CS294-43: Visual Object and Activity Recognition

Prof. Trevor Darrell
Spring 2009

March 3rd, 2009

Last Week– **Voting, Hashing, and Random Forest techniques**

- B. Leibe, A. Leonardis, and B. Schiele, "An implicit shape model for combined object categorization and segmentation," In ECCV workshop on statistical learning in computer vision 2006, pp. 508-524.

- A. Frome, Y. Singer, F. Sha, and J. Malik, "Learning globally-consistent local distance functions for shape-based image retrieval and classification," in Proceedings of IEEE 11th International Conference on Computer Vision, 2007, pp. 1-8.

- J. Shotton, M. Johnson, and R. Cipolla, "Semantic texton forests for image categorization and segmentation," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

- P. Jain, B. Kulis, and K. Grauman, "Fast image search for learned metrics," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, 2008, pp. 1-8

- M. Ozuysal, P. Fua, and V. Lepetit, "Fast keypoint recognition in ten lines of code," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007, pp. 1-8.

- A. Torralba, R. Fergus, and Y. Weiss, "Small codes and large image databases for recognition," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8.

Today – **(More) Discriminative approaches (SVM, HCRF)**

- Classic SVM on “bags of features”:

  C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in ECCV International Workshop on Statistical Learning in Computer Vision, 2004.

- ISM + SVM + Local Kernels:

  M. Fritz; B. Leibe; B. Caputo; B. Schiele: Integrating Representative and Discriminant Models for Object Category Detection, ICCV'05, Beijing, China, 2005 *[M. Fritz]*

- Local SVM:

  H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2126-2136. *[M. Maire]*

- “Latent” SVM with deformable parts:

  P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008., June 2008.

- Hidden Conditional Random Fields:

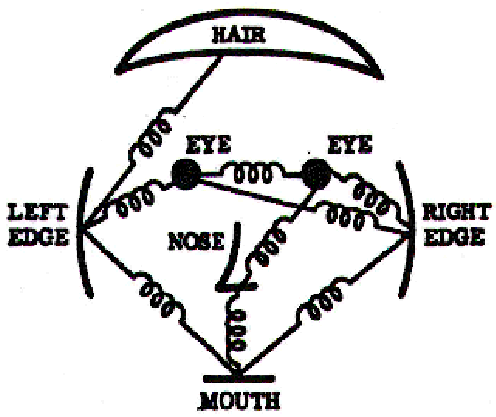
  Y. Wang and G. Mori, “Learning a Discriminative Hidden Part Model for Human Action Recognition”, Advances in Neural Information Processing Systems (NIPS), 2008

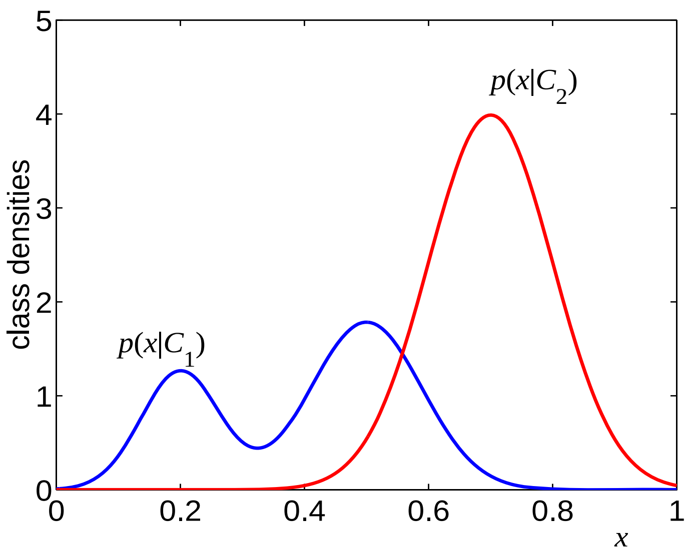
But first…

Some hints from Subransu Maji's latest work on discriminative voting….

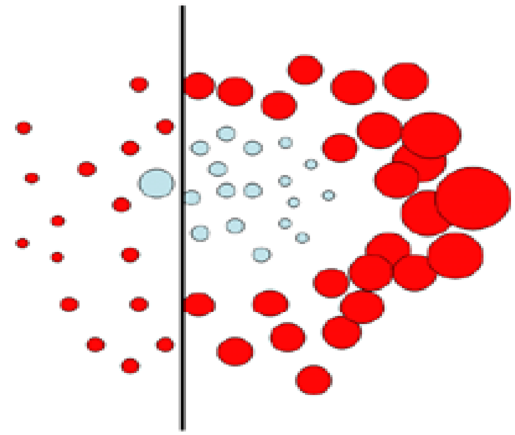
Generative vs. Discriminative

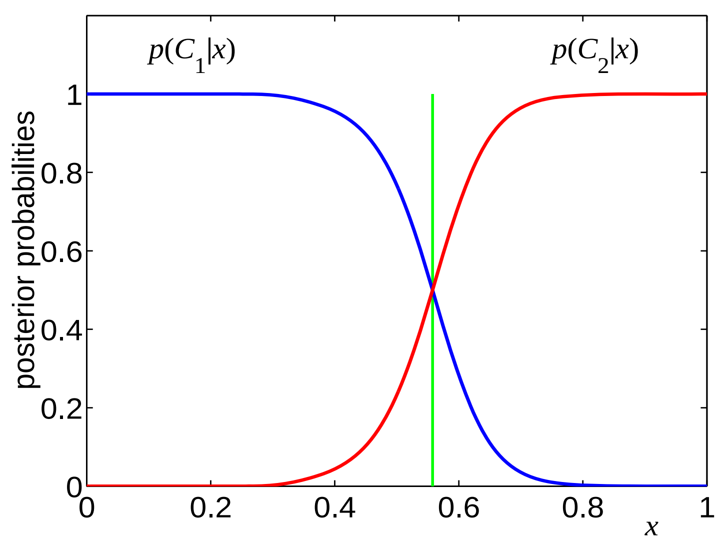
“Model the world”





“Model the decision”





Today – **Discriminative approaches**

- **C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in ECCV International Workshop on Statistical Learning in Computer Vision, 2004.**

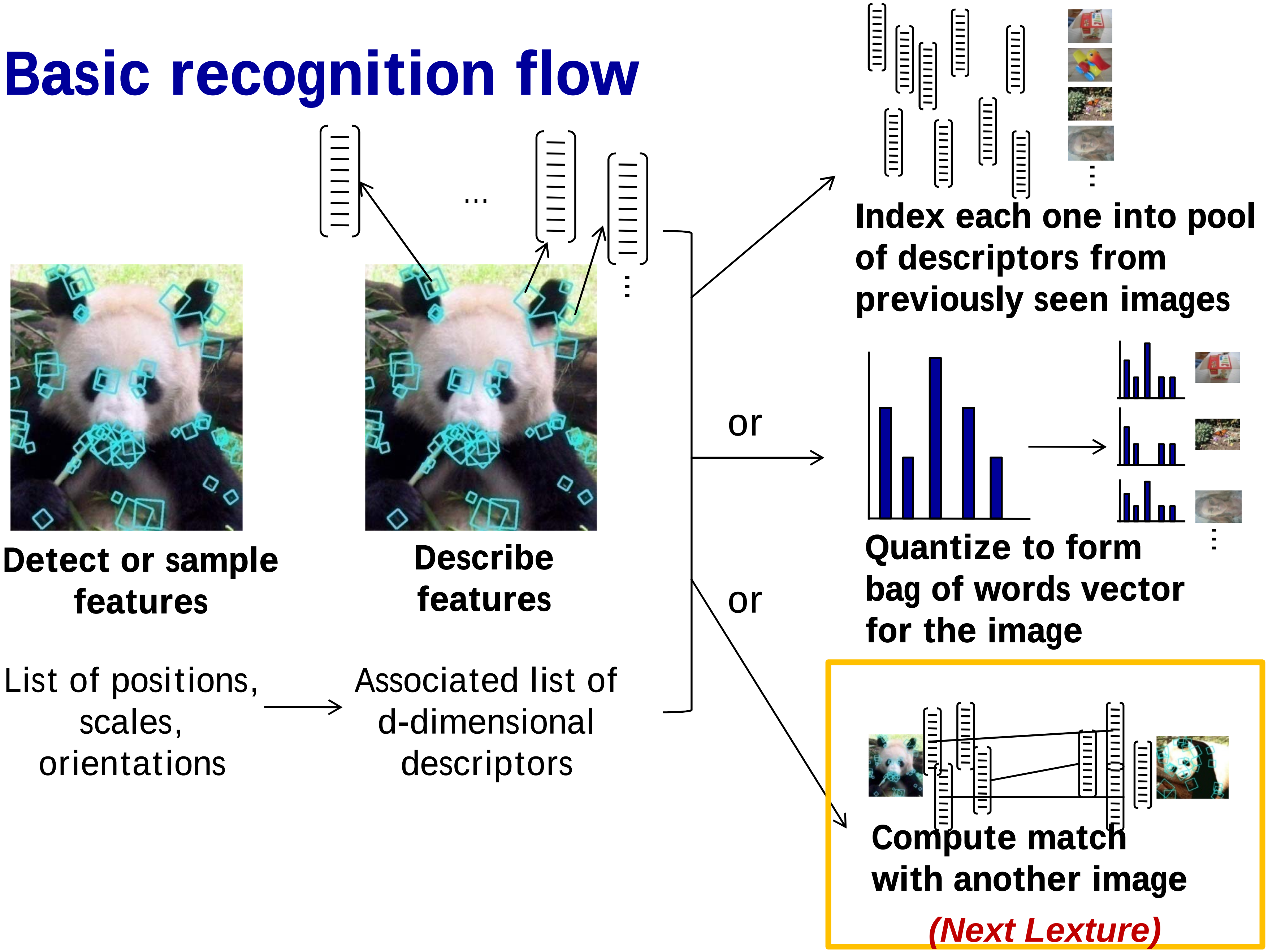
- M. Fritz; B. Leibe; B. Caputo; B. Schiele: Integrating Representative and Discriminant Models for Object Category Detection, ICCV'05, Beijing, China, 2005

- H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2126-2136.

- P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008., June 2008**.**

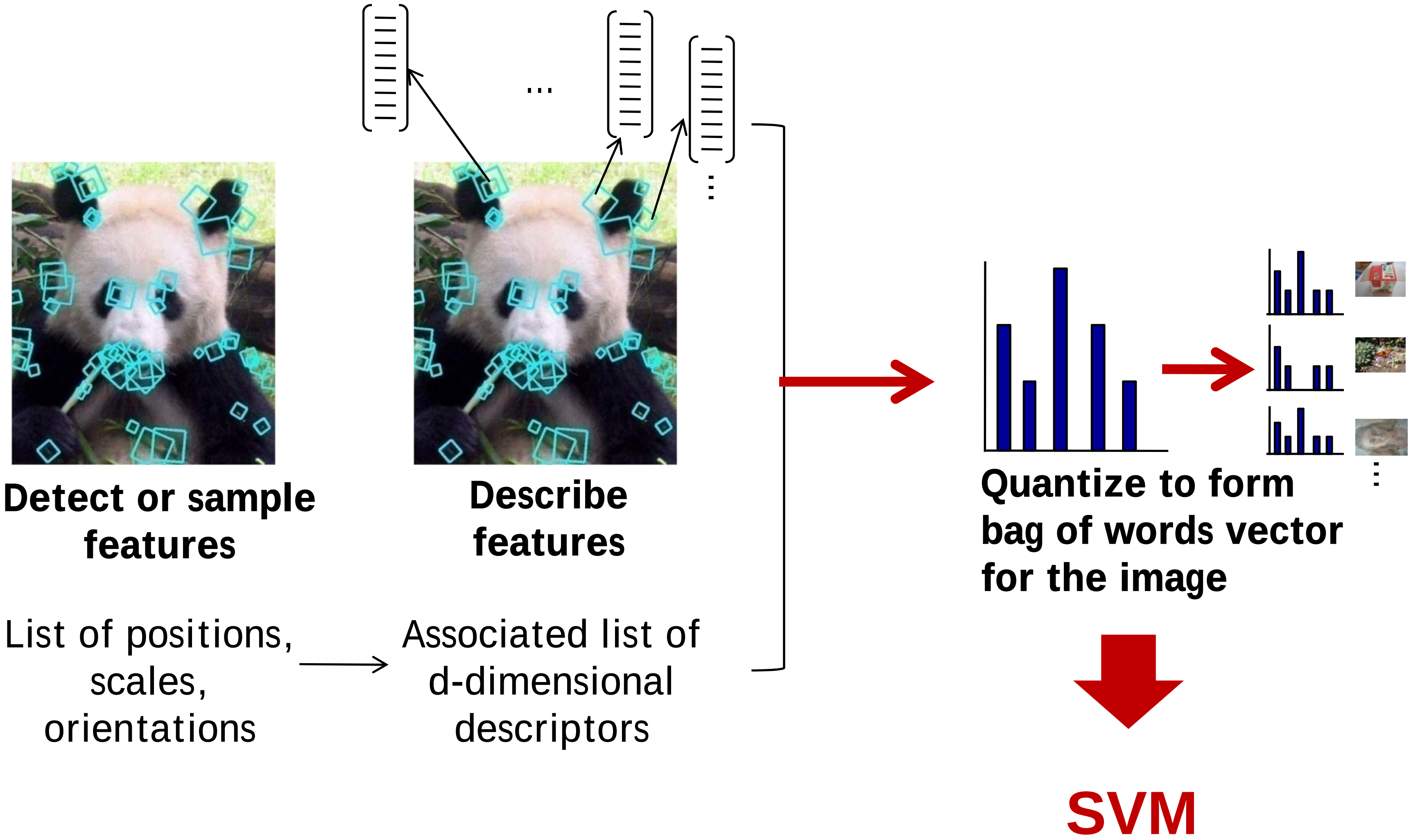
- Y. Wang and G. Mori, "Learning a Discriminative Hidden Part Model for Human Action Recognition", Advances in Neural Information Processing Systems (NIPS), 2008

Basic recognition flow



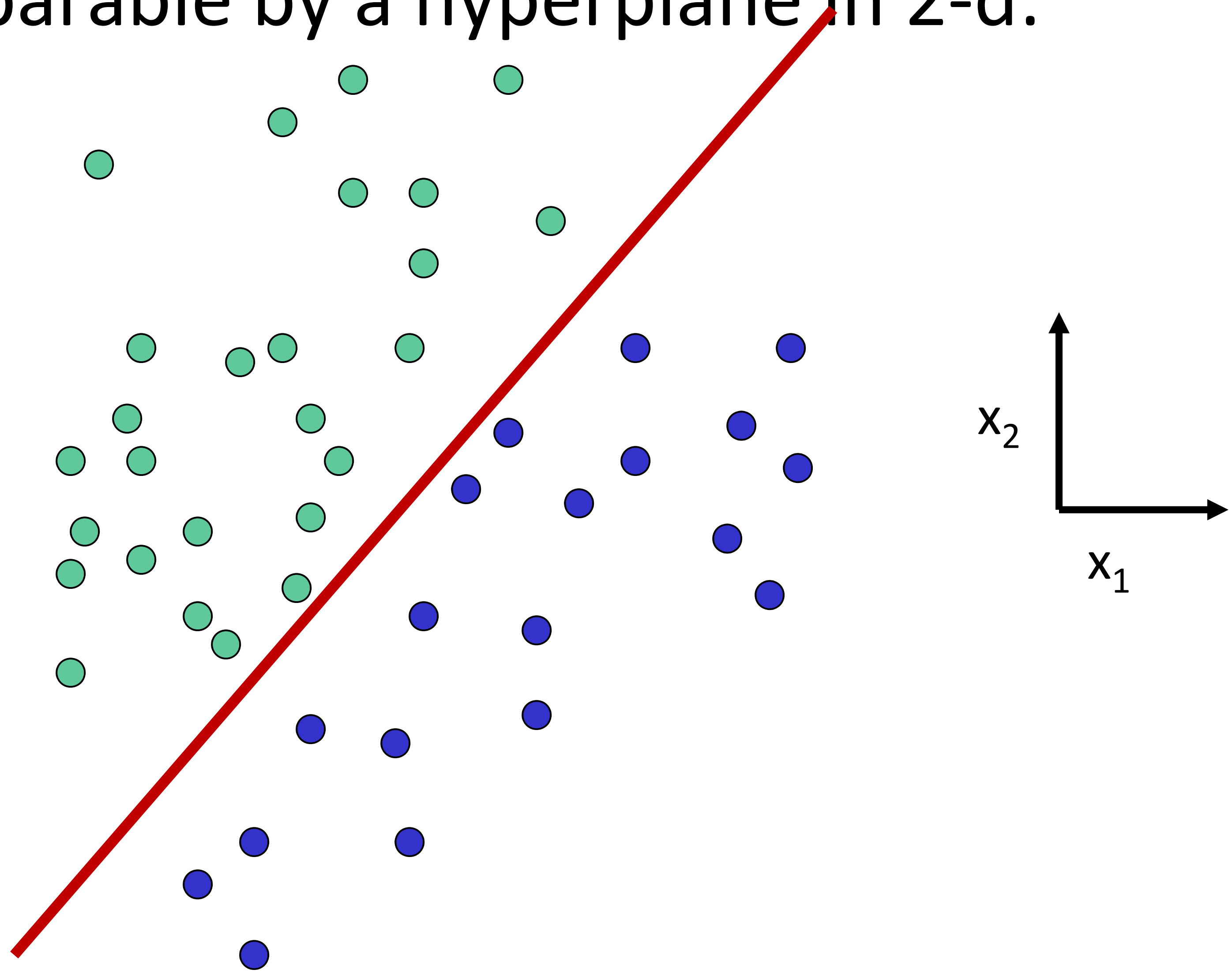
K. Grauman, B. Leibe

Dance et al.

