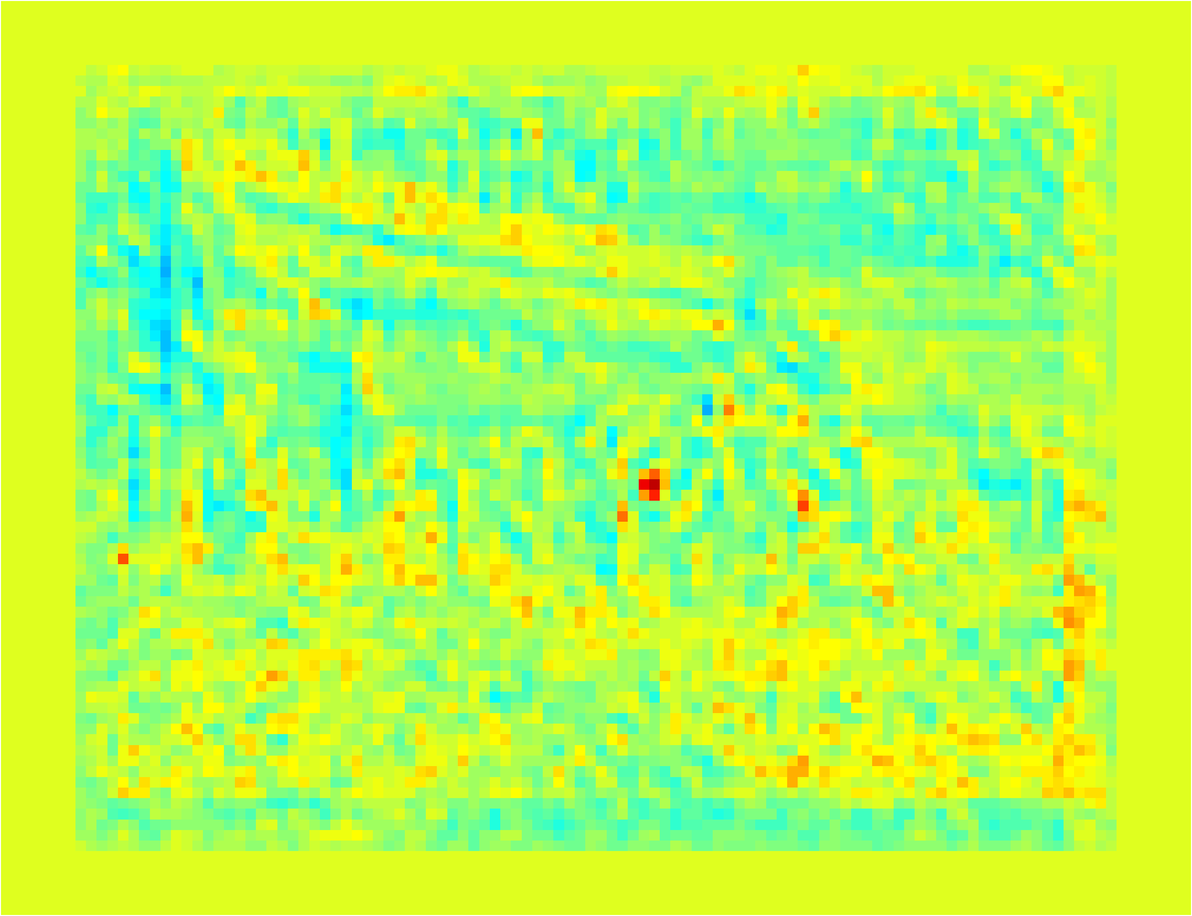
I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Bound demonstration: $\hat{s}[x]$

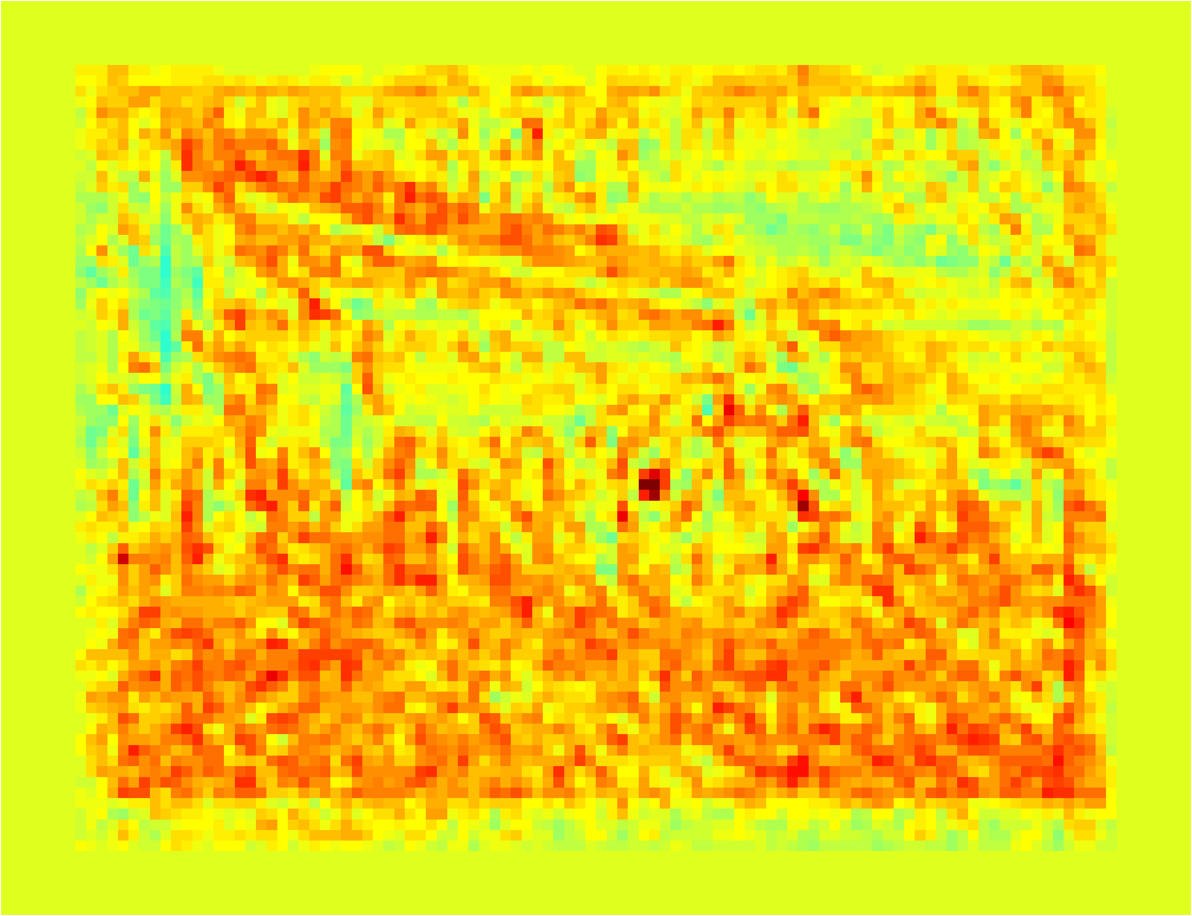


I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Bound demonstration: $\underline{s}[x],\ p_e = .05$

