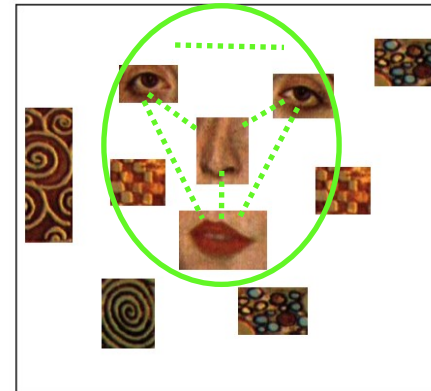
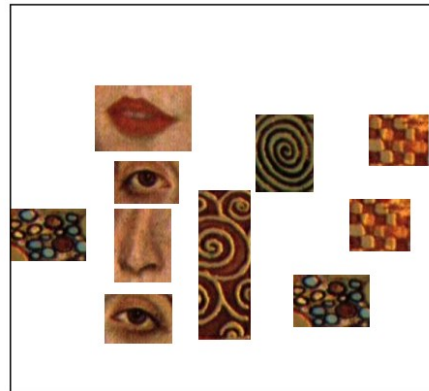
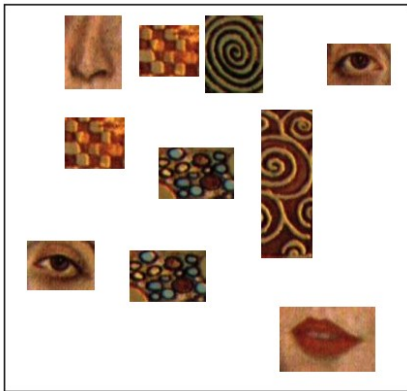


Extending bag of words with locality information

Weakness of the BoW model

- With the bag of words model there is no rigorous geometric information of the objects components. Instead it's intuitive that objects are made of parts... and relations between parts are really informative!

An example ("face" detection):

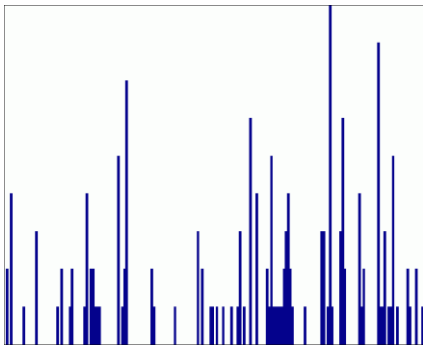


All have equal probability for BoW model

- Segmentation and localization unclear

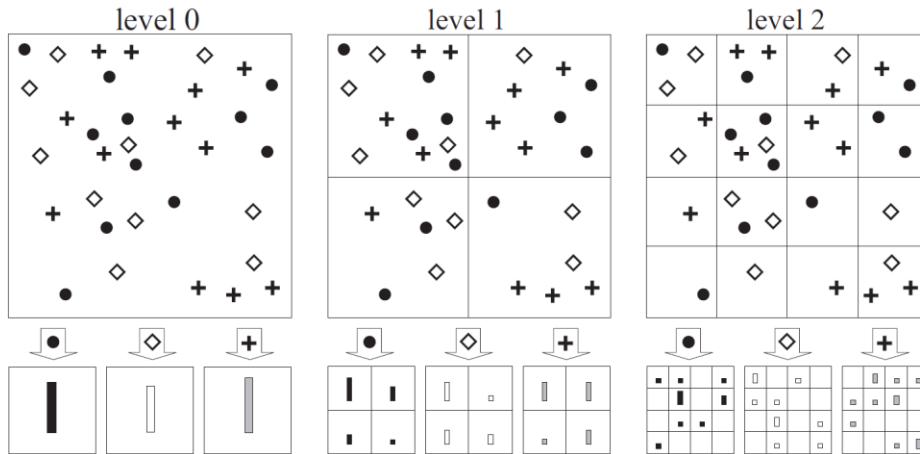
Spatial pyramid representation

- Spatial pyramid is an extension of a bag of features representation that accounts for spatial information of image features
- With BoW there is a locally orderless representation at several levels of resolution