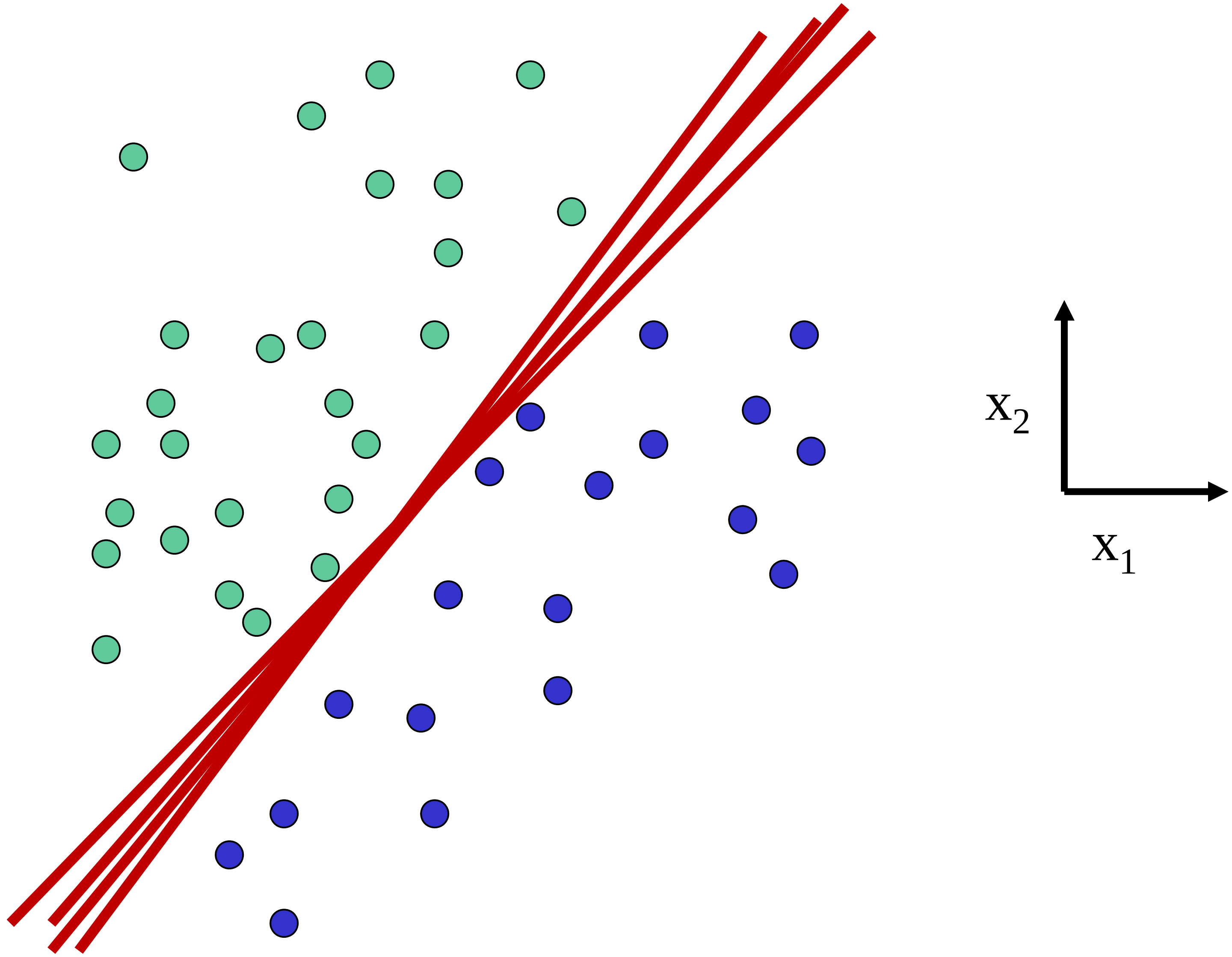


...
Detect or sample features
Describe features
List of positions, scales, orientations
Associated list of d-dimensional descriptors
Quantize to form bag of words vector for the image
SVM

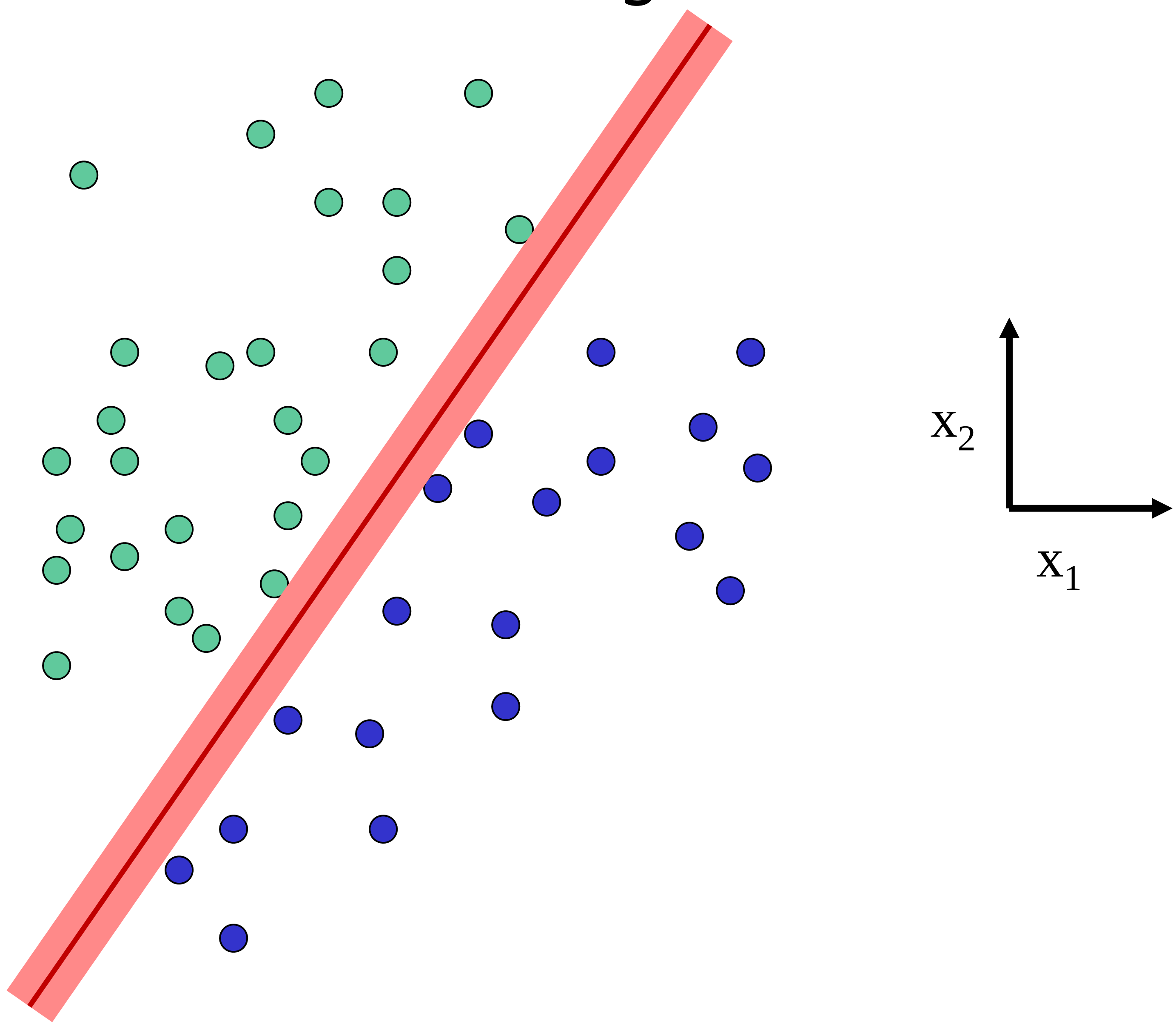
SVM Review…
Separable by a hyperplane in 2-d:



Which one?



Maximum *Margin*:



Linear SVM Classifier

Data: $\{\mathbf{x}_i , y_i\}$ $i = 1,2,3 \ldots N$ $y_i = \{-1,+1\}$

Discriminant: $f(\mathbf{x}) = (\mathbf{w} \mathbf{.} \mathbf{x} + b) > 0$

minimize $|| \mathbf{w} ||$

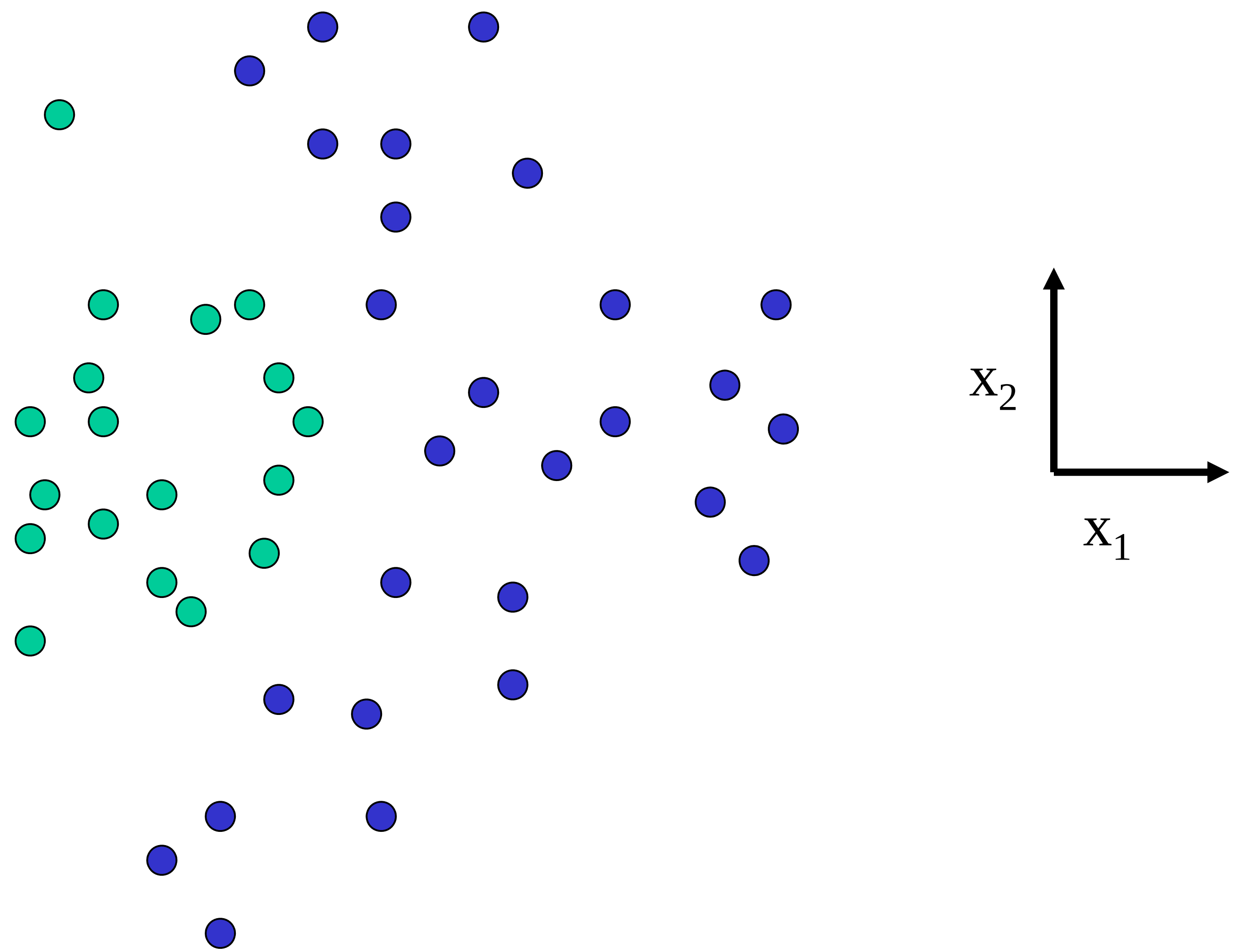
subject to $y_i (\mathbf{w} \mathbf{.} \mathbf{x}_i + b) > 1$ for all i

Solution: QP gives $\{\alpha_i\}$

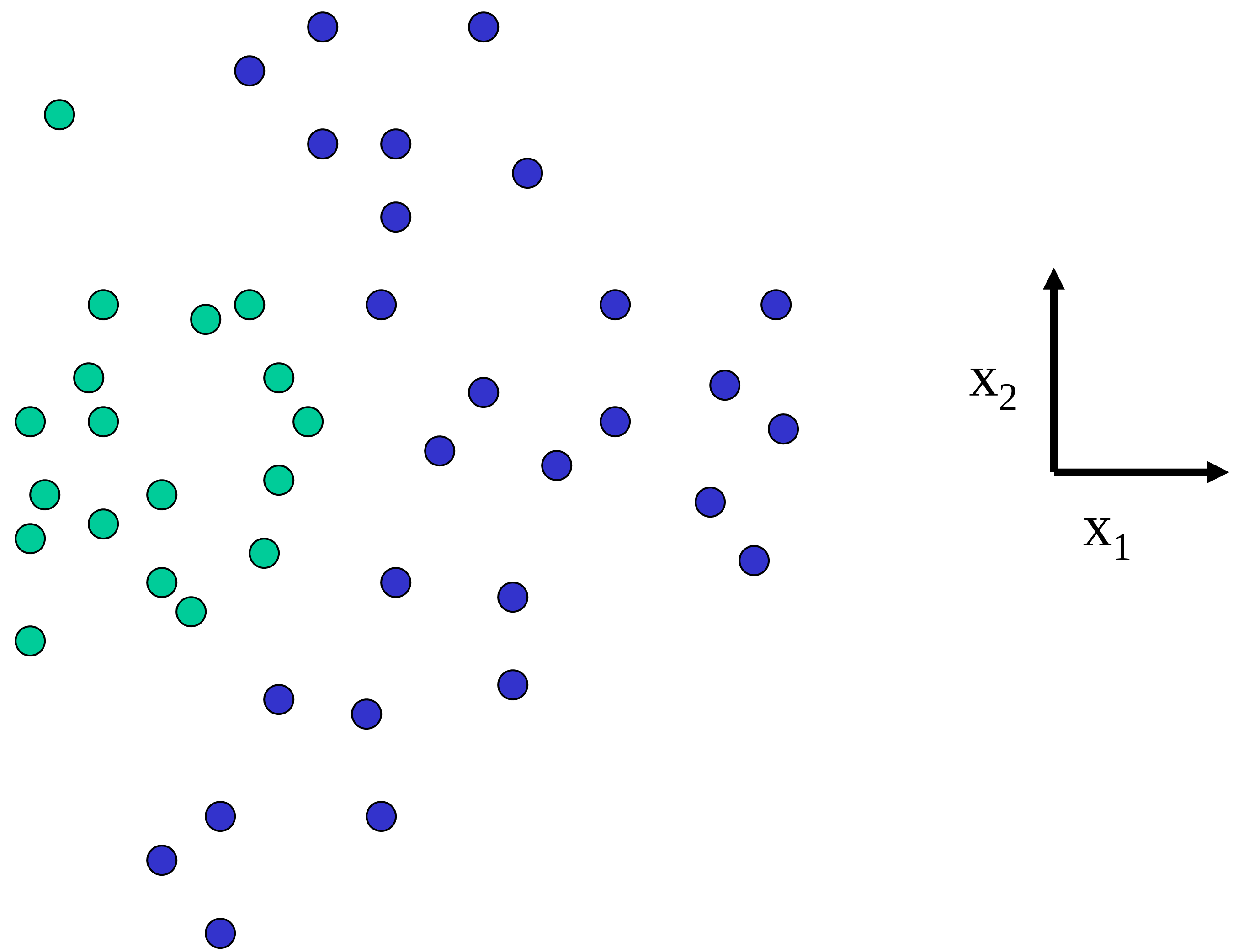
$\mathbf{w}_{opt} = \Sigma \alpha_i y_i \mathbf{x}_i$

