

COMPUTER VISION 2023

LAB 5

TABLE OF CONTENTS

Bag of words	2
Bag of visual words	2
Bag of visual words for image classification	3
Bag of visual words coding steps	5
Reference.....	5

BAG OF WORDS

- The bag-of-words model is a simplifying representation used in natural language processing and information retrieval (IR). In this model, a text (such as a sentence or a document) is represented as the bag (multiset) of its words, disregarding grammar and even word order but keeping multiplicity.
- The bag-of-words model is commonly used in methods of document classification where the (frequency of) occurrence of each word is used as a feature for training a classifier.

VISUAL BAG OF WORDS

- In **computer vision**, the **bag-of-words** model (BoW model) can be applied to image classification, by treating image features as words. In document classification, a bag of words is a sparse vector of occurrence counts of words; that is, a sparse histogram over the vocabulary. In computer vision, a bag of visual words is a vector of occurrence counts of a vocabulary of local image features.



face, flowers, building

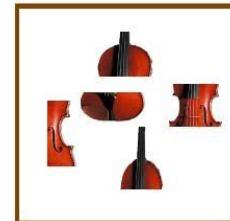
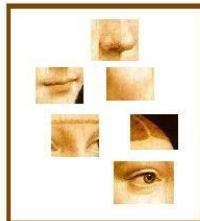
- Bag-of-words models have been useful in matching an image to a large database of object instances.



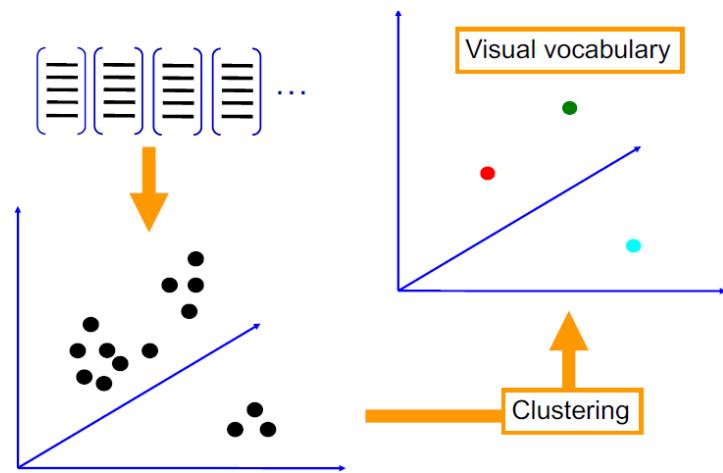
VISUAL BAG OF WORDS FOR IMAGE CLASSIFICATION

- First, take a bunch of images, extract features, and build up a “dictionary” or “visual vocabulary” – a list of common features.
- Given a new image, extract features and build a histogram – for each feature, find the closest visual word in the dictionary.
- Bag of features : outline

1. Extract the common features from the dataset (HARRIS, SIFT, or SURF)



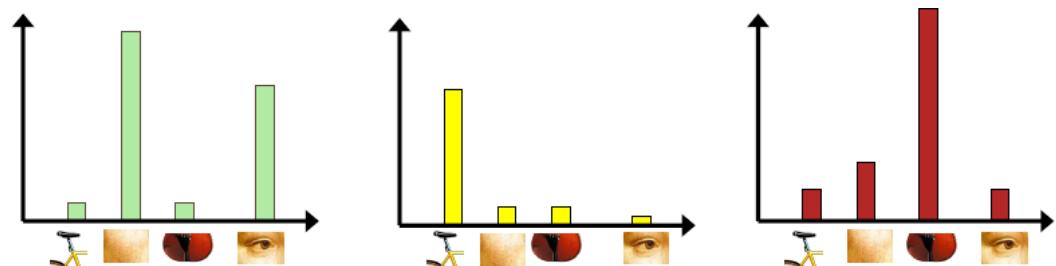
2. Learn visual vocabulary (Clustering)



3. Image representation



4. Image classification (Given the bag-of-features representations of images from different classes, how do we learn a model for distinguishing them).



BAG OF VISUAL WORDS CODING STEPS FOR IMAGE CLASSIFICATION

- Read the training images & their labels.
- Compute the features of each image using SIFT.
- Group the descriptors of the images to **n clusters** using K-mean algorithm (Each **cluster** denotes a particular **visual word** and **every image** can be represented as a combination of multiple visual words (**Histogram**)).
- Generate a sparse histogram that contains the frequency of occurrence of each visual word (assign each visual word to one of the n clusters).
- Thus, the vocabulary comprises of a set of histograms of encompassing all descriptions for all images.
- Normalize the histograms of the vocabularies.
- Train a classifier on these histograms.

What should I do in the testing phase ??

REFERENCE

- [1] https://en.wikipedia.org/wiki/Bag-of-words_model_in_computer_vision
- [2] http://vision.stanford.edu/teaching/cs131_fall1718/files/14_Bow_bayes.pdf
- [3] <https://medium.com/analytics-vidhya/bag-of-visual-words-bag-of-features-9a2f7aec7866>
- [4] <https://www.analyticsvidhya.com/blog/2019/09/feature-engineering-images-introduction-hog-feature-descriptor/>