# Approximate Correspondences in High Dimensions

Kristen Grauman[1,2] and Trevor Darrell[1]

[1]CSAIL, Massachusetts Institute of Technology

[2]Department of Computer Sciences, University of Texas-Austin

## Problem

The correspondence between sets of local feature vectors is often a good measure of similarity, but it is computationally expensive.


Accuracy of existing matching approximations declines linearly with the feature dimension.

## Our approach

- Form multi-resolution decomposition of the feature space to efficiently count "implicit" matches without directly comparing features
- Exploit structure in feature space when placing partitions in order to fully leverage their grouping power

→

- Approximate partial matching
- Linear-time match
- Mercer kernel
- Accurate for feature dimensions > 100

## Optimal partial match

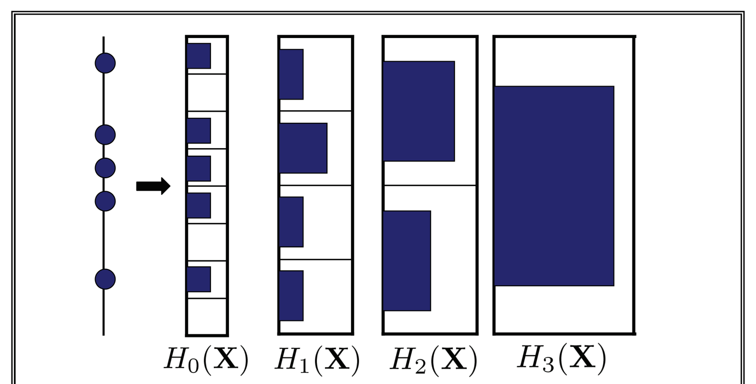
$\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_m\},\ \mathbf{Y} = \{\mathbf{y}_1, \ldots, \mathbf{y}_n\} \qquad \mathbf{x}_i, \mathbf{y}_i \in \Re^d, m \le n$

$$\pi^* = \underset{\pi}{\operatorname{argmin}} \sum_{\mathbf{x}_i \in \mathbf{X}} ||\mathbf{x}_i - \mathbf{y}_{\pi_i}||_2 \longrightarrow \boxed{O(n^3)\text{ time}}$$

## The Pyramid Match

[Grauman and Darrell, ICCV 2005]

In $\boxed{O(nL)\text{ time}}$, approximate the optimal partial matching cost: use multi-resolution histograms to count matches that are possible within a discrete set of distances.