$f(\mathbf{x}) = \Sigma \alpha_i y_i (\mathbf{x}_i . \mathbf{x}) + b$

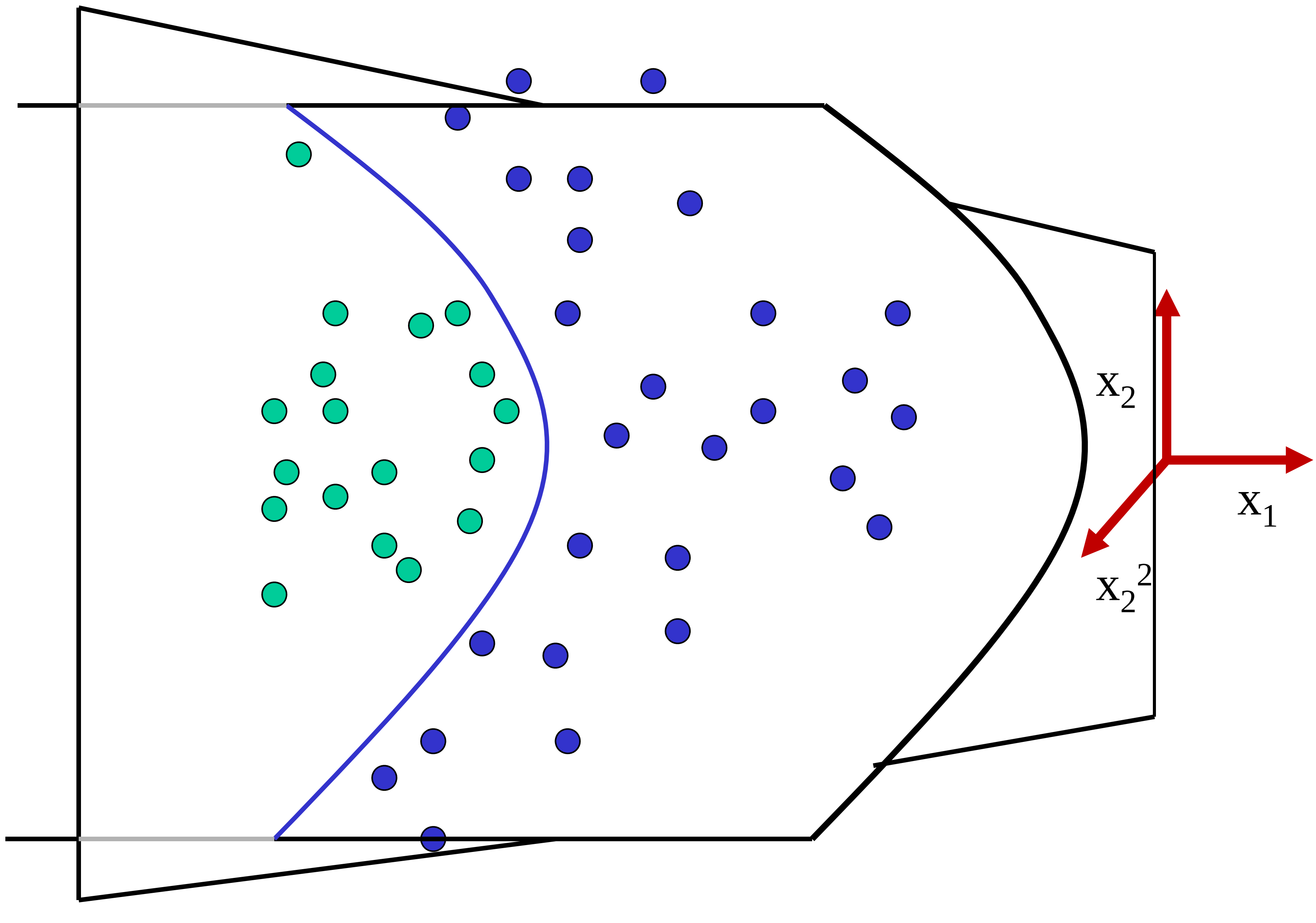
Non-separable by a hyperplane in 2-d



Non-separable by a hyperplane in 2-d



Separable by a hyperplane in 3-d



Embedding

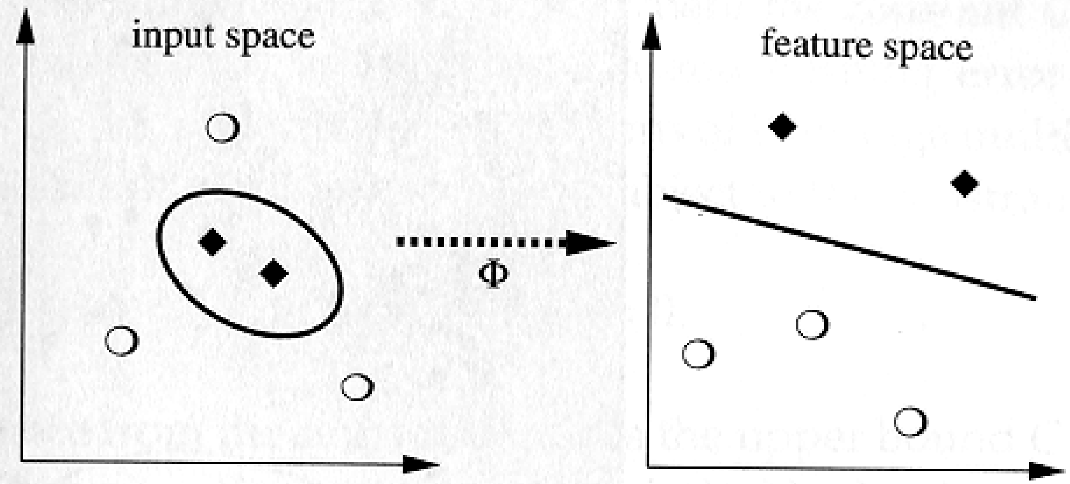


Figure 1.6 The idea of SVMs: map the training data into a higher-dimensional feature space via Φ, and construct a separating hyperplane with maximum margin there. This yields a nonlinear decision boundary in input space. By the use of a kernel function (1.2), it is possible to compute the separating hyperplane without explicitly carrying out the map into the feature space.

Learning with Kernels, Scholkopf and Smola, 2002

Kernels

- linear classifier:

$$f(x) = \mathbf{sign}(w^T x + b)$$

- Kernel classifier:

$$K(u, v) = \Phi(u) \cdot \Phi(v)$$

$$f(x) = \mathbf{sign}\left(\sum_i y_i \alpha_i K(x, x_i) + b\right).$$

[Dance et al.]

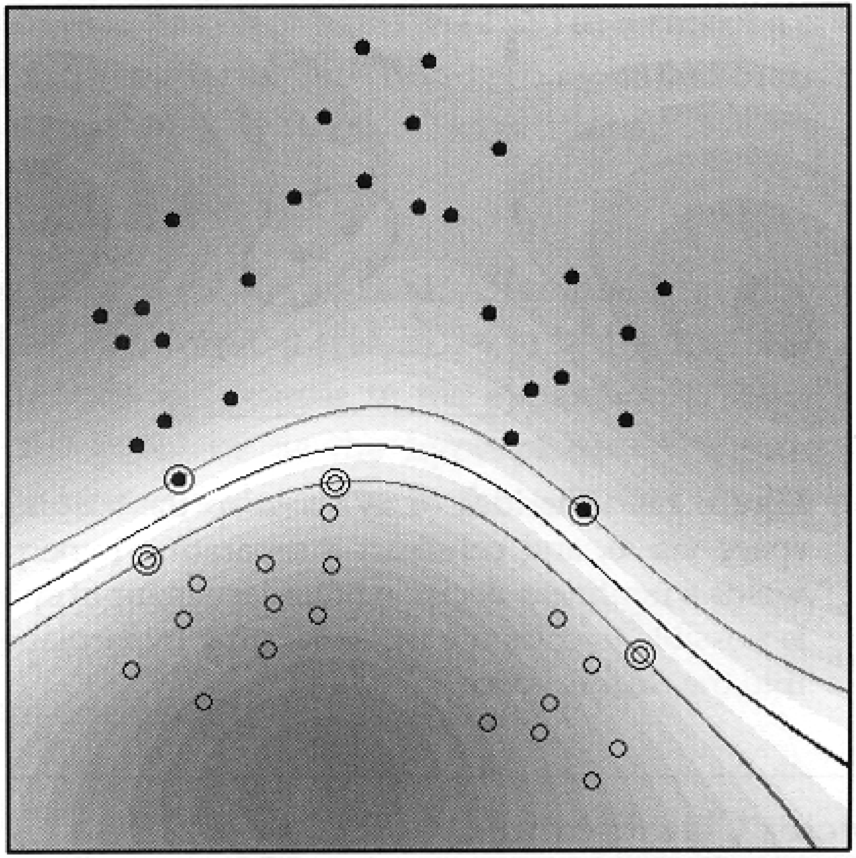


Figure 1.7 Example of an SV classifier found using a radial basis function kernel $k(x, x') = \exp(-\|x - x'\|^2)$ (here, the input space is $\mathcal{X} = [-1, 1]^2$). Circles and disks are two classes of training examples; the middle line is the decision surface; the outer lines precisely meet the constraint (1.25). Note that the SVs found by the algorithm (marked by extra circles) are not centers of clusters, but examples which are critical for the given classification task. Gray values code $|\sum_{i=1}^{m} y_i \alpha_i k(x, x_i) + b|$, the modulus of the argument of the decision function (1.35). The top and the bottom lines indicate places where it takes the value 1 (from [471]).

Learning with Kernels, Scholkopf and Smola, 2002

Example Kernel functions

- Polynomials
- Gaussians
- Sigmoids
- Radial basis functions
- **Local feature Kernels (c.f. Fritz et al. and correspondence Kernels in next lecture)**

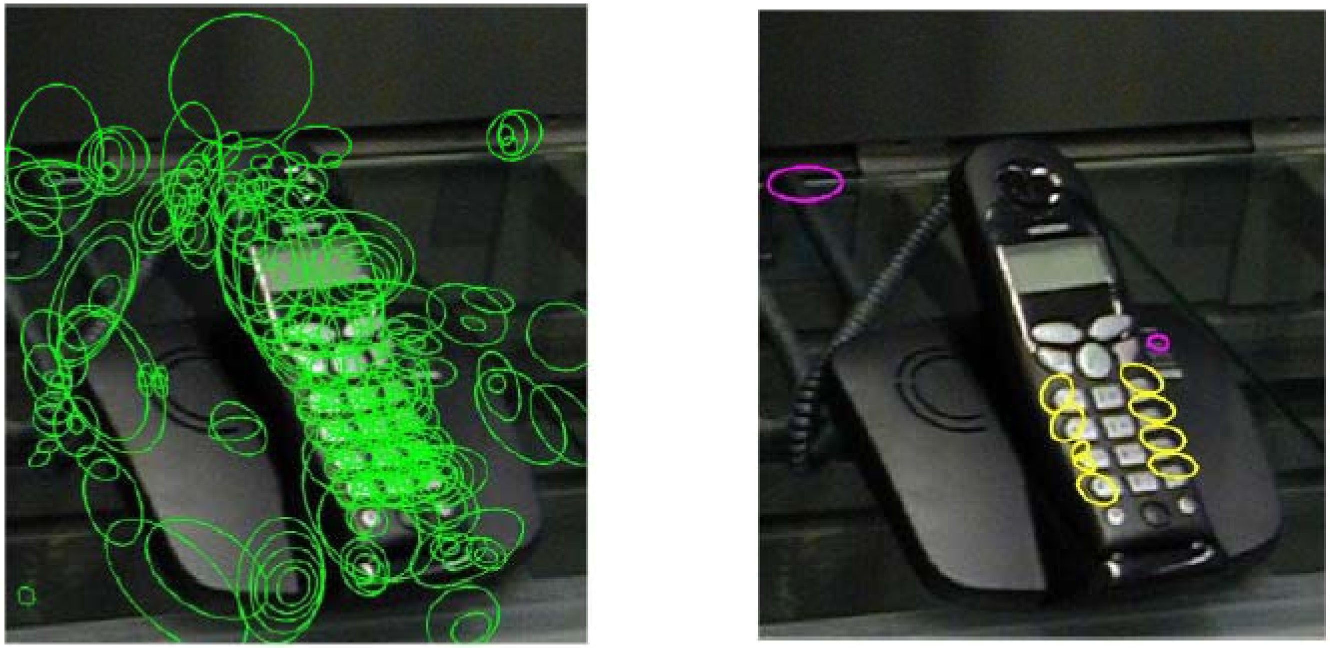
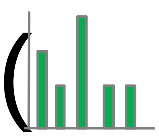
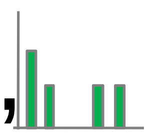
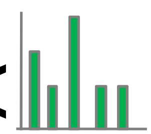
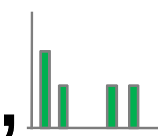


Fig. 4. *Left* all patches detected for this image. *Right* patches from two selected clusters occurring in this image (yellow and magenta ellipses).

Tried linear, quadratic, cubic; linear had best performance….

K( , ) = K( , ) = <  ,  >

[Dance et al.]

Fig. 5. Images correctly classified containing multiple objects of the same category.

SVM:

Table 2. Confusion matrix and mean rank for SVM (*k=1000, linear kernel)*.

True classes →	*faces*	*buildings*	*trees*	*cars*	*phones*	*bikes*	*books*
faces	**98**	14	10	10	34	0	13
buildings	1	**63**	3	0	3	1	6
trees	1	10	**81**	1	0	6	0
cars	0	1	1	**85**	5	0	5
phones	0	5	4	3	**55**	2	3
bikes	0	4	1	0	1	**91**	0
books	0	3	0	1	2	0	**73**
Mean ranks	1.04	1.77	1.28	1.30	1.83	1.09	1.39

Naïve Bayes:

Table 1. Confusion matrix and the mean rank for the best vocabulary (*k=1000)*.

True classes →	*faces*	*buildings*	*trees*	*cars*	*phones*	*bikes*	*books*
faces	**76**	4	2	3	4	4	13
buildings	2	**44**	5	0	5	1	3
trees	3	2	**80**	0	0	5	0
cars	4	1	0	**75**	3	1	4
phones	9	15	1	16	**70**	14	11
bikes	2	15	12	0	8	**73**	0
books	4	19	0	6	7	2	**69**
Mean ranks	1.49	1.88	1.33	1.33	1.63	1.57	1.57

[Dance et al.]

Today – **Discriminative approaches**

- C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in ECCV International Workshop on Statistical Learning in Computer Vision, 2004.
- **M. Fritz; B. Leibe; B. Caputo; B. Schiele: Integrating Representative and Discriminant Models for Object Category Detection, ICCV'05, Beijing, China, 2005**
- H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2126-2136.
- P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008., June 2008**.**
- Y. Wang and G. Mori, “Learning a Discriminative Hidden Part Model for Human Action Recognition”, Advances in Neural Information Processing Systems (NIPS), 2008

Today – **Discriminative approaches**

- C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in ECCV International Workshop on Statistical Learning in Computer Vision, 2004.

- M. Fritz; B. Leibe; B. Caputo; B. Schiele: Integrating Representative and Discriminant Models for Object Category Detection, ICCV'05, Beijing, China, 2005

- **H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2126-2136**.

- P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008., June 2008**.**

- Y. Wang and G. Mori, “Learning a Discriminative Hidden Part Model for Human Action Recognition”, Advances in Neural Information Processing Systems (NIPS), 2008

Today – **Discriminative approaches**

- C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in ECCV International Workshop on Statistical Learning in Computer Vision, 2004.

- M. Fritz; B. Leibe; B. Caputo; B. Schiele: Integrating Representative and Discriminant Models for Object Category Detection, ICCV'05, Beijing, China, 2005

- H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2126-2136.

