CS294-43: Visual Object and Activity Recognition

Prof. Trevor Darrell

Spring 2009

March 17th, 2009

Last Lecture – **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:

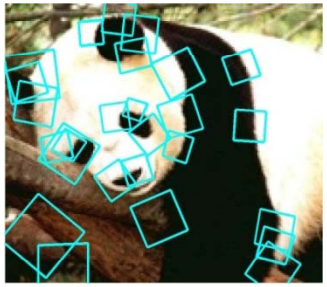
  Y. Wang and G. Mori, “Learning a Discriminative Hidden Part Model for Human Action Recognition”, Advances in Neural Information Processing Systems (NIPS), 2008

Today – **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. Kin]***
- 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
- 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]***

Comparing sets of local features

$\mathbf{X} = \{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m\}$



$\mathbf{Y} = \{\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_n\}$

Previous strategies:

- Match features individually, vote on small sets to verify

  [Schmid, Lowe, Tuytelaars et al.]

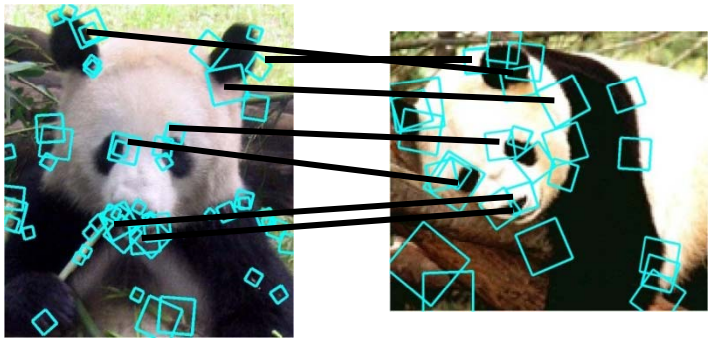
- Explicit search for one-to-one correspondences

  [Rubner et al., Belongie et al., Gold & Rangarajan, Wallraven & Caputo, Berg et al., Zhang et al.,…]

- Compare frequencies of prototype features

  [Csurka et al., Sivic & Zisserman, Lazebnik & Ponce]

Partially matching sets of features



$\mathbf{X} = \{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m\}$ $\mathbf{Y} = \{\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_n\}$

Optimal match: $O(m^3)$
Greedy match: $O(m^2 \log m)$
Pyramid match: O(m)

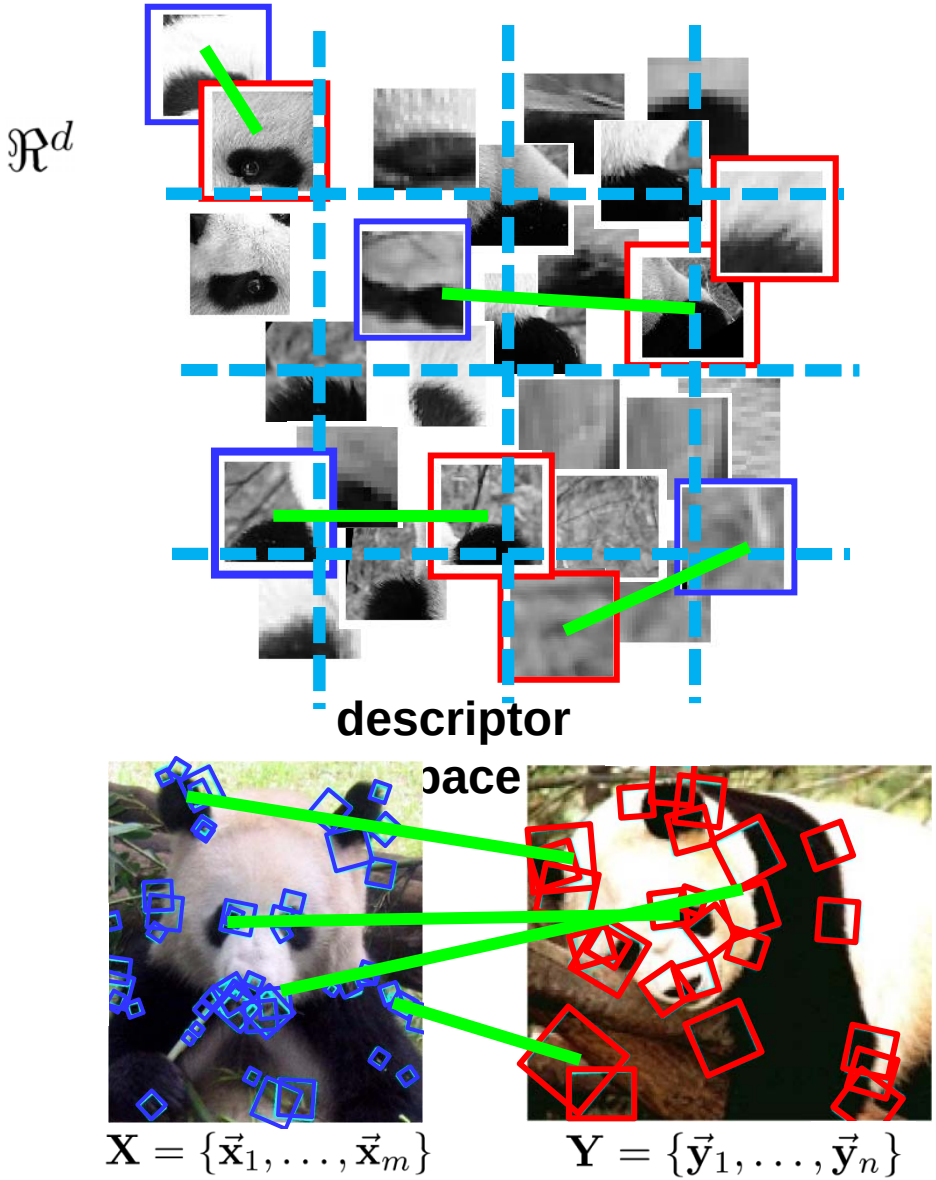
(**m**=num pts)

$$\min_{\pi:\mathbf{X}\to\mathbf{Y}} \sum_{\mathbf{x}_i \in \mathbf{X}} ||\mathbf{x}_i - \pi(\mathbf{x}_i)||$$

mate matching kernel that makes it practical to compare large sets of features based on their partial correspondences.

