Image Classification using BoVW

A Brief Introduction

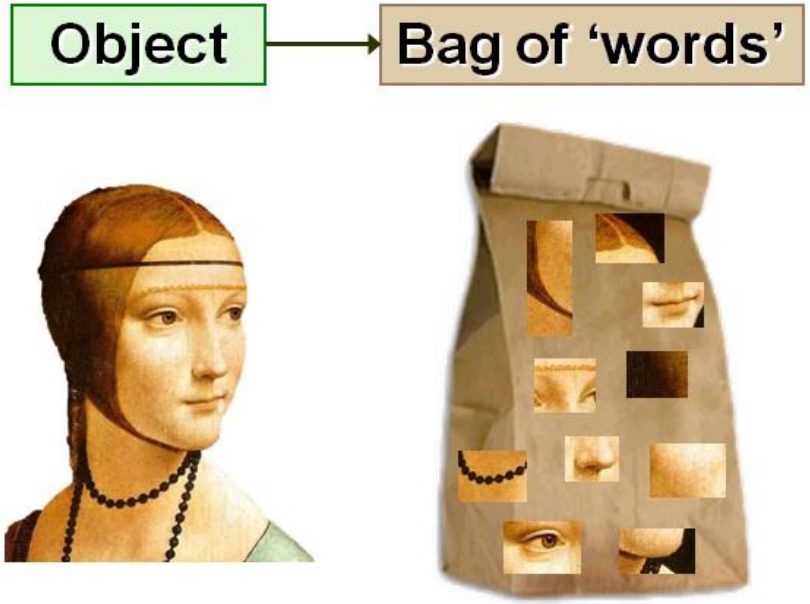
FH Salzburg

Bag of Visual Words

Bag of Visual Words

FH Salzburg

- Commonly used in image classification
- Concept adapted from NLP‘s bag of words
  - Counting appearance of each word in document
  - Generating frequency histograms
  - Each document is treated as a bag of word
- BoVW uses image features as „words“



Feature Extraction - Idea

FH Salzburg

Feature Extraction - Solution

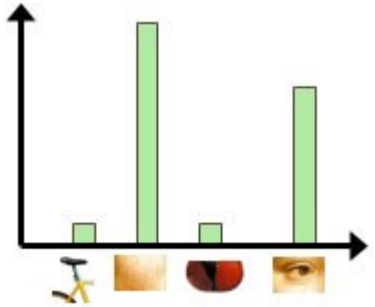
FH Salzburg

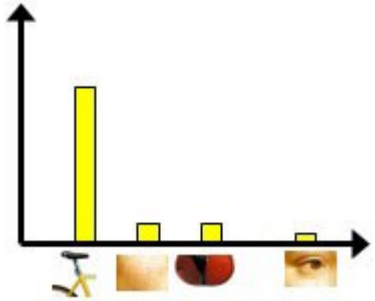
- Simple Features (not good enough for us)
  - Image Brightness
  - Grayscale Histograms
  - RGB Histograms
  - Etc.
- Commonly used feature extraction
  - Harris Corner Detection
  - SIFT (Scale-Invariant Feature Transform)
  - SURF (Speeded-Up Robust Features)
  - FAST (Features from Accelerated Segment Test)
  - Etc.

Histogram of Visual Words

Histogram of Visual Words - Idea

FH Salzburg

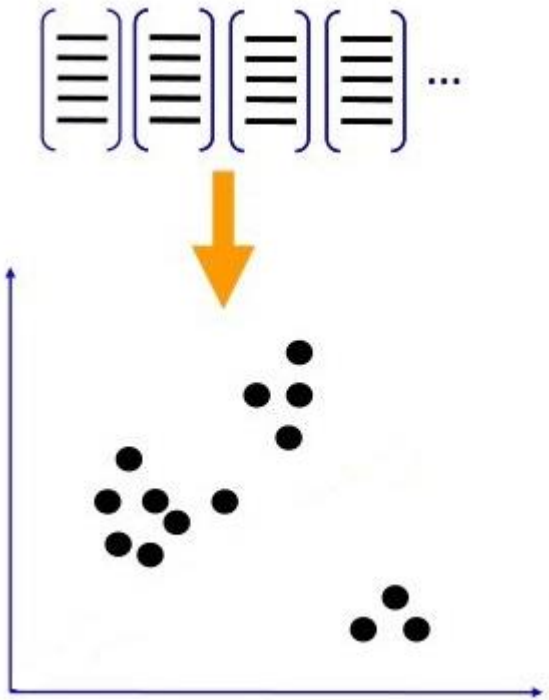




Histogram of Visual Words - How?

FH Salzburg

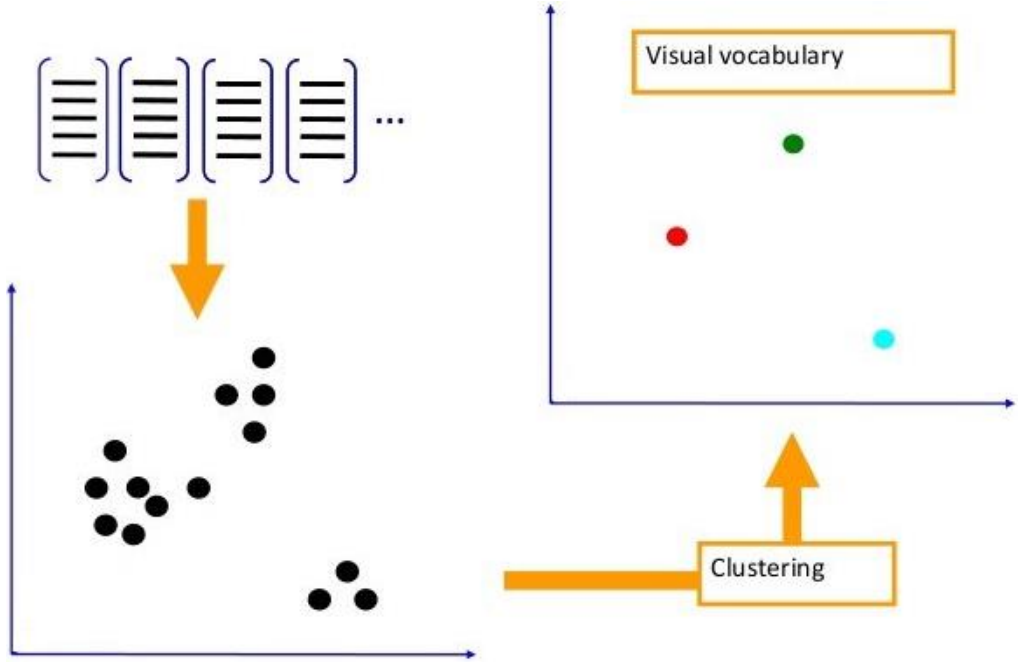
- Extract local features from your training set using SIFT
- Extracted descriptors build point clouds in a hyperdimensional space
- Use K-Means to find clusters (quantization of the feature space)



Histogram of Visual Words - How?

FH Salzburg

- Center points found during clustering are our visual words
- Go through the training set and assign every descriptor to one cluster
- Count the appearance of each cluster in every image



Classification

Classification - How?

FH Salzburg

- Training:
  - Use the histograms of visual words of each image to train your model
- Testing:
  - Extract local features from your test set using SIFT
  - Create a histogram of visual words for each image from the test dataset
  - Use the computed histograms for prediction