Spatial pyramid formation

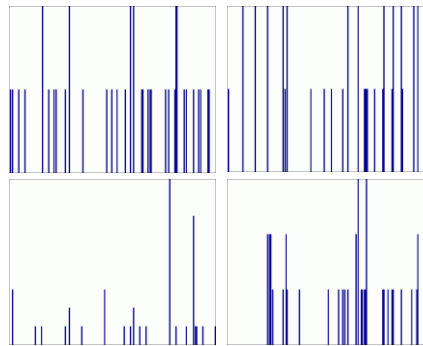
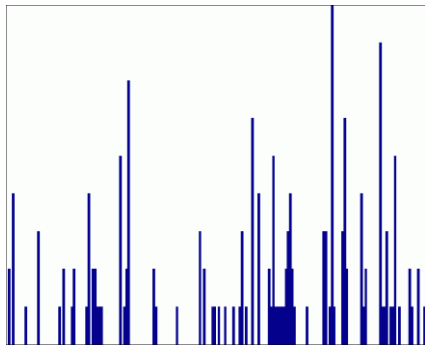
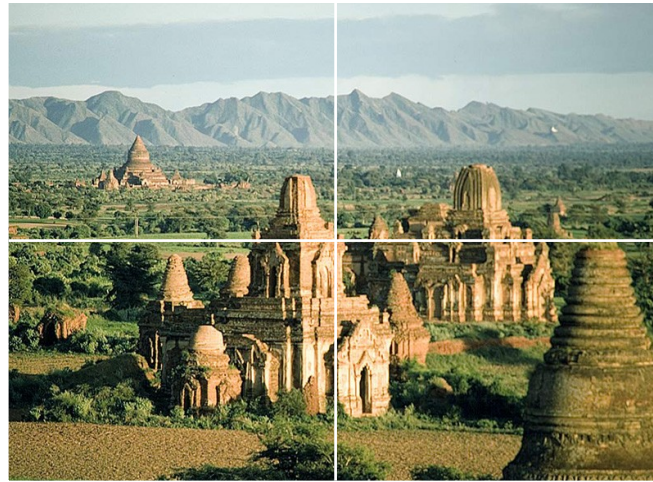
- Spatial pyramid formation:
 - Partition the image recursively.
 - Accumulate visual word counts separately



Toy example with a 3-words dictionary and three levels.

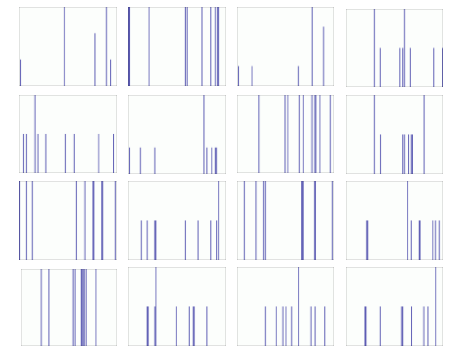
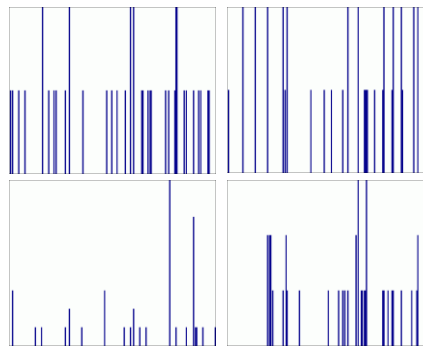
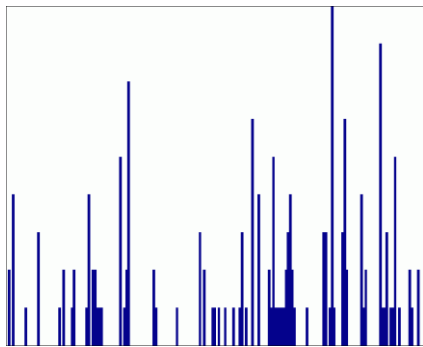
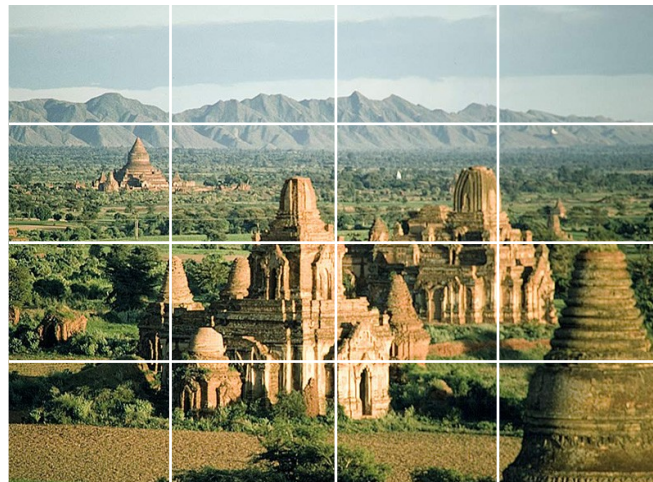
Spatial pyramid representation