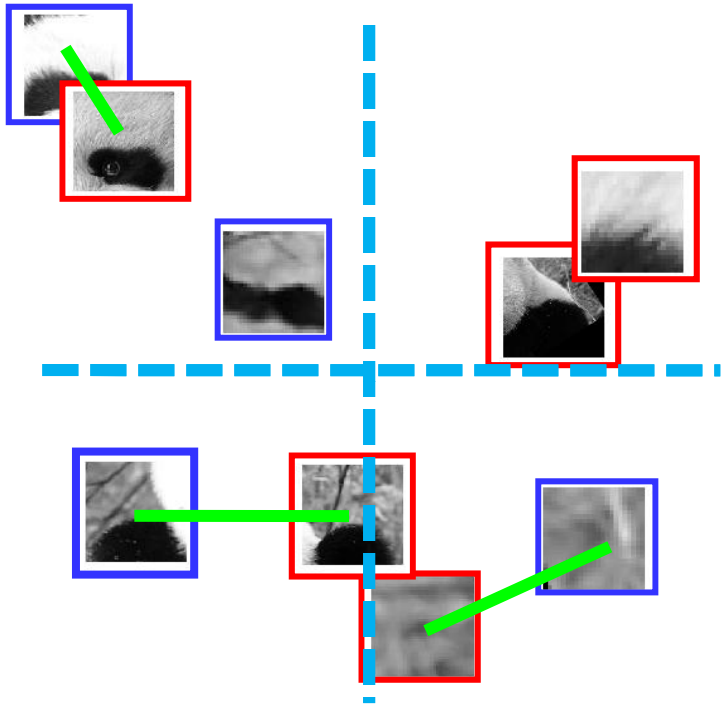
[Previous work: Indyk & Thaper, Bartal, Charikar, Agarwal & Varadarajan, …]

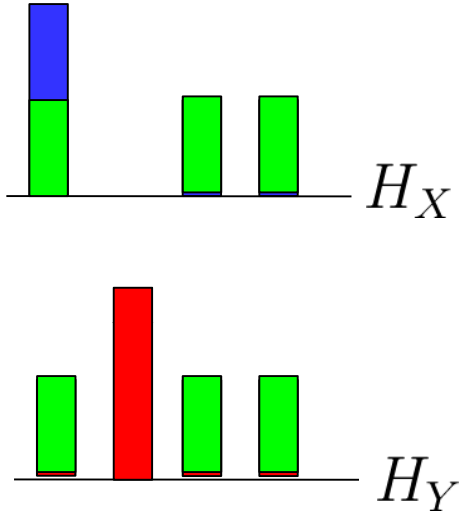
Pyramid match: main idea



Feature space partitions serve to “match” the local descriptors within successively wider regions.

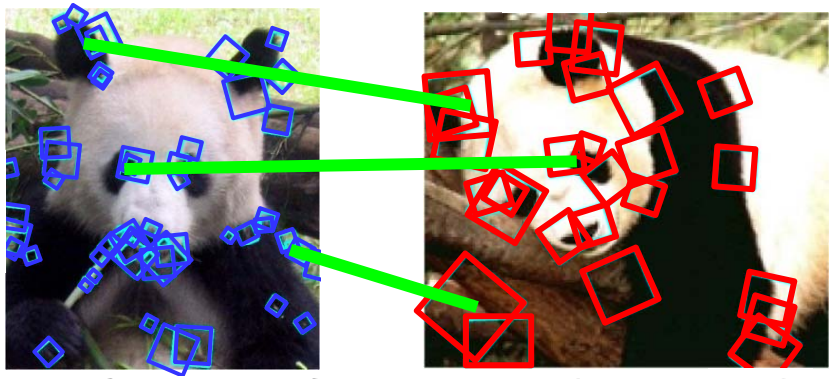
Pyramid match: main idea





$$\mathcal{I}(H_X, H_Y) = \sum_j \min\left(H_X(j), H_Y(j)\right)$$

$$= 3$$



Histogram intersection counts number of possible matches at a given partitioning.

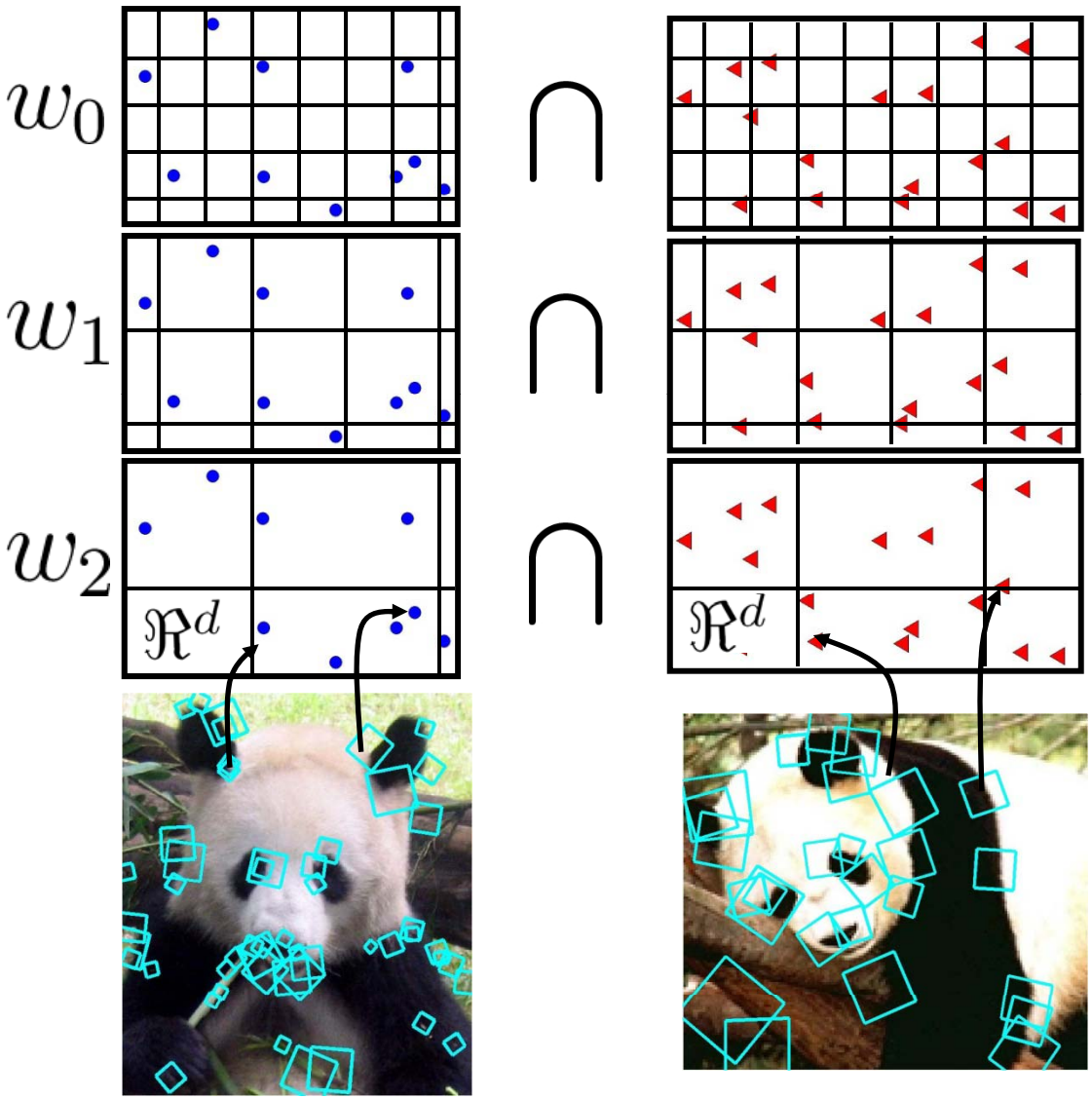
Pyramid match kernel

$$K_\Delta(X,Y) = \sum_{i=0}^{L} \underbrace{2^{-i}}_{\text{measures difficulty of a match at level } i} \underbrace{\mathcal{I}\left(H_X^{(i)}, H_Y^{(i)}\right) - \mathcal{I}\left(H_X^{(i-1)}, H_Y^{(i-1)}\right)}_{\text{number of newly matched pairs at level } i}$$

