Representations for Object Categorisation

CS 783
Visual Recognition

Last class

Problem

- Given: Set of positive training images that contain images from a particular object class

- And a set of negative images that do not contain the particular object class

- Predict given a test image whether the image contains the class or not

Last class

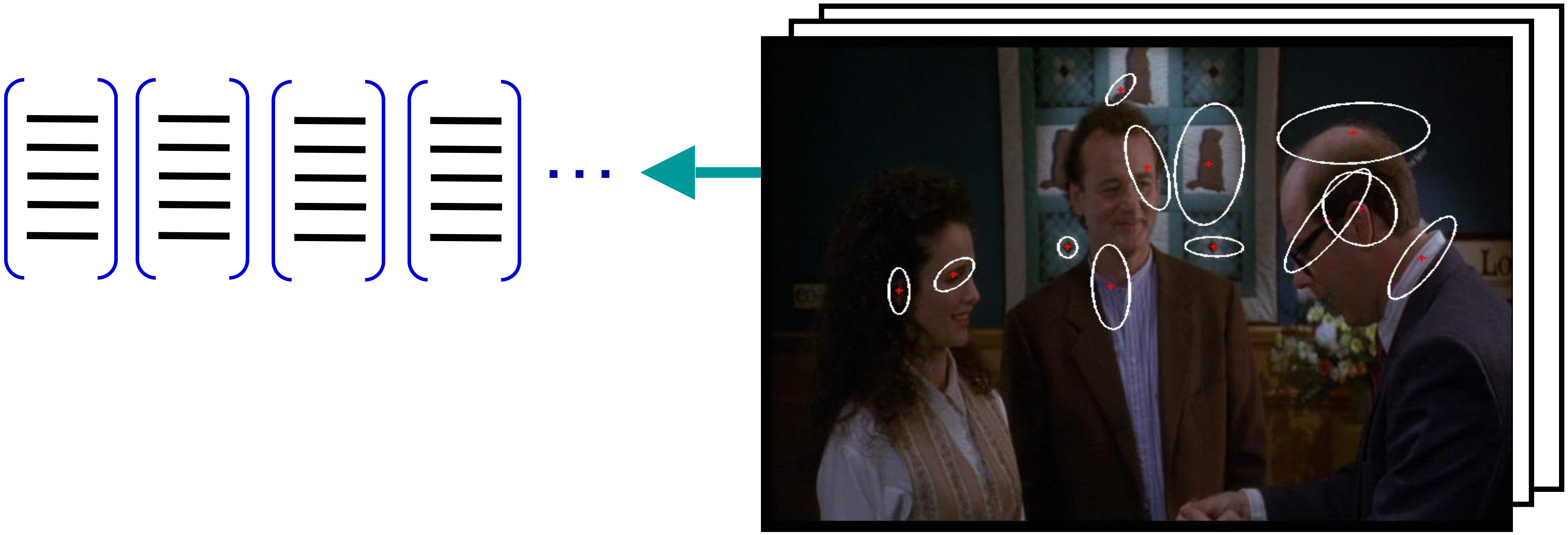
Object → **Bag of 'words'**

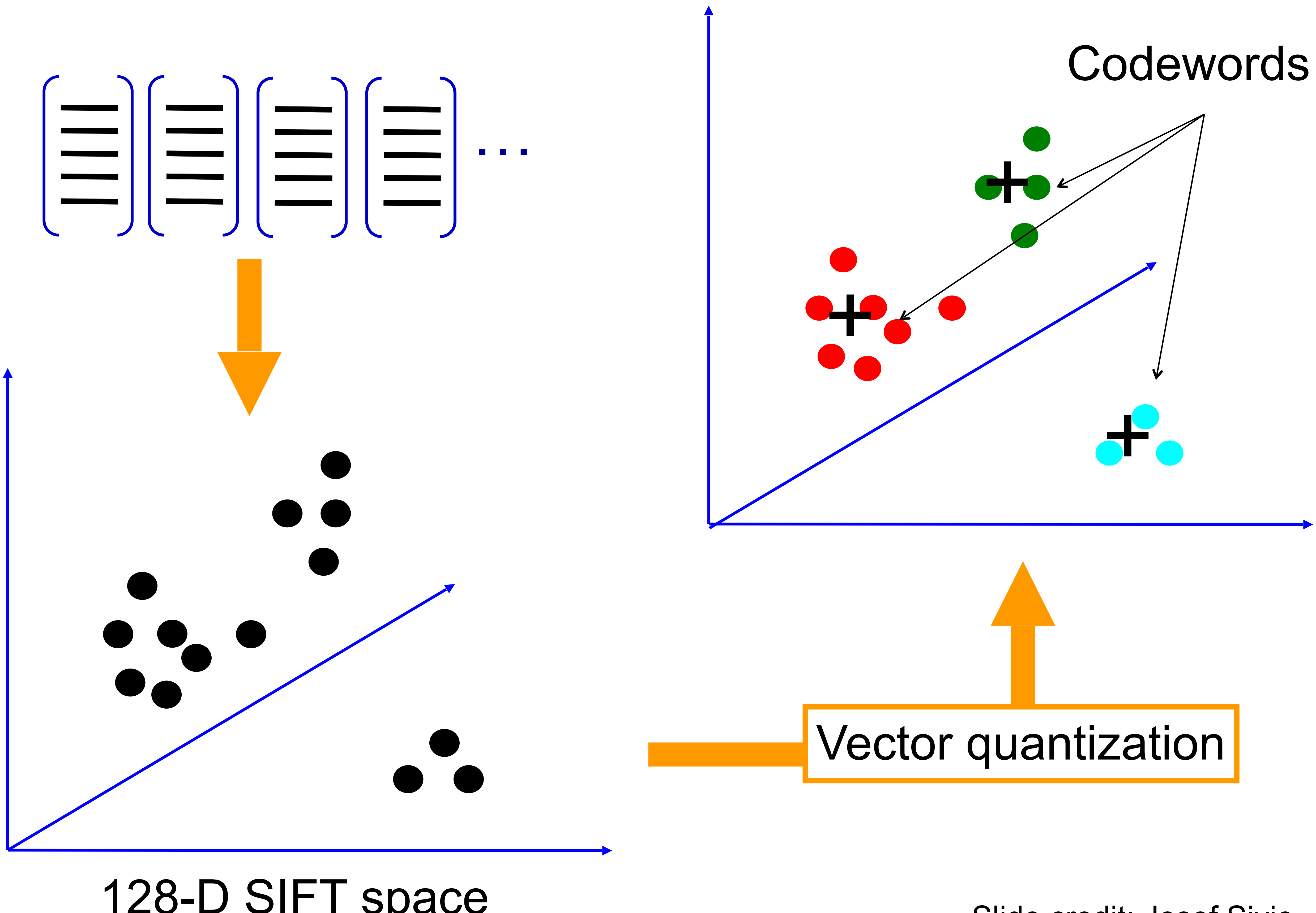
Slide credit: Li Fei Fei

1.Feature detection and representation

- Regular grid
  - Vogel et al. 2003
  - Fei-Fei et al. 2005
- Interest point detector
  - Csurka et al. 2004
  - Fei-Fei et al. 2005
  - Sivic et al. 2005
- Other methods
  - Random sampling (Ullman et al. 2002)