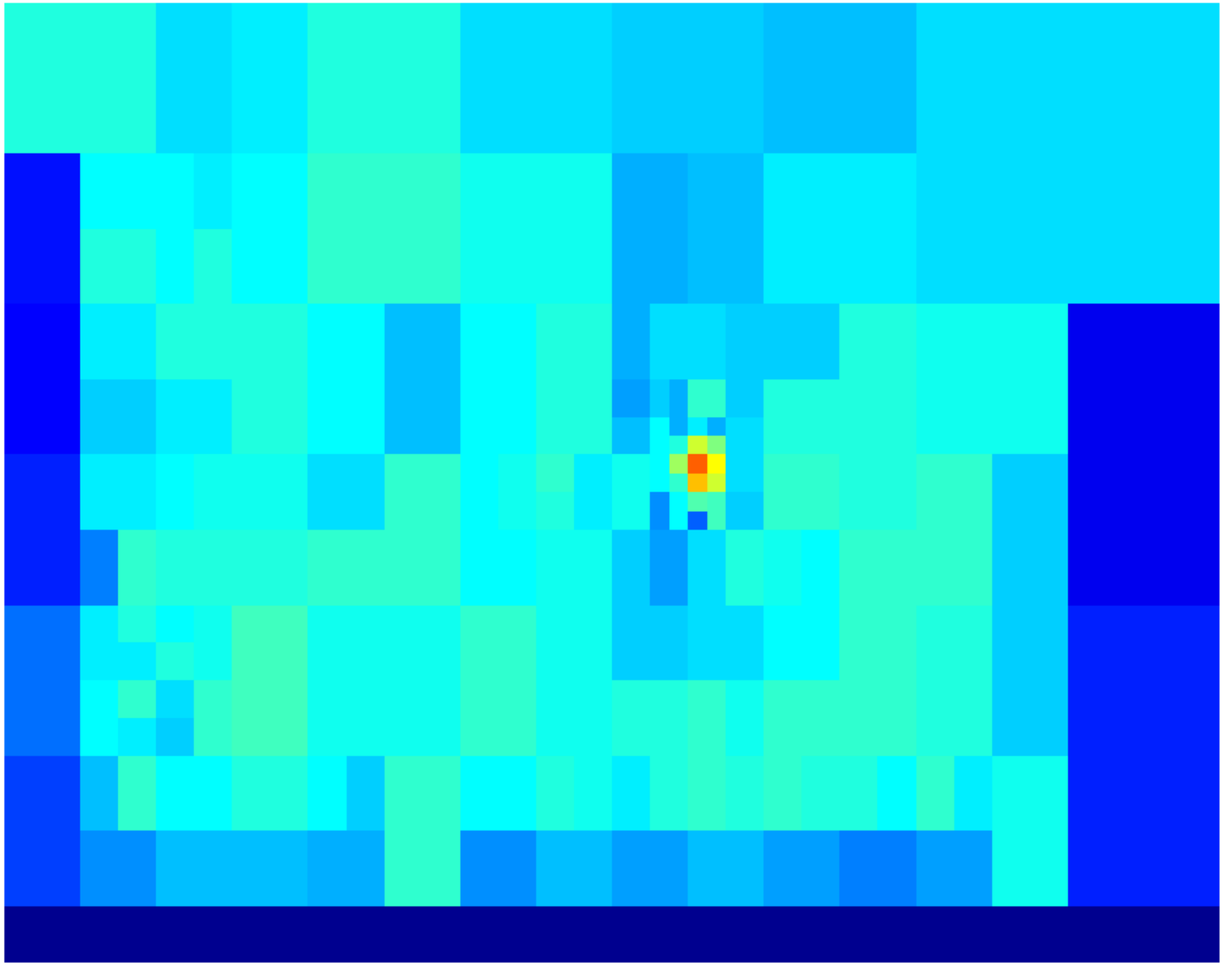


Bound demonstration: $\overline{s}[x],\ p_e = .05$

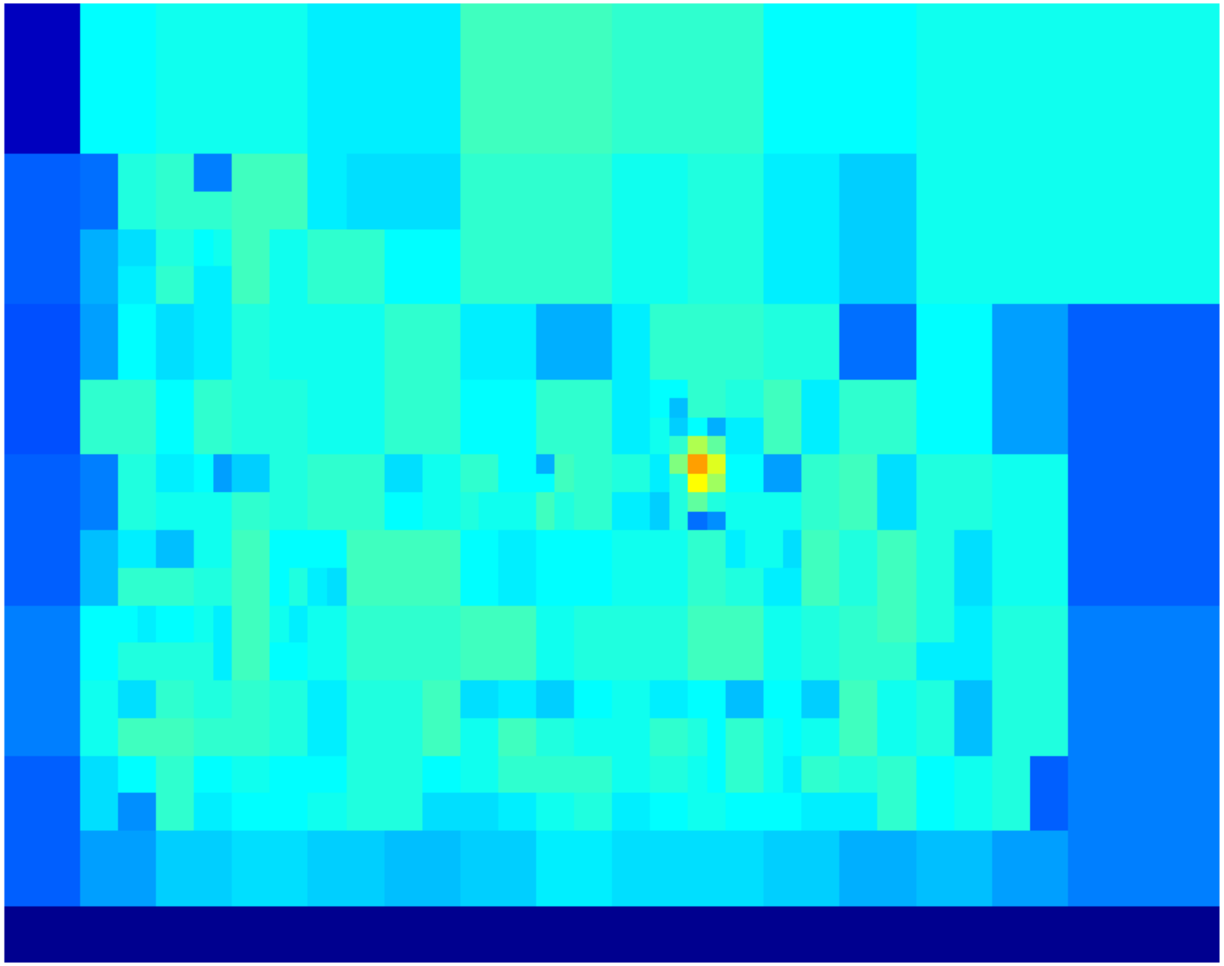


I. K., *Bounding Part Scores for Rapid Object Detection with DPMs*, PnA, 2012

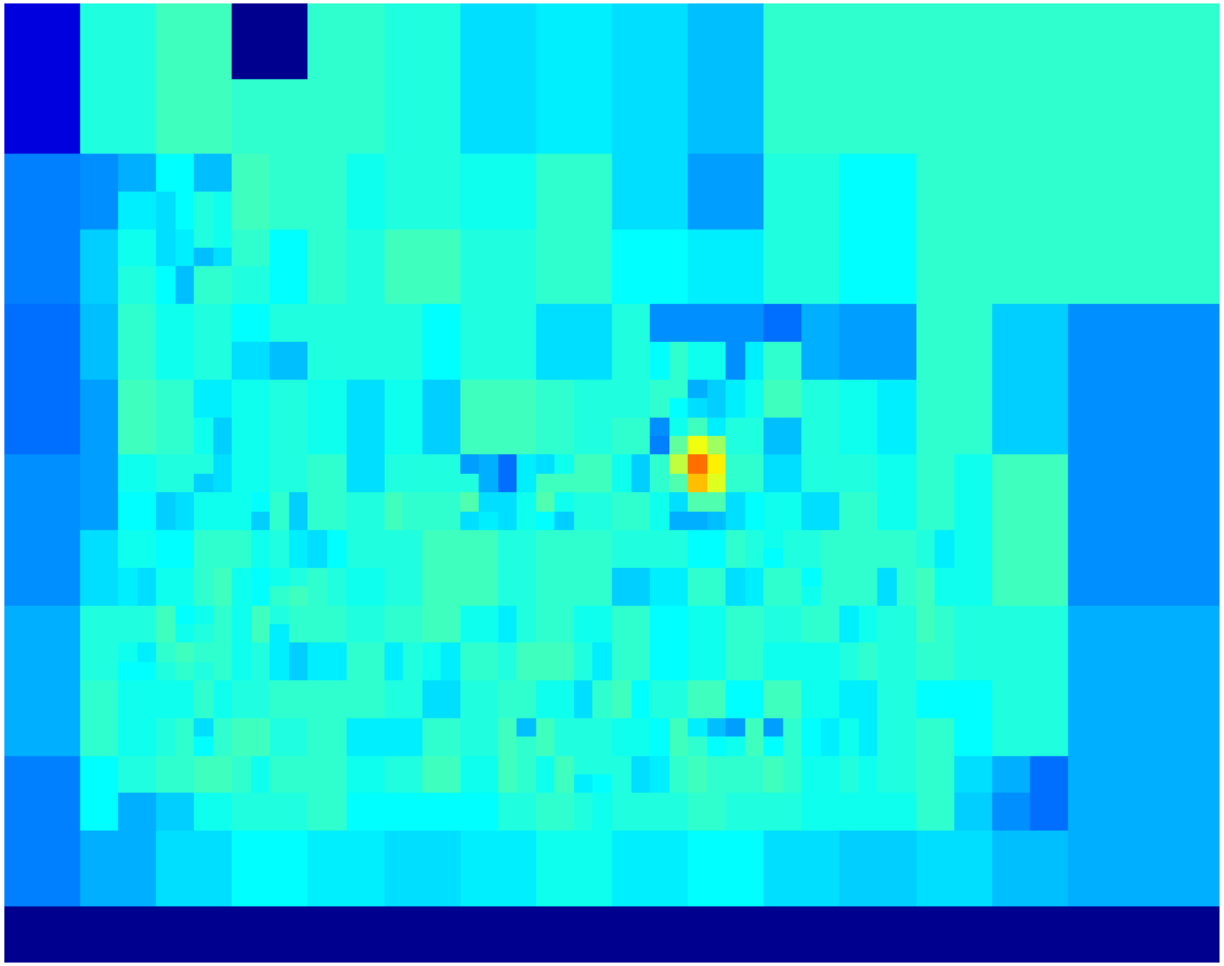
DTBB results: exact part scores



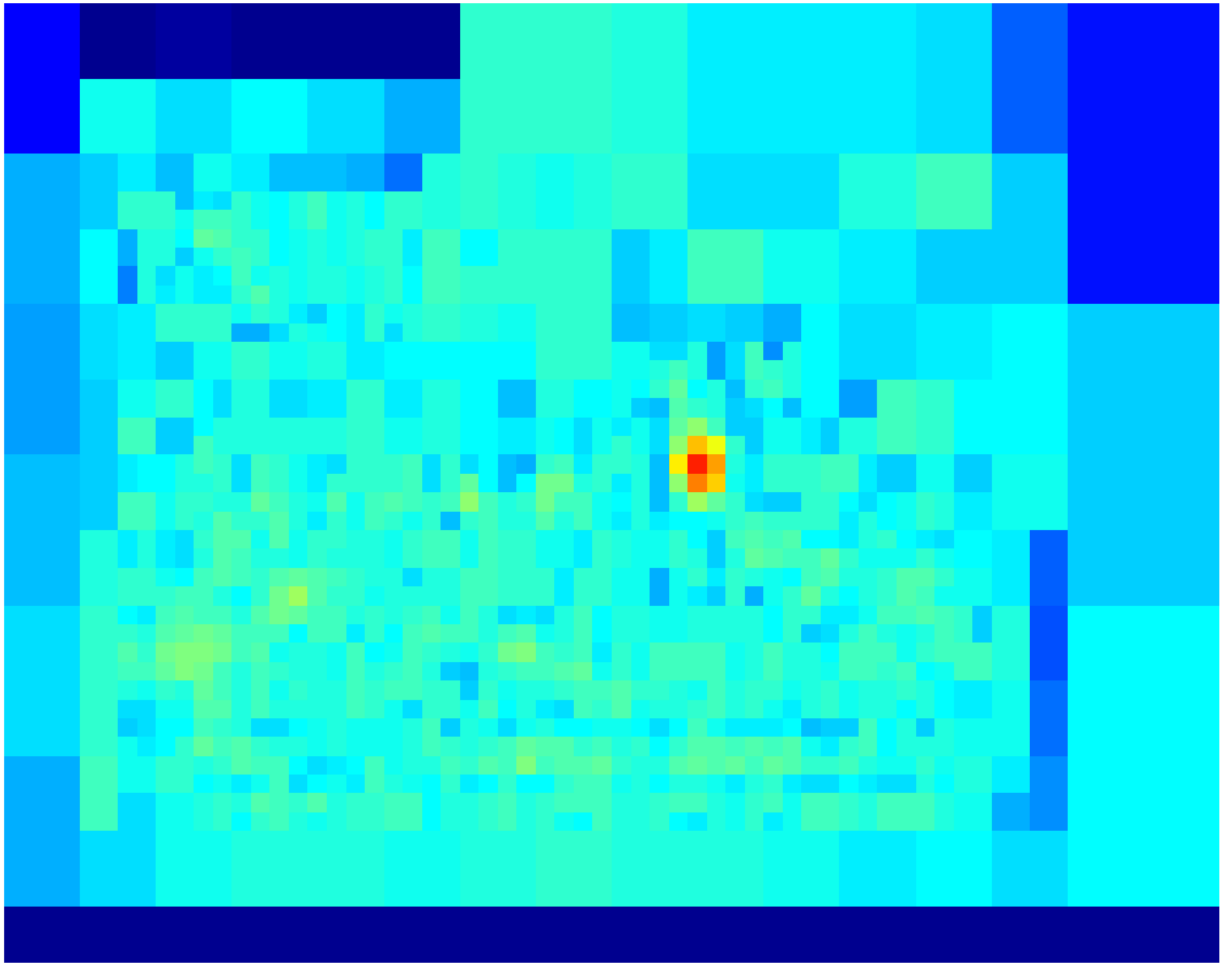
DTBB results, part score bounds @ $p_e = 0.2$



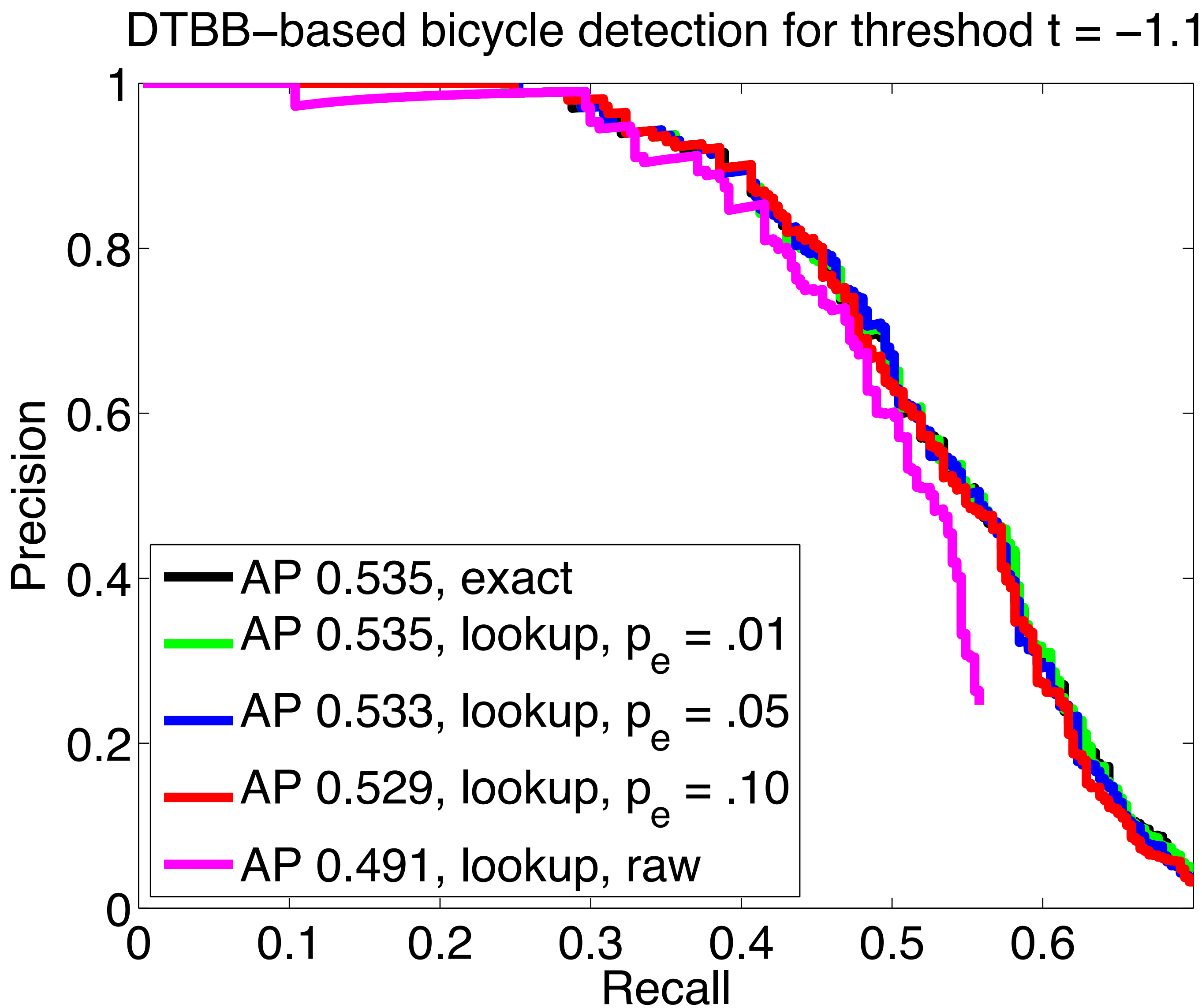
DTBB results, part score bounds @ $p_e = 0.1$



DTBB results, part score bounds @ $p_e = 0.05$



Impact on performance



I. K., Bounding Part Scores for Rapid Object Detection with DPMs, PnA, 2012

Speedup results

	GDTs	DTBB	$p_e = 0.05$	$p_e = 0.01$
Part terms	8.35 ± 0.77	1.69 ± 0.18	0.69 ± 0.03	0.69 ± 0.06

Detection with Cascade DPMs (C-DPMs)

$$S_0(x) = 0,\ \mathcal{I}_0 = [1, N] \times [1, M]$$

$$S_k(x) = S_{k-1}(x) + \max_{x'} \left(U_p(x') + B_p(x', x)\right)$$