- For similarity, weights inversely proportional to bin size (or may be learned)
- Normalize these kernel values to avoid favoring large sets

[Grauman & Darrell, ICCV 2005]

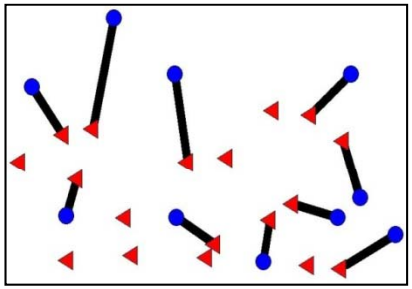
Pyramid match kernel



$\mathbf{X} = \{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m\}$ $\mathbf{Y} = \{\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_n\}$

Optimal match: $O(m^3)$

Pyramid match: O(mL)



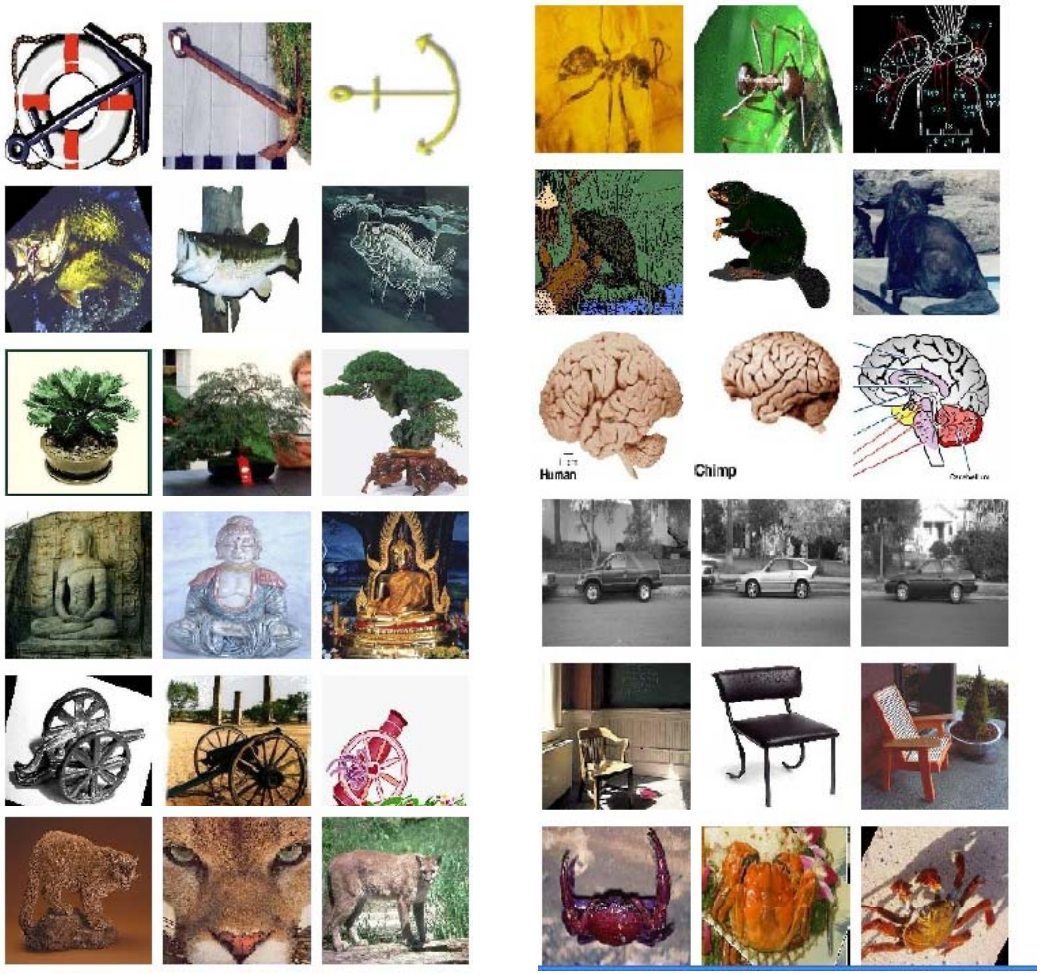
optimal partial matching

Highlights of the pyramid match

- Linear time complexity
- Formal bounds on expected error
- Mercer kernel
- Data-driven partitions allow accurate matches even in high-dim. feature spaces
- Strong performance on benchmark object recognition datasets

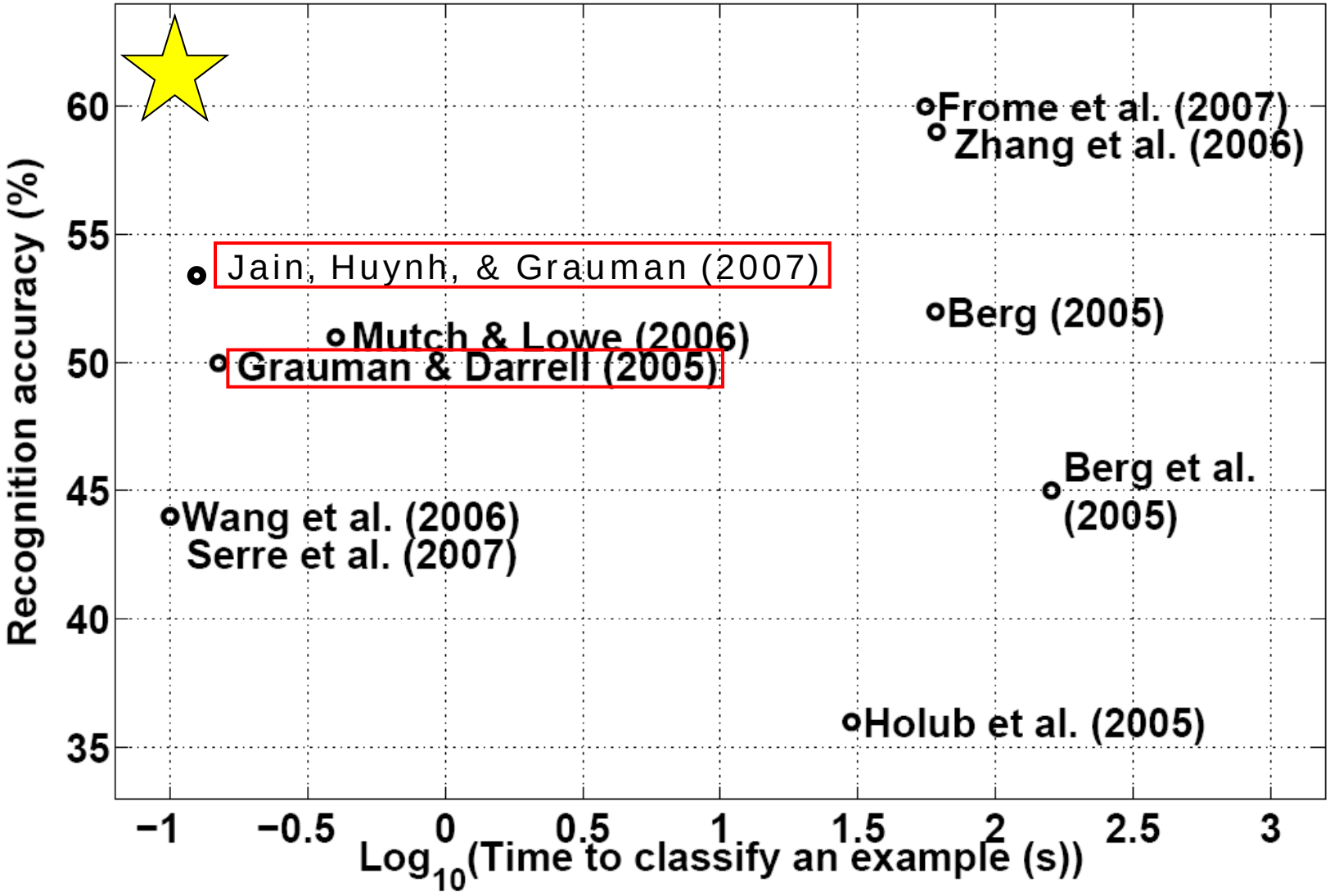
Recognition results:
Caltech-101 dataset

