Practice 8: PHOW

Oscar Francisco Trujillo Puentes
Universidad de los Andes
Carrera 1 # 19-27
of.trujillo10@uniandes.edu.co

Abstract

1. Introduction

Recognizing an object or characteristics of a specific group within an image, and being able to assign it to a specific category, is one of the most studied areas of artificial vision and is defined as classification. The classification seeks to assign a category-specific label to an image according to the objects present in it.

Dedicated to the importance of classifying a group of images, algorithms and databases have been developed that allow approaching the solution of the problem. In this way, the Caltech 101 databases are used in the present laboratory, which contains 101 different categories of images and a sub-sample of ImageNet, with 200 different categories.

In addition, the PHOW algorithm was used, with which a series of histograms is obtained in the present information of the image as is the case of the characteristics of each object present within the image. This algorithm was performed in matlab with the help of the VLfeat library.

Next, the methodology made with materials and an explanation of the function is presented, then the results obtained in each of the databases varying the parameters, then an analysis of the results obtained and ends with a series of conclusions.

2. Methodology

2.1. Caltech 101

Caltech 101 is a database built by by Fei-Fei Li, Marco Andreetto, and Marc 'Aurelio Ranzato in 2003. This base consist of 9146 images of objects divide in 101 categories. Each categorie has between 40 and 800 images (the most common amount is 50 images); therefore, it is unbalanced. Finally the size of each images is 300 x 200 pixels. For this lab, the entire database was used. [2]

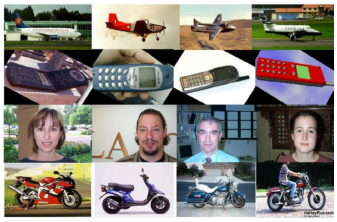

Figure 1. Examples of database Caltech 101

2.2. ImageNet

ImageNet project is a large batabase with different visual elements presented in CVPR 2009 by researchers from the Computer Science department at Princeton University. This database consist of 14 million of images, hand annotated to indicate the objects in the picture; moreover, some images has delimiter boxes in the objects; for example, presents anotations as 'this image has a tiger' or 'this tiger hasn't a tiger'. For this lab, a sample of the database with 200 categories was used. [1]

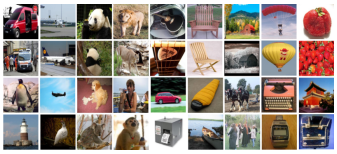

Figure 2. Examples of database ImageNet

2.3. PHOW

Pyramid Histogram of Visual Words (PHOW) is an extension to the bag of words model, the features images are treated as words. The idea is generate a sparse vector of the frequency of image features and search the repeated. The difference of PHOW with BoW is that PHOW take the spacial information of the image. [3]