$$\mathcal{I}_k = \{x \in \mathcal{I}_{k-1} : S_{k-1}(x) \geq \theta_k\}$$

Felzenszwalb, Girschick, et al: use PCA-projection of $\mathbf{h}, \mathbf{w}$

Our work: use quick upper bounds, thresholds for full HOG

	GDTs	C-DPM	$p_e = 0.05$	$p_e = 0.01$
$\theta = -0.5$	8.95 ± 0.82	0.56 ± 0.07	0.19 ± 0.03	0.23 ± 0.04
$\theta = -0.7$	8.95 ± 0.82	0.72 ± 0.09	0.29 ± 0.04	0.36 ± 0.06
$\theta = -1.0$	8.95 ± 0.82	1.04 ± 0.16	0.51 ± 0.10	1.07 ± 0.29

Future work

Deformable Models: fast and accurate

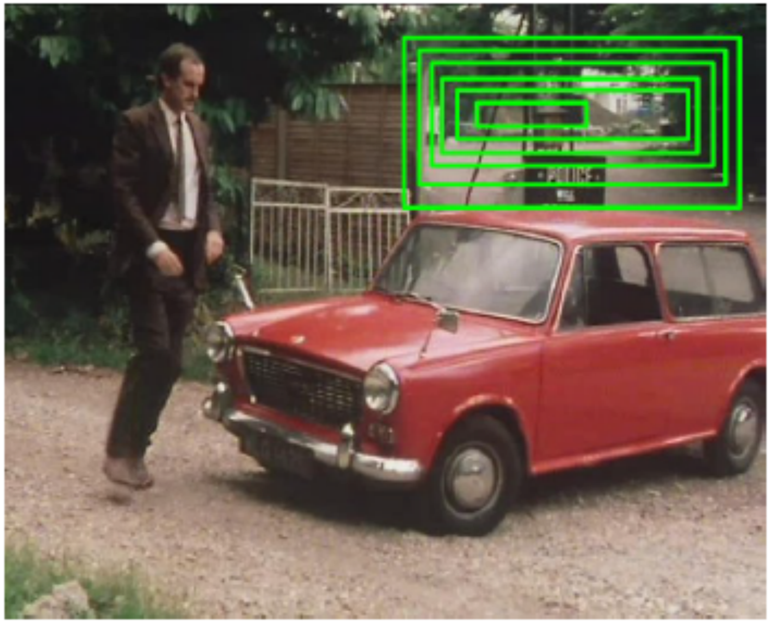
BB: theory-grounded blend of optimization and low-level processing

On-going work

- Part sharing
- Tighter bounds, cascades

Beyond sliding windows

Problem: Exhaustive evaluation of $\operatorname{argmax}_{B \in \mathcal{B}} f_I(B)$ is too slow.
Solution: Use the problem's *geometric structure*.