- 101 categories
  40-800 images per class

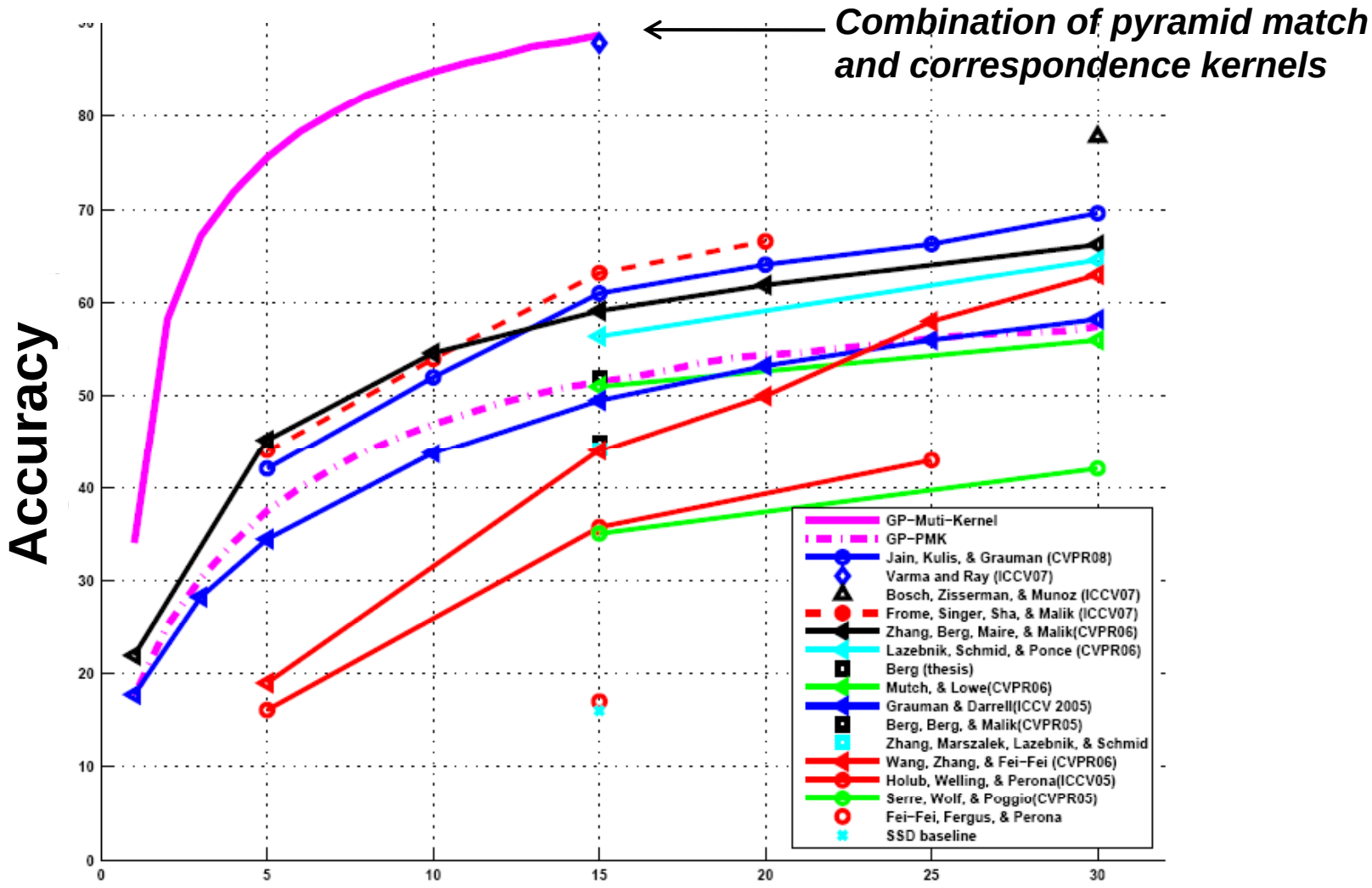


Data provided by Fei-Fei, Fergus, and Perona

Recognition results:
Caltech-101 dataset

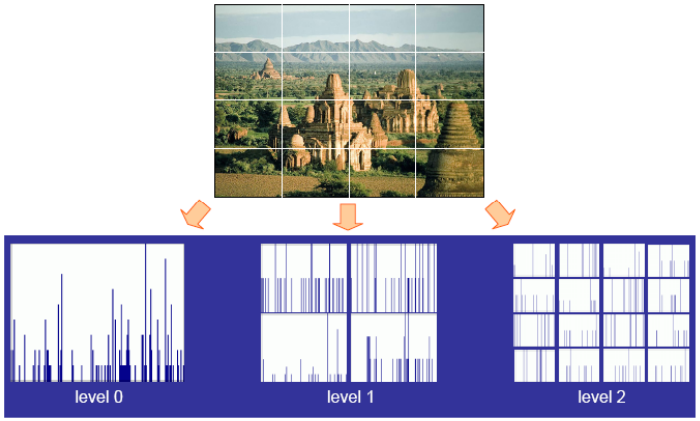


Recognition results:
Caltech-101 dataset



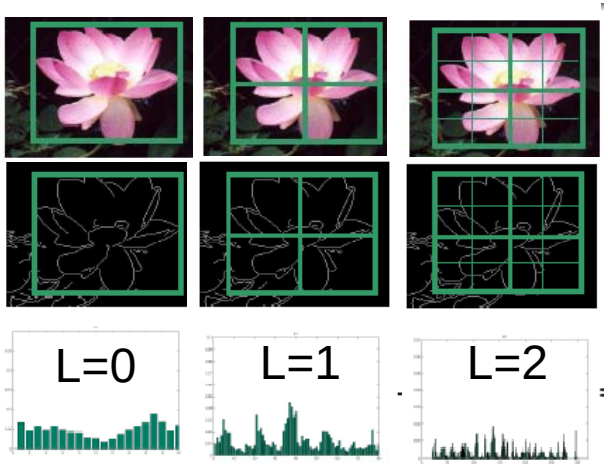
[Kapoor et al. IJCV 2009]

Pyramid match kernel: examples of extensions and applications by other groups



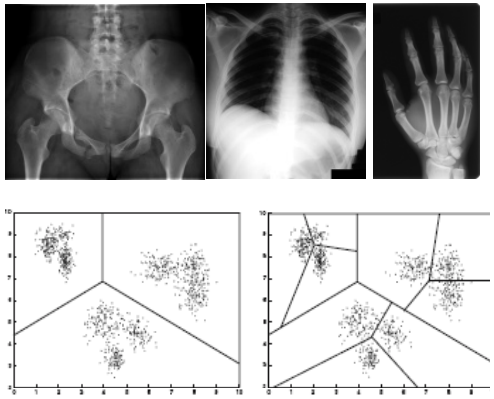
Spatial Pyramid Match Kernel
Lazebnik, Schmid, Ponce, 2006.