- **P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008., June 2008.**

- Y. Wang and G. Mori, “Learning a Discriminative Hidden Part Model for Human Action Recognition”, Advances in Neural Information Processing Systems (NIPS), 2008

Discriminatively Trained Mixtures of Deformable Part Models

Pedro Felzenszwalb and Ross Girshick
University of Chicago

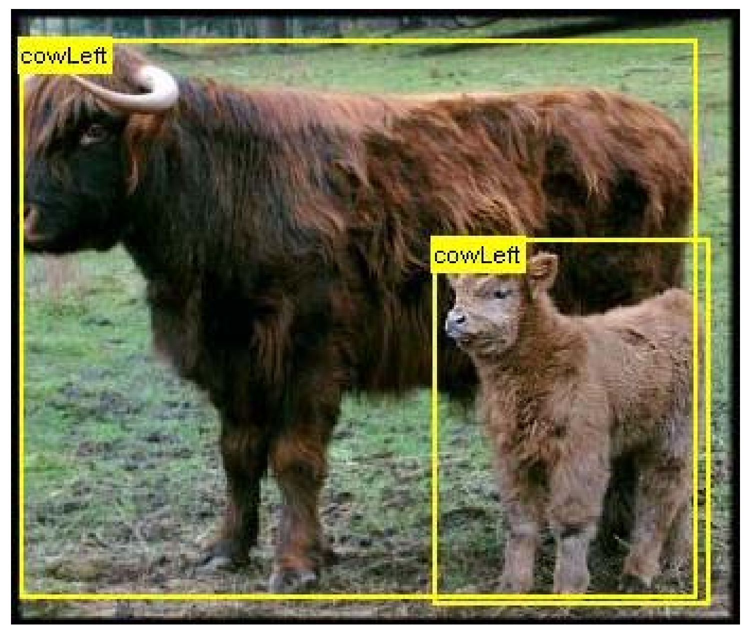
David McAllester
Toyota Technological Institute at Chicago

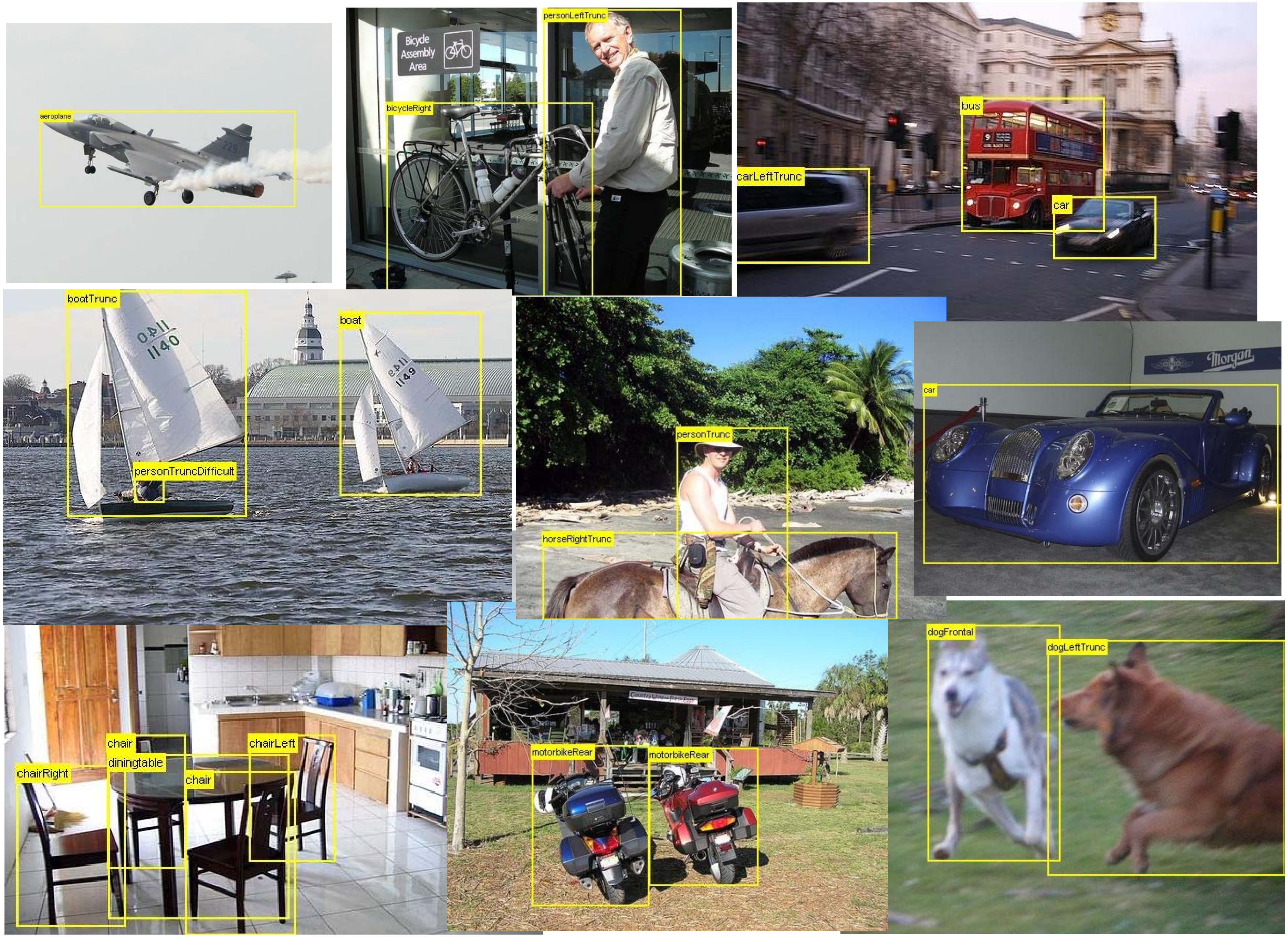
Deva Ramanan
UC Irvine

http://www.cs.uchicago.edu/~pff/latent

PASCAL Challenge

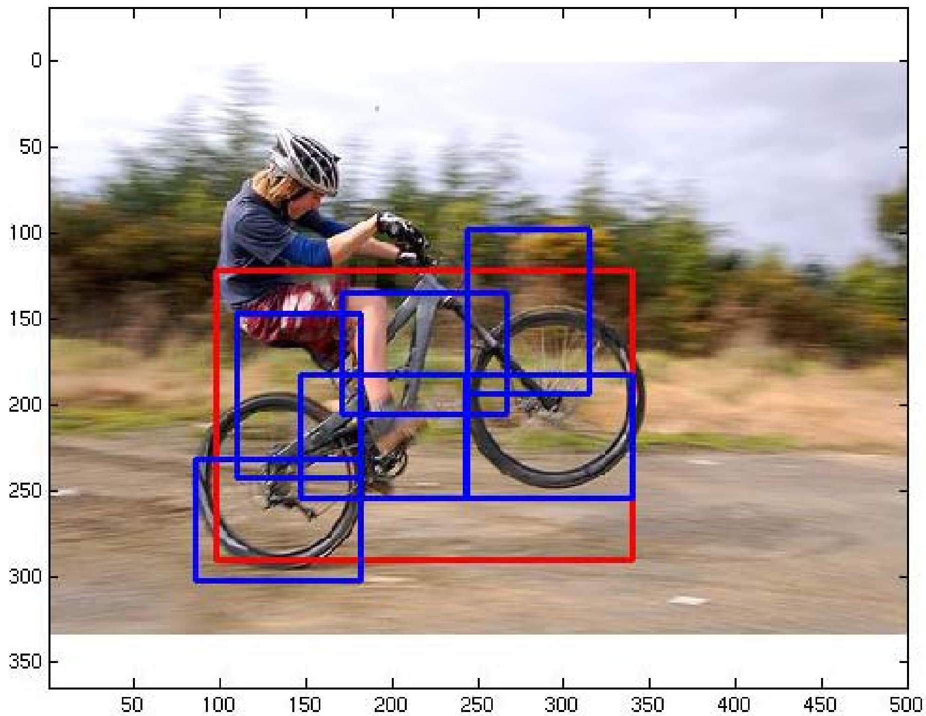
- ~10,000 images, with ~25,000 target objects.
  - Objects from 20 categories (person, car, bicycle, cow, table...).
  - Objects are annotated with labeled bounding boxes.

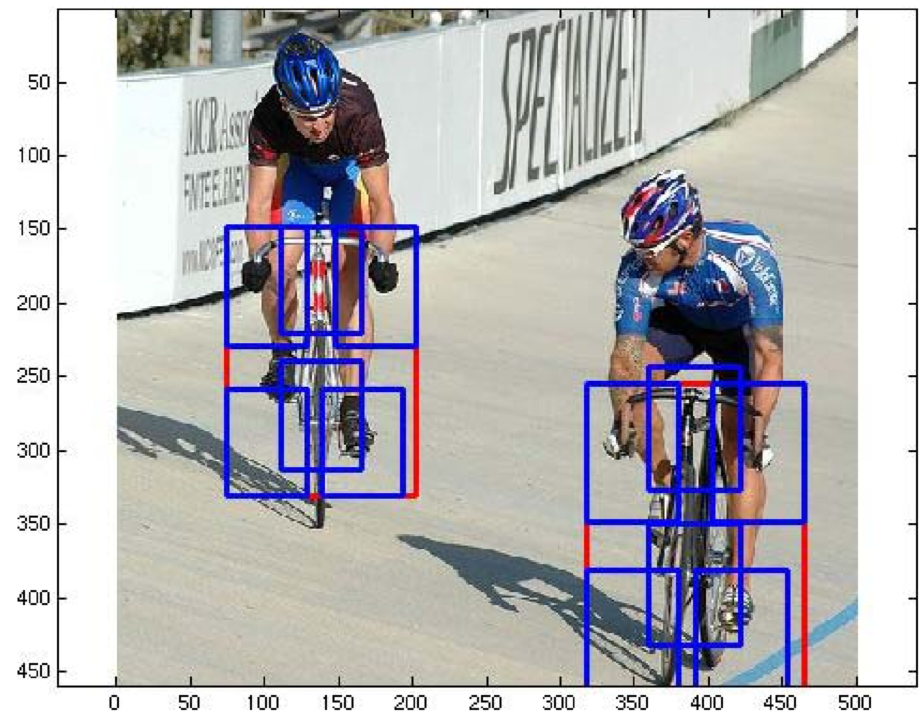




aeroplane
Bicycle Assembly Area
personLeftTrunc
bicycleRight
bus
carLeftTrunc
car
boatTrunc
1140
personTruncDifficult
boat
1149
personTrunc
horseRightTrunc
Morgan
car
chair
chairLeft
diningtable
chairRight
chair
motorbikeRear
motorbikeRear
dogFrontal
dogLeftTrunc

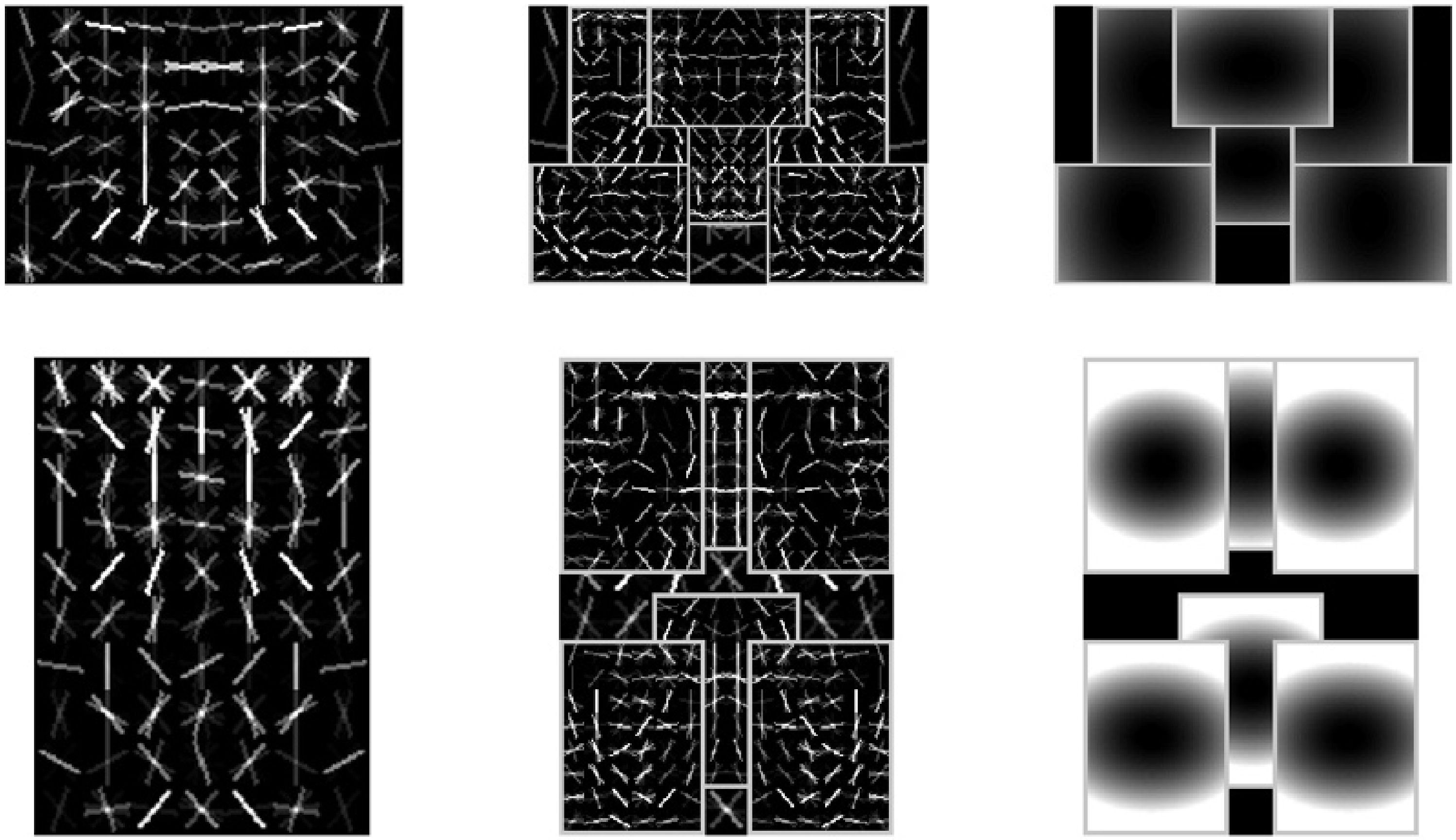
Model Overview





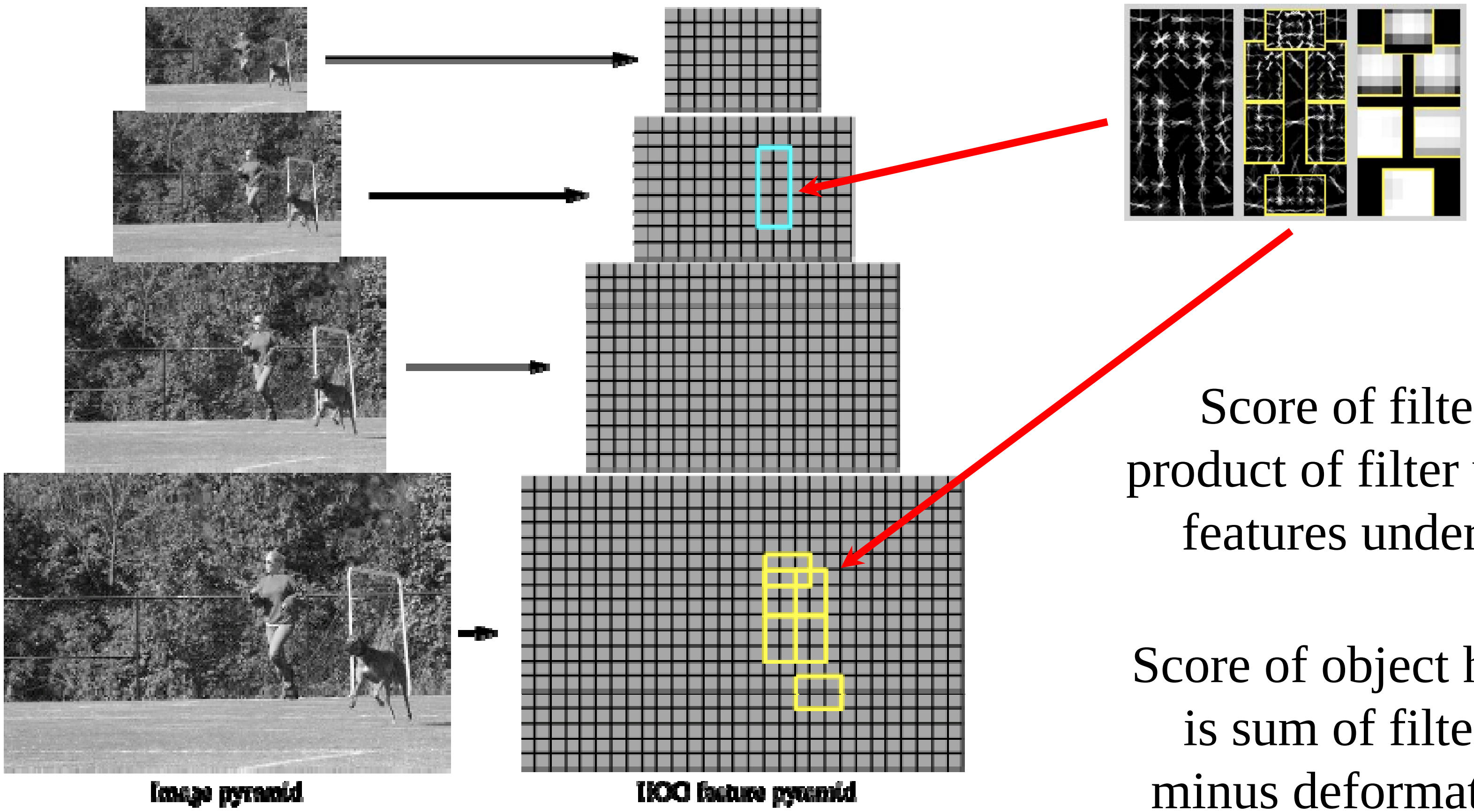
- Mixture of deformable part models (pictorial structures)
- Each component has global template + deformable parts
- Fully trained from bounding boxes alone

2 component bicycle model



root filters coarse resolution | part filters finer resolution | deformation models

