Laboratory 11 “Computational Vision”

Laboratory 11: Bag of words (BOW) for object recognition

In this laboratory, we will practice how we can recognize objects in images using the Bag-of-words (BOW) method. The main issues are:

1) BOW features
2) Vocabulary
3) Spatial histograms
4) Support Vector Machines

   To complete the laboratory it is necessary to know and apply the basic concepts on BOW for Object recognition. For this, refer to the material of the theoretical slides.

11.1 Object recognition by BOW. The VLFeat library contains a code for BOW object recognition called phow_caltech101.m (`vlfeat-0.9.20/apps)`. The goal of this exercise is to explore how the BOW method works and how it depends on its parameters.

a) Download the Caltech101 database from Campus Virtual and store it in a subdirectory named data. Run step by step and explore the BOW object recognition algorithm given by phow_caltech101.m. Assure that you are working with a subset of categories and that the model is trained and tested. Add a function to visualize the images in a green framework if the classification is correct and in a red framework if it does not (see Fig.1).

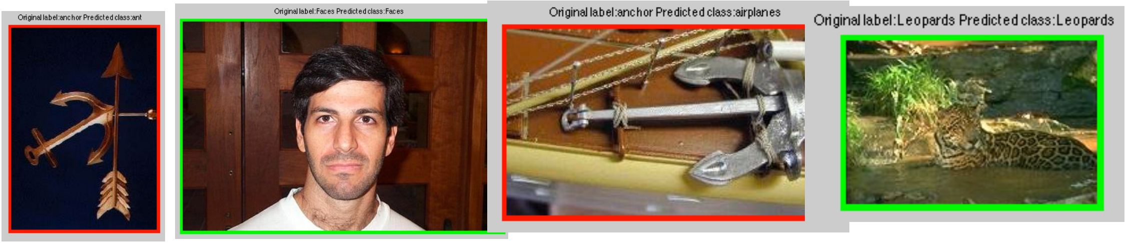
Original label:anchor Predicted class:ant
Original label:Faces Predicted class:Faces
Original label:anchor Predicted class:airplanes
Original label:Leopards Predicted class:Leopards

Fig.1 Correct (in green) and wrong (in red) object recognition

b) What are the PHOW descriptors? What does the Sizes parameter mean? What does happen if the Sizes parameter is augmented? What does the Step parameter mean? What does happen if the Step parameter is augmented?

c) What are the words in the algorithm? How are they extracted? What is their dimension? How their number does affect the accuracy of the results?

d) To construct the vocabulary, the method uses the k-means algorithm. Explore and explain how the k-means is applied, which parameters are used, what is there meaning and what are the advantages and the limitations of the method.

http://www.vlfeat.org/applications/apps.html#apps.caltech-101

e) The function getImageDescriptor gives the spatial histogram of the image. What is the spatial histogram? What dimension does it have? What happens if we augment model.numSpatialX from 2 to 4?

f) The SVM classification with homogeneous kernels is performed by the vl_svmtrain function that needs the homogeneous kernel map obtained by the vl_homkermap function. What are the parameters obtained after the training process? What dimension do they have? How do you apply them to obtain the score of the final classification?

Note: Interested students about the calculation of the kernel map can refer to the paper: A. Vedaldi and A. Zisserman `Efficient Additive Kernels via Explicit Feature Maps', Proc. CVPR, 2010.

g) Is the algorithm invariant to rotation of the images? Is the algorithm invariant to rescaling of the images?

h) Compute how the F-score changes when the number of categories augment (e.g. from 5 to 10 to 15 to 100).

i) Compute the F-score when different parameters of the algorithm are used.

Help: The F-score is given by the averaged F-score of each class. An F-score of a binary classification is given by the formula:

F1=2*(precision*recall)/(precision+recall), where

Precision=TP/(TP+FP), Recall=TP/(TP+FN), TP – true positives, TN – true negatives, FP – false positives, FN – false negatives.

Delivery of the laboratory

Deadline: 10 January at 23:50h.

This part is to be delivered by a stored in a compressed file named "names_surnames_P11.zip" containing: - .m files with the created functions and a pdf file explaining the implementation and the performance of the algorithms and the answers raised in the exercises.

Note: If you don’t have it installed, you can download the VLFeat from Campus Virtual or http://www.vlfeat.org/index.html. To install it in Matlab, you should run:

```
% Install the VLFeat toolbox
srcPath=cd('toolbox');
vl_setup;
cd(srcPath);
close all;
```

http://www.vlfeat.org/applications/apps.html#apps.caltech-101