Beyond bags of Features

Spatial Pyramid Matching for Recognizing Natural Scene Categories

Camille Schreck, Romain Vavassori

Ensimag

December 14, 2012

Introduction

Overall objective : semantic categorisation.

- Use spatial information.
- Global representation.

State of art

Bags of feature

- Images described as an orderless collection of features.
- Good performances.
- Do not use the information about the spatial layout of the features.

Multiresolution histograms

Subsampling an image and compute a global histogram at each level.

Generalized histograms to locally orderless images

For each Gaussian aperture at a given location and scale, the locally orderless image returns the histogram of image features aggregated over that aperture.

Pyramid Match Kernel

Goal

Find the approximate correspondance between 2 set of vectors, X and Y, in a *d*-dimentional feature space.

Idea

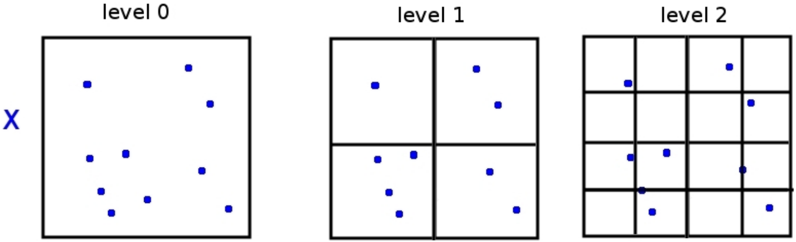
- correpondances are computed at different levels of resolution.
- at each level, a finer grid is set on the space.
- 2 vectors are said to match if they are on the same cell.
- take the weighted sum of all the matches.

Subdivisions of the feature space

We compute matches at different level of resolution 0,..,L.

At the level of resolution l

- the grid is divided in 2^l along each dimension.
- the grid has $D = 2^{dl}$ cells where d is the number of dimensions.



Histograms intersection

Histograms of X and Y

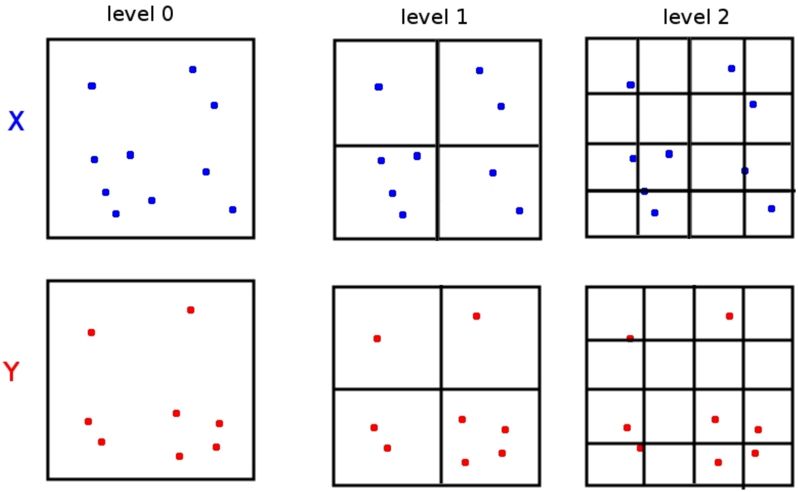
H^l_X and H^l_Y are the histograms of X and Y at level l
where $H^l_X(i)$ is the number of vectors of X in the *ith* cell of the grid.

Histogram Intersection function

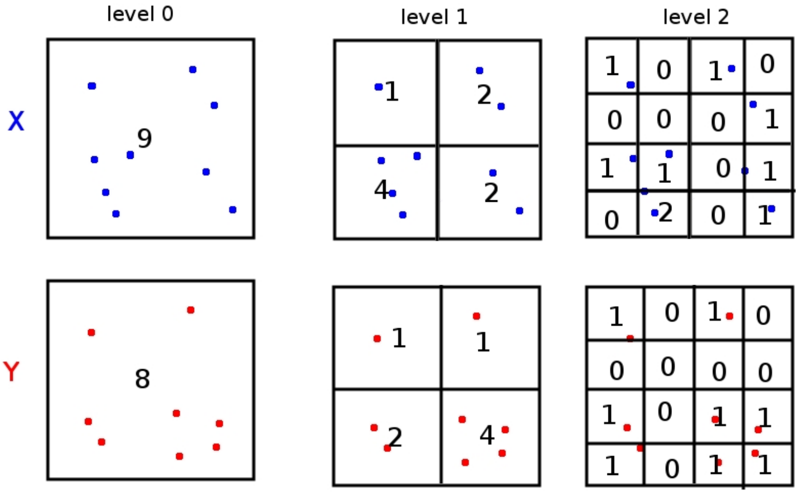
Give the number of matches at level l :