Scene recognition



Representing Shape with a Pyramid Kernel
Bosch & Zisserman, 2007.

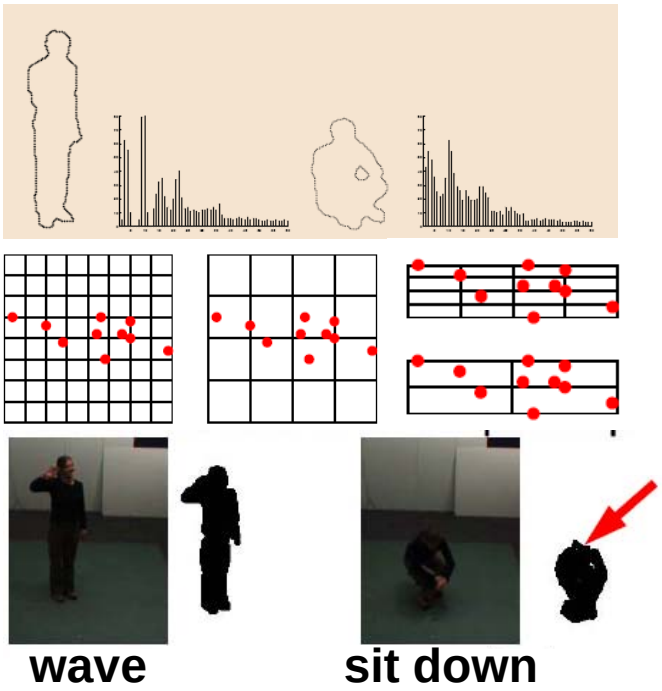
Shape representation



Dual-space Pyramid Matching
Hu et al., 2007.

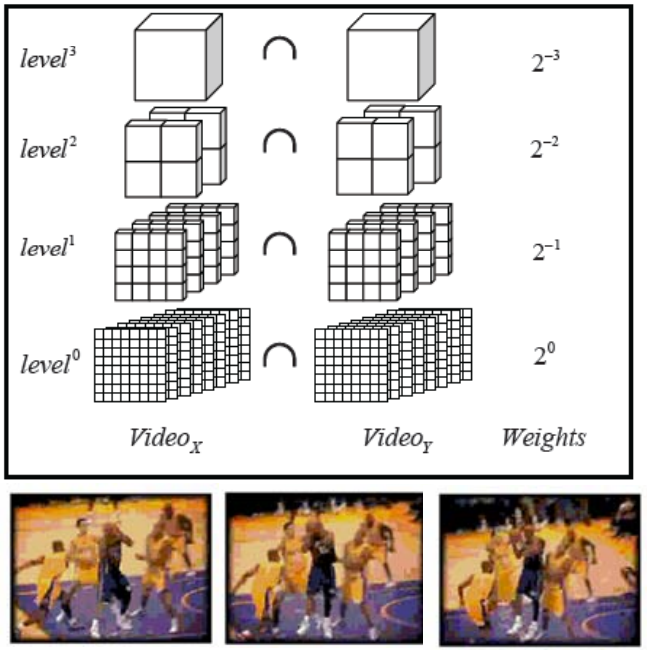
Medical image classification

Pyramid match kernel: examples of extensions and applications by other groups



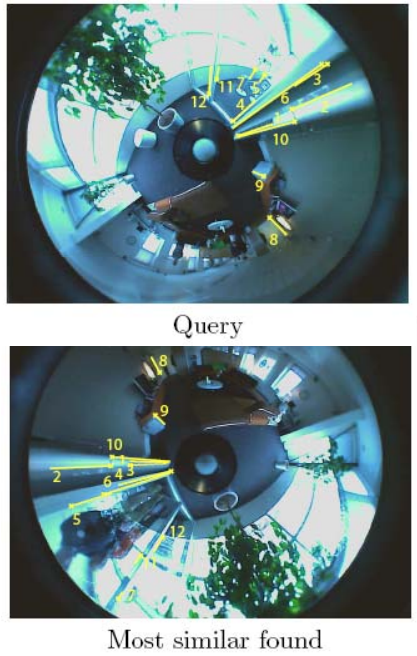
Single View Human Action Recognition using Key Pose Matching, Lv & Nevatia, 2007.

Action recognition



Spatio-temporal Pyramid Matching for Sports Videos, Choi et al., 2008.

Video indexing

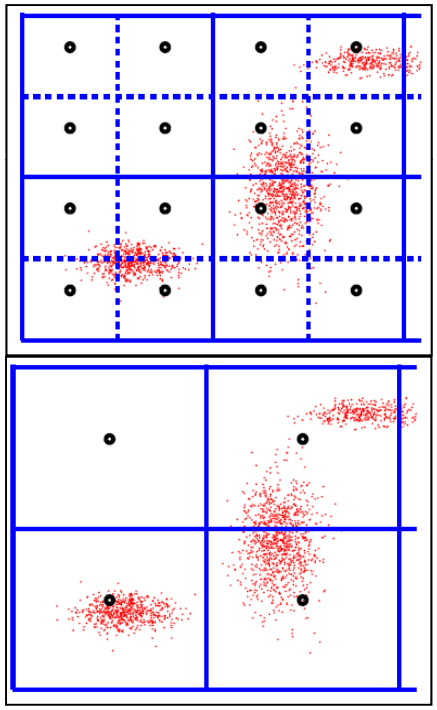


From Omnidirectional Images to Hierarchical Localization, Murillo et al. 2007.

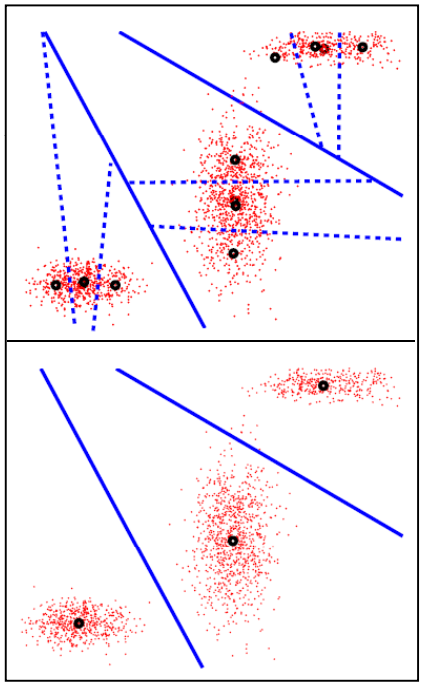
Robot localization

Vocabulary-guided pyramid match

Uniform bins



***Vocabulary-guided* bins**



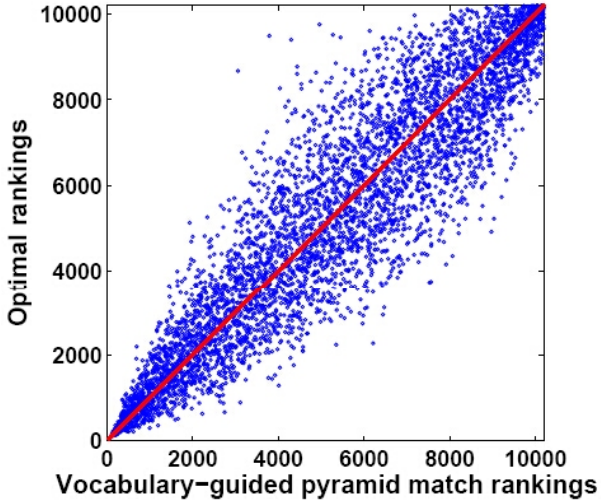
[Grauman & Darrell, NIPS 2006]

