Visual Category Recognition using Information-Theoretic Co-clustering

Ashish Gupta and Richard Bowden

University of Surrey

a.gupta@surrey.ac.uk

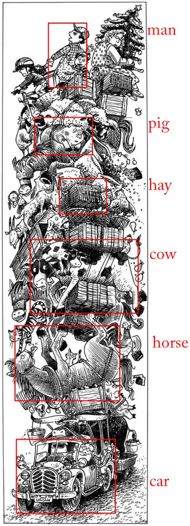
Feb.25,2012

Contents

- Introduction: Visual Categorization
- Current practice: Bag-of-Features Model
- Problem: Semantic scatter
- Idea: Visual Topic $\leftarrow \sum$ Visual Words
- Approach: Information-Theoretic Co-Clustering
- Solution: Bag-of-Topics Model
- Experiments: multiple Datasets and Dictionary sizes
- Summary

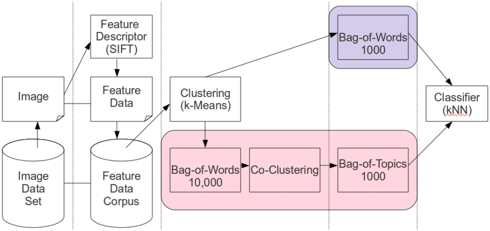
UNIVERSITY OF SURREY

Visual Category Recognition



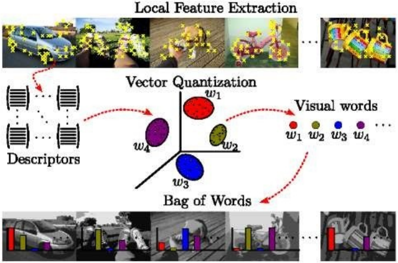
Definition

Detect presence of an instance of a visual category in an image.



Visual Categorization Pipeline

Bag-of-Features



Visual Word

Representative feature vector (generally centroid) of each cluster.

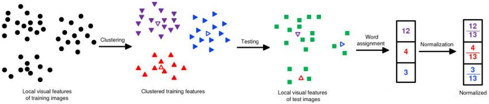
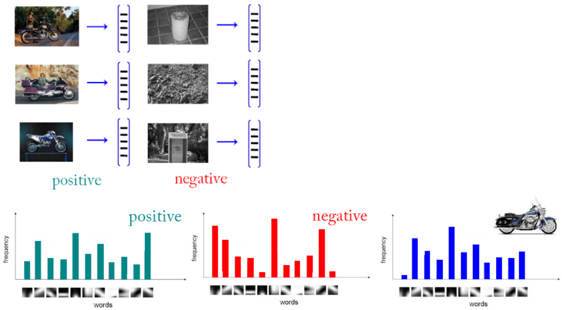


Image Histogram

Histogram of assignments of image feature vectors to visual words.

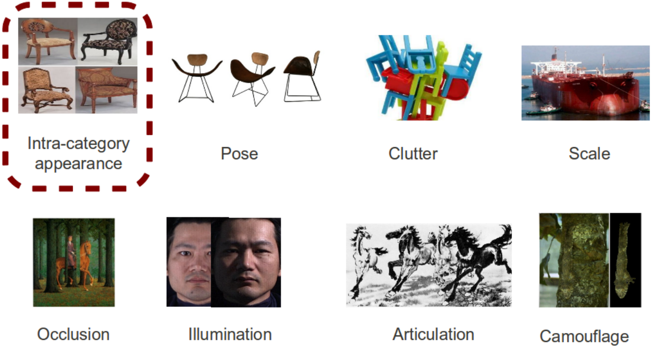
Discriminative Dictionary



Basic Intuition

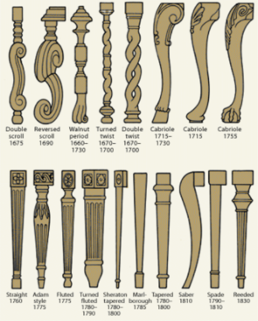
If normalized image histogram of test sample is closer to trained positive histogram then image is classified as positive (motocycle).

Challenges



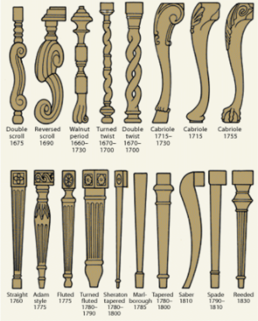
Several variations in visual category appearance render category recognition very difficult.