By performing spatial partitions and taking histograms for each of them



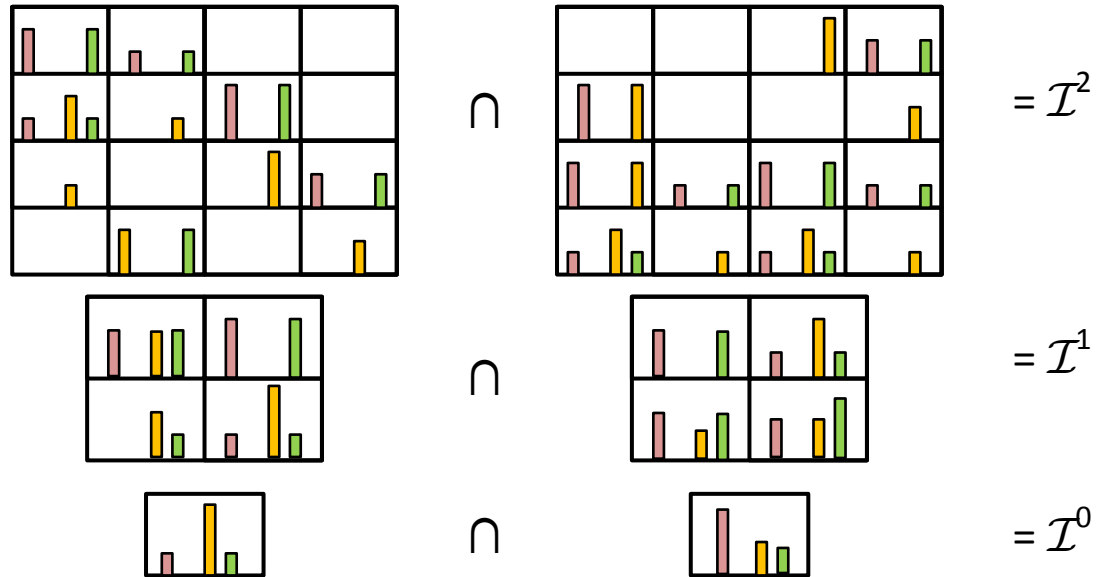
Spatial pyramid representation

By performing spatial partitions and taking histograms for each of them



Spatial pyramid matching

- Once spatial pyramids have been formed, matching can be performed following a strategy inspired by pyramid matching kernel [Grauman&Darrell]. In this case at each level of the pyramid the number of elements is the same.



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

- At each level spatial configuration detail importance is increased.
- Matches are only counted once
- Level 0 equals standard Bag-of-Words