Object Hypothesis



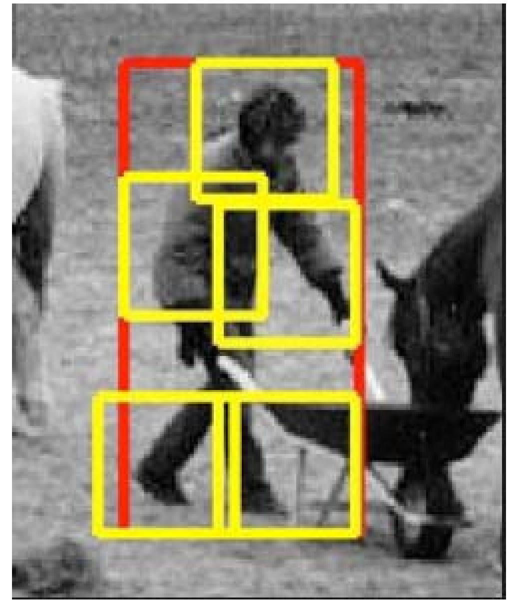
Score of filter is dot product of filter with HOG features underneath it

Score of object hypothesis is sum of filter scores minus deformation costs

Multiscale model captures features at two resolutions

Model

$$f_w(x) = w \cdot \Phi(x)$$



$$f_w(x) = \max_z w \cdot \Phi(x, z)$$

Z = vector of part offsets

$\Phi(x, z)$= vector of HOG features (from root filter & appropriate part sub-windows) and part offsets

Latent SVM

$$f_w(x) = \max_z w \cdot \Phi(x, z)$$

Linear in w if z is fixed

Training data: $(x_1, y_1), \ldots, (x_n, y_n)$ with $y_i \in \{-1, 1\}$

Learning: find w such that $y_i f_w(x_i) > 0$

$$w^* = \underset{w}{\operatorname{argmin}}\, \lambda ||w||^2 + \sum_{i=1}^{n} \max(0, 1 - y_i f_w(x_i))$$

Regularization

Hinge loss

Latent SVM training

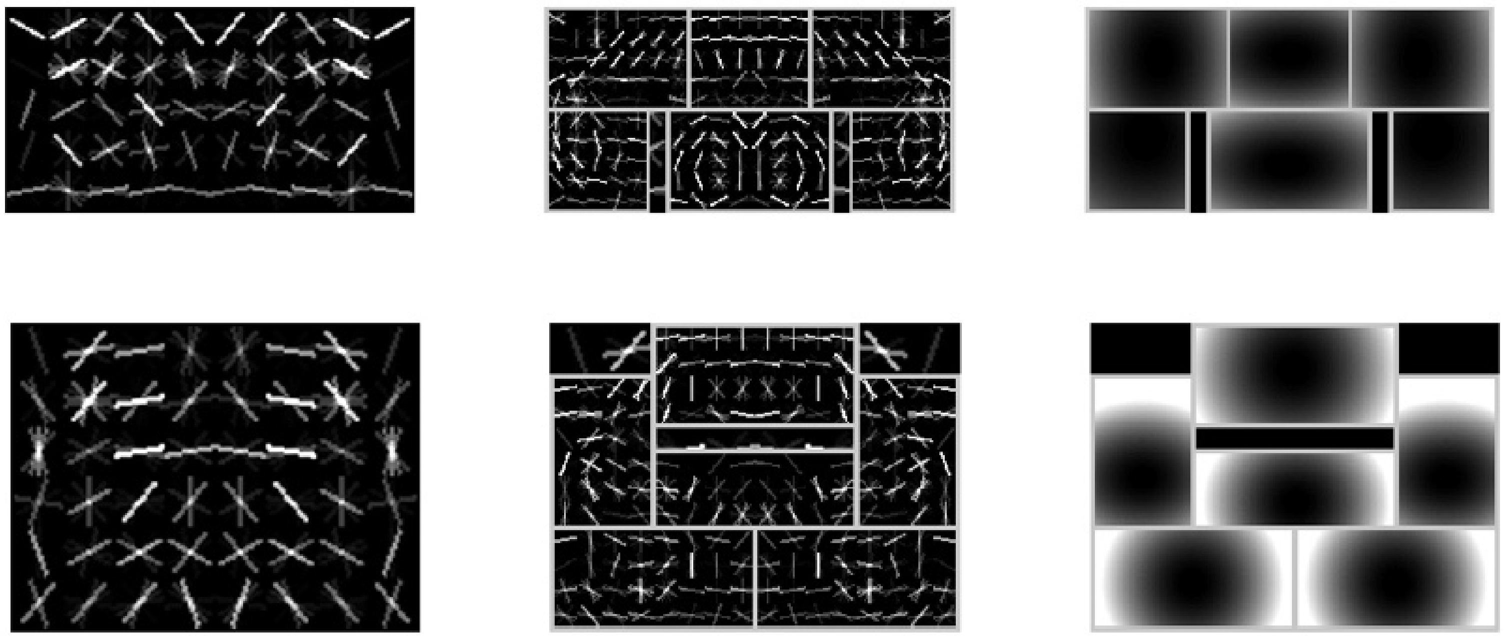
$$w^* = \underset{w}{\operatorname{argmin}}\, \lambda||w||^2 + \sum_{i=1}^{n} \max(0, 1 - y_i f_w(x_i))$$

- Non-convex optimization
- Huge number of negative examples
- Convex if we fix z for positive examples
- Optimization:
  - Initialize w and iterate:
    - Pick best z for each positive example
    - Optimize w via gradient descent with data mining

Initializing w

- For k component mixture model:
- Split examples into k sets based on bounding box aspect ratio
- Learn k root filters using standard SVM
  - Training data: warped positive examples and random windows from negative images (Dalal & Triggs)
- Initialize parts by selecting patches from root filters
  - Subwindows with strong coefficients
  - Interpolate to get higher resolution filters
  - Initialize spatial model using fixed spring constants

Car model

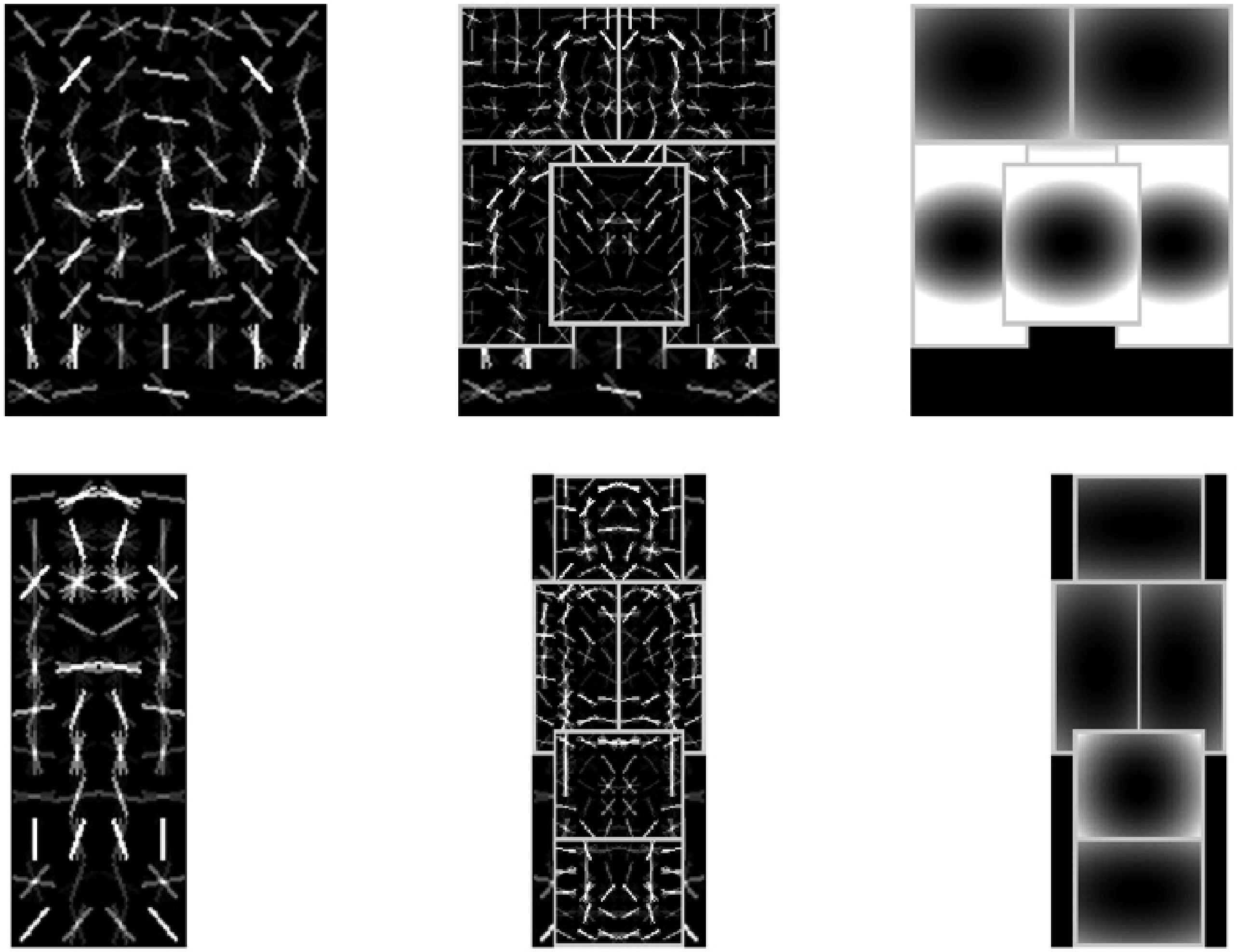


root filters
coarse resolution

part filters
finer resolution

deformation
models

Person model

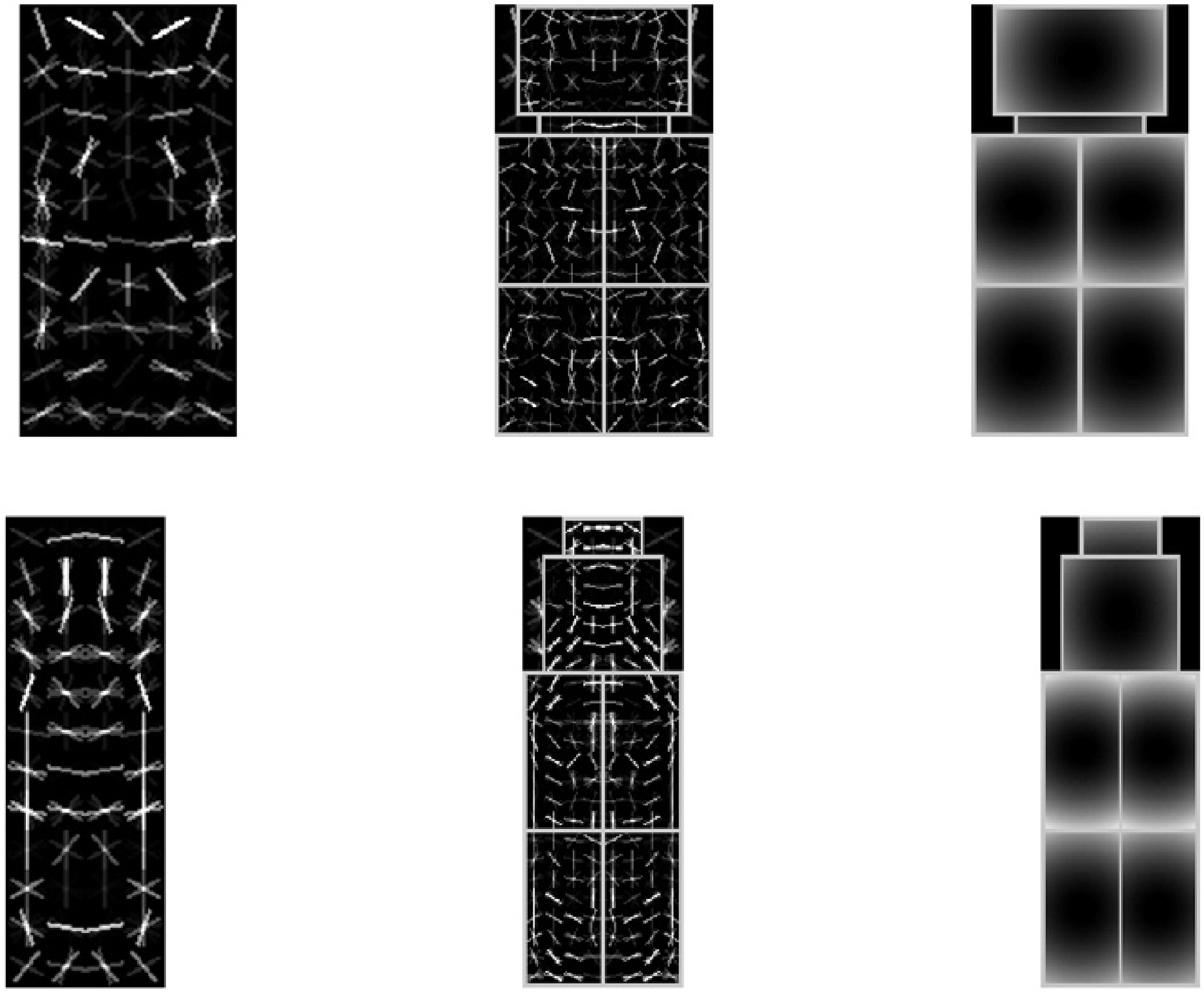


root filters
coarse resolution

part filters
finer resolution

deformation
models

Bottle model

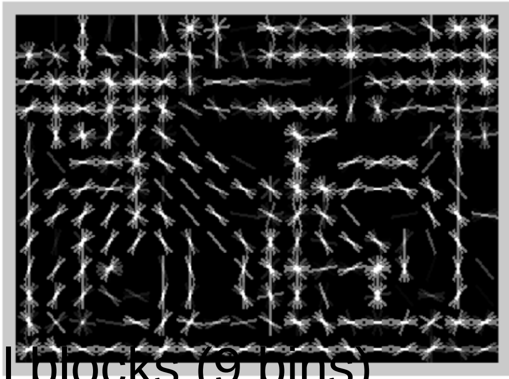


root filters
coarse resolution

part filters
finer resolution

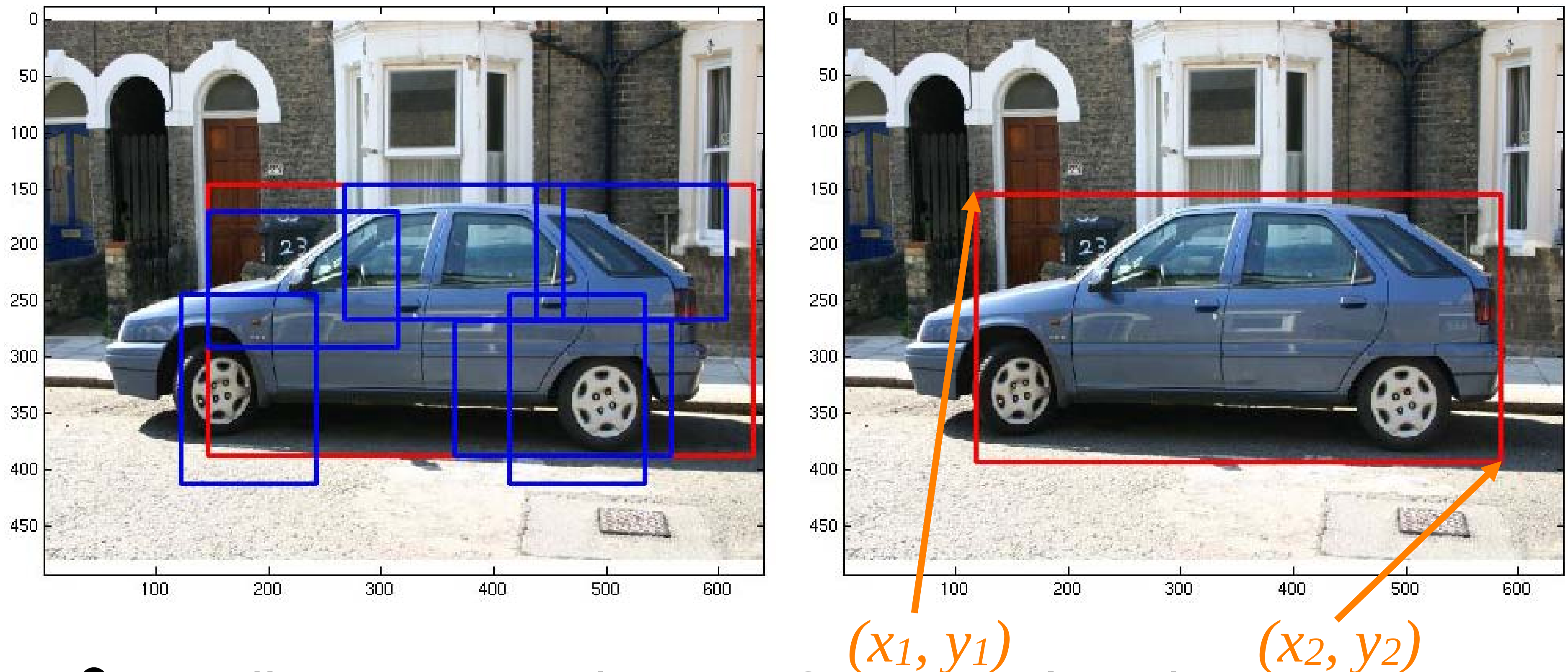
deformation
models

Histogram of Gradient (HOG) features



- Dalal & Triggs:
  - Histogram gradient orientations in 8x8 pixel blocks (9 bins)
  - Normalize with respect to 4 different neighborhoods and truncate
  - 9 orientations * 4 normalizations = 36 features per block
- PCA gives ~10 features that capture all information
  - Fewer parameters, speeds up convolution, but costly projection at runtime
- Analytic projection: spans PCA subspace and easy to compute
  - 9 orientations + 4 normalizations = 13 features
- We also use 2*9 contrast sensitive features for 31 features total