- Similar boxes have similar scores.
- Calculate scores for *sets of boxes* jointly (upper bound).
- If no element can contain the object, discard the set.
- Else, split the set into smaller parts and re-check, etc.

$\Rightarrow$ efficient branch & bound algorithm

Chebyshev inequality-II

For a weighted sum of i.i.d. zero-mean random variables:

$$X' = \sum_{k=1}^{K} w_k X_k$$

with probability of error smaller than p_e :

$$X' \in \left[-\sqrt{\frac{\left(\sum_k w_k^2\right) E\{X^2\}}{p_e}}, \sqrt{\frac{\left(\sum_k w_k^2\right) E\{X^2\}}{p_e}} \right]$$

Chebyshev inequality for cell-level error

$$e[y] = \sum_{f=1}^{F} \mathbf{e}_y[f]\mathbf{w}_y[f]$$

with probability of error smaller than p_e :

$$e_y \in \left[-\sqrt{\frac{\|\mathbf{w}[y]\|^2\|\mathbf{e}[y]\|^2}{p_e F}}, \sqrt{\frac{\|\mathbf{w}[y]\|^2\|\mathbf{e}[y]\|^2}{p_e F}}\right]$$

Chebyshev inequality for part-level error

$$\boxed{\epsilon = \hat{s} - s} = \sum_y e[y]$$

with probability of error smaller than p_e :

$$\epsilon \in \left[-\sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[y]\|^2}{p_e F}}, \sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[y]\|^2}{p_e F}} \right]$$

→ with probability of error smaller than p_e:

$$s \in \left[\hat{s} - \sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[y]\|^2}{p_e F}}, \hat{s} + \sqrt{\frac{\sum_y \|\mathbf{w}[y]\|^2 \|\mathbf{e}[y]\|^2}{p_e F}} \right]$$

Bounding the DPM cost function

Goal: $$\overline{S}(X) \geq S(X) = \max_{x \in X} S(x) \quad \forall X, \quad \overline{S}(\{x\}) = S(x)$$

where $$S(x) = \sum_{p=1}^{P} \underbrace{\max_{x'} [U_p(x') + B_p(x, x')]}_{m_p(x)}$$

Since $$\max_{x \in X} \sum_{p=1}^{P} m_p(x) \leq \sum_{p=1}^{P} \max_{x \in X} m_p(x)$$

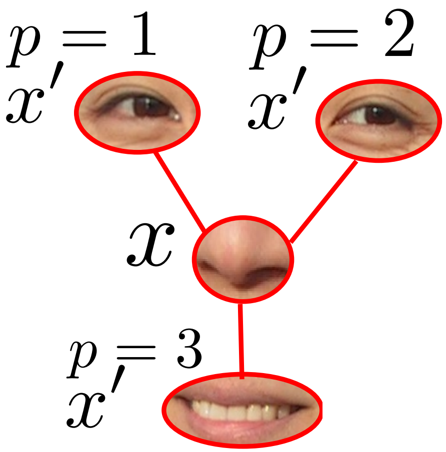
$$\max_{x \in X} \max_{x' \in X'} [U_p(x') + B_p(x, x')] \leq \underbrace{\max_{x' \in X'} U_p(x') + \max_{x \in X, x' \in X'} B_p(x, x')}_{\overline{m}_p(X)}$$

$$\max_{x \in X} S(x) \leq \sum_{p=1}^{P} \overline{m}_p(X) \doteq \overline{S}(X)$$

I. K., Rapid DPM detection using Dual Tree Branch-and-Bound, NIPS, 2011/INRIA TR 2012

Object detection with DPMs

Score for object hypothesis: $$S(x) = \sum_{p=1}^{P} \max_{x'} [U_p(x') + B_p(x, x')]$$



For any candidate object location, x , gather support from all parts, p

For every part p, maximize over all part locations, x’

Dual Recursion

Goal: keep the computation of $\overline{m}(X) = \max_d \max_s \overline{\mu}_d^s$ able

Source-/ Domain-tree: KD-trees for source/domain points, respectively.

Key idea: entertain list of s that can possibly contribute to any d

$$\mathcal{S}_d = \{s_i\}$$

Originally: d is Domain-tree root, its supporter is Source-tree root

$$\mathcal{S}_0 = \{0\}$$

When splitting (branching) d, split also its supporters

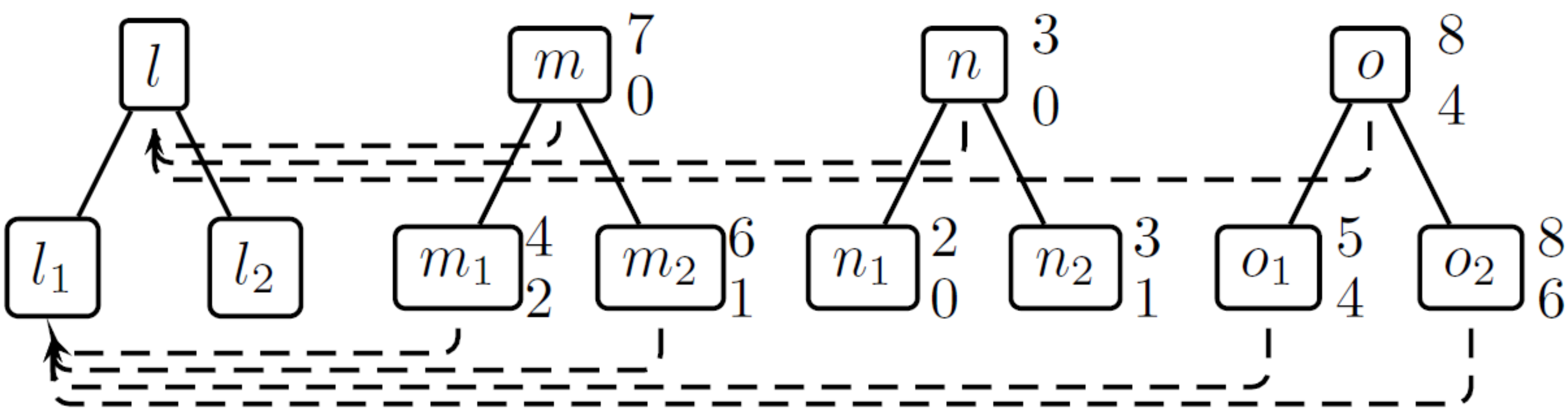
Prune supporters based on upper and lower bounds

Supporter pruning

$\mathcal{S}_l = \{m, n, o\}$ source nodes m, n, o upport’ domain node l

Recurse: $\mathcal{S}_{\mathrm{ch}(l)} \subset \{\mathrm{ch}(m), \mathrm{ch}(n), \mathrm{ch}(o)\}$

Pruning criterion: $\overline{\mu}^l_d < \max_{j \in \mathcal{S}_d} \underline{\lambda}^j_d$