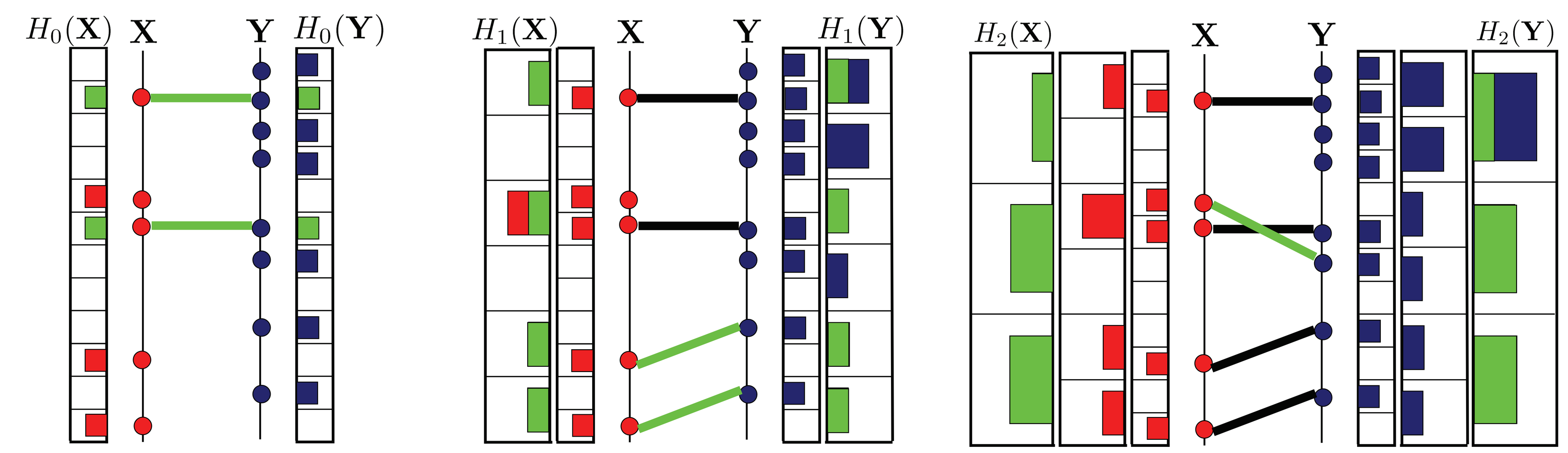



$\Psi(\mathbf{X}) = [H_0(\mathbf{X}), \ldots, H_L(\mathbf{X})]$

*set of features → histogram pyramid*

**Pyramid match cost:**

$$\mathcal{C}_{PM}\left(\Psi(\mathbf{X}), \Psi(\mathbf{Y})\right) = \sum_{i=1}^{L} \underbrace{2^i}_{\substack{\text{Weight according}\\ \text{to bin size } (w_i)}} \underbrace{\left(\mathcal{I}(H_i(\mathbf{X}), H_i(\mathbf{Y})) - \mathcal{I}(H_{i-1}(\mathbf{X}), H_{i-1}(\mathbf{Y}))\right)}_{\substack{\text{Number of new matches at level } i \text{ counted by difference}\\ \text{in histogram intersections across levels } (N_i)}}$$



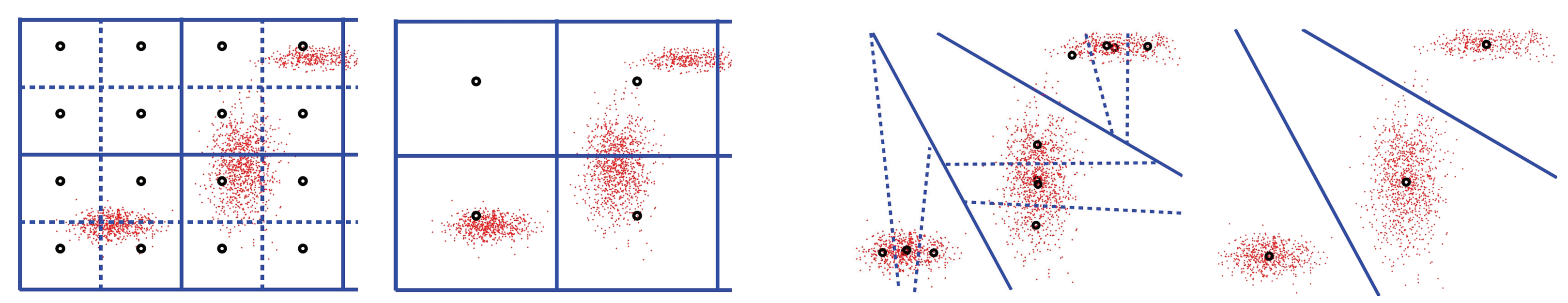

$\mathcal{I}\left(H_0(\mathbf{X}), H_0(\mathbf{Y})\right) = 2 \rightarrow N_0 = 2,\ w_0 = 1$

$\mathcal{I}\left(H_1(\mathbf{X}), H_1(\mathbf{Y})\right) = 4 \rightarrow N_1 = 4 - 2 = 2,\ w_1 = 2$

$\mathcal{I}\left(H_2(\mathbf{X}), H_2(\mathbf{Y})\right) = 5 \rightarrow N_2 = 5 - 4 = 1,\ w_2 = 4$

***No explicit search for matches!***

## The Vocabulary-Guided Pyramid Match



*Uniform bins* *Vocabulary-guided bins*

Uniformly shaped bins result in decreased matching accuracy for high-dimensional features… ⟹ **Tune pyramid partitions to the feature distribution**