Bounding box prediction

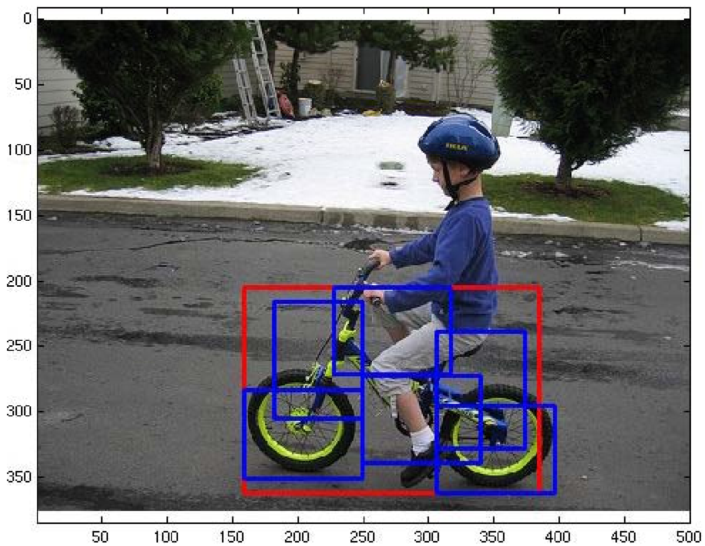


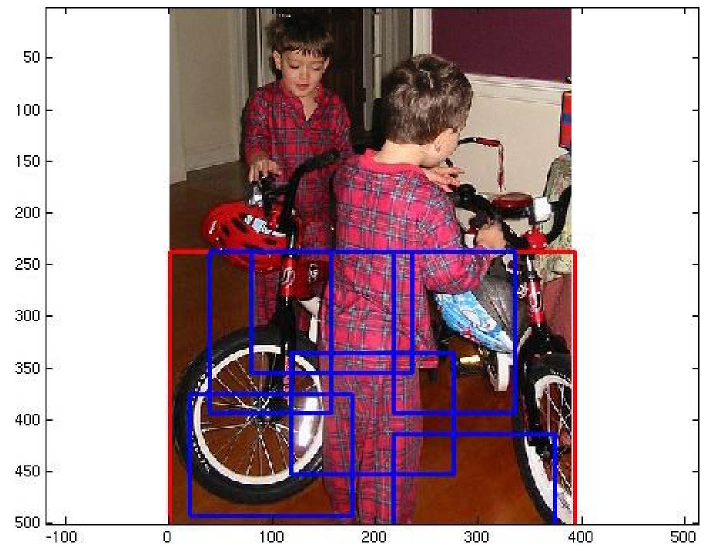
- predict (x_1, y_1) and (x_2, y_2) from part locations
- linear function trained using least-squares regression

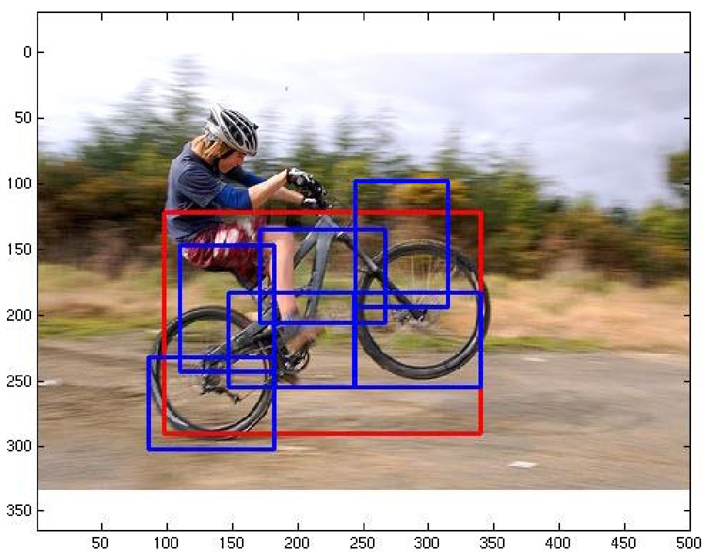
Context rescoring

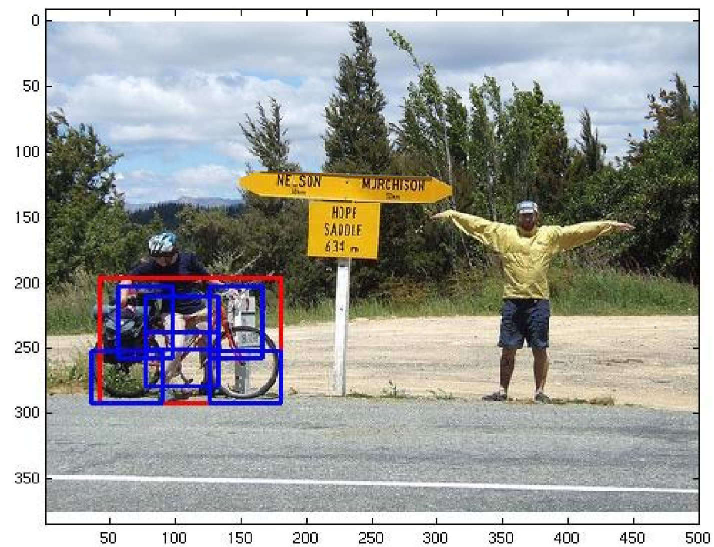
- Rescore a detection using “context” defined by all detections
- Let v_i be the max score of detector for class i in the image
- Let s be the score of a particular detection
- Let $(x_1,y_1), (x_2,y_2)$ be normalized bounding box coordinates
- $f = (s, x_1, y_1, x_2, y_2, v_1, v_2\ldots, v_{20})$
- Train class specific classifier
  - f is positive example if true positive detection
  - f is negative example if false positive detection

Bicycle detection

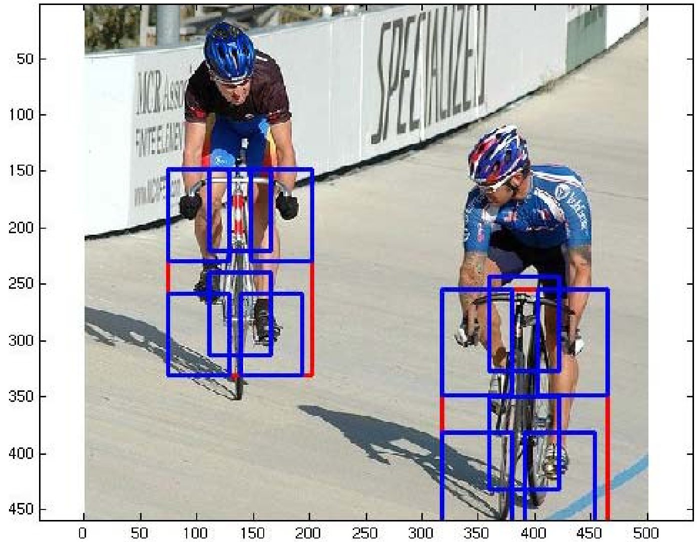


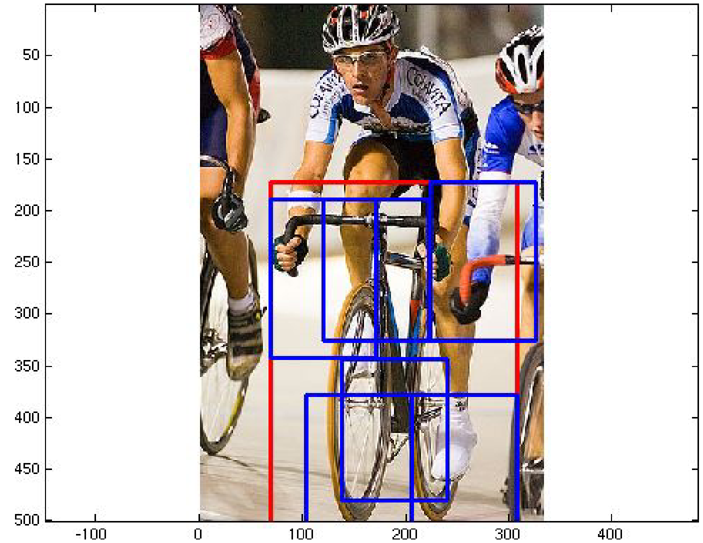




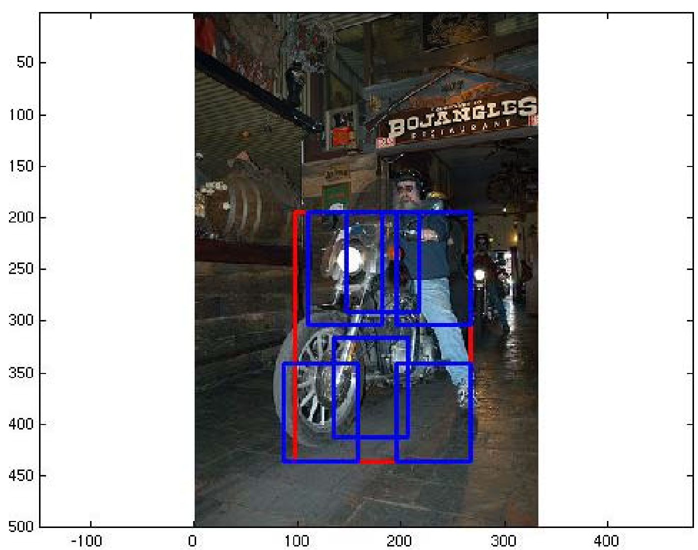


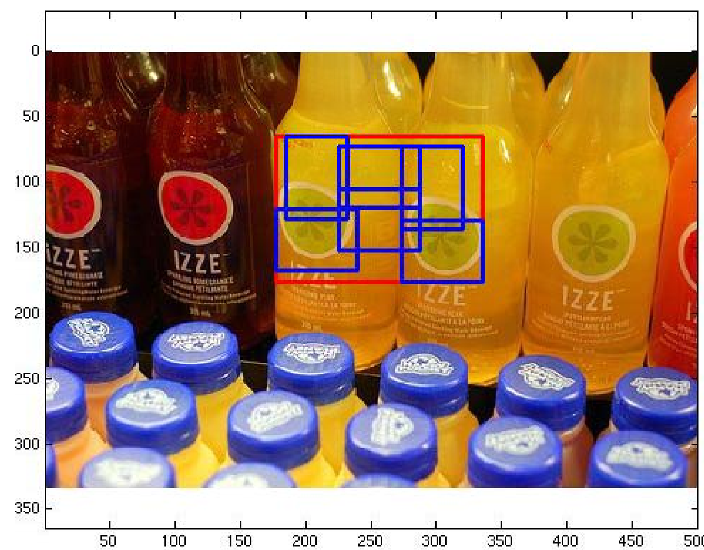
More bicycles



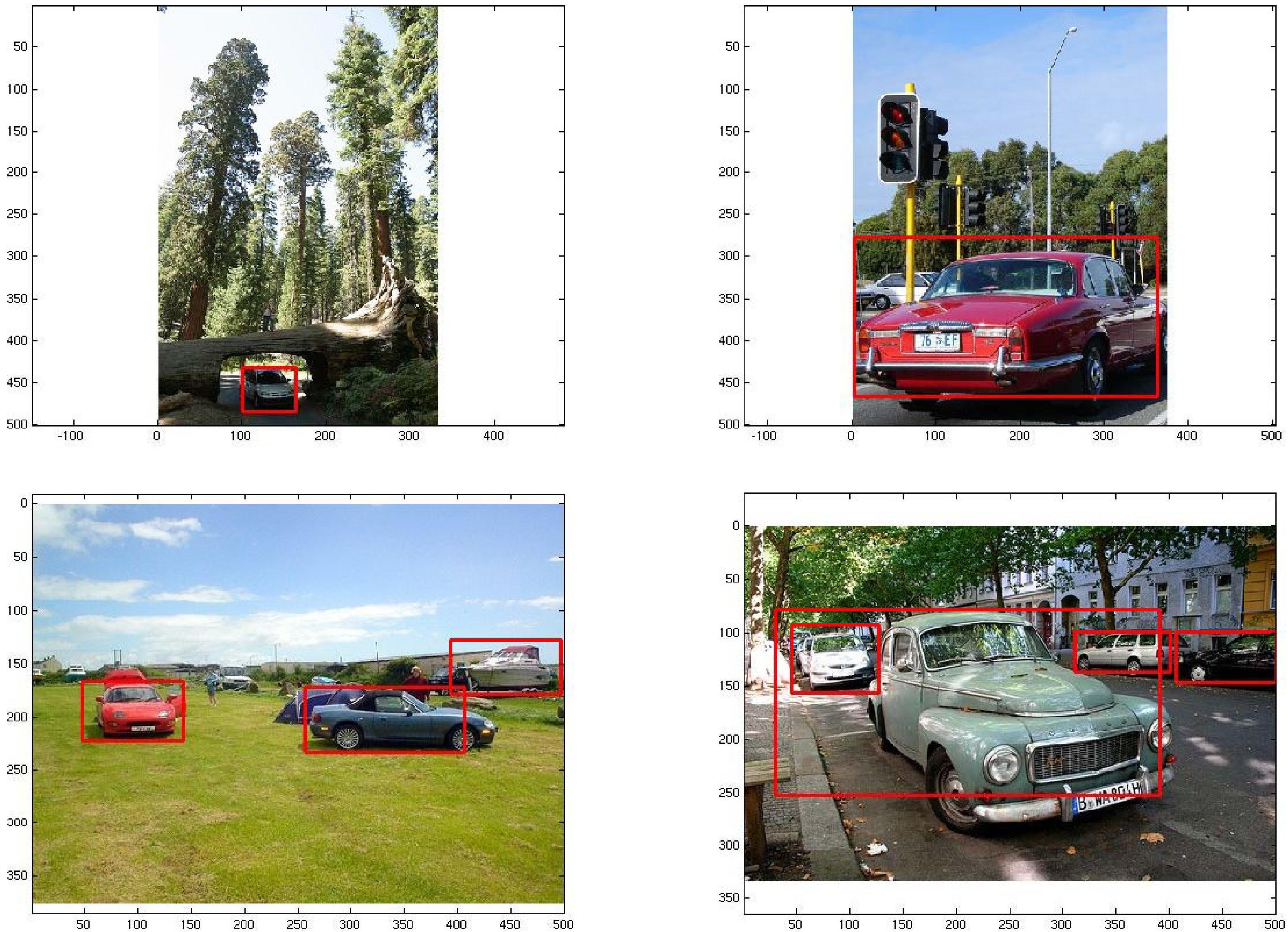


False positives

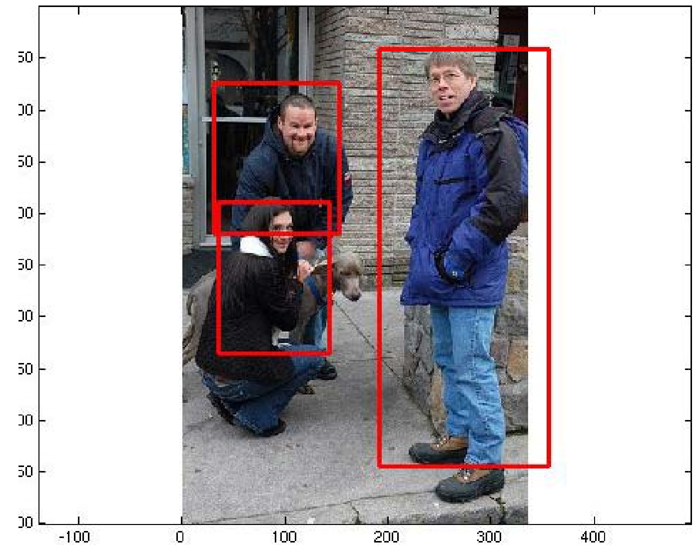




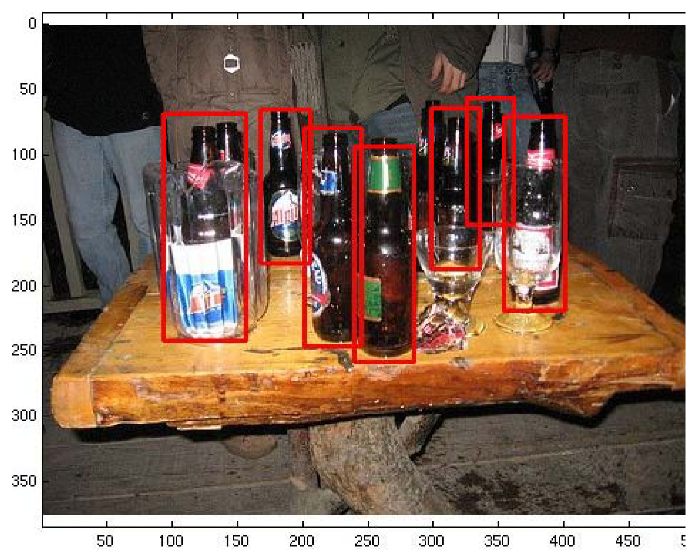
Car


50 100 150 200 250 300 350 400 450 500
-100 0 100 200 300 400
50 100 150 200 250 300 350 400 450 500
-100 0 100 200 300 400 500
0 50 100 150 200 250 300 350
50 100 150 200 250 300 350 400 450 500
0 50 100 150 200 250 300 350
50 100 150 200 250 300 350 400 450 500

Person



Bottle



Horse

Code

Source code for the system and models trained on PASCAL 2006, 2007 and 2008 data are available here:

http://www.cs.uchicago.edu/~pff/latent

Today – **Discriminative approaches**

- C. Dance, J. Willamowski, L. Fan, C. Bray, and G. Csurka, "Visual categorization with bags of keypoints," in ECCV International Workshop on Statistical Learning in Computer Vision, 2004.