Part-based models for objects and actions

Part-based object detection

Part-based action recognition

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

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

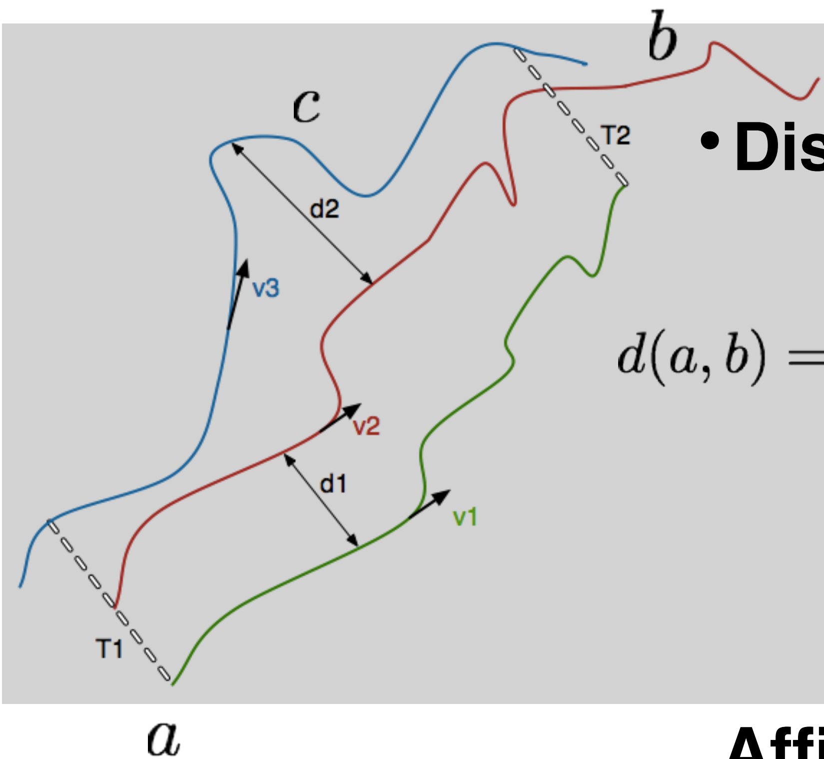
Discovery of Salient Regions

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

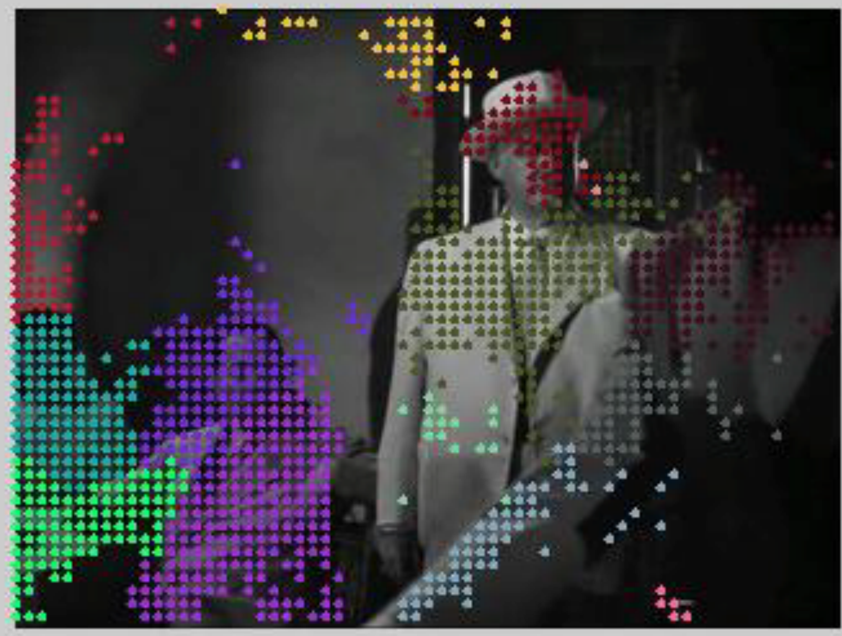
Trajectory Groups: $\{G_k\},\ k = 1,\ldots,N$

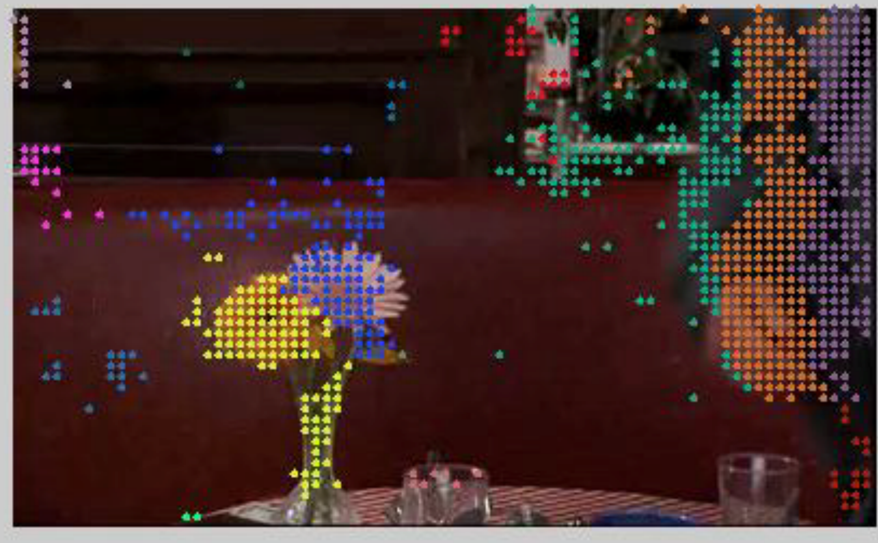
(Brox and Malik, ECCV' 10)

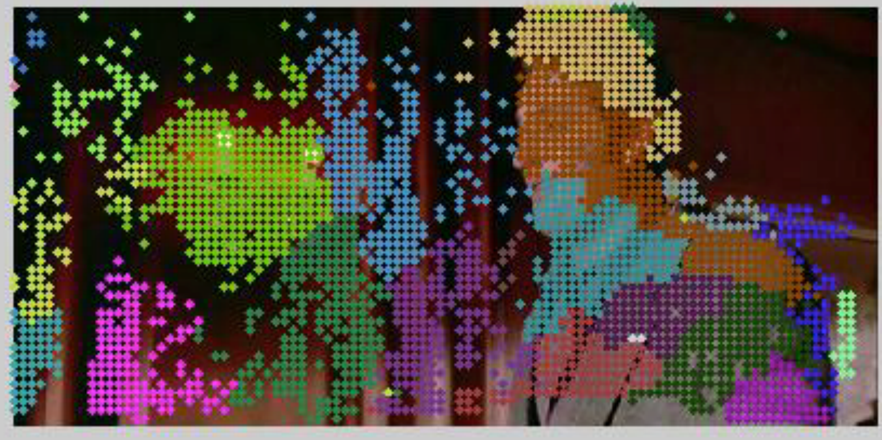
Part-based models for objects and actions

Part-based object detection

Part-based action recognition



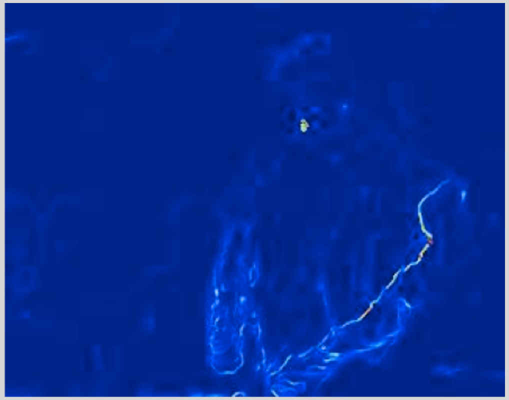





Trajectory Group Descriptor

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

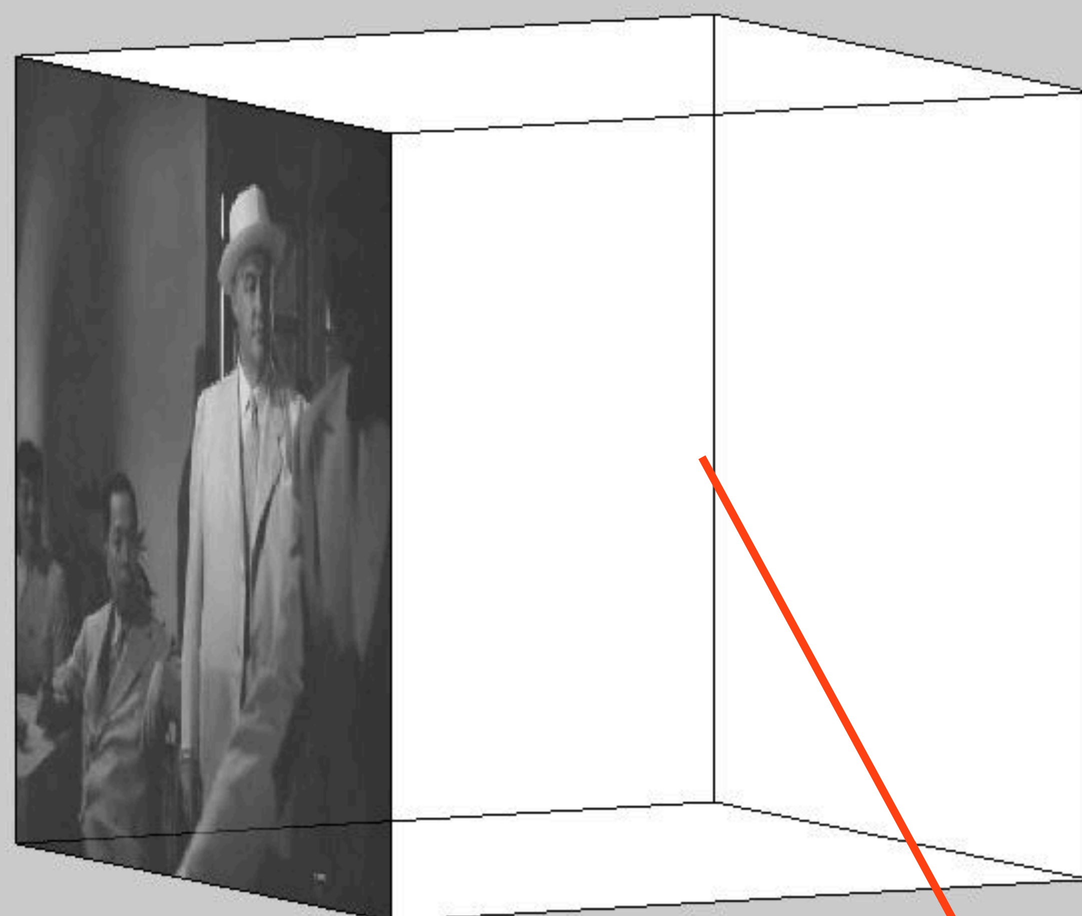
Optical Flow



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



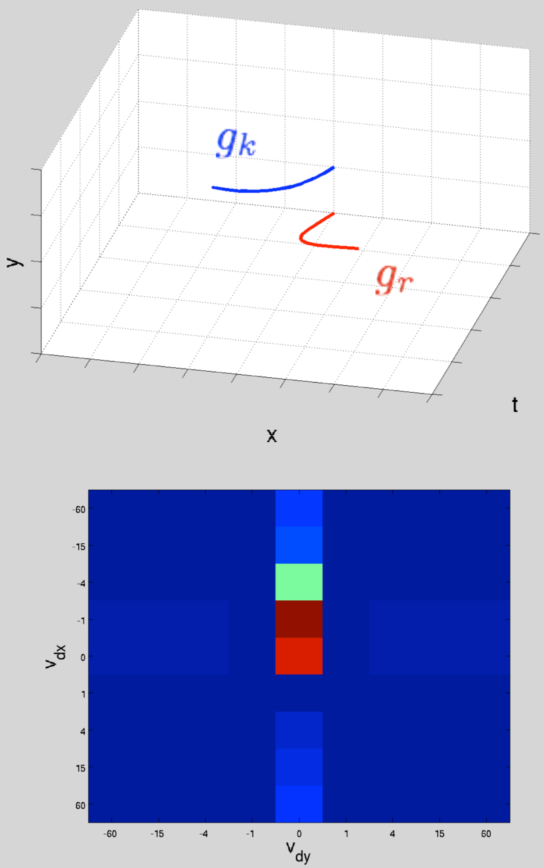
- Histogram of dense descriptors: h_k
- 3D “average” trajectory: g_k