- Tune pyramid partitions to the feature distribution
- Accurate for $d > 100$
- Requires initial corpus of features to determine pyramid structure
- Small cost increase over uniform bins: *kL* distances against bin centers to insert points

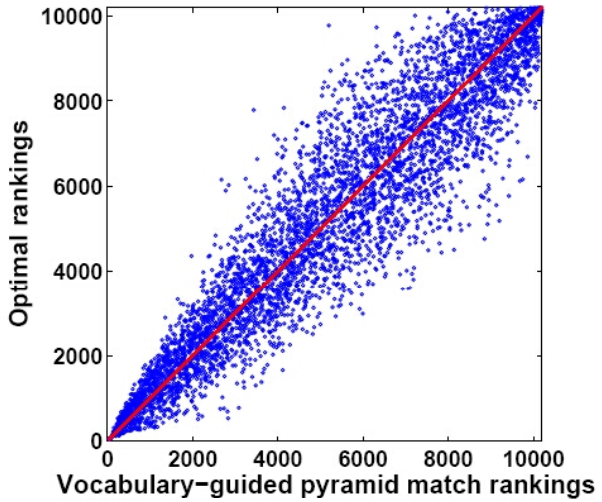
Approximation quality

ETH-80 images, sets of SIFT features

***d*=8**

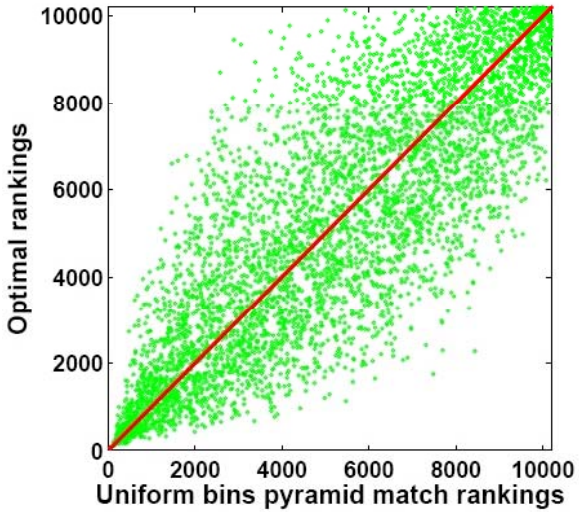


***d*=128**

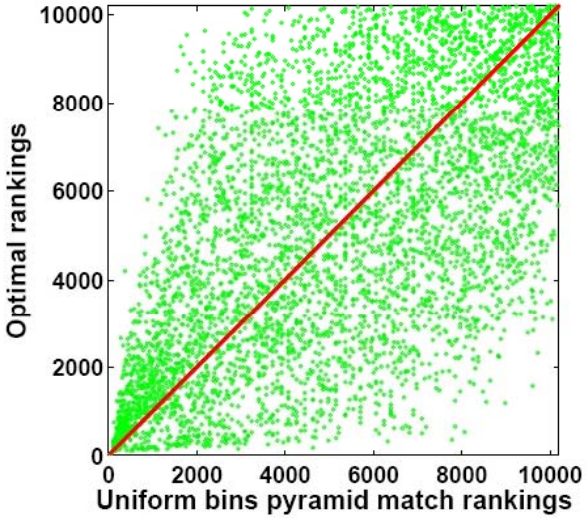


Vocabulary-guided pyramid match

***d*=8**



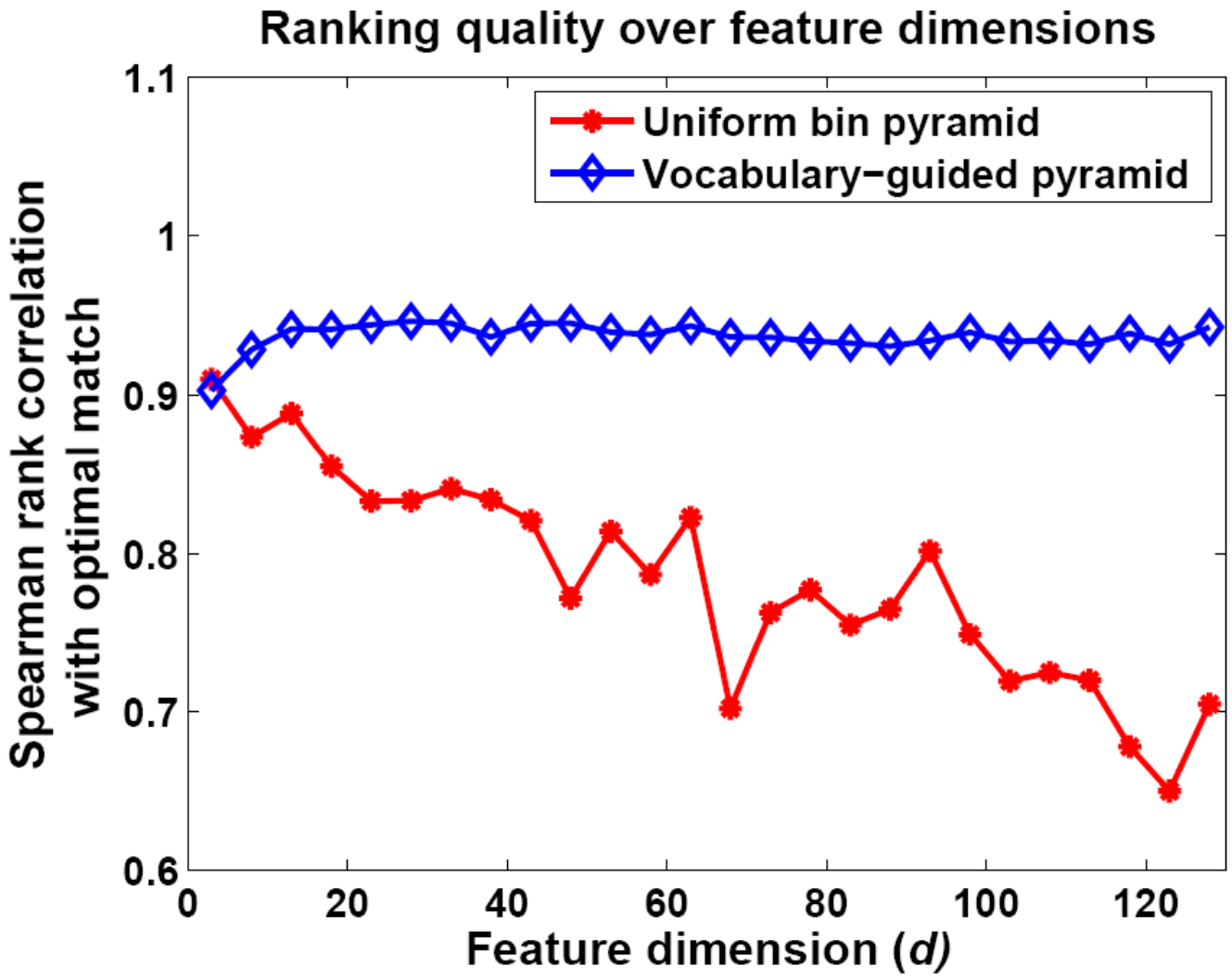
***d*=128**



Uniform bin pyramid match

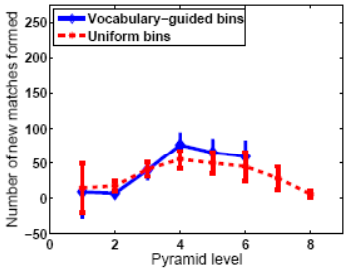
Approximation quality