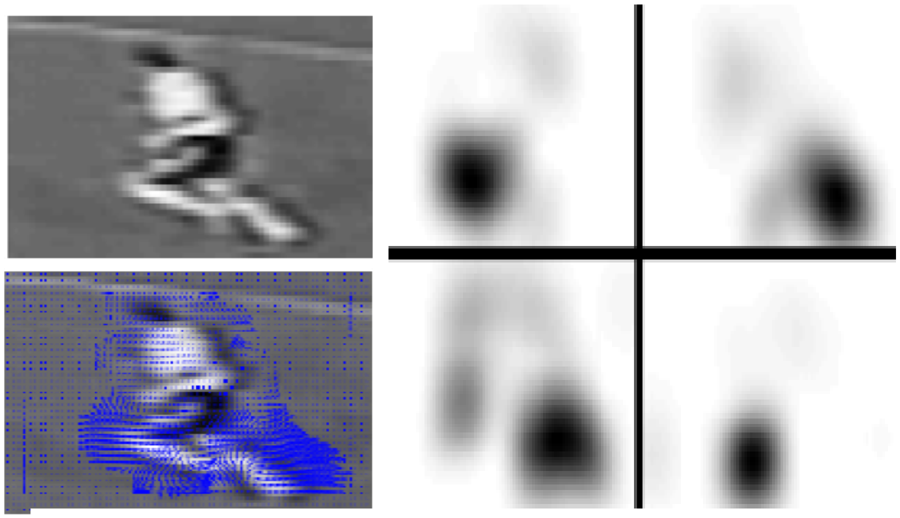
- M. Fritz; B. Leibe; B. Caputo; B. Schiele: Integrating Representative and Discriminant Models for Object Category Detection, ICCV'05, Beijing, China, 2005

- H. Zhang, A. C. Berg, M. Maire, and J. Malik, "Svm-knn: Discriminative nearest neighbor classification for visual category recognition," in CVPR '06: Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Washington, DC, USA: IEEE Computer Society, 2006, pp. 2126-2136.

- P. Felzenszwalb, D. Mcallester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in IEEE International Conference on Computer Vision and Pattern Recognition (CVPR) Anchorage, Alaska, June 2008., June 2008.

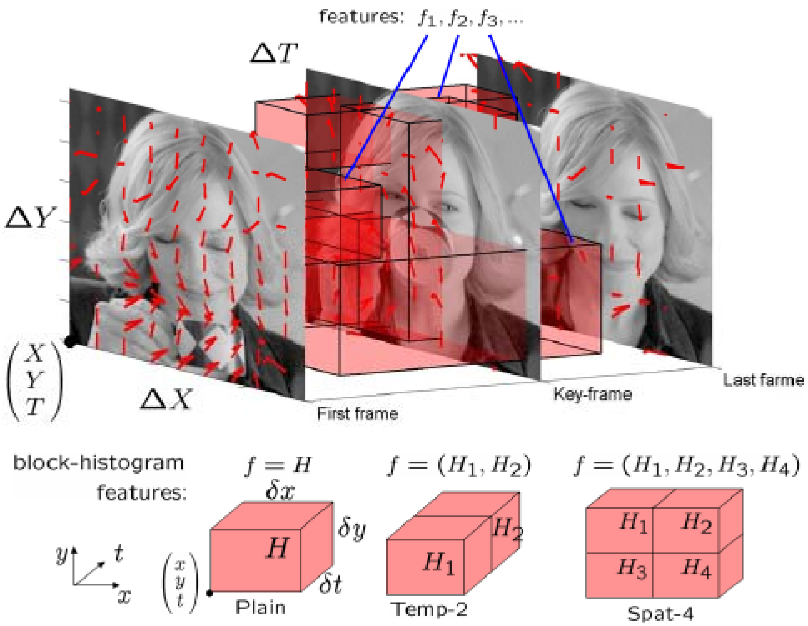
- **Y. Wang and G. Mori, "Learning a Discriminative Hidden Part Model for Human Action Recognition", Advances in Neural Information Processing Systems (NIPS), 2008**

Previous Work



Large-scale feature

[e.g. Efros et al., ICCV03]



Local patches

[e.g. Laptev & Perez, ICCV07]

Wang and Mori NIPS 2008

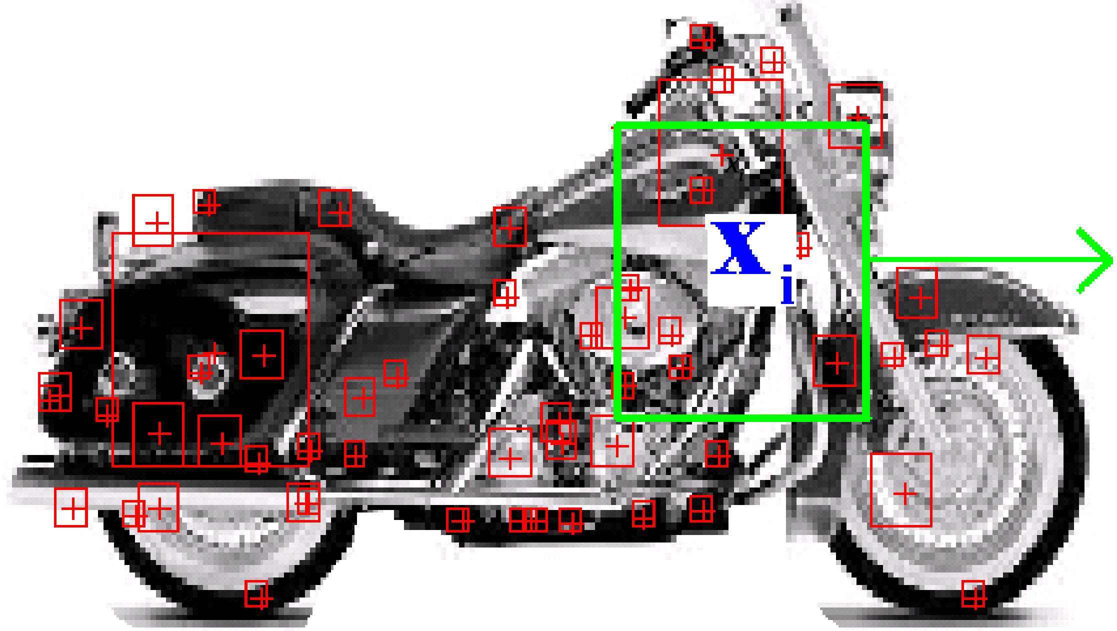
Large vs. Small Scale Features

Wang and Mori NIPS 2008: Explore Hidden-state conditional random field model integrating local features and global template

Wang and Mori NIPS 2008

CRF Part Based Models

Given *n* pairs ($\mathbf{x}_i$, y_i), learn a model that maps images to object categorie (where $\mathbf{x}_i$ is an image, y_i is an object category).



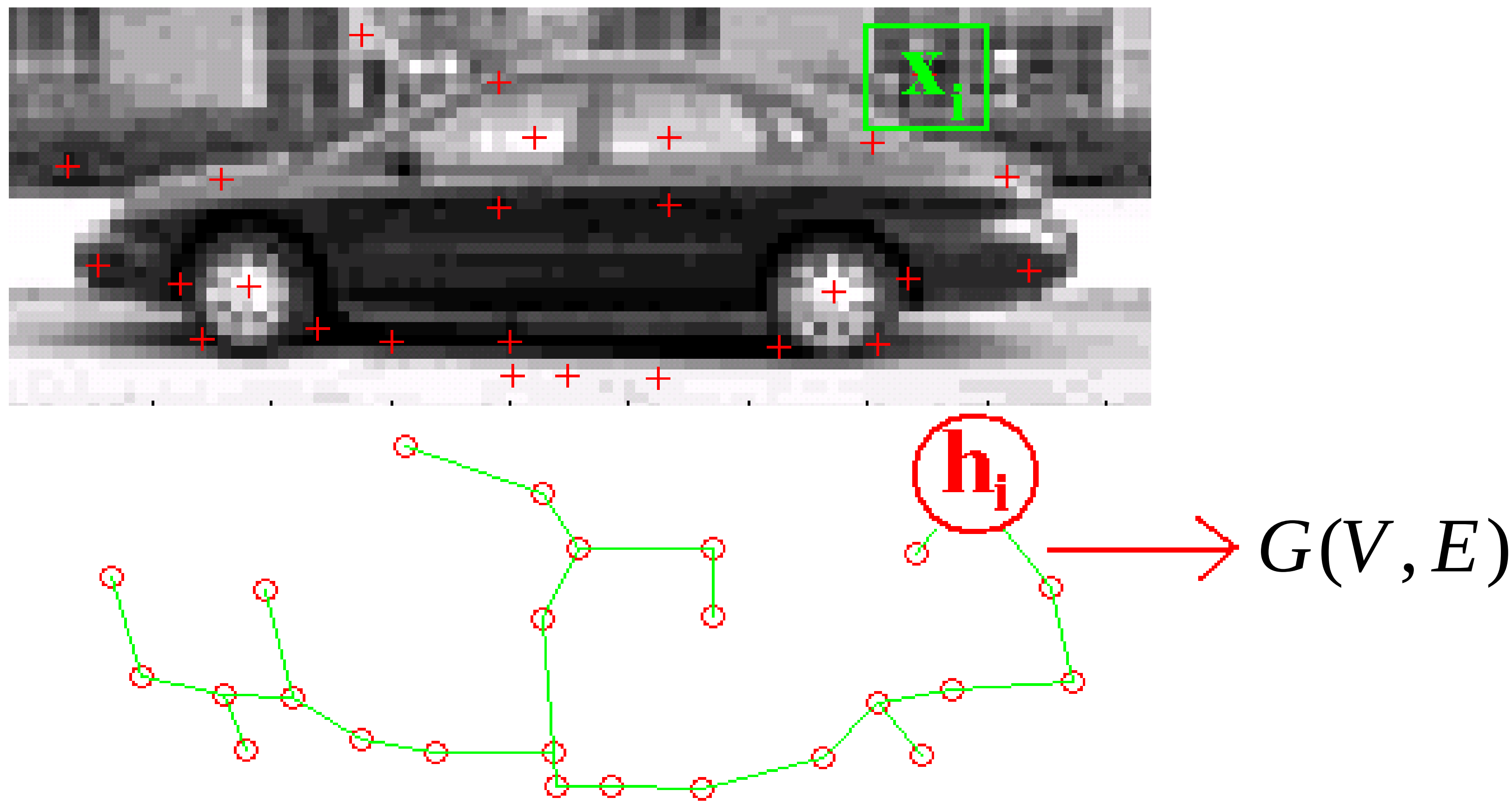
$$\phi(x_i) \in R^d$$

$$\mathbf{x}_i \rightarrow \{\phi(x_1), \ldots \phi(x_m)\}$$

Quattoni et al. 2004, 2007 develop Hidden-state CRF model for category recognition: capture inter-part dependencies with a hidden (or 'latent') part label…

Quattoni et al. 2004:

Graph Structure



Quattoni et al. 2004:

CRFs with hidden variables for Object Recognition

We introduce a hidden variable: $\mathbf{h}_i = \{h_1 \ldots h_m\}$, $h_j \in \mathbf{H}$ and define the conditional model:

$$P(y, \mathbf{h} \mid \mathbf{x}; \theta) = \frac{e^{\psi(y, \mathbf{h}, \mathbf{x}; \theta)}}{\sum_{y', \mathbf{h}} e^{\psi(y', \mathbf{h}, \mathbf{x}; \theta)}}$$

$$P(y \mid \mathbf{x}; \theta) = \sum_{\mathbf{h}} P(y, \mathbf{h} \mid \mathbf{x}; \theta) = \frac{\sum_{\mathbf{h}} e^{\psi(y, \mathbf{h}, \mathbf{x}; \theta)}}{\sum_{y', \mathbf{h}} e^{\psi(y', \mathbf{h}, \mathbf{x}; \theta)}}$$

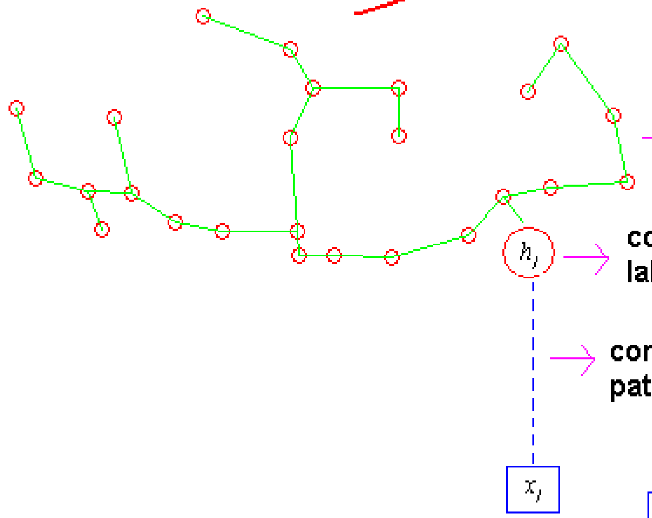
$\psi(y, \mathbf{h}, \mathbf{x}; \theta) \rightarrow$ Maps a configuration to $\Re$

Quattoni et al. 2004:

Potentials

$G(V,E)$ is a minimum spanning tree.