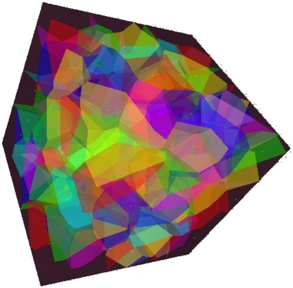
Intra-category appearance variation



- Small variations in instances of object part causes associated descriptors to get scattered in feature space.

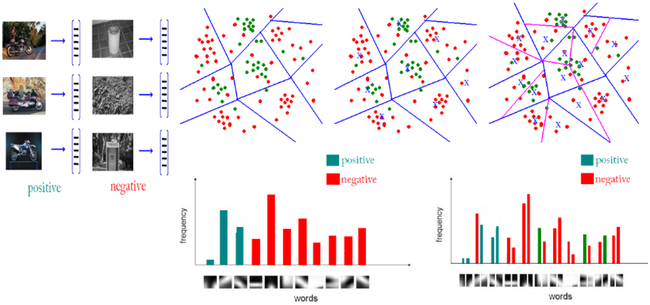
Semantic Scatter





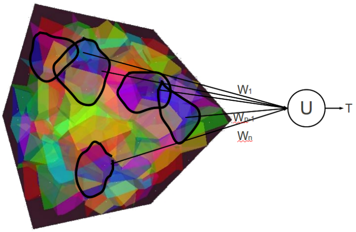
- Small variations in instances of object part causes associated descriptors to get scattered in feature space.
- Descriptors of different categories are inter-leaved in feature space (local patch decriptors have low specificity).

Towards Discriminative Dictionary



- Increase dictionary size to split up interleaved feature vectors.
- Dictionary learning is unsupervised $\Rightarrow$ discriminative dictionary not assured.

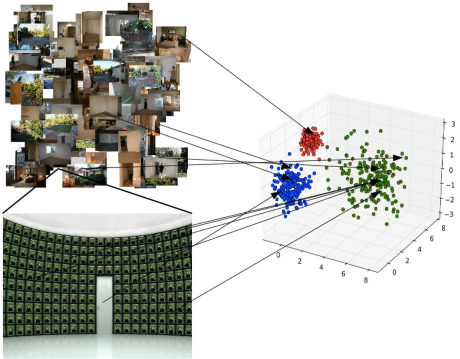
Visual Topic $\leftarrow \sum$ Visual Words



Idea

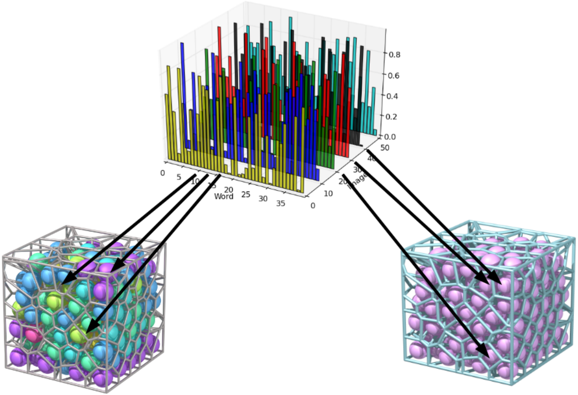
- Combine visual words which are semantically related and create a visual topic.
- Dictionary of visual topics should be more robust to intra-category variation.

Visual Word Distribution



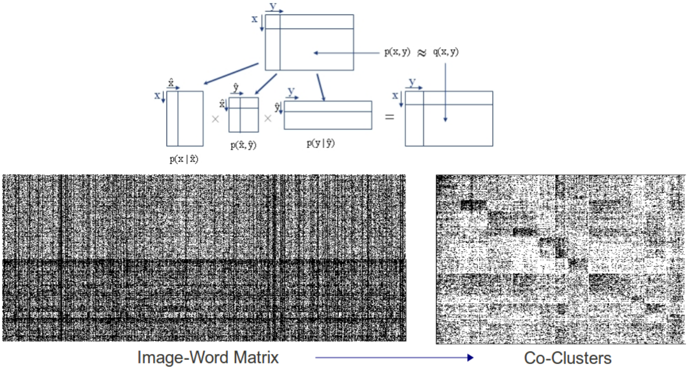
- Visual word selection is based entirely on distribution density of feature vectors independent of image source.
- Distribution across images important for selection of visual word.