ETH-80 images, sets of SIFT features

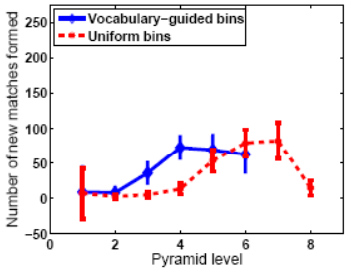


Bin structure and match counts

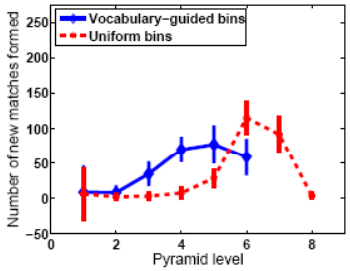
d=3



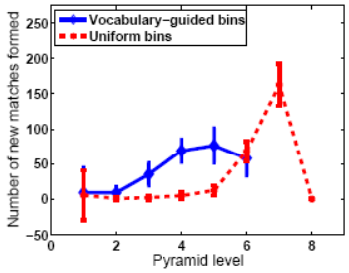
d=8



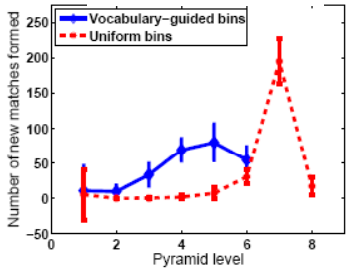
d=13



d=23



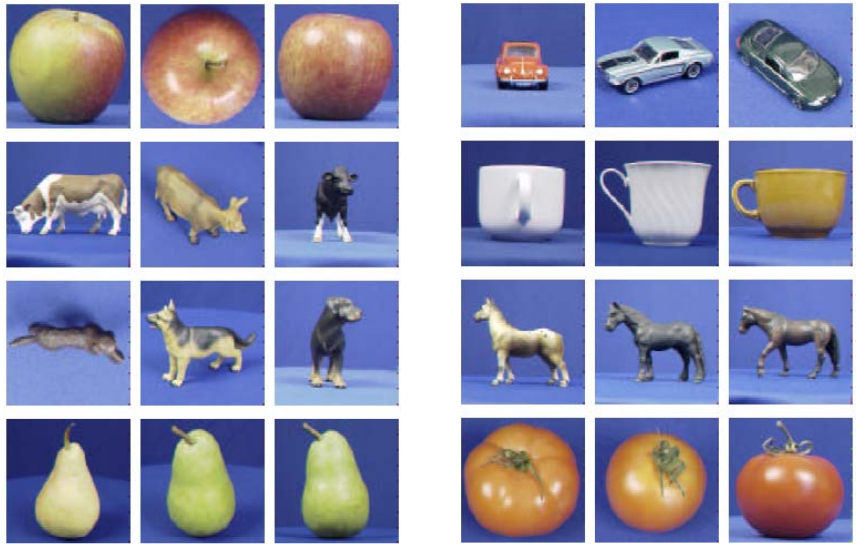
d=128



Data-dependent bins allow more gradual distance ranges in high dimensions

Recognition on the ETH-80

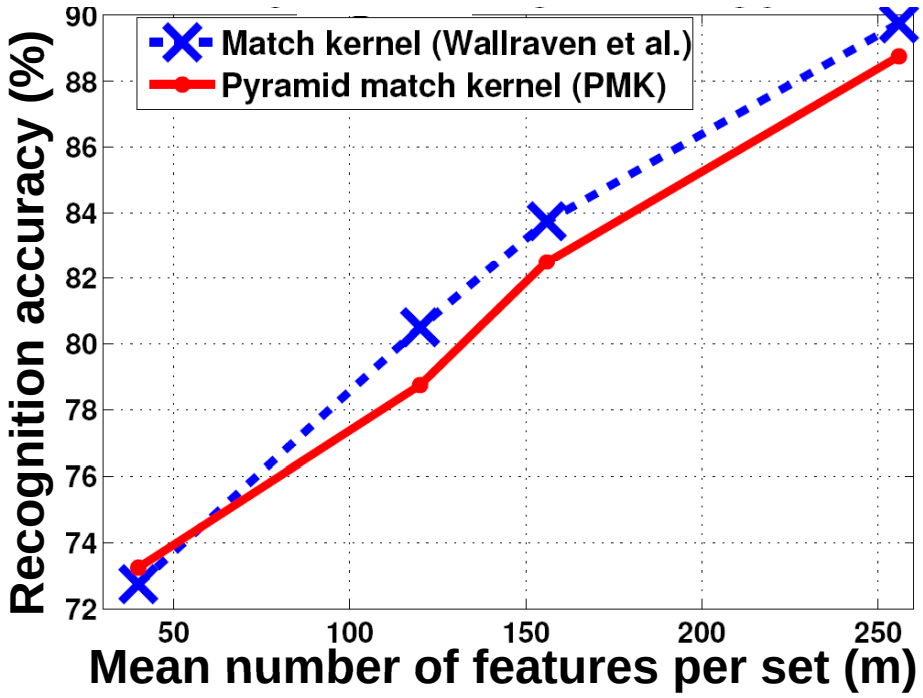
- **ETH-80 data set**
  8 categories
  50 images each
- **One-vs-all SVM with PMK**
- **Features:**
  - Harris detector
    (vary *m* with saliency threshold)
  - PCA-SIFT descriptor