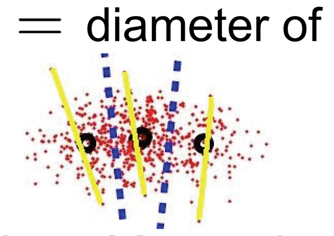
- Hierarchical *k*-means over corpus of features
- Record diameters of the irregularly shaped cells

**Vocabulary-guided (VG) pyramid match cost:**

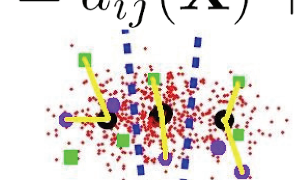
$$\mathcal{C}_{VG-PM}\left(\Psi(\mathbf{X}), \Psi(\mathbf{Y})\right)$$

$$= \sum_{i=0}^{L-1} \sum_{j=1}^{k^i} w_{ij} \underbrace{\left[ \overbrace{\min\left(n_{ij}(\mathbf{X}), n_{ij}(\mathbf{Y})\right)}^{\text{Number of matches in bin i,j}} - \overbrace{\sum_{h=1}^{k} \min\left(c_h\left(n_{ij}(\mathbf{X})\right), c_h\left(n_{ij}(\mathbf{Y})\right)\right)}^{\text{Number of matches in bin i,j's children}} \right]}_{\text{Number of new matches for } j^{th} \text{ bin at } i^{th} \text{ level}}$$

Weighting options: $(1)\, w_{ij} =$ diameter of cell i,j $\qquad (2)\, w_{ij} = d_{ij}(\mathbf{X}) + d_{ij}(\mathbf{Y})$



