Image Classification Analysis with Bag of Word Model

Lorenzo Cioni
lore.cioni@gmail.com

Saverio Meucci
s.meucci91@gmail.com

Abstract

In this paper, we evaluated the performance of Nearest Neighbour and Support Vector Machine as images classifier, using Bag of Word model. Different metric functions and different kernel functions were taken into account. With regard to features quantization, we have focused our attention to soft-assignment and truncated soft-assignment, in comparison to hard-assignment.

Keywords - Bag of Words (BoW), SIFT, SVM, Nearest Neighbours, Hard-Soft Assignment.

1. Introduction

In computer vision, the bag-of-words model is a well established image classification method. Comparing two image by matching each local features of the images is computationally intensive; not only, different images have different number of local features. BoW model simplified this process by building a vocabulary of visual-words, called a codebook. This way, an image descriptor is a histogram built by counting how many codewords are present in an image.

Given this model for image classification, we have compared how different classifiers performs. We have studied the Nearest-Neighbours classifier, comparing the results using both the euclidean distance and the chi-square distance. Also, we have examined the Support Vector Machine classifier, implementing different kernel functions; in particular, the linear kernel, the intersection kernel, the chi-square kernel, and the radial basis functions kernel have been implemented. In evaluating the performance of a classifiers, many parameters, that can vary the results, must be taken into account. For instance, we have also compared the results varying parameters like the density or sparseness of the key-points in each image, the dimension of the vocabulary and the dimension of the training set used to train the classifiers.

Another aspect that can influence the performance of a classifier, using the BoW model, is choosing between hard-assignment and soft-assignment. Briefly, with hard-assignment each local features of an image is assigned to one codeword; with soft-assignment each local features is assigned to every codeword of the vocabulary, but with some weigths that give a measure of how strong is each assignment. We give more details in a later section of this paper.

An image descriptor is computed by following this work flow:

1. given an image, dense or sparse local features are extracted and described using SIFT descriptor;
2. using every SIFT descriptor of all images, a vocabulary of visual words is built by the clustering algorithm k-means; the centroids of each clusters will be the codewords of the vocabulary;
3. given an image, for each SIFT descriptor, features quantization is performed using hard-assignmend or soft-assignment, based on the computed codebook;
4. finally, for each image, a histogram is built, with each bin stating how much of the corresponding codeword is present in an image.

Once the histograms are computed for each images of a given dataset, the image descriptors are ready for the classifiers and we can proceed with a training phase.

In the following sections we give a better insight behind features extraction, vocabulary creation, and features quantization.

2. Feature extraction

which describes the features employed and their properties.

- features globali non precise per la classificazione - uso features locali; come le estraggo? - sparse, uso un keypoints detector come harris laplace - dense, prendo keypoints uniformi sull'immagine - Multi-Scale Dense Sampling, prendo keypoints densi a scale diverse dell'immagine. - estratti i keypoints uso sift descriptor per descrivere ogni features locale dell'immagine - perch sift? propriet di invarianza, cosa prendono in considerazione di un'immagine.

- per ogni immagine estraggo i keypoint; posso farlo in modo sparso, ovvero uso un keypoints detection, oppure

in modo denso, ovvero prendo i keypoints uniformemente distribuiti nell'immagini. Terza opzione multiscale sparse etc. - dopo aver estratto i keypoints uso SIFT descriptor per descrivere ogni local features di un'immagine. - breve descrizione sift, propriet di invarianza ecc.

3. Image representation

4. Image classification

5. Experimental results

6. Conclusions