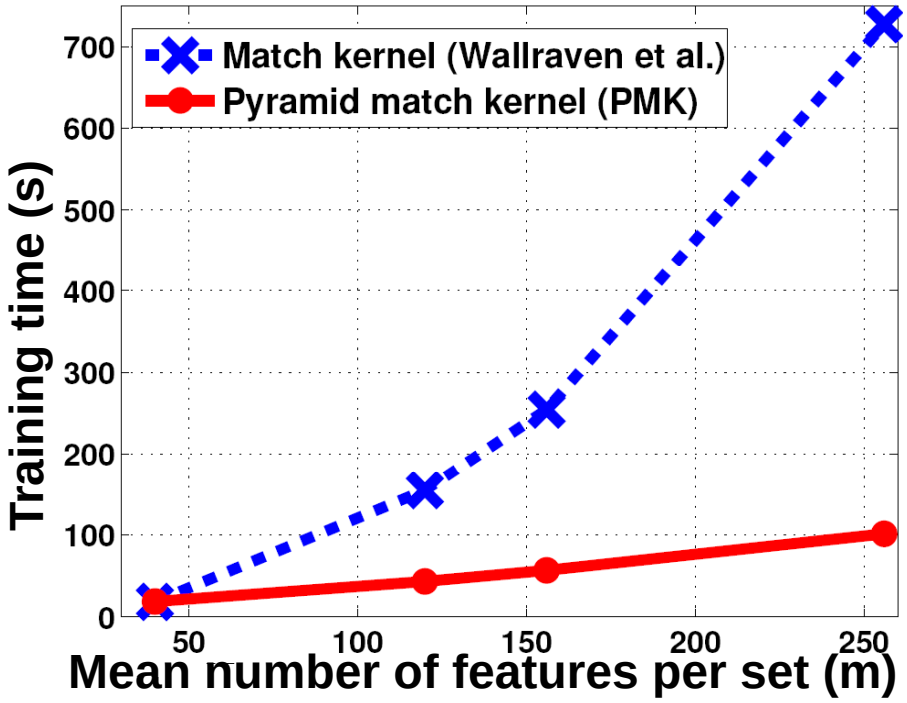


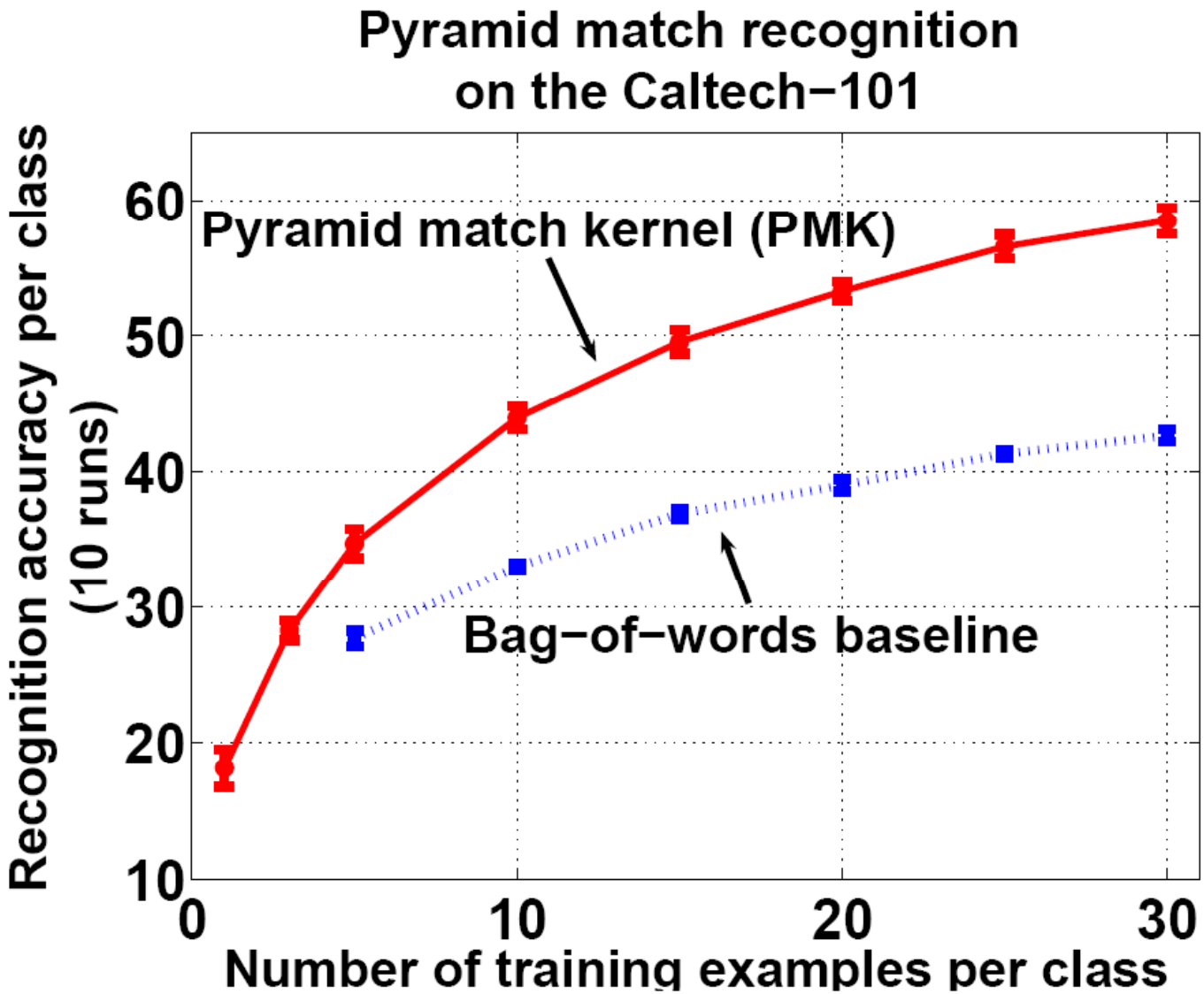
Data provided by Bastian Leibe and Bernt Schiele

Recognition on the ETH-80

Kernel	Complexity
Match [Wallraven et al.]	$O(dm^2)$
Pyramid match	$O(dm)$






Pyramid match recognition
on the Caltech−101
Recognition accuracy per class
(10 runs)
60
50
40
30
20
10
0
10
20
30
Number of training examples per class
Pyramid match kernel (PMK)
Bag−of−words baseline

Today – **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. Kin]***
- 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
- 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]***

Next Lecture – Category Discovery from the Web

- R. Fergus, L. Fei-Fei, P. Perona, and A. Zisserman, "Learning object categories from google's image search," ICCV vol. 2, 2005, pp. 1816-1823 Vol. 2. Available: http://dx.doi.org/10.1109/ICCV.2005.142
- L.-J. Li, G. Wang, and L. Fei-Fei, "Optimol: automatic online picture collection via incremental model learning," in Computer Vision and Pattern Recognition, 2007. CVPR '07. IEEE Conference on, 2007, pp. 1-8. Available: http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=4270073
- F. Schroff, A. Criminisi, and A. Zisserman, "Harvesting image databases from the web," in Computer Vision, 2007. ICCV 2007. IEEE 11th International Conference on, 2007, pp. 1-8. Available: http://dx.doi.org/10.1109/ICCV.2007.4409099
- K. Saenko and T. Darrell, "Unsupervised Learning of Visual Sense Models for Polysemous Words". Proc. NIPS, December 2008, Vancouver, Canada. http://people.csail.mit.edu/saenko/saenko_nips08.pdf
- T. Berg and D. Forsyth, "Animals on the Web". In Proceedings of the 2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, Washington, DC, 1463-1470. Available: http://dx.doi.org/10.1109/CVPR.2006.57