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

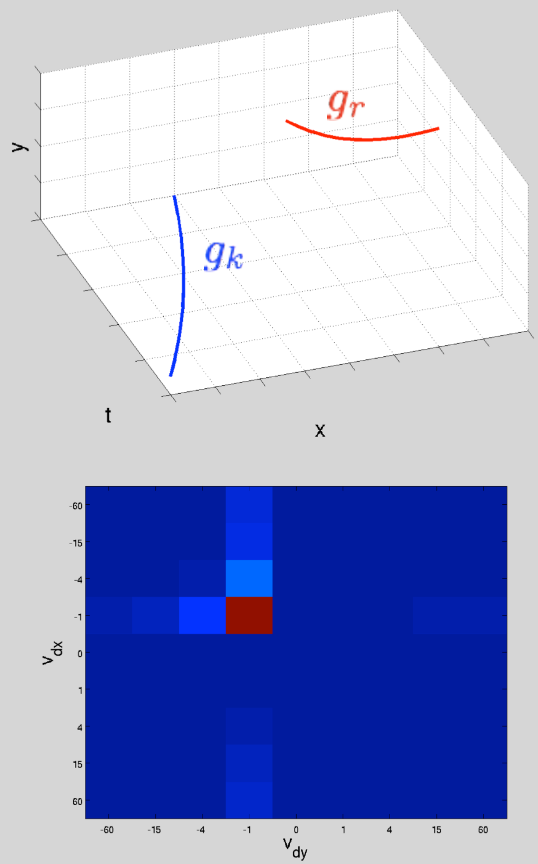
Pairwise Relationships

- **Characterize the change of the relative position of two “average” trajectories**

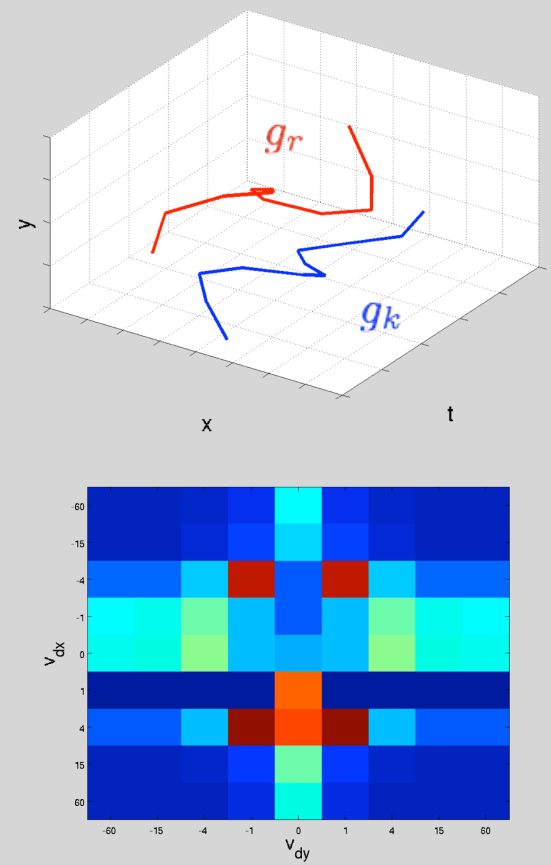
Converging in x dimension



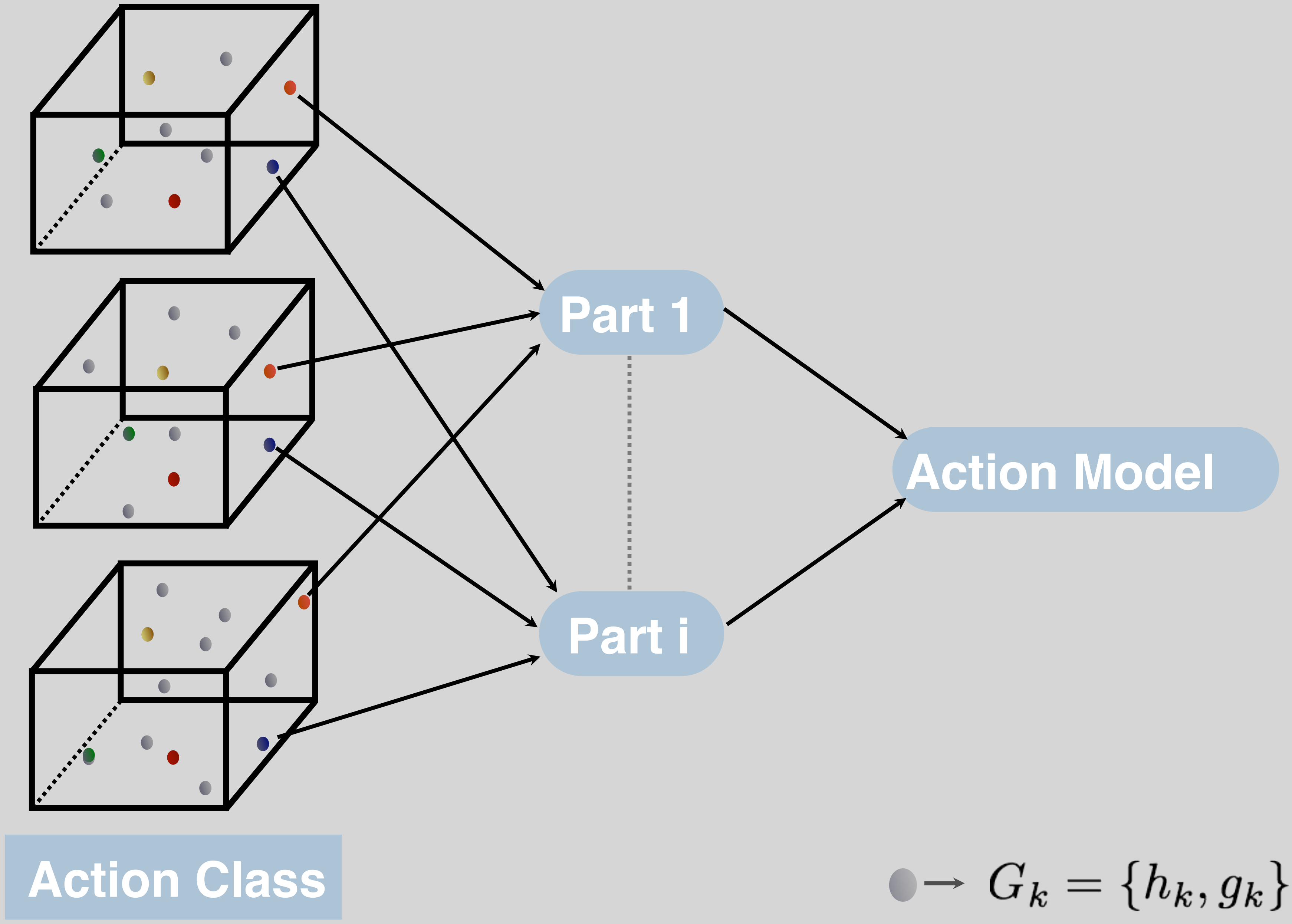
Slowing Converging in both dimensions



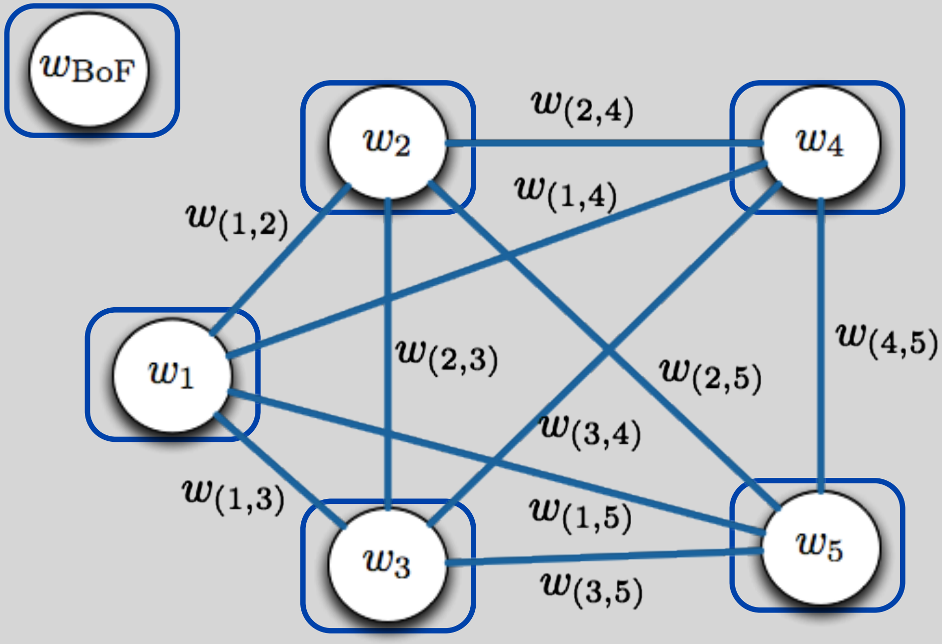
Converging and Diverging



Mid-level Representation



Action Model



- BoF Term
- Unary Terms
- Pairwise Terms

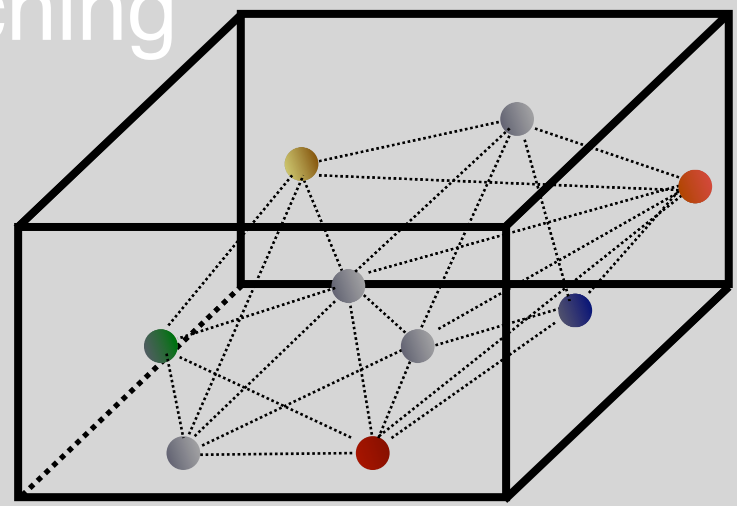
$$\text{score}(P,\sigma;\mathbf{w}) = \boxed{\langle w_0, h_{\text{BoF}}\rangle} + \boxed{\sum_{i=1}^{5}\langle w_i, h_{p_i}\rangle} + \boxed{\sum_{i=1}^{5}\sum_{j=i+1}^{5}\langle w_{i,j}, \psi(g_{p_i}, g_{p_j}, \sigma)\rangle} = \langle \mathbf{w}, \Phi(x,P,\sigma)\rangle$$