$$\mathcal{I}(H^l_X, H^l_Y) = \sum_{i=1}^{D} min(H^l_X(i), H^l_Y(i))$$

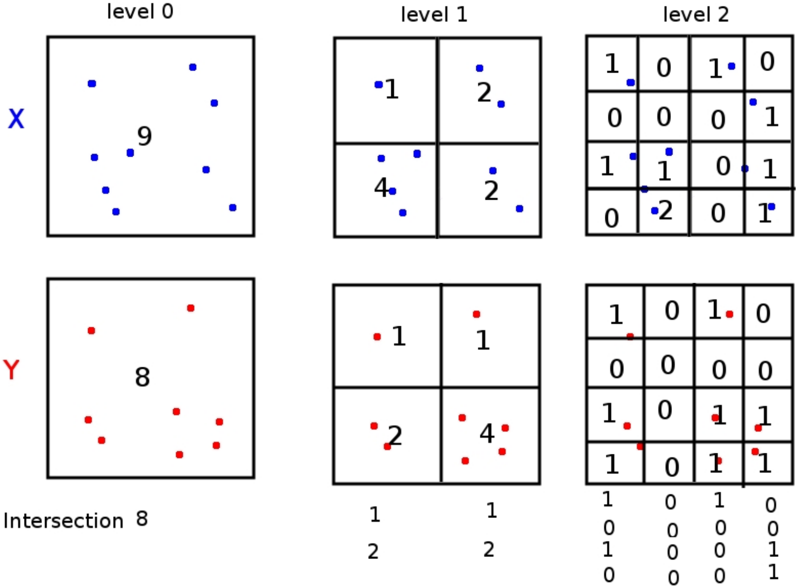
Histograms intersection



Histograms intersection



Histograms intersection



Computation of the kernel

All matches found at level $l+1$ are found also at level l.
$\rightarrow$ Number of new matches at level l is given by $\mathcal{I}^l - \mathcal{I}^{l+1}$. We sum all the matches weighted by $\frac{1}{2^{L-l}}$. The more the grid is coarse, the less the matches are weighted.

pyramid match kernel

Mercer kernel :

$$\kappa^l(X, Y) = \mathcal{I}^L + \sum_{l=0}^{L-1} \frac{1}{2^{L-l}} (\mathcal{I}^l - \mathcal{I}^{l-1})$$

"Orthogonal" approach

Matching of two collection of features in a high-dimensional appearance space

- quantize all feature vectors into M discrete types, giving M channels.
- perform pyramid matching in the 2-dimensional image space for each channel $m = 1..M$.

Assumption

Only features of the same type m can be matched to one another.

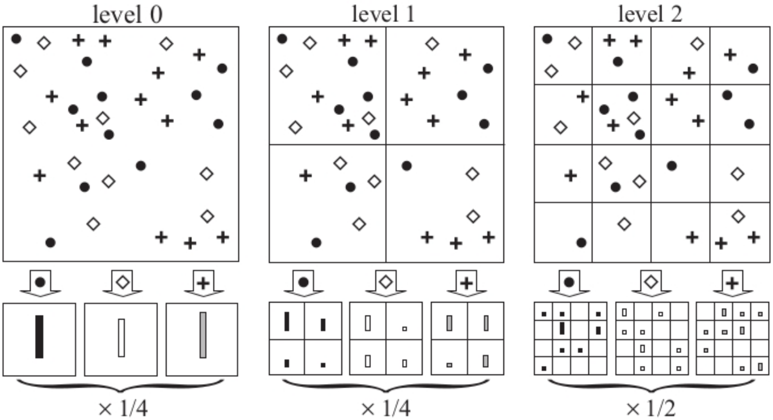
Final kernel is the sum of separate channel kernels

$$K^L(X, Y) = \sum_{m=1}^{M} \kappa^L(X_m, Y_m)$$

where X_m and Y_m are the coordinates of features of type m found in the respective image.

K_L can be computed as the intersection of the histograms obtain by concatenating the histograms of each channel.

Example



Two kinds of features for the experiments

Weak features

Edge points at two scales and eight orientations.
$\rightarrow M = 16$ channels.

Strong features

SIFT descriptors of 16x16 pixels.
k-mean clustering to get a visual vocabulary.
In the experiments vocalubary size is $M = 200$ or $M = 400$.

Summary of the method

- Clustering features from a training set.
- Computation of the “description” of a query image.
- Comparison with the description of each image in test set.

1. State of art

2. Method
   - Pyramids
   - Spatial Matching Scheme
   - Features extraction
   - Summary

3. Results
   - Scene Category Recognition
   - Caltech-101
   - Graz Dataset

4. Conclusion

Scene Category Recognition

	Weak features ($M = 16$)		Strong features ($M = 200$)		Strong features ($M = 400$)	
L	Single-level	Pyramid	Single-level	Pyramid	Single-level	Pyramid
0 (1×1)	45.3 ±0.5		72.2 ±0.6		74.8 ±0.3	
1 (2×2)	53.6 ±0.3	56.2 ±0.6	77.9 ±0.6	79.0 ±0.5	78.8 ±0.4	80.1 ±0.5
2 (4×4)	61.7 ±0.6	64.7 ±0.7	79.4 ±0.3	**81.1** ±0.3	79.7 ±0.5	**81.4** ±0.5
3 (8×8)	63.3 ±0.8	**66.8** ±0.6	77.2 ±0.4	80.7 ±0.3	77.2 ±0.5	81.1 ±0.6

Example

Caltech-101

	Weak features		Strong features (200)	
L	Single-level	Pyramid	Single-level	Pyramid
0	15.5 ±0.9		41.2 ±1.2	
1	31.4 ±1.2	32.8 ±1.3	55.9 ±0.9	57.0 ±0.8
2	47.2 ±1.1	49.3 ±1.4	63.6 ±0.9	**64.6** ±0.8
3	52.2 ±0.8	**54.0** ±1.1	60.3 ±0.9	64.6 ±0.7

Graz Dataset

Class	$L = 0$	$L = 2$	Opelt [14]	Zhang [25]
Bikes	82.4 $\pm$2.0	86.3 $\pm$2.5	86.5	92.0
People	79.5 $\pm$2.3	82.3 $\pm$3.1	80.8	88.0

Conclusion

- “holistic” approach for categorisation.
- Simple method.
- Gives better results than bag-of-features.