*admits a Mercer kernel*



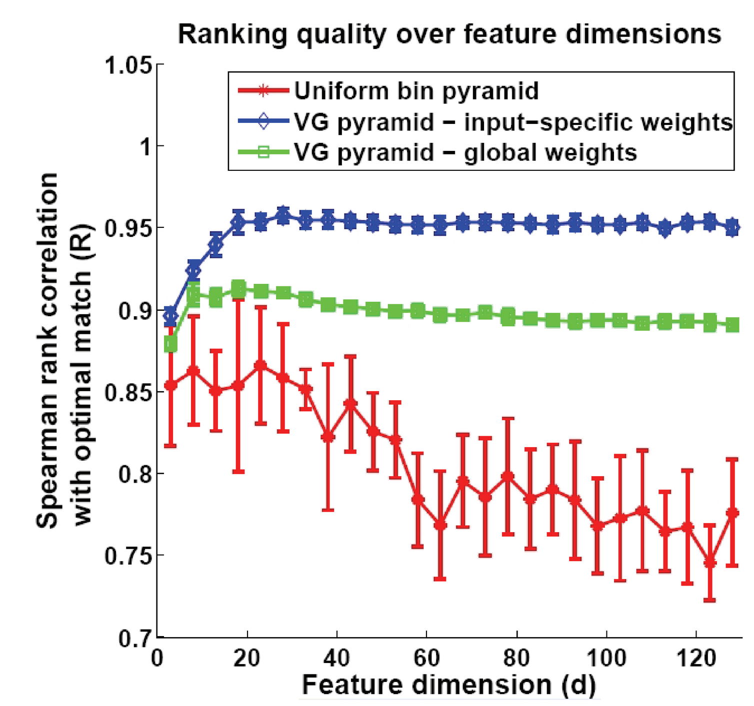
*input-specific upper bound*

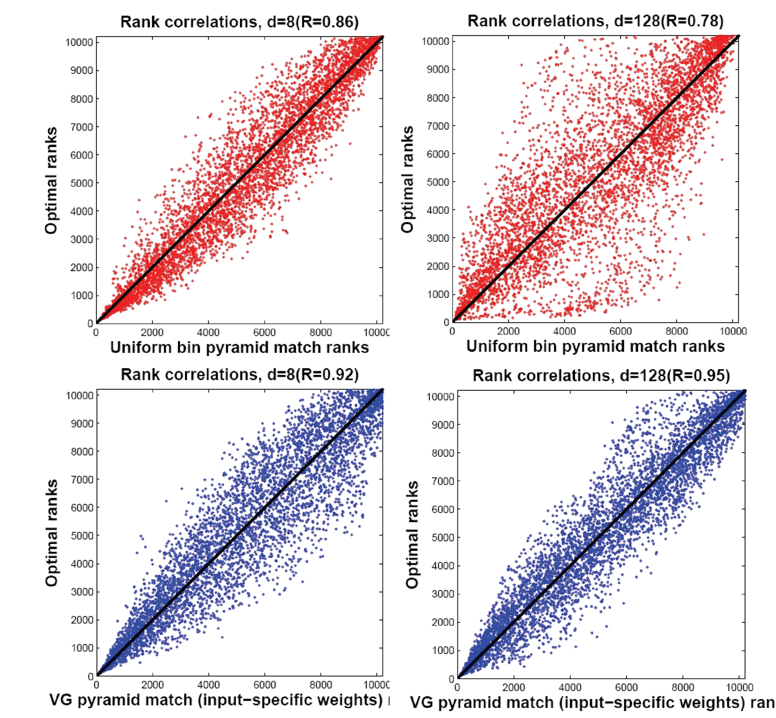
- VG pyramid structure stored once in $O(k^L)$
- Histograms stored sparsely in $O(nL)$ entries
- Inserting point sets into histograms adds $O(kL)$ time
- Match time still only $\boxed{O(nL)}$

## Results

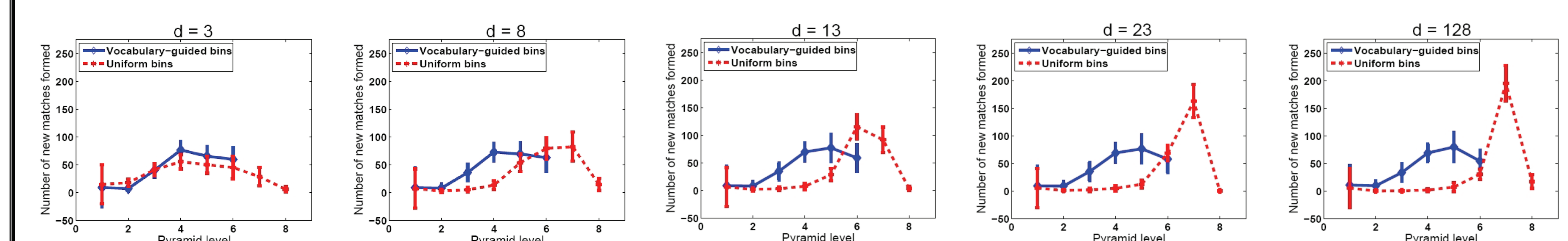
VG pyramids' matching scores consistently highly correlated with the optimal matching, even for high dimensional features.