Image-Word Distribution

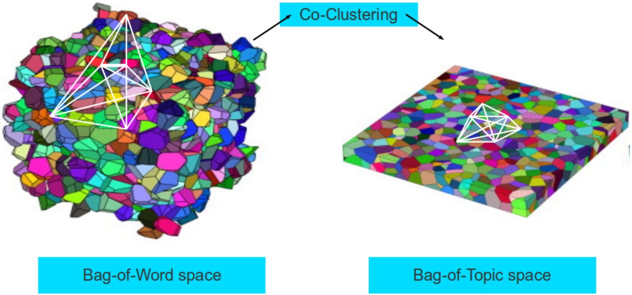


- Word distribution across images $\propto$ semantic equivalence.
- Histogram distribution across words $\propto$ image similarity.

Co-Clustering



- Formulate the image-word matrix as a joint probability distribution.
- Optimal co-clustering maximizes mutual information between clustered random variables.

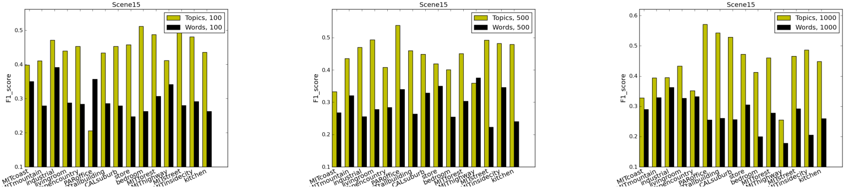


- Image histogram feature vectors in high-dimensional visual words space are projected to lower dimensional visual topic space.
- The distance between feature vectors from the same category is reduced.

Experiment

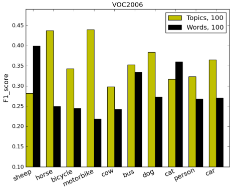
- Feature Descriptor : SIFT : Affine invariant local image patch descriptor
- Data set:
  - Scene 15 : 15 visual categories of natural indoor and outdoor scenes. Each category has about 200 to 400 images and the entire dataset has 4485 images.
  - Pascal VOC2006 : It has 10 visual categories with about 175 to 650 images per category. There are a total of 5304 images.
  - Pascal VOC2007 : It has 20 visual categories. Each category contains images ranging from 100 to 2000, with 9963 images in all.
  - Pascal VOC2010 : It has 20 visual categories with 300 to 3500 images per category. There are a total of 21738 images.
- Classifier : k-NN

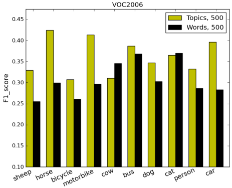
Scene-15 Dataset

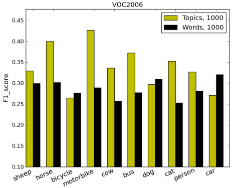


F1-score across categories for 100, 500, and 1000 visual Topics.

PASCAL VOC2006 Dataset

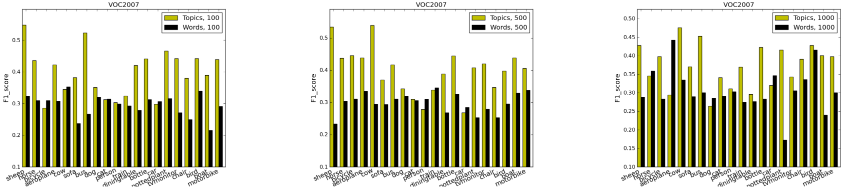






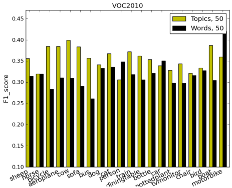
F1-score across categories for 100, 500, and 1000 visual Topics.

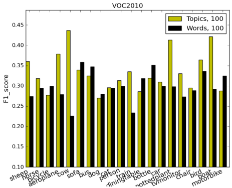
PASCAL VOC2007 Dataset

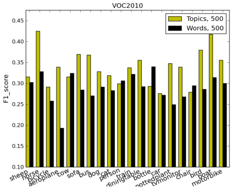


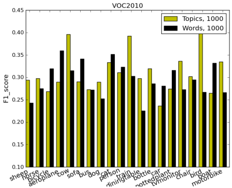
F1-score across categories for 100, 500, and 1000 visual Topics.

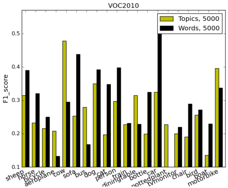
10,000 words $\rightarrow$ n Topics. Appropriate number of Topics?











Large dictionary becomes category dependent.

UNIVERSITY OF SURREY

Summary

Bag-of-Features in limited: unsupervised clustering.

Significant intra-category appearance variation: semantic scatter.

Visual Topic $\leftarrow \sum$ Visual Word.

Co-clustering Image-Word distribution.

Semantic dimensionality reduction.

END