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

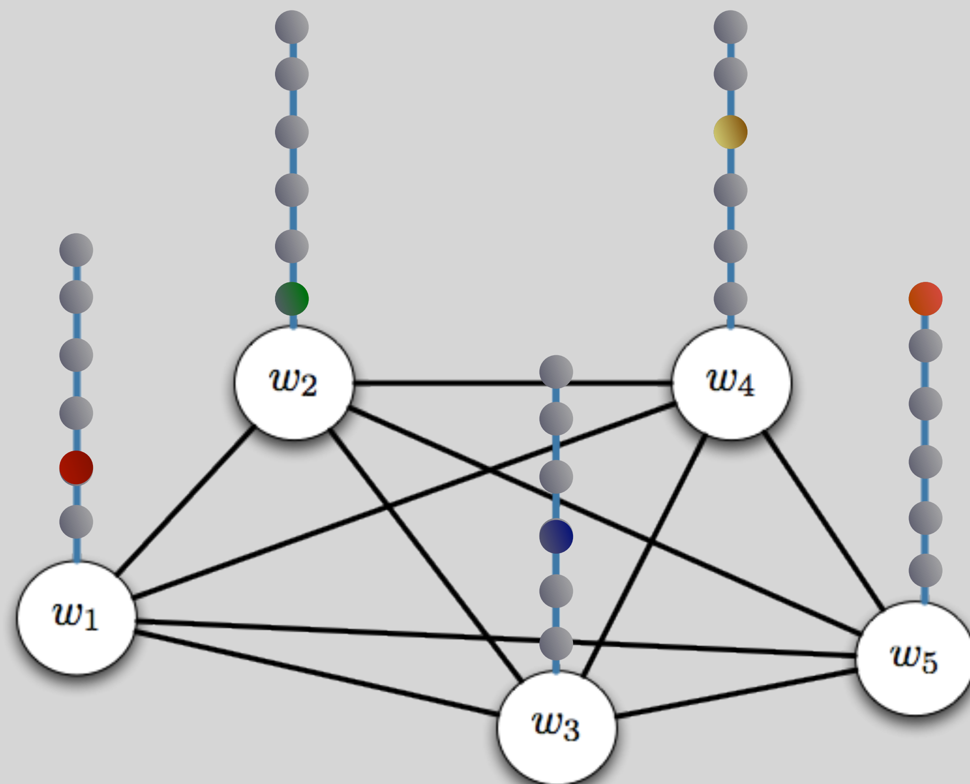
Subgraph Matching



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^*:

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

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

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

Ranking SVM

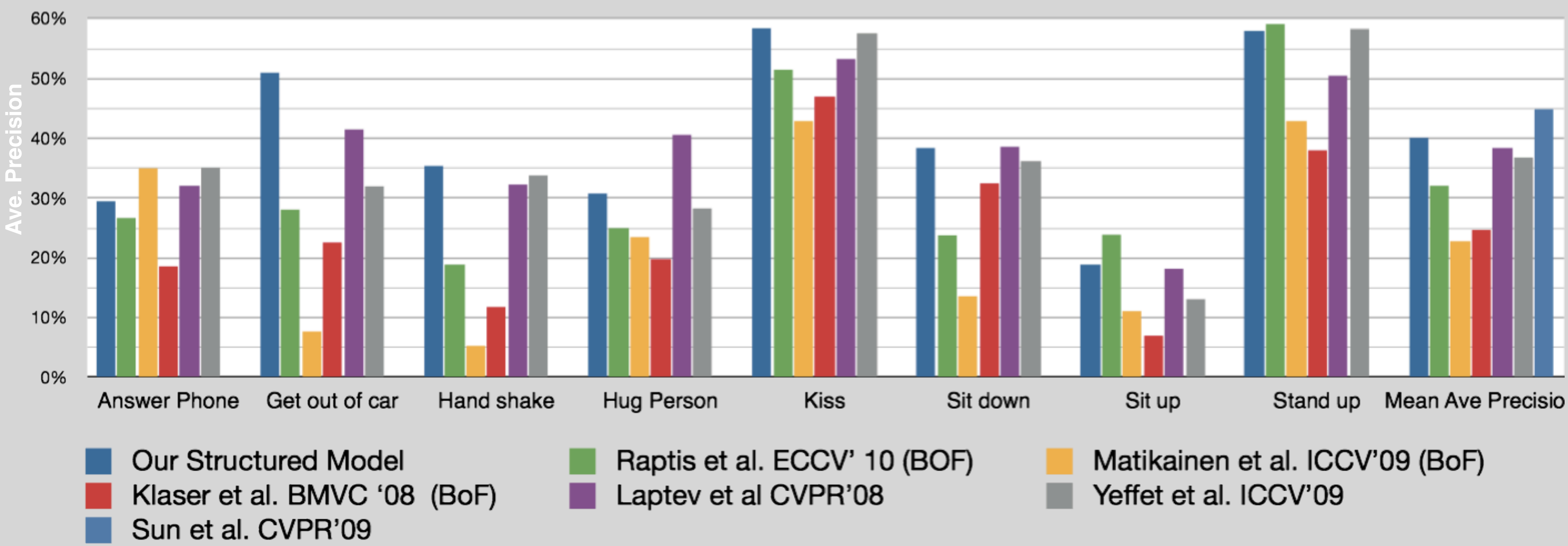
Non-convex optimization -- CCCP: (Yuille and Rangarajan 2003)

Initialization

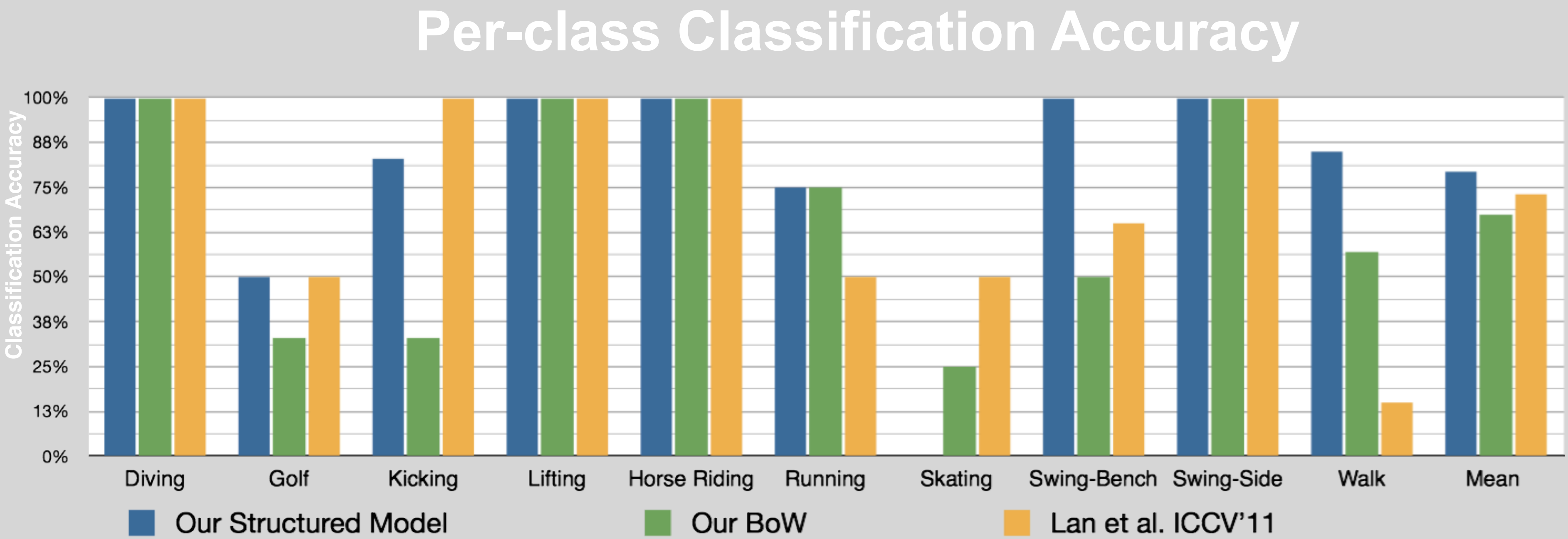
- **Weights of pairwise terms set to zero**
- **Unary terms weights:**
    - **Set equal to the center of a cluster produced by K-means on the descriptors of the positive training videos**