Weight (i, j)= distance between patches x_i and x_j



compatibility between a pair of part labels and a category

compatibility between a part label and a category

compatibility between a patch and a part label

$$f^2(j,k,y,h_j,h_k,\mathbf{x};\theta) = \theta(y,h_j,h_k)$$

$$f^1(j,y,h_j,\mathbf{x};\theta) = \phi(x_j)\bullet\theta(h_j) + \theta(y,h_j)$$

obtained with Lowe's detector

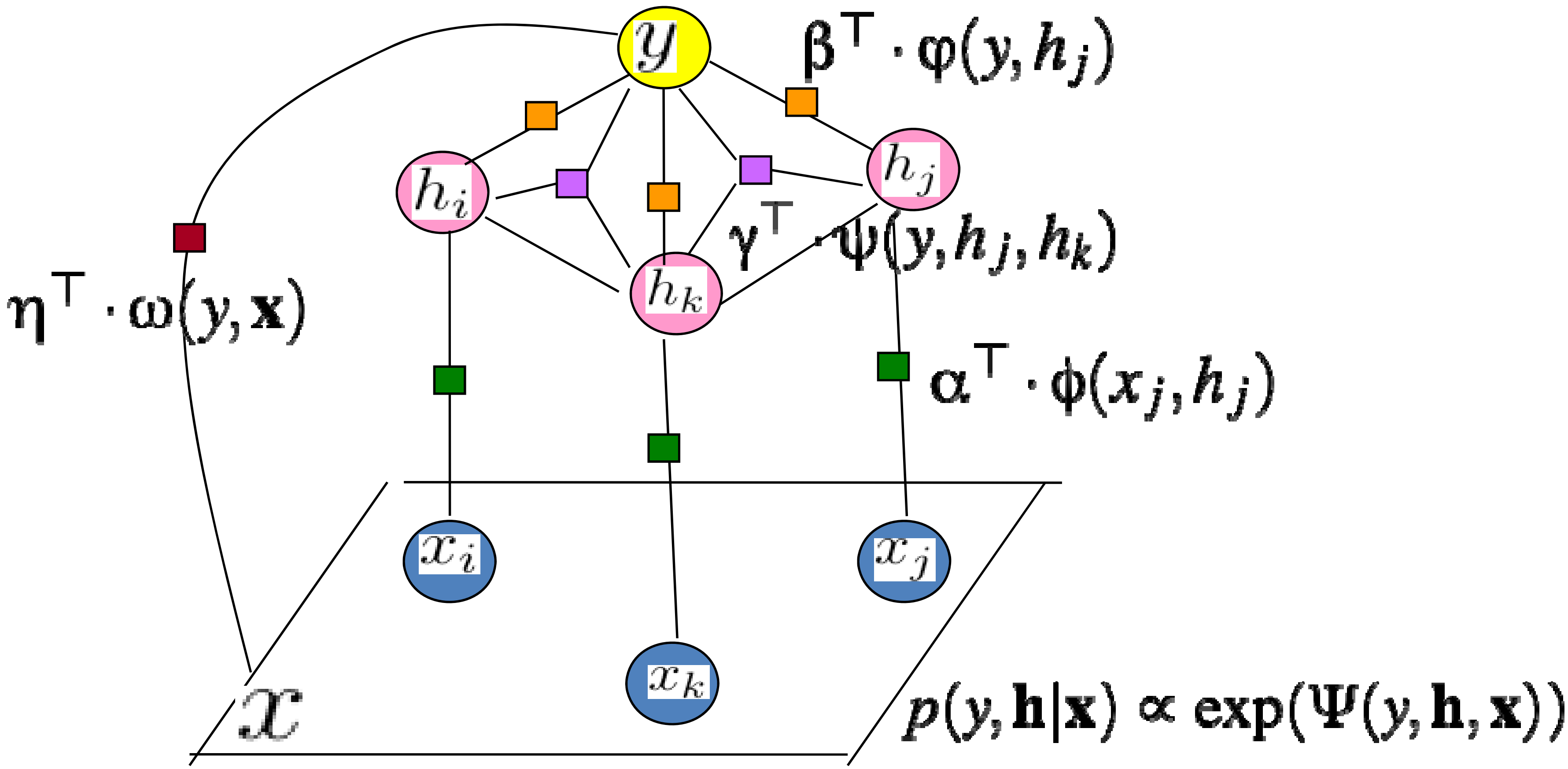
SIFT features + relative location and scale

$$\psi(y,\mathbf{h},\mathbf{x};\theta) = \sum_{j\in V} f^1(j,y,h_j,\mathbf{x};\theta) + \sum_{(j,k)\in E} f^2(j,k,y,h_j,h_k,\mathbf{x};\theta)$$

Wang and Mori, NIPS ‘08

- extend Quattoni et al. to include a global descriptor
- develop an efficient initialization scheme similar to Felzenswab et al.
- apply to activities using local and global spatio-temporal features…

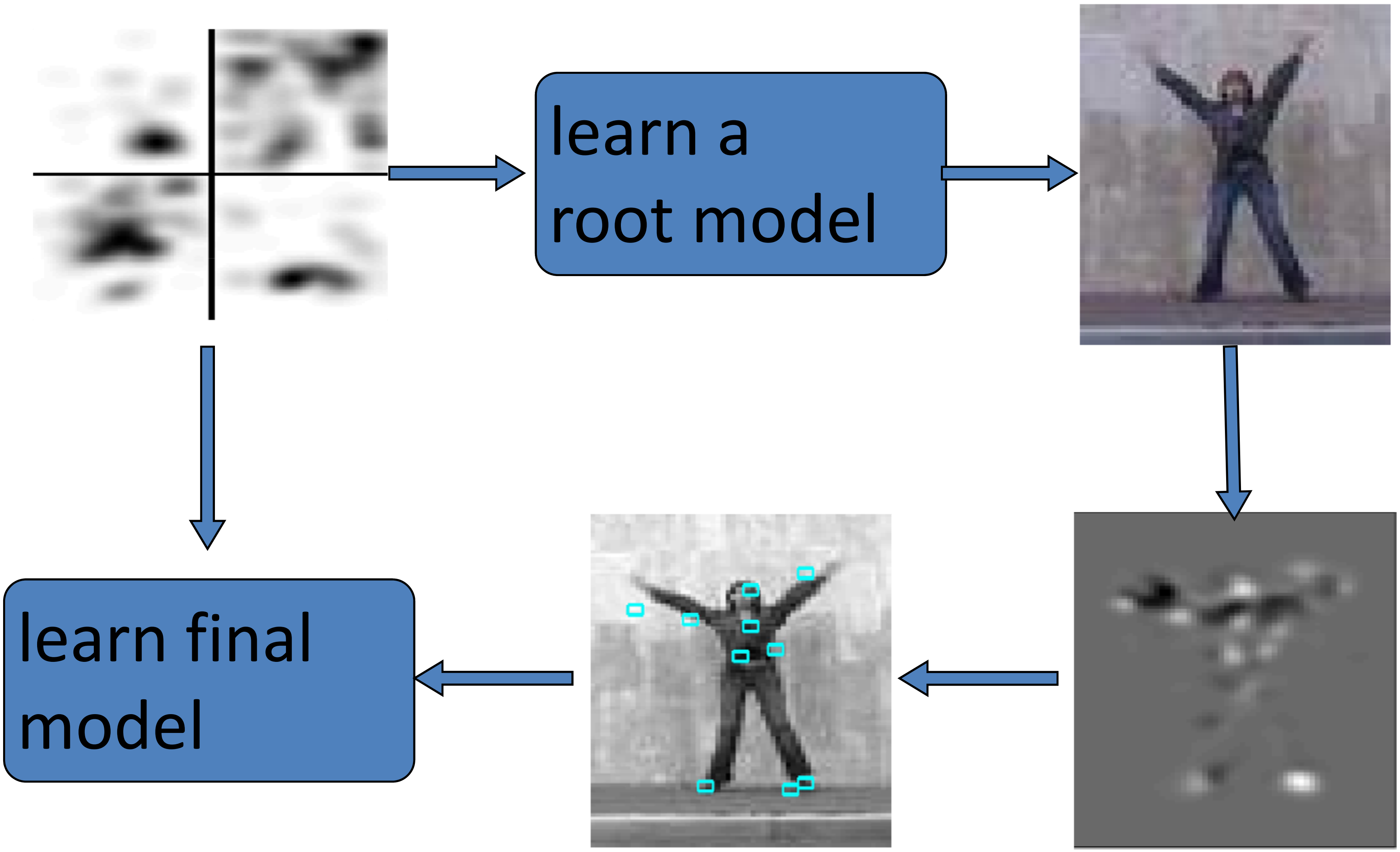
Hidden Conditional Random Field with a global feature:



$$\ell = \sum_t \log p(y^t|\mathbf{x}^t) = \sum_t \log\Big(\boxed{\sum_{\mathbf{h}} p(y^t, \mathbf{h}|\mathbf{x}^t)}\Big)$$

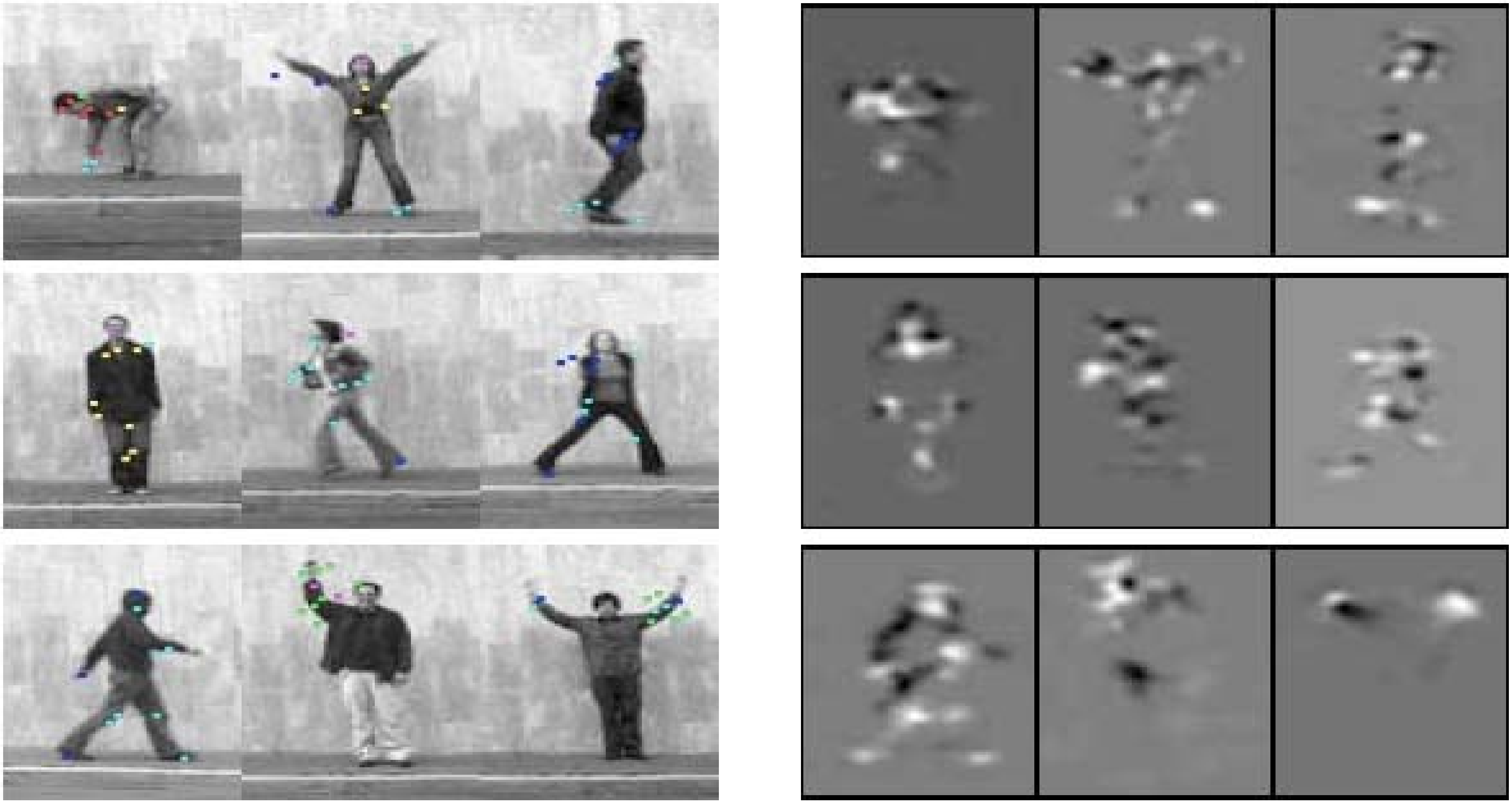
Wang and Mori, NIPS '08

Learning a HCRF Model



Wang and Mori, NIPS '08

Visualization of Learned Model



Wang and Mori, NIPS '08

Results: Weizmann dataset

	bend	jack	jump	pjump	run	side	walk	wave1	wave2
bend	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
jack	0.02	0.93	0.01	0.02	0.00	0.00	0.00	0.00	0.01
jump	0.01	0.03	0.74	0.00	0.06	0.02	0.12	0.02	0.00
pjump	0.01	0.00	0.00	0.99	0.00	0.00	0.00	0.00	0.00
run	0.00	0.05	0.00	0.00	0.72	0.06	0.17	0.00	0.00
side	0.00	0.01	0.07	0.00	0.02	0.73	0.17	0.00	0.00
walk	0.00	0.00	0.01	0.00	0.05	0.06	0.88	0.00	0.00
wave1	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.99	0.00
wave2	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Wang and Mori, NIPS '08

Wang and Mori, CVPR'09:

Max-Margin Hidden Conditional Random Fields for Human Action Recognition

Yang Wang and Greg Mori
TR 2008-21
School of Computing Science
Simon Fraser University
{ywang12,mori}@cs.sfu.ca

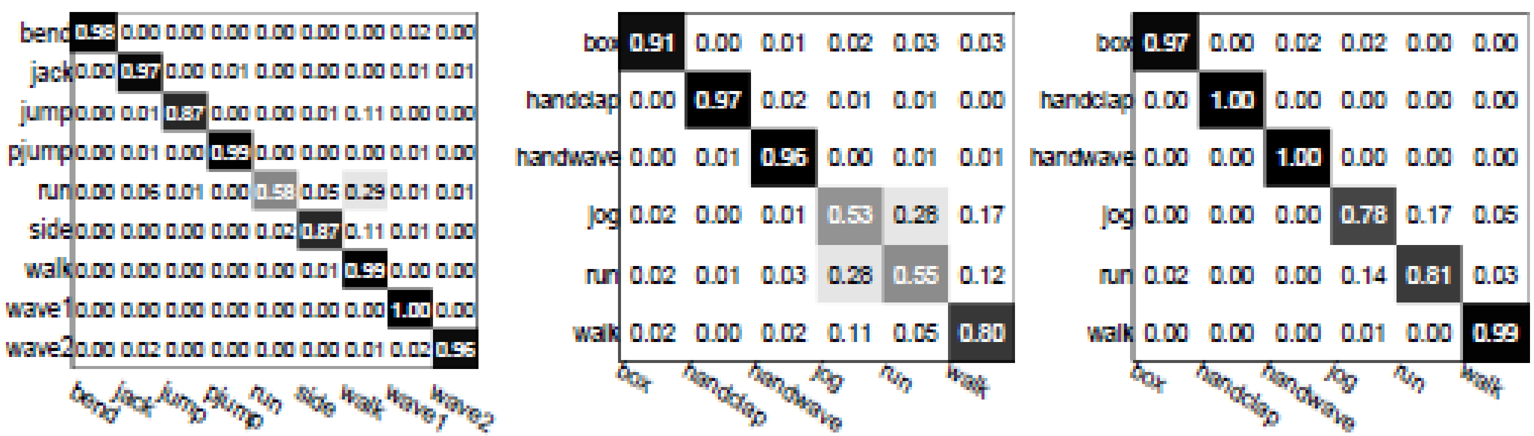
Abstract

We present a new method for classification with structured latent variables. Our model is formulated using the max-margin formalism in the discriminative learning literature. We propose an efficient learning algorithm based on the cutting plane method and decomposed dual optimization. We apply our model to the problem of recognizing human actions from video sequences, where we model a human action as a global root template and a constellation of several "parts". We show that our model outperforms another similar method that uses hidden conditional random fields, and is comparable to other state-of-the-art approaches. More importantly, our proposed work is quite general and can potentially be applied in a wide variety of vision problems that involve various complex, interdependent latent structures.

Wang and Mori, CVPR’09:

- Max-margin version extension of NIPS’08
- Similar to LSVMs:
  - semi-convex
  - hinge-loss
- But:
  - inherently multi-class
  - inter-part constraints
  - does not explicitly solve for latent part position

Wang and Mori, CVPR'09:



(a) Weizmann: per-frame (b) KTH: per-frame (c) KTH: per-video

Figure 2. Confusion matrices of classification results on Weizmann and KTH dataset. The confusion matrix of per-video classification on the Weizmann dataset is not shown, since it is simply a perfect diagonal matrix.

Wang and Mori, CVPR'09:

method	per-frame	per-video	per-cube
Our method	**0.9311**	**1.0000**	N/A
Wang & Mori [31]	0.9029	0.9722	N/A
Jhuang et al. [12]	N/A	0.9880	N/A
Niebles & Fei-Fei [20]	0.5500	0.7280	N/A
Blank et al. [3]	N/A	N/A	0.9964

Table 2. Comparison of classification accuracy with previous work on the Weizmann dataset.

method	$\|\mathcal{H}\|$=6	$\|\mathcal{H}\|$=10	$\|\mathcal{H}\|$=20
HCRF	0.6633 0.7855	0.6698 0.8760	0.6444 0.7512
Our approach	**0.7064** **0.8475**	**0.7853** **0.9251**	**0.7486** **0.8966**

Table 3. Comparison of our approach with the HCRF model on the KTH dataset. The first number in each cell is the accuracy of per-frame classification. The second number is the accuracy of per-video classification.

method	accuracy
Our method	**0.9251**
Liu & Shah [18]	**0.9416**
Jhuang et al. [12]	0.9170
Wang & Mori [31]	0.8760
Niebles et al. [21]	0.8150
Dollár et al. [8]	0.8117
Schuldt et al. [25]	0.7172

Table 4. Comparison of per-video classification accuracy with previous approaches on the KTH dataset.

Mar 17th – **Correspondence and Pyramid-based techniques**

- C. Berg, T. L. Berg, and J. Malik, "Shape matching and object recognition using low distortion correspondences," in CVPR '05: Proceedings of the 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)
- K. Grauman and T. Darrell, "The pyramid match kernel: discriminative classification with sets of image features," ICCV, vol. 2, 2005, pp. 1458-1465 Vol. 2 ***[F. Grabler]***
- S. Lazebnik, C. Schmid, and J. Ponce, "Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories," CVPR, vol. 2, 2006, pp. 2169-2178 ***[L. Bourdev]***
- S. Maji, A. C. Berg, and J. Malik, "Classification using intersection kernel support vector machines is efficient," in Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008, pp. 1-8. ***[S. Maji]***
- K. Grauman and T. Darrell, "Approximate correspondences in high dimensions," in In NIPS, vol. 2006.
- A. Bosch, A. Zisserman, and X. Munoz, "Representing shape with a spatial pyramid kernel," in CIVR '07: Proceedings of the 6th ACM international conference on Image and video retrieval ***[D. Bellugi]***