*(ETH-80 image data, SIFT features, k=10, L=5, results from 10 runs)*

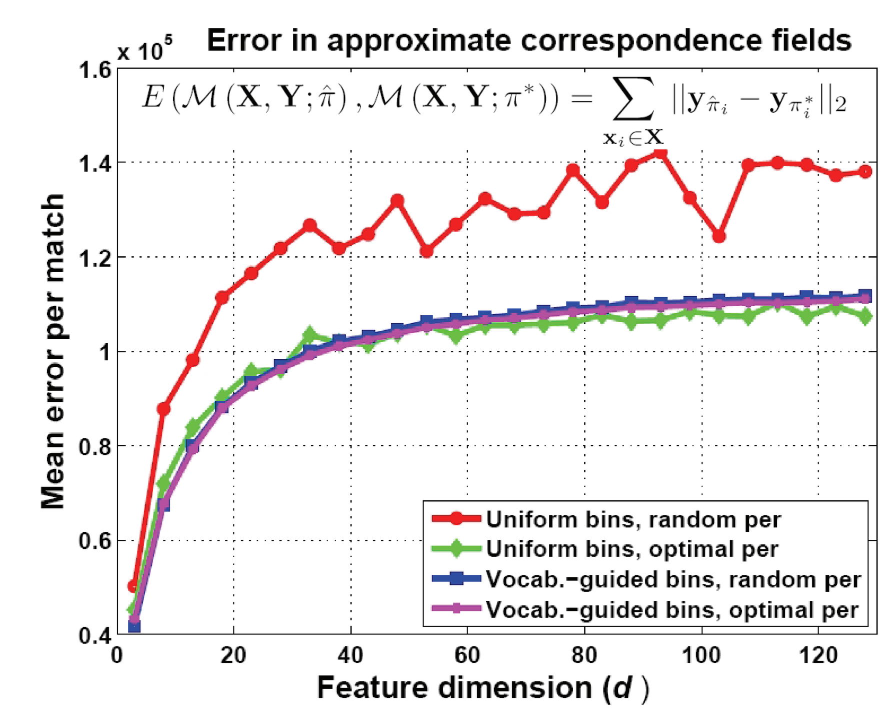


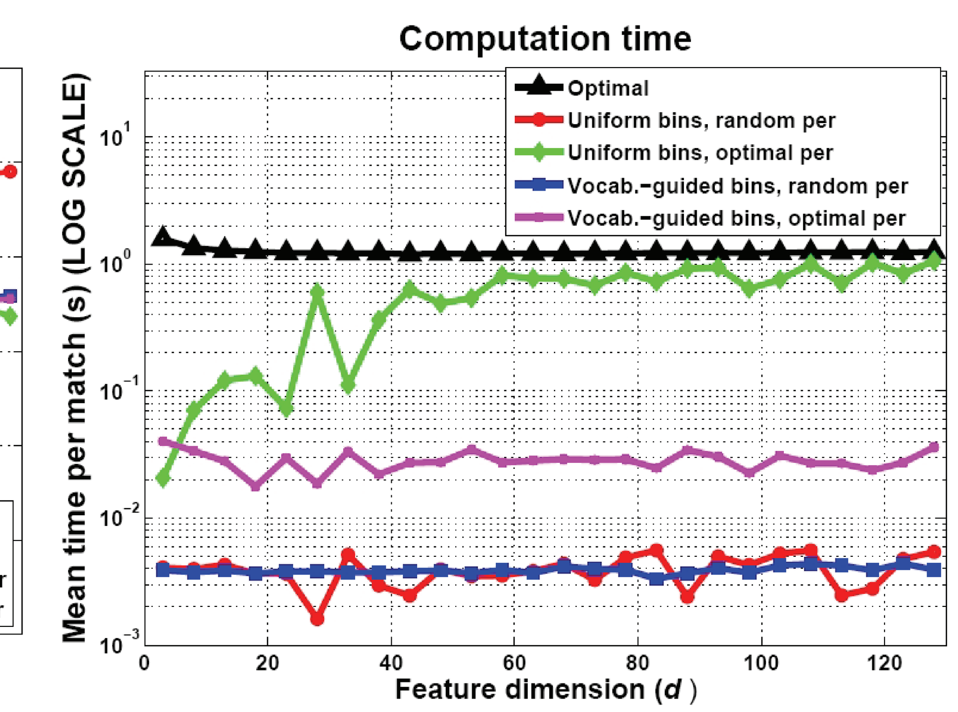




Data-dependent pyramid structure allows more gradual distance ranges.




Explicit correspondence fields are more accurate and faster to compute.




Improved object recognition when used as a kernel in an SVM.

*(Caltech-4 data set, Harris and MSER-detected SIFT features)*

| Pyramid matching method | Mean recognition rate/class (*d*=128/*d*=10) | Time/match (s) (*d*=128/*d*=10) |
|---|---|---|
| Vocabulary-guided bins | **99.0 / 97.7** | 6.1e-4/6.2e-4 |
| Uniform bins | 64.9 / 96.5 | 1.5e-3 / 5.7e-4 |

## Future work

- Sub-linear time PM hashing (ongoing)
- Distortion bounds for the VG-PM?
- Learning weights on pyramid bins
- Beyond geometric vocabularies