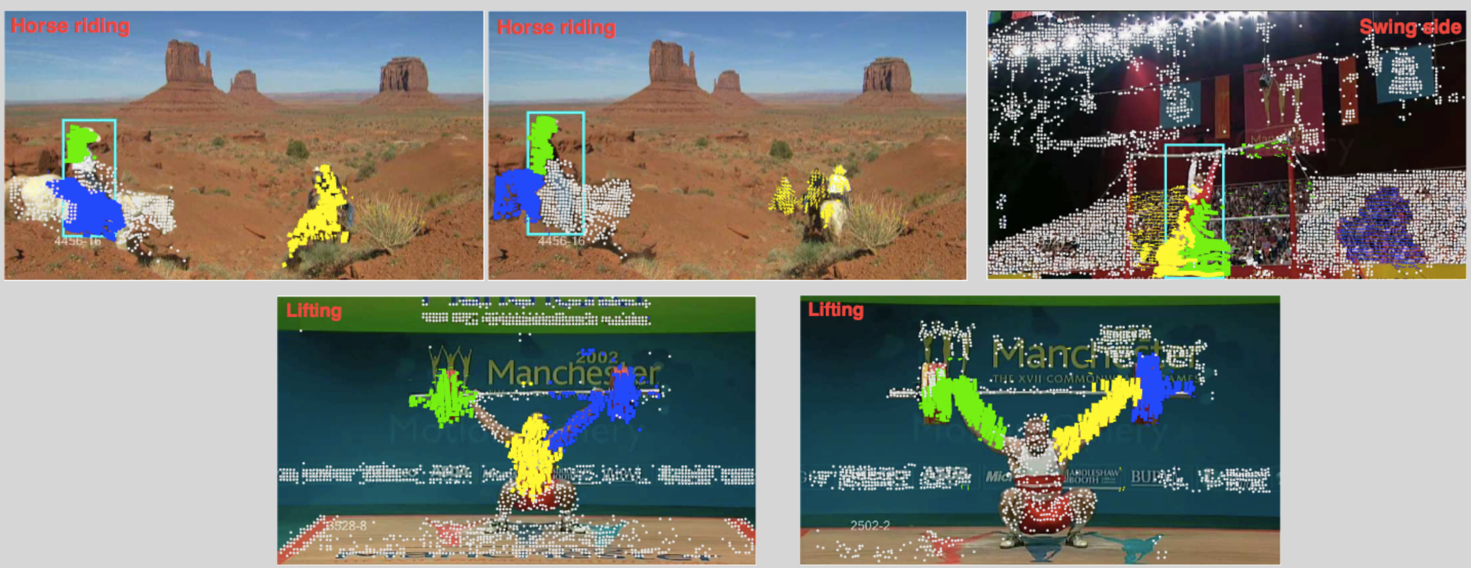
- **Weak Annotations with Bounding Boxes**
    - **Restrict the selections of parts to groups intersecting the bounding boxes**

HOHA Dataset

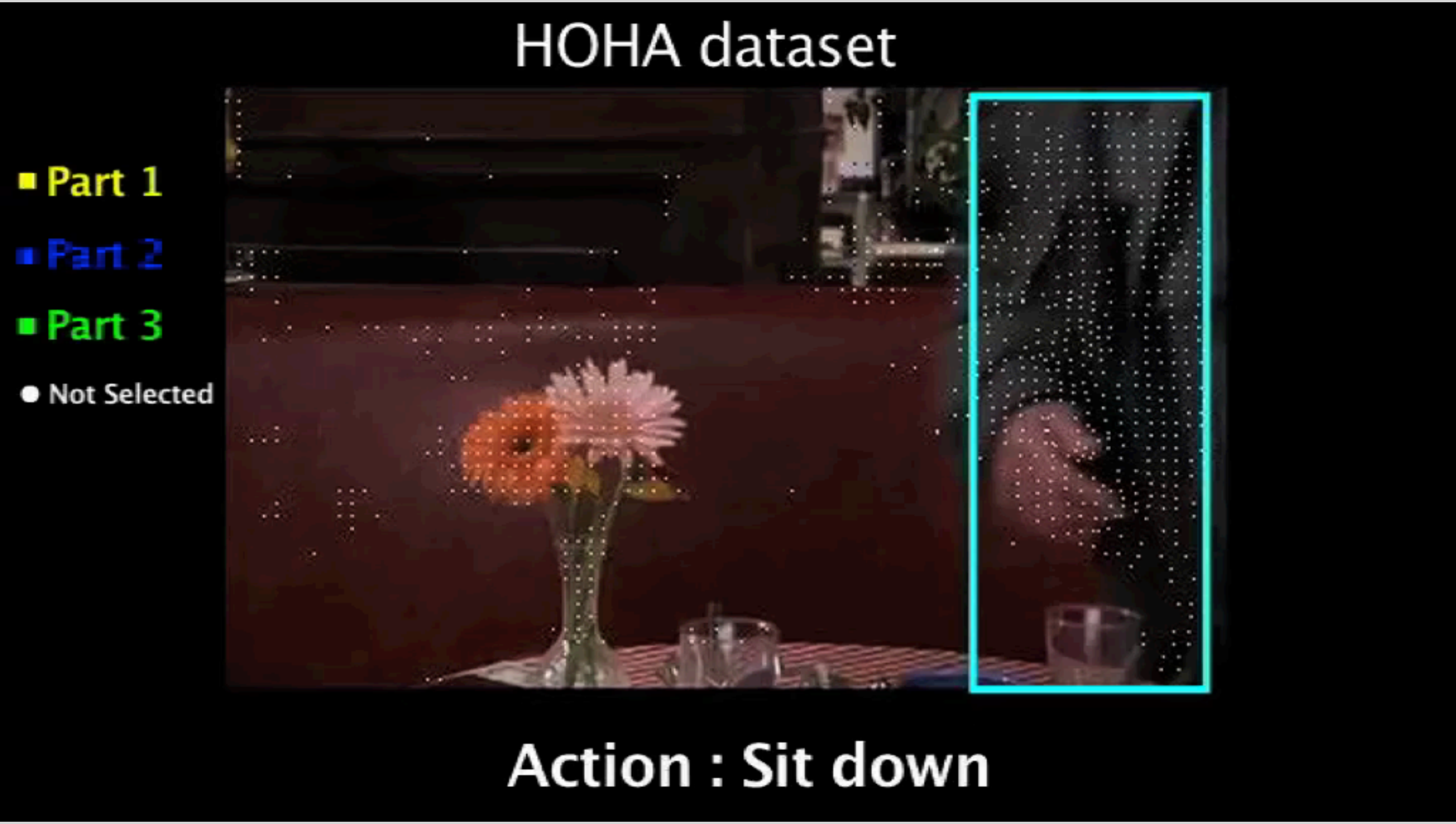


UCF Sports Dataset



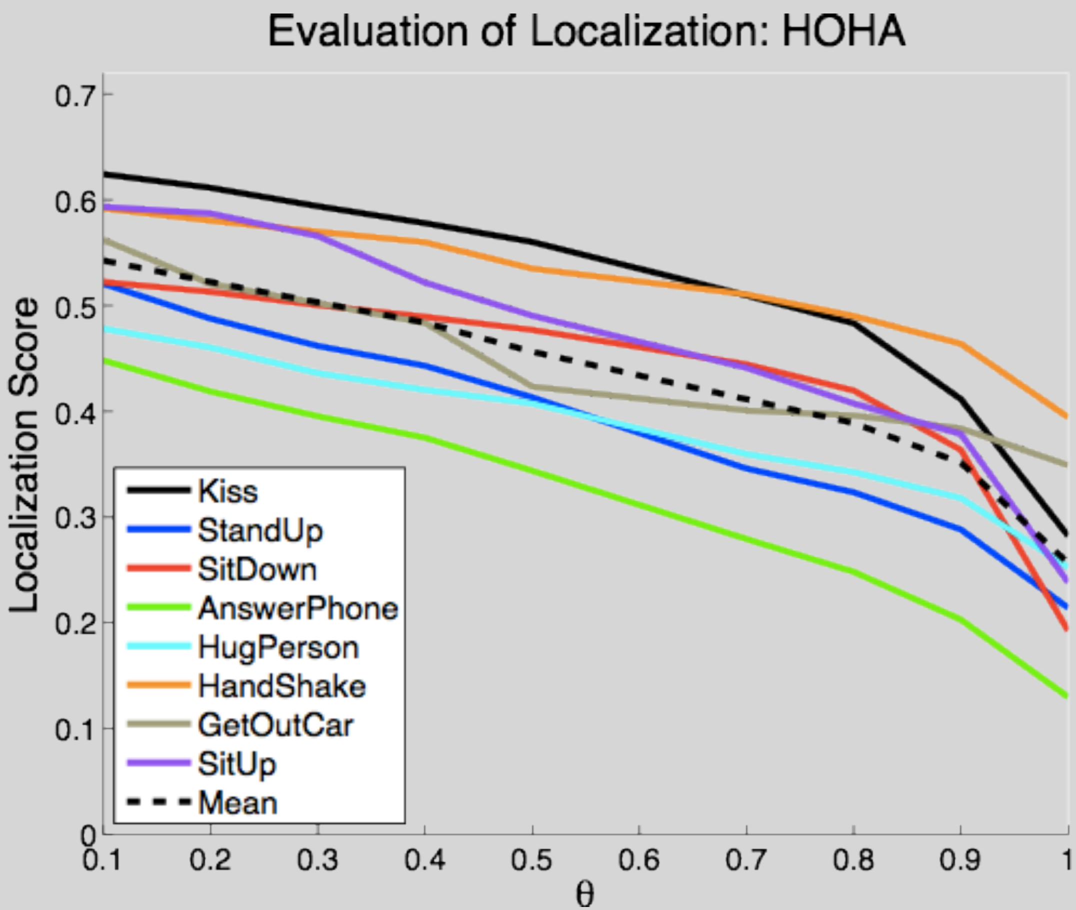


Test Examples



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

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