1.Feature detection and representation

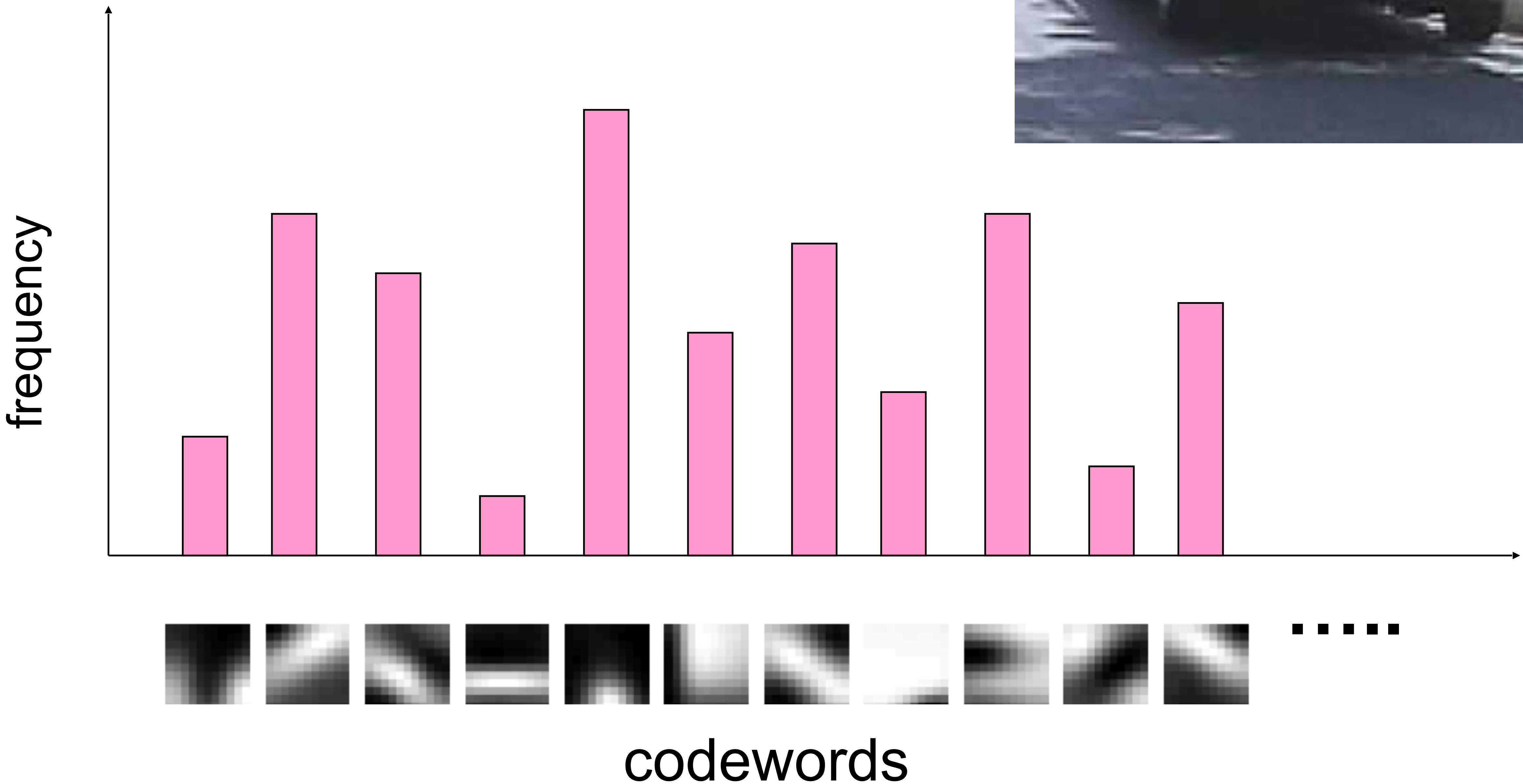




Slide credit: Josef Sivic

Image representation

Histogram of features
assigned to each cluster



Slide credit: Li Fei Fei

Soft-Margin SVM

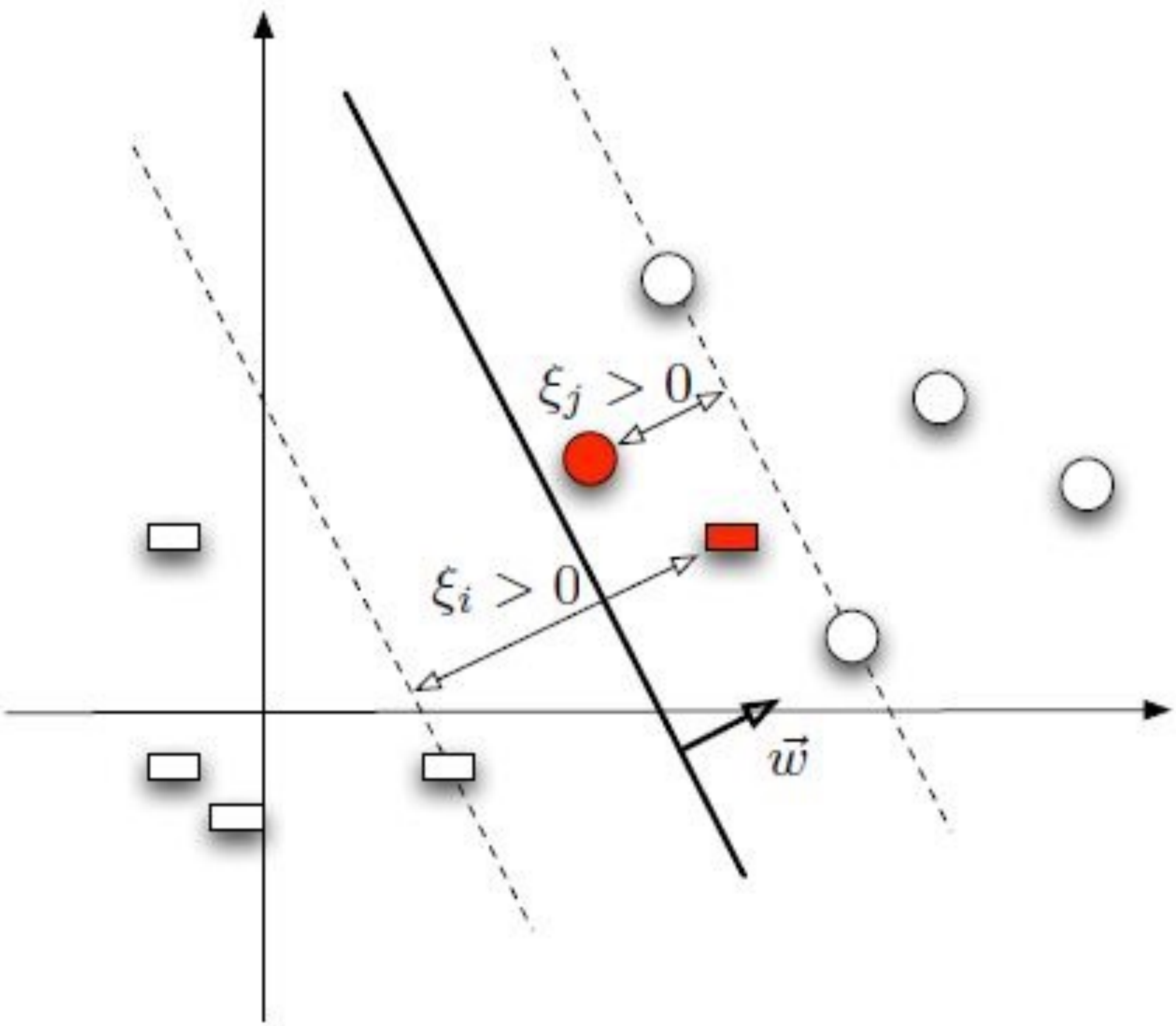
$$\min_{w,\xi} \frac{\|w\|^2}{2} + C \sum_i \xi_i$$

s.t.

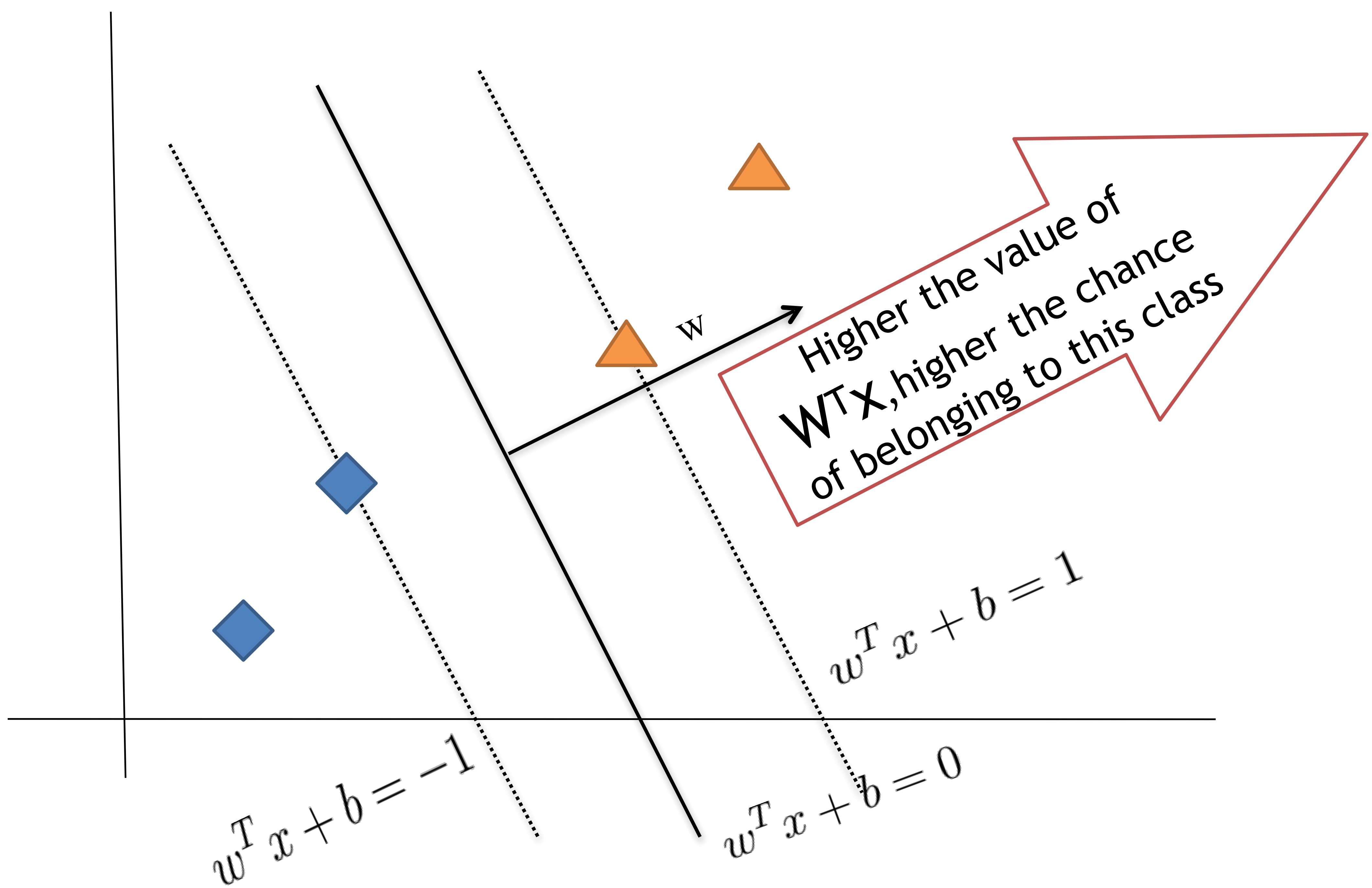
$$if \quad y_i = Y_1, \quad w^T x_i + b \geq 1 - \xi_i$$
$$if \quad y_i = Y_2, \quad w^T x_i + b \leq -(1 - \xi_i)$$

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

Soft-Margin SVM



Score


w
Higher the value of
$W^T x$, higher the chance
of belonging to this class
$w^T x + b = 1$
$w^T x + b = 0$
$w^T x + b = -1$

Multi-Class SVM

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

s.t.

$$w_{Y_i} \cdot x_i - w_{\hat{Y} \neq Y_i} \cdot x_i \geqslant 1 - \xi_i, \forall\ (x_i, Y_i)$$

$$\forall i, \hat{y} \neq y_i, \xi_i \geqslant 0$$

here, $$w = \begin{pmatrix} w_{Y_1} \\ w_{Y_2} \\ w_{Y_3} \end{pmatrix}$$

Multi-Class SVM

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

s.t.

$$\boldsymbol{w} \cdot \boldsymbol{\Phi}(x_i, y_i) - \boldsymbol{w} \cdot \boldsymbol{\Phi}(x_i, \bar{y}) \geq \Delta(y_i, \bar{y}) - \xi_i$$

$$\forall i, \hat{y} \neq y_i, \xi_i \geqslant 0$$

$$w = \begin{pmatrix} w_{Y_1} \\ w_{Y_2} \\ . \\ . \\ . \\ w_{Y_{k-1}} \\ w_{Y_k} \end{pmatrix} \quad \Phi(x, y) = \begin{pmatrix} 0 \\ . \\ . \\ x \\ . \\ . \\ 0 \end{pmatrix}$$

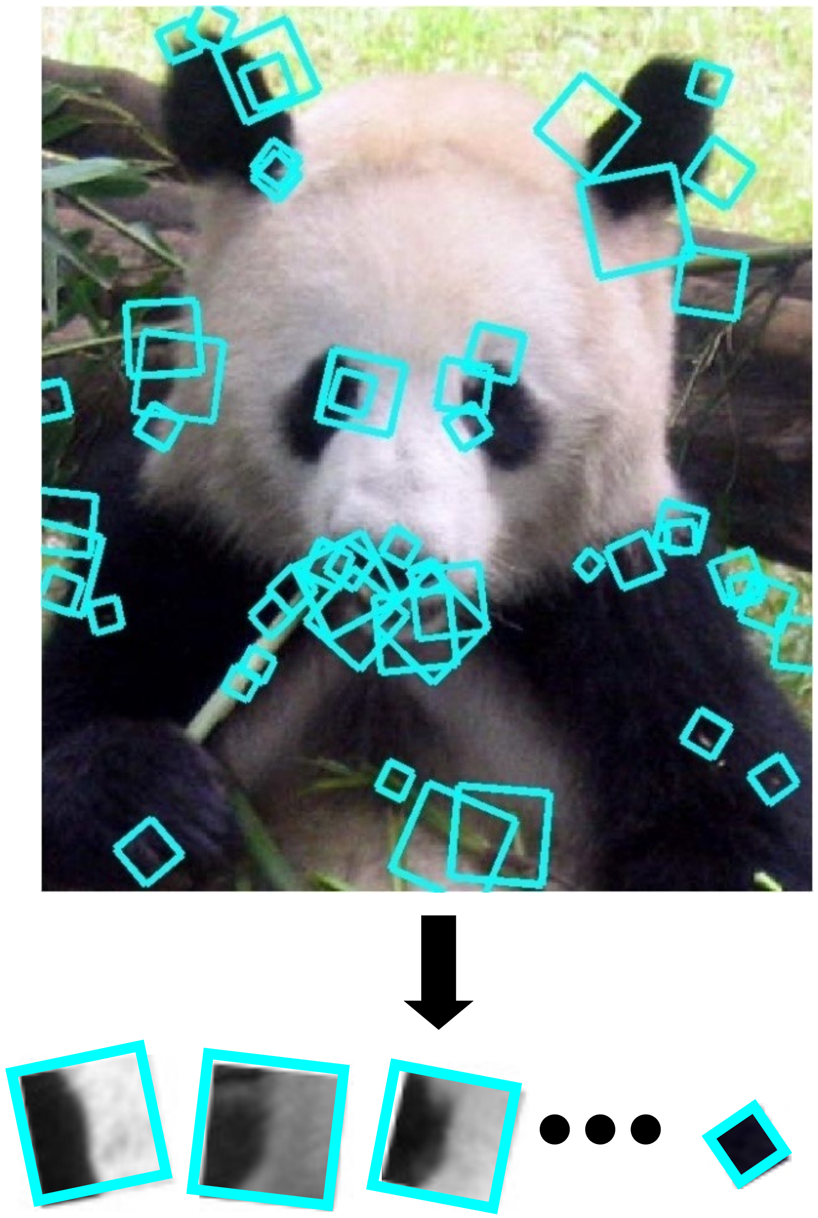
Feature Matching

- In the BoW framework, we obtain a BoW representation to match all images
- We use a support vector machine framework for training to classify the various categories, either binary or multi-class classification
- Prediction: Use trained model to predict the class of a particular test example

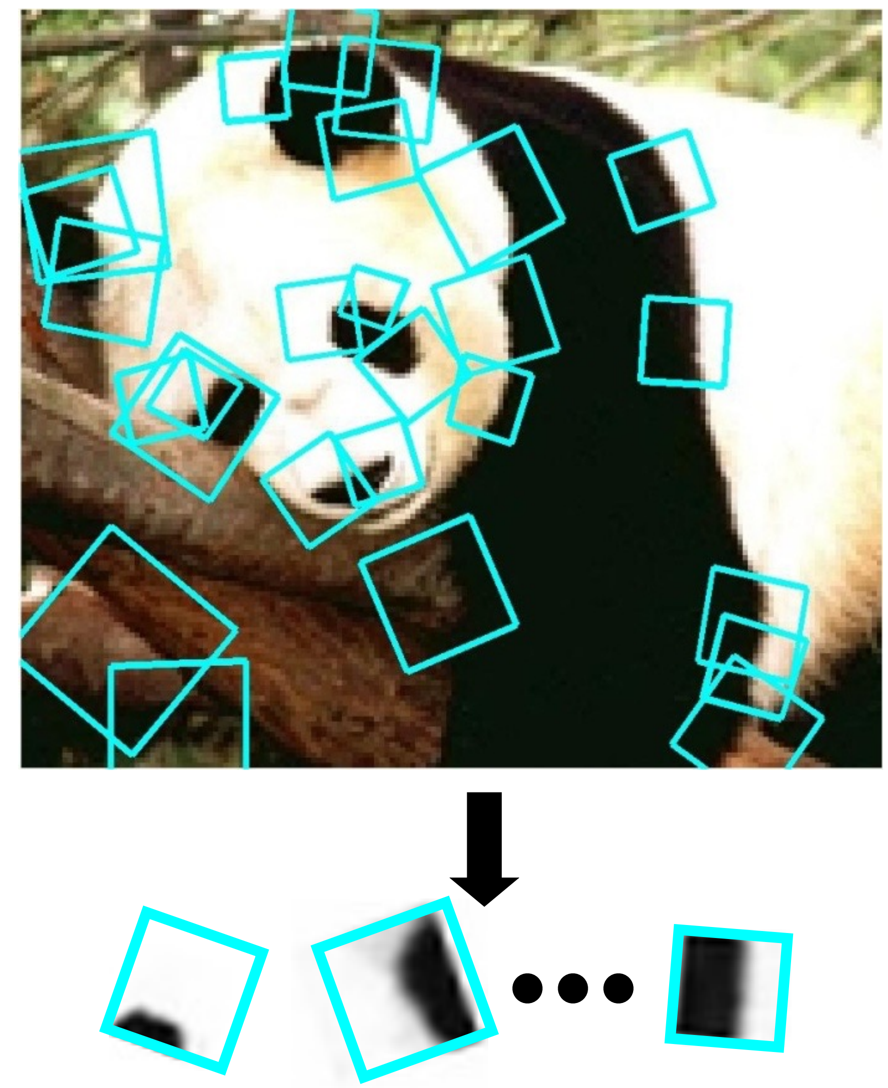
Best matching approach?

- Given an image obtain the exact correspondence between the parts of an image between a test example and training example

Sets of features



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

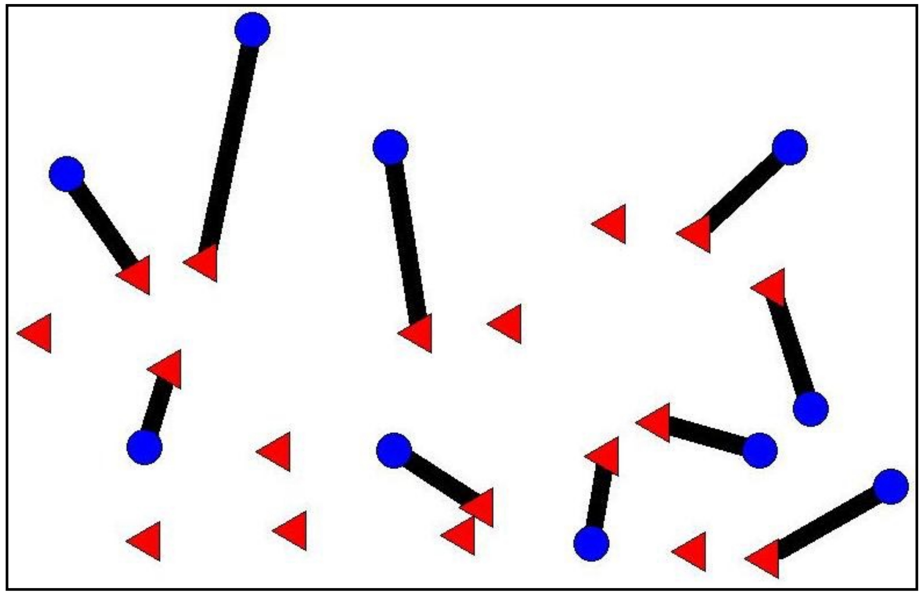


$$\mathbf{Y} = \{\vec{\mathbf{y}}_1, \ldots, \vec{\mathbf{y}}_n\}$$

Slide credit
Kristen Grauman

Partial matching for sets of features

Compare sets by computing a *partial matching* between their features.



Robust to clutter, segmentation errors, occlusion…

Slide credit
Kristen Grauman

Pyramid match overview

Pyramid match kernel measures similarity of a partial matching between two sets:

Place multi-dimensional, multi-resolution grid over point sets

Consider points matched at finest resolution where they fall into same grid cell

Approximate similarity between matched points with worst case similarity at given level

No explicit search for matches!

Slide credit
Kristen Grauman

Pyramid match kernel

Number of newly matched pairs at level *i*

Approximate partial match similarity

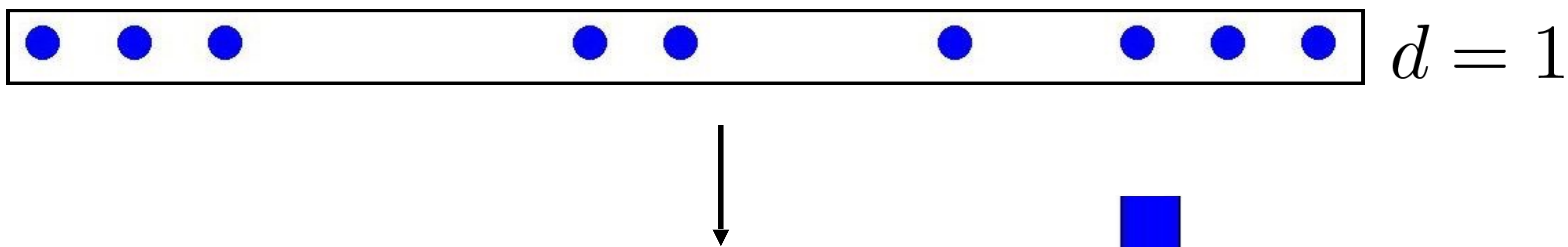
$$K_\Delta = \sum_{i=0}^{L} w_i N_i$$

Measure of difficulty of a match at level *i*

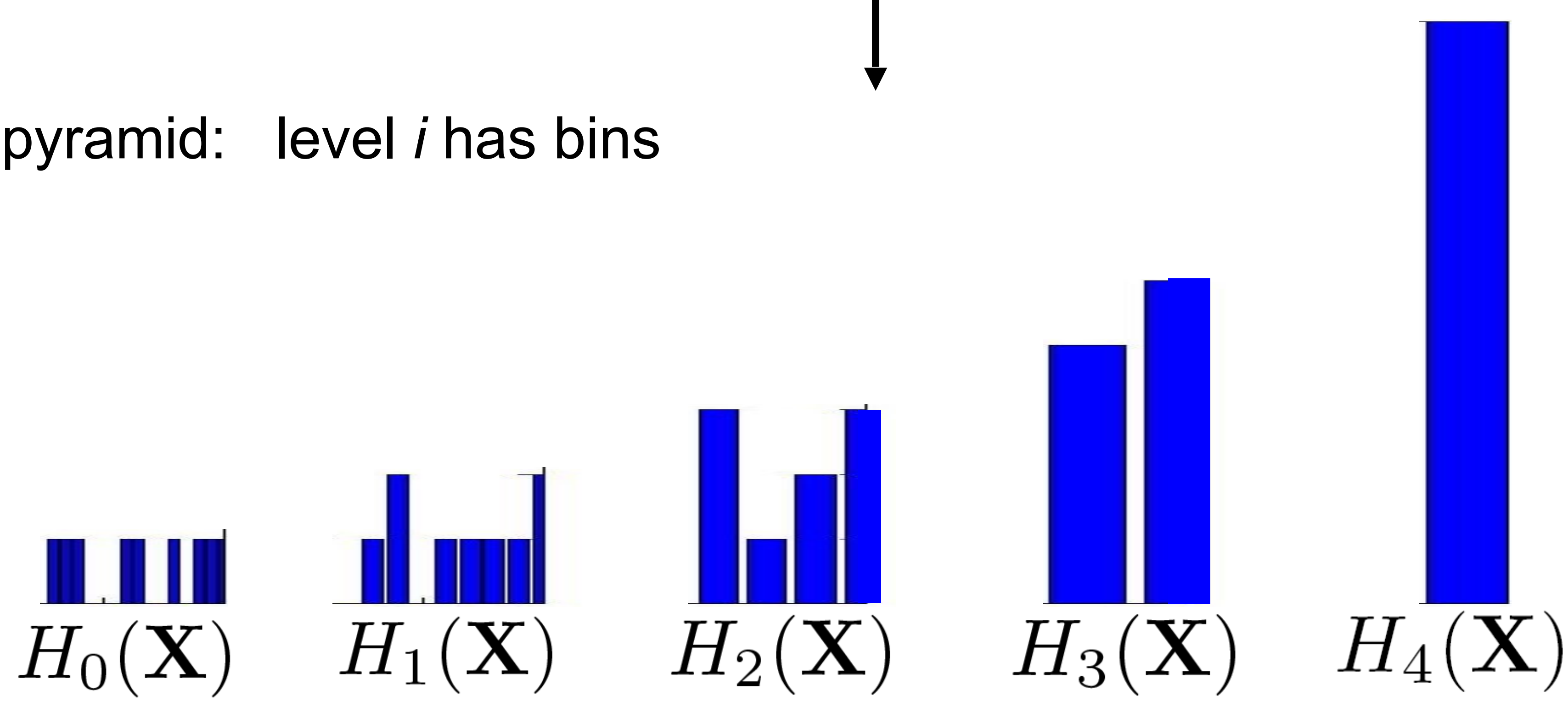
Slide credit
Kristen Grauman

Feature extraction

$$\mathbf{X} = \{\vec{\mathbf{x}}_1, \ldots, \vec{\mathbf{x}}_m\}, \qquad \vec{\mathbf{x}}_i \in \Re^d$$



Histogram pyramid: level *i* has bins of size 2^i



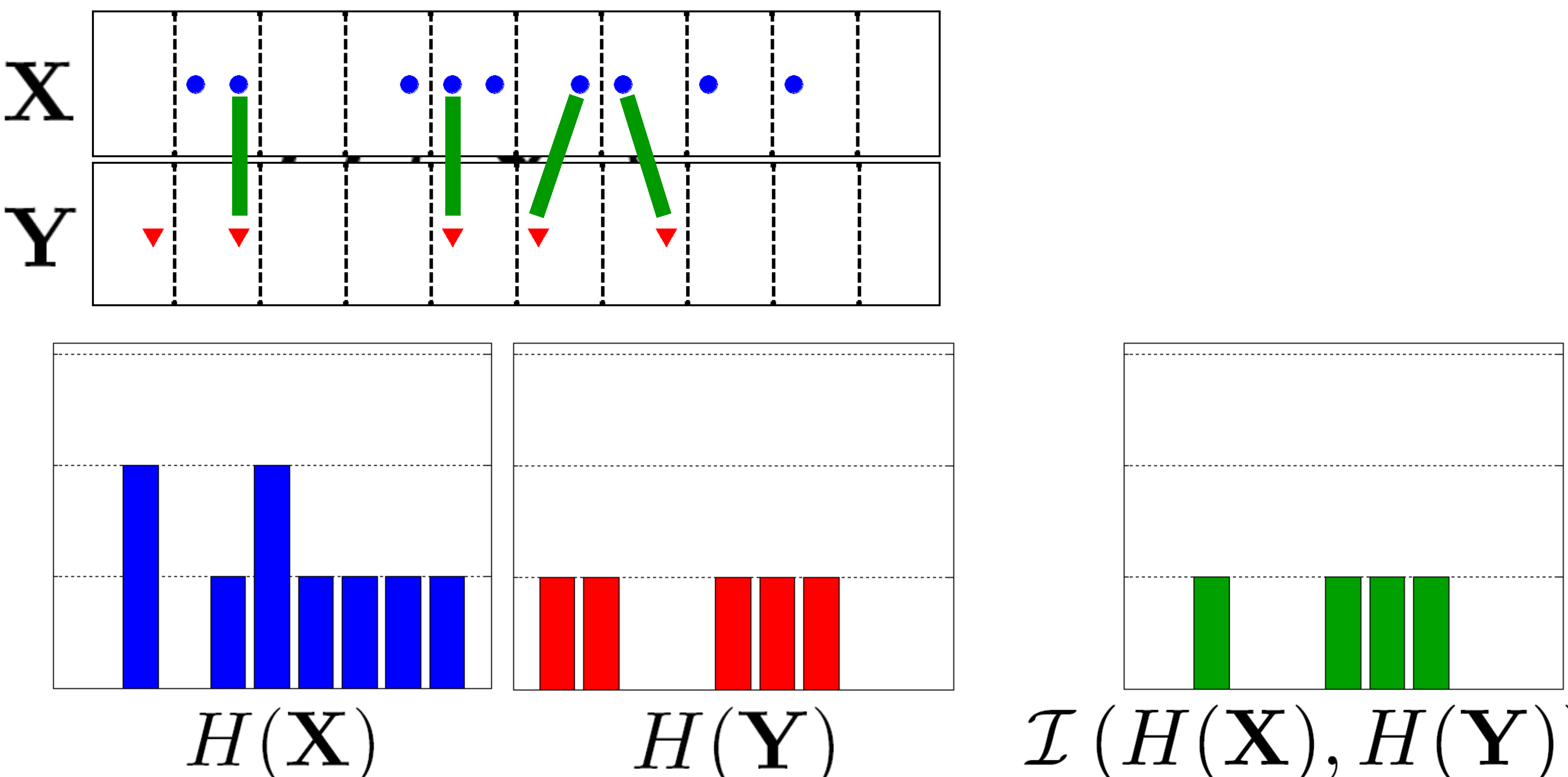
$$\Psi(\mathbf{X}) = [H_0(\mathbf{X}), \ldots, H_L(\mathbf{X})]$$

Slide credit
Kristen Grauman

Counting matches

Histogram intersection

$$\mathcal{I}\left(H(\mathbf{X}), H(\mathbf{Y})\right) = \sum_{j=1}^{r} \min\left(H(\mathbf{X})_j, H(\mathbf{Y})_j\right)$$



$$\mathcal{I}\left(H(\mathbf{X}), H(\mathbf{Y})\right) = 4$$

Slide credit Kristen Grauman

Counting new matches

Histogram intersection

$$\mathcal{I}\left(H(\mathbf{X}), H(\mathbf{Y})\right) = \sum_{j=1}^{r} \min\left(H(\mathbf{X})_j, H(\mathbf{Y})_j\right)$$

$$\boxed{N_i = \overbrace{\mathcal{I}\left(H_i(\mathbf{X}), H_i(\mathbf{Y})\right)}^{\text{matches at this level}} - \overbrace{\mathcal{I}\left(H_{i-1}(\mathbf{X}), H_{i-1}(\mathbf{Y})\right)}^{\text{matches at previous level}}}$$

Difference in histogram intersections across levels counts *number of new pairs* matched

Slide credit
Kristen Grauman

Pyramid match kernel

$$K_\Delta\left(\overbrace{\Psi(\mathbf{X}),\Psi(\mathbf{Y})}^{\text{histogram pyramids}}\right) =$$

$$\sum_{i=0}^{L}\underbrace{\frac{1}{2^i}}_{\substack{\text{measure of difficulty of a}\\ \text{match at level } i}}\left(\underbrace{\mathcal{I}\left(H_i(\mathbf{X}),H_i(\mathbf{Y})\right)-\mathcal{I}(H_{i-1}(\mathbf{X}),H_{i-1}(\mathbf{Y}))}_{\text{number of newly matched pairs at level } i}\right)$$

- Weights inversely proportional to bin size
- Normalize kernel values to avoid favoring large sets

Slide credit
Kristen Grauman

Efficiency

For sets with *m* features of dimension *d*, and pyramids with *L* levels, computational complexity of

Pyramid match kernel:

$$O(dmL)$$

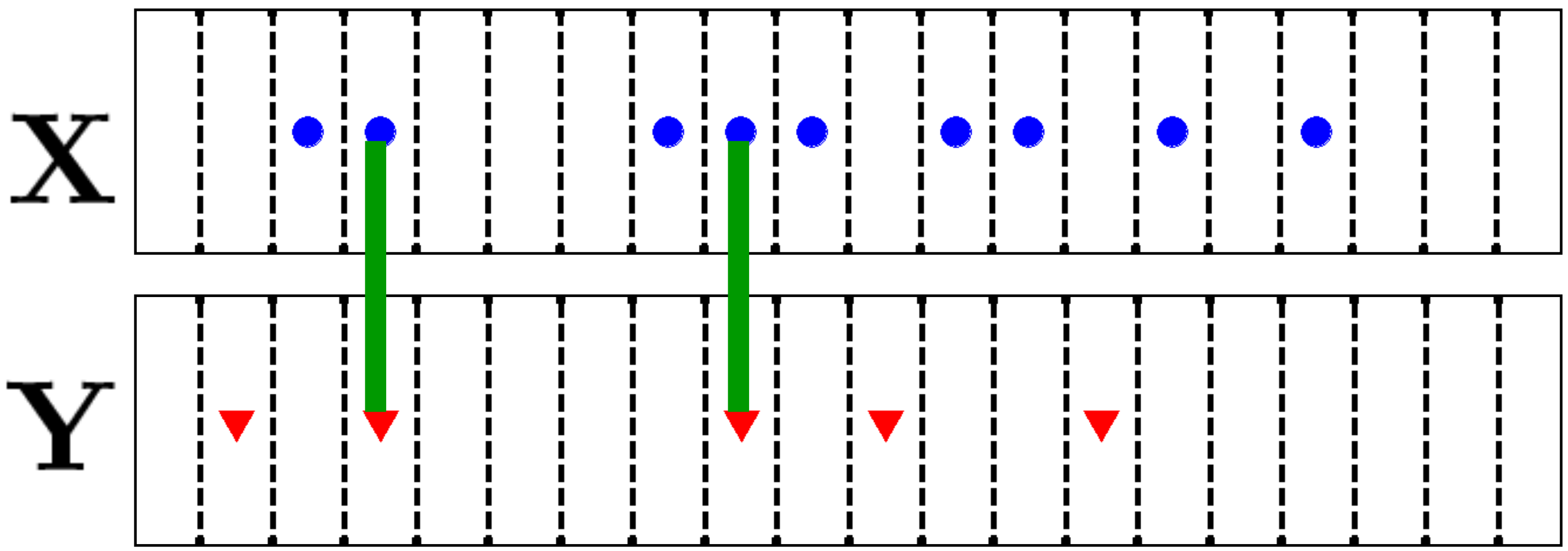
Existing set kernel approaches:

$$O(dm^3) \quad \text{or} \quad O(dm^2)$$

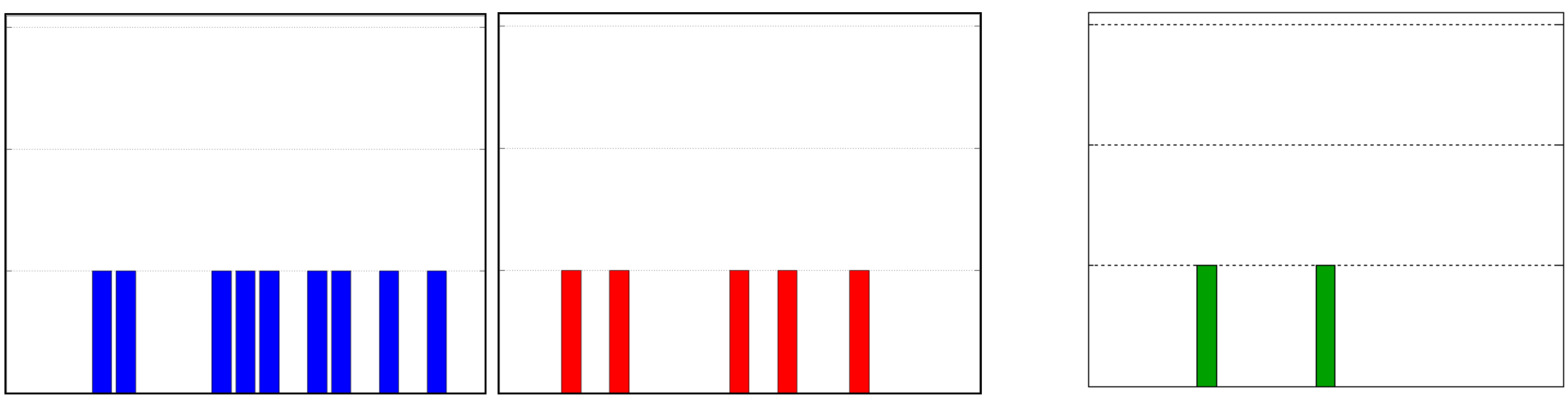
Slide credit
Kristen Grauman

Example pyramid match

Level 0



$$N_0 = 2$$
$$w_0 = 1$$

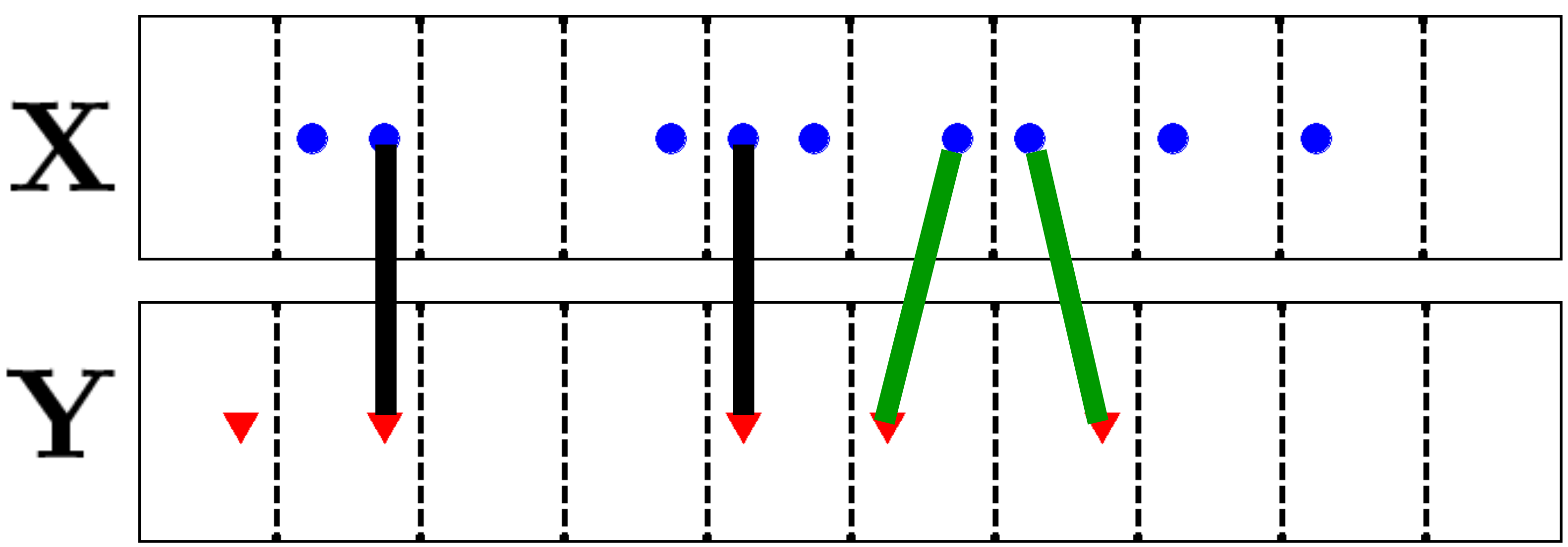


$H_0(\mathbf{X})$ $H_0(\mathbf{Y})$ $\mathcal{I}_0 = 2$

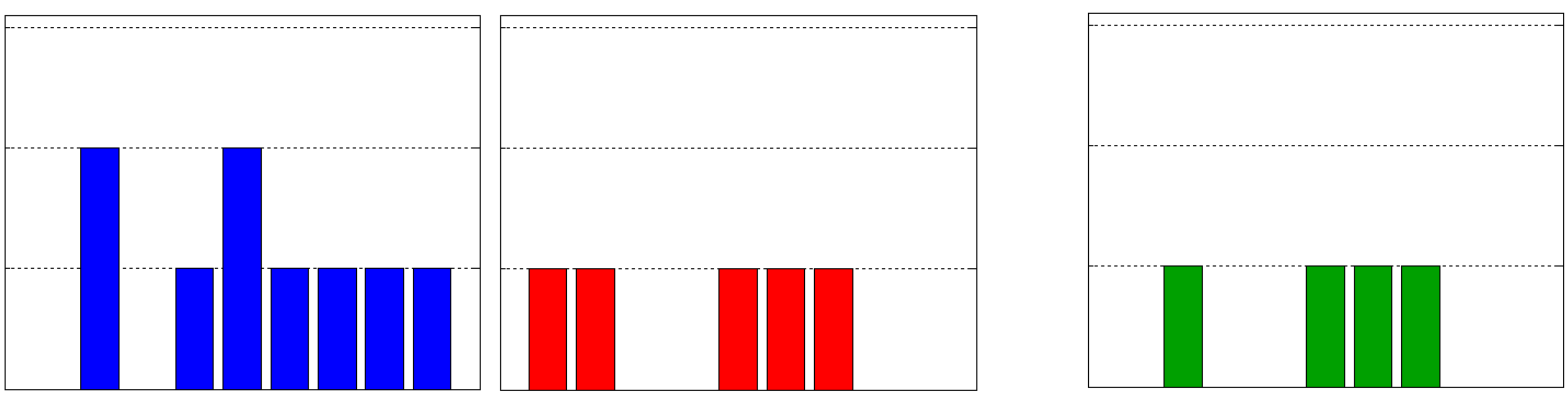
Slide credit
Kristen Grauman

Example pyramid match

Level 1



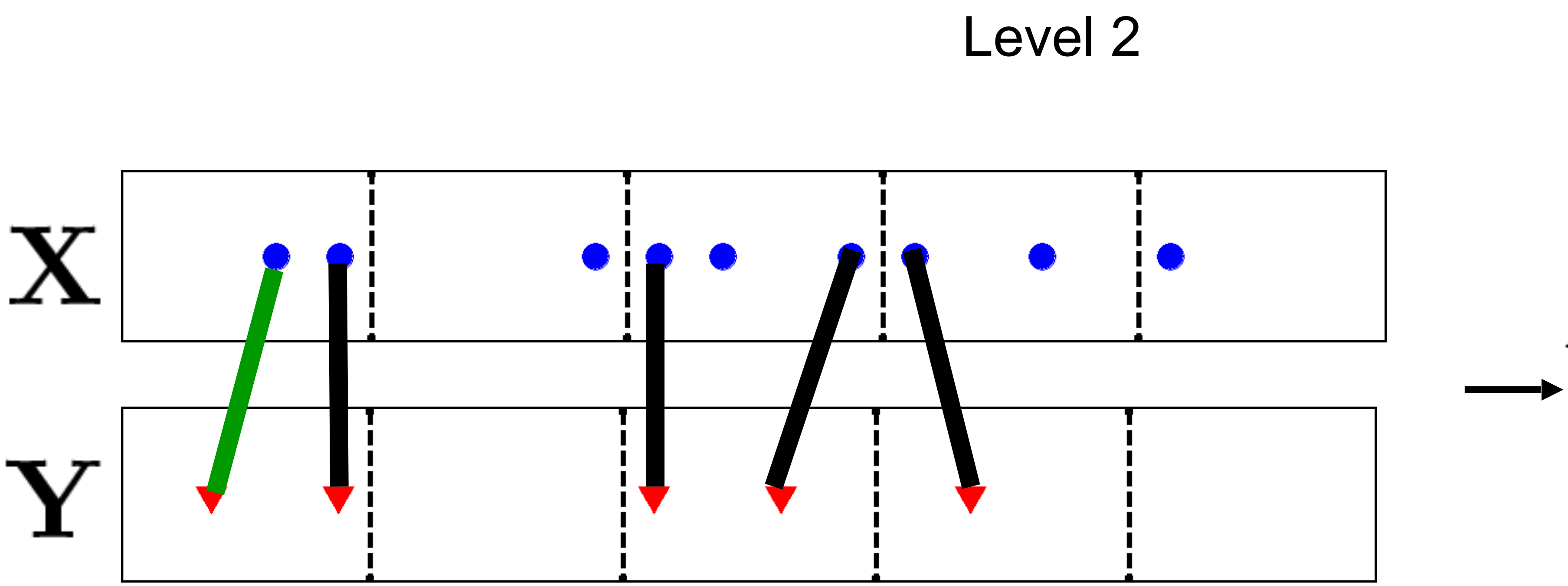
$$N_1 = 4 - 2 = 2$$
$$w_1 = \frac{1}{2}$$



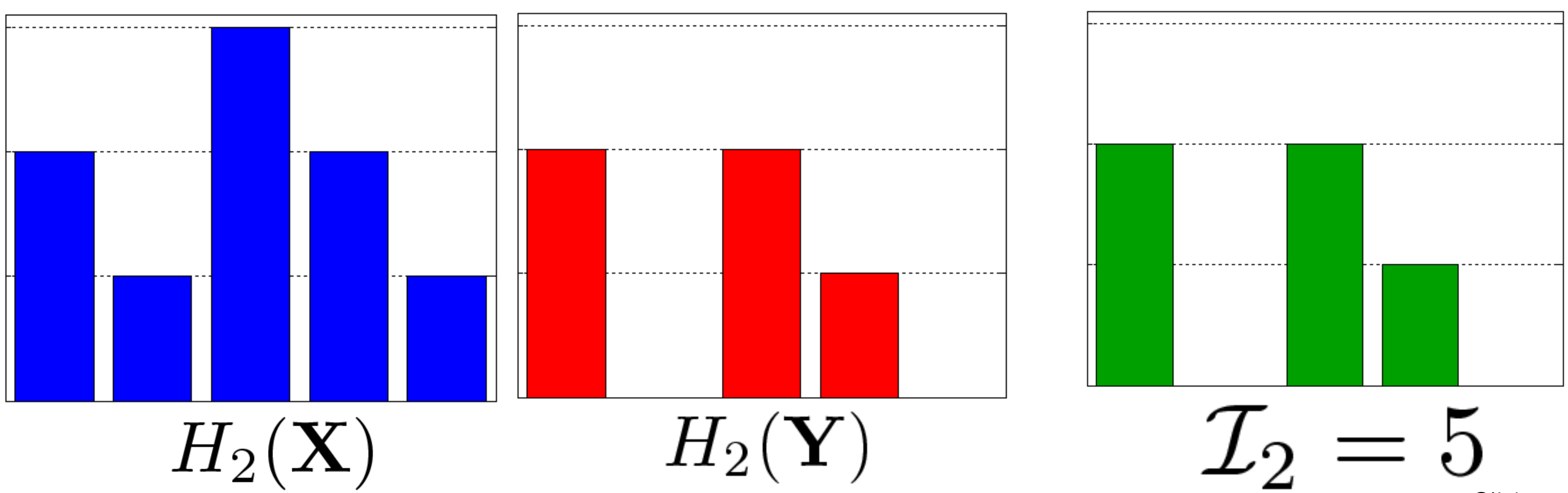
$H_1(\mathbf{X})$ $H_1(\mathbf{Y})$ $\mathcal{I}_1 = 4$

Slide credit
Kristen Grauman

Example pyramid match



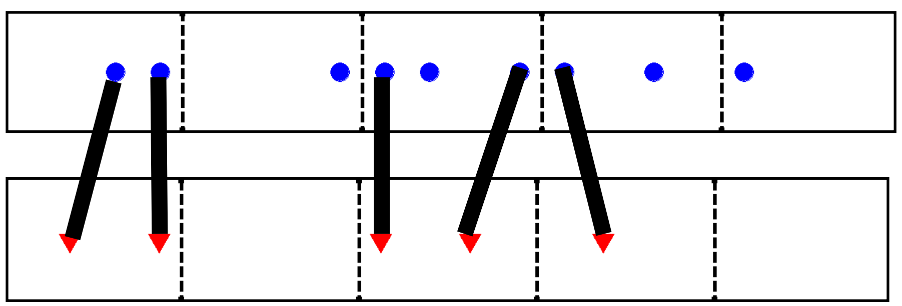
$$N_2 = 5 - 4 = 1$$
$$w_2 = \frac{1}{4}$$



Slide credit Kristen Grauman

Example pyramid match

pyramid match

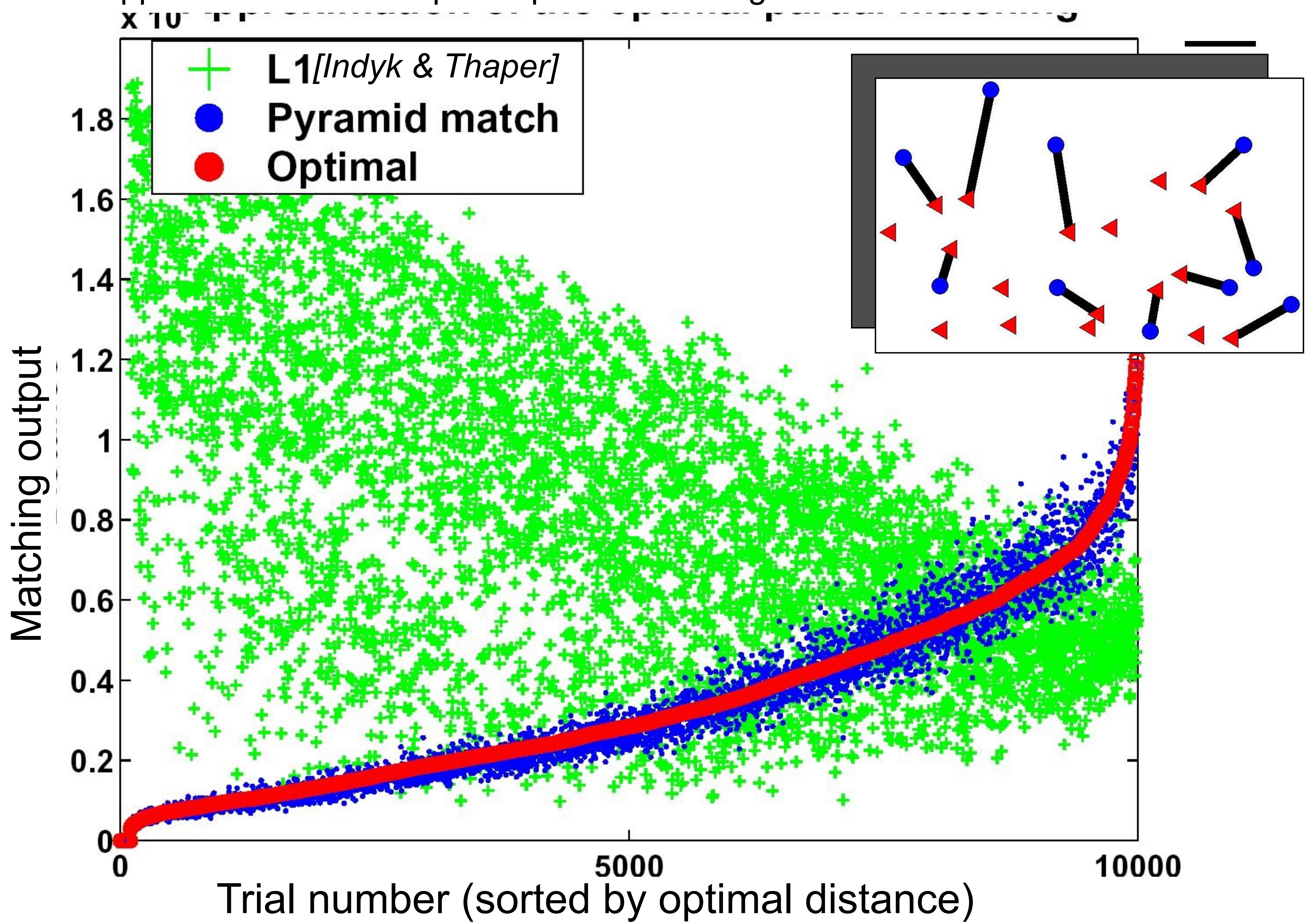


$$K_\Delta = \sum_{i=0}^{L} w_i N_i$$

$$= 1(2) + \frac{1}{2}(2) + \frac{1}{4}(1) = 3.25$$

Slide credit
Kristen Grauman

Approximation of the optimal partial matching



100 sets with 2D points, cardinalities vary between 5 and 100

Slide credit
Kristen Grauman

Building a classifier

Train SVM by computing kernel values between all labeled training examples

Classify novel examples by computing kernel values against support vectors

One-versus-all for multi-class classification

Convergence is guaranteed since pyramid match kernel is positive-definite.

Slide credit
Kristen Grauman

Object recognition results

ETH-80 database 8 object classes

Features:

- Harris detector
- PCA-SIFT descriptor, *d*=10

Kernel	**Complexity**	**Recognition rate**
Match *[Wallraven et al.]*	$O(dm^2)$	84%
Bhattacharyya affinity *[Kondor & Jebara]*	$O(dm^3)$	85%
Pyramid match	$O(dmL)$	84%

Eichhorn and Chapelle 2004

Slide credit Kristen Grauman

Object recognition results

Caltech objects database 101 object classes

Features:

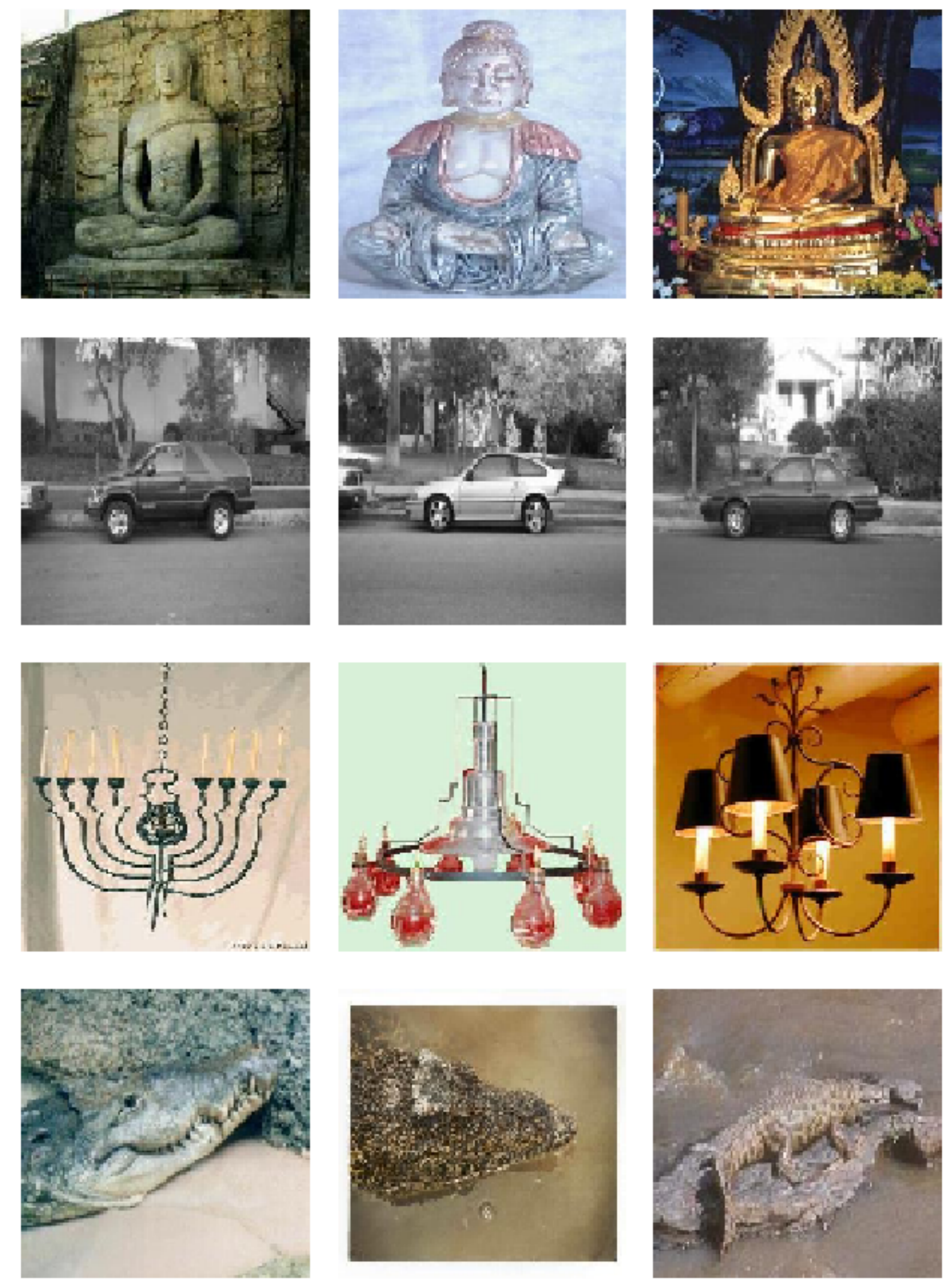
- SIFT detector
- PCA-SIFT descriptor, d=10

30 training images / class

43% recognition rate

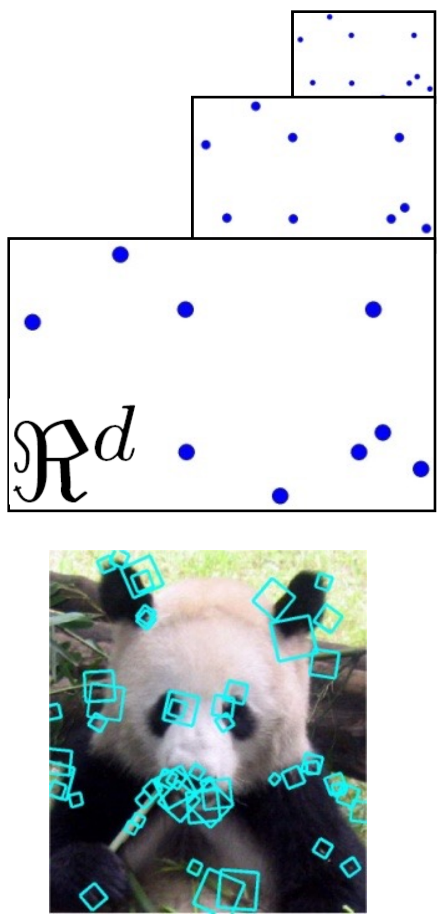
(1% chance performance)

0.002 seconds per match

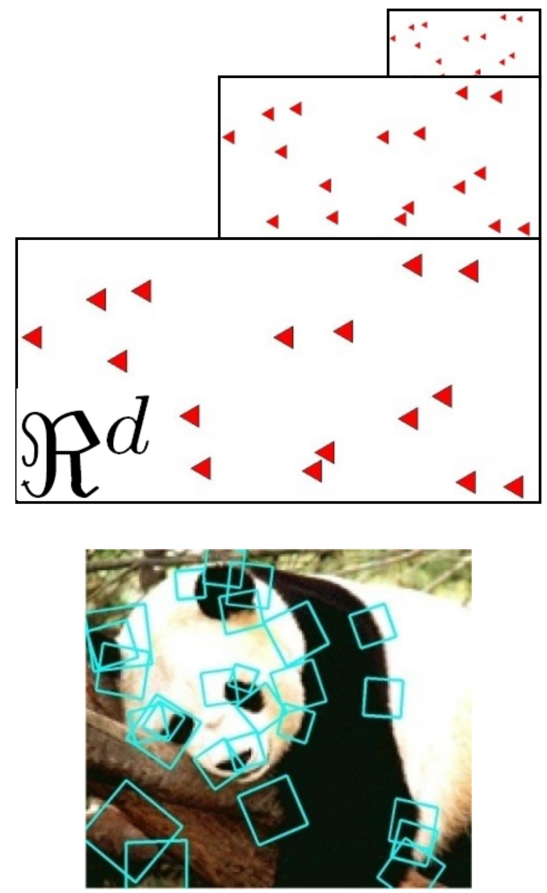


Slide credit
Kristen Grauman

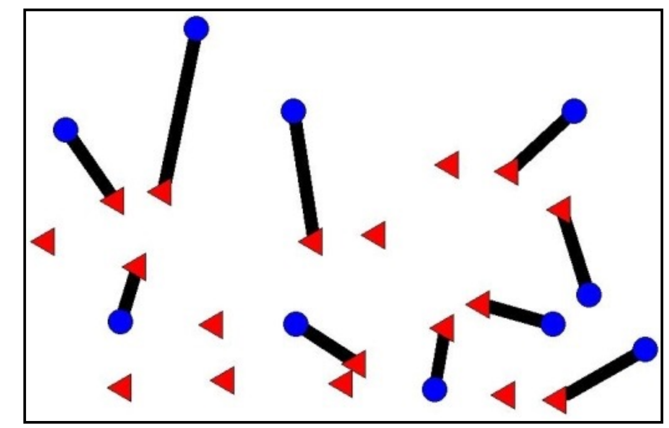
Summary: Pyramid match kernel



$\cap$



$\approx$

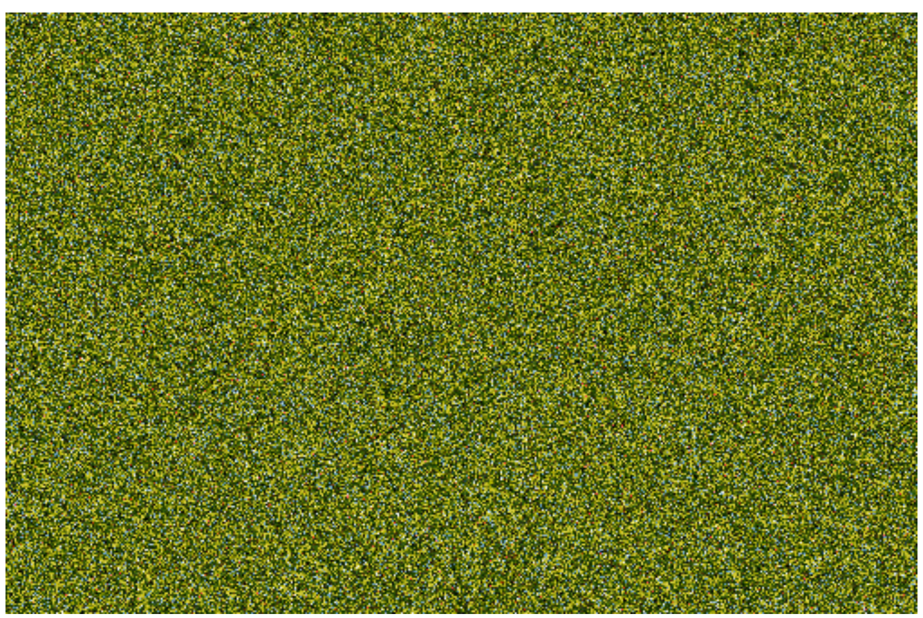


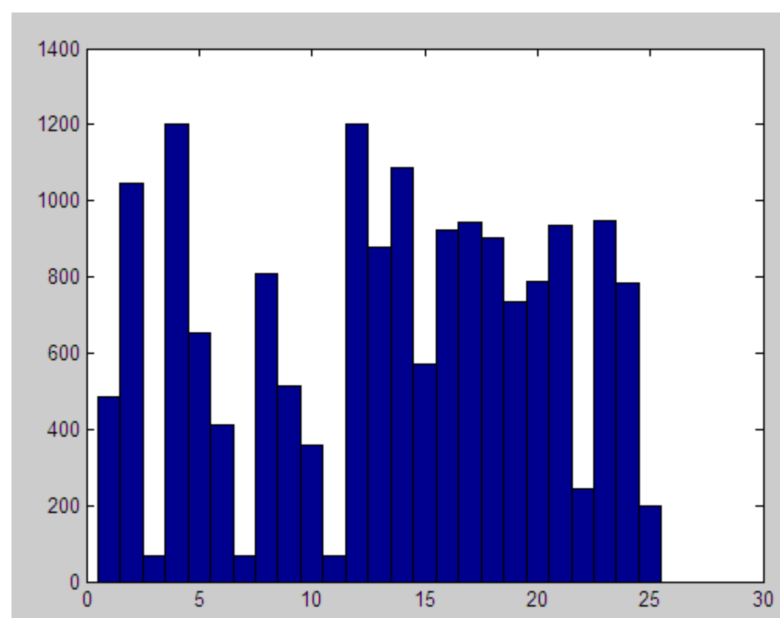
optimal partial matching between sets of features

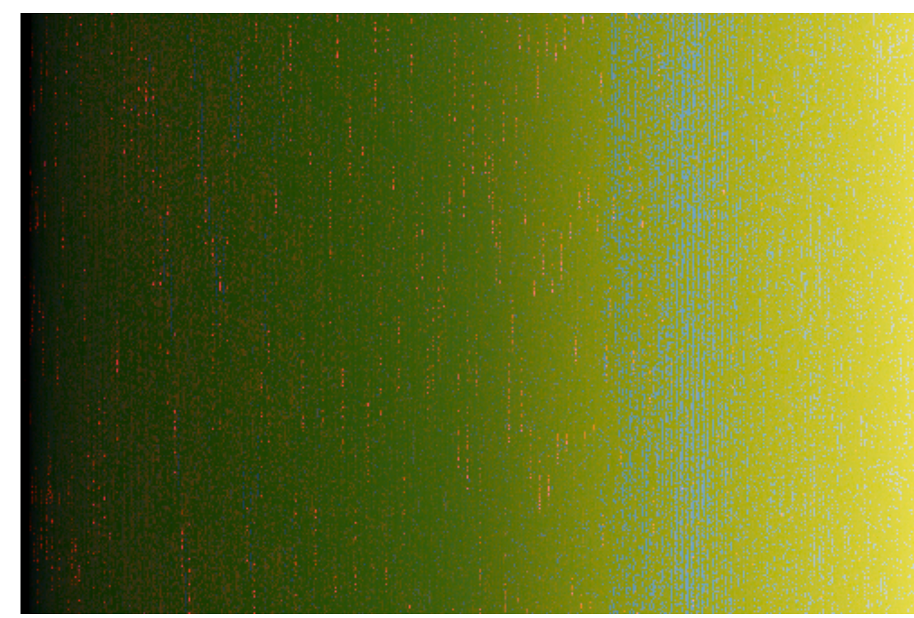
$$K_\Delta\left(\Psi(\mathbf{X}),\Psi(\mathbf{Y})\right) = \sum_{i=0}^{L} \underbrace{\frac{1}{2^i}}_{\text{difficulty of a match at level i}} \Big(\underbrace{\mathcal{I}\left(H_i(\mathbf{X}),H_i(\mathbf{Y})\right) - \mathcal{I}\left(H_{i-1}(\mathbf{X}),H_{i-1}(\mathbf{Y})\right)}_{\text{number of new matches at level i}} \Big)$$

Slide credit Kristen Grauman

But what about layout?







All of these images have the same color histogram

Beyond Bags of Features: Spatial Pyramid Matching for Recognizing Natural Scene Categories

To appear in CVPR 2006

Svetlana Lazebnik (slazebni@uiuc.edu)

Beckman Institute, University of Illinois at Urbana-Champaign

Cordelia Schmid (cordelia.schmid@inrialpes.fr)

INRIA Rhône-Alpes, France

Jean Ponce (ponce@di.ens.fr)

Ecole Normale Supérieure, France

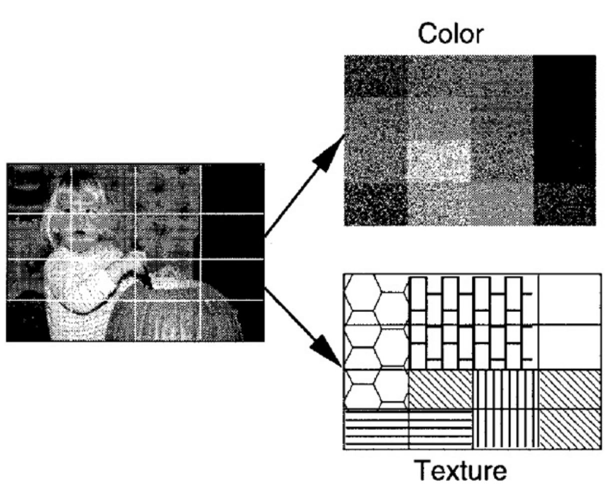
http://www-cvr.ai.uiuc.edu/ponce_grp

Slide credit Svetlana Lazebnik

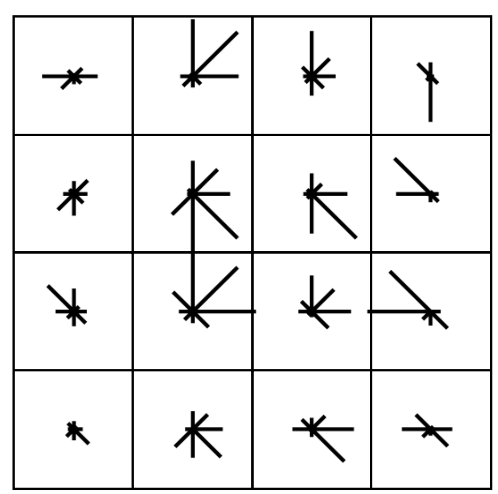
Overview

- A "pre-attentive" approach: recognize the scene as a whole without examining its constituent objects   Biederman (1988), Thorpe et al. (1996), Fei-Fei et al. (2002), Renninger & Malik (2004
- Inspiration: *locally orderless images*   Koenderink & Van Doorn (1999)


- Previous work: "subdivide-and-disorder" strategy



Szummer & Picard (1997)



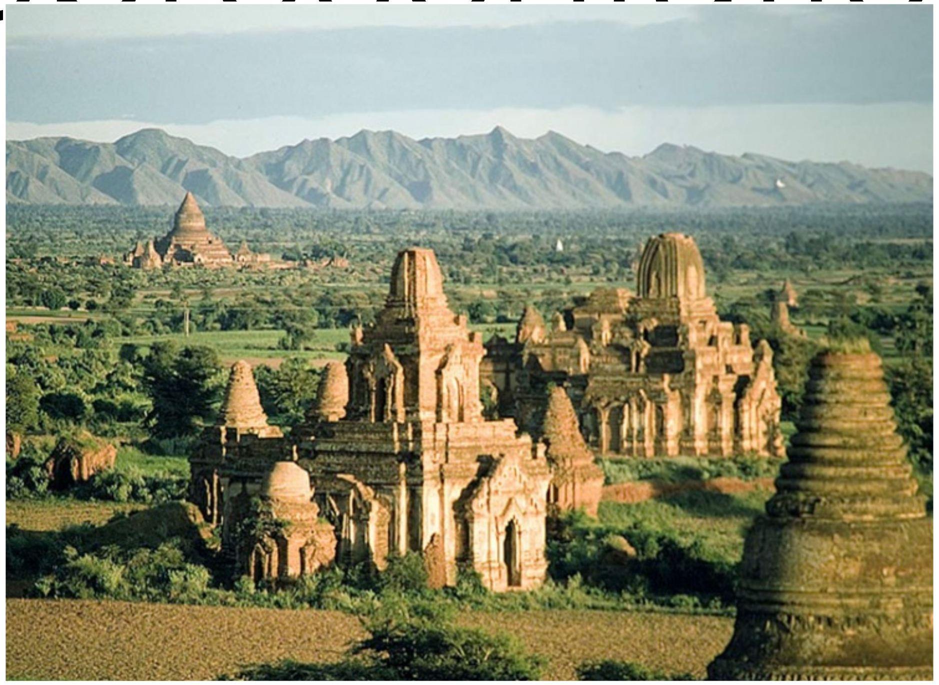
SIFT: Lowe (1999, 2004)

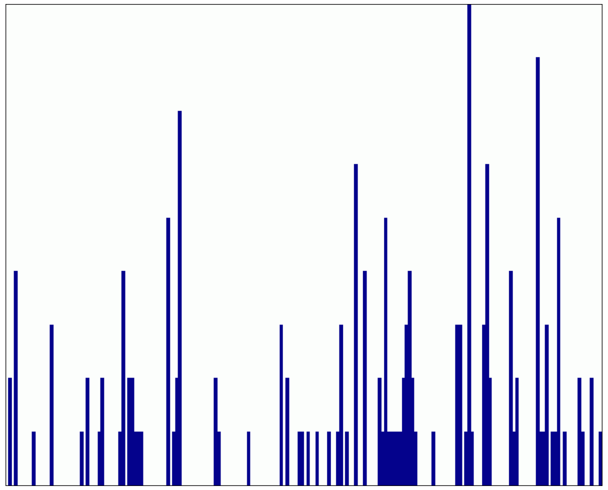
Gist: Torralba et al. (2003)

Slide credit
Svetlana
Lazebnik

Spatial pyramid

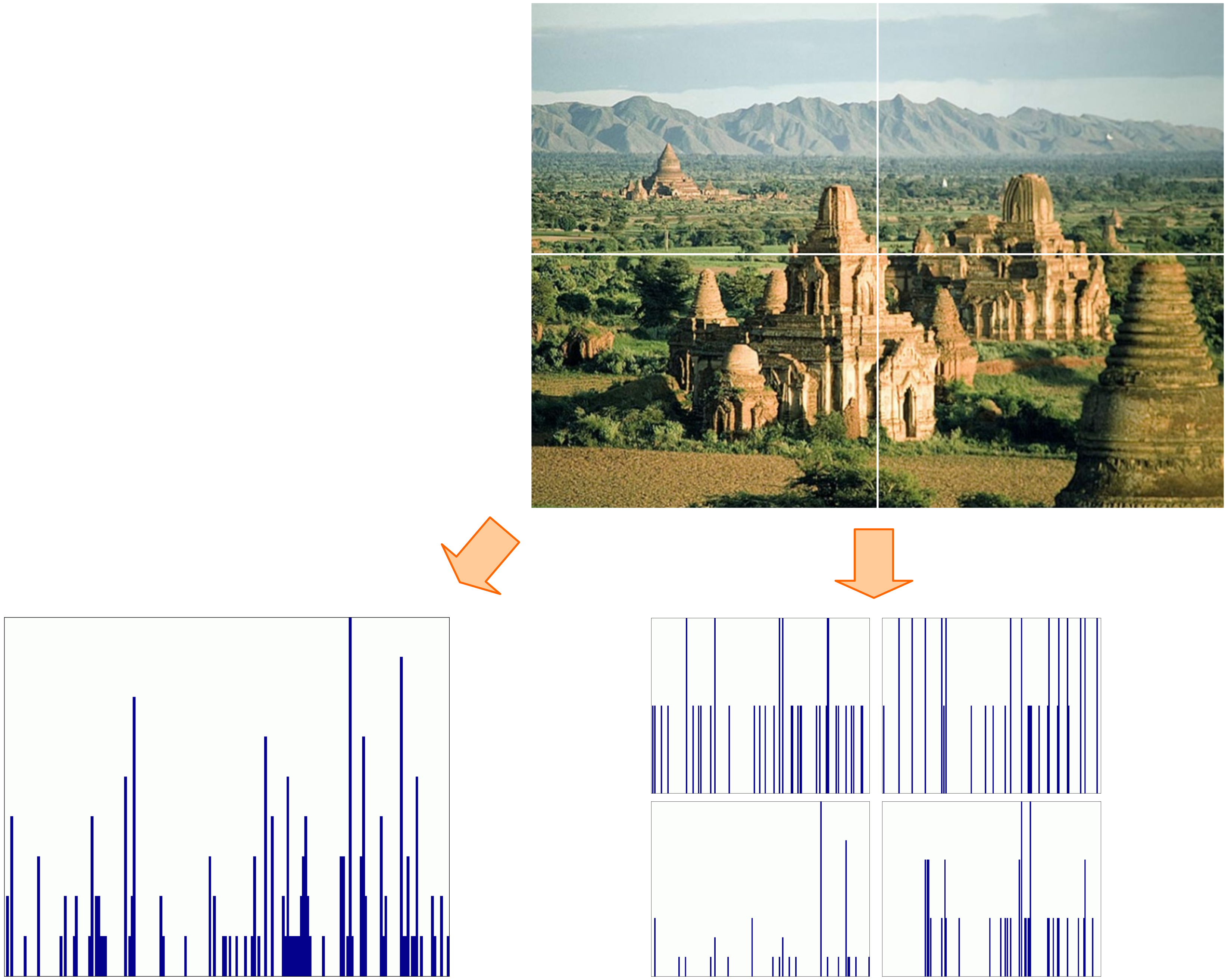
Spatial pyramid representation





Lazebnik, Schmid & Ponce (CVPR 2006)

Spatial pyramid representation



Lazebnik, Schmid & Ponce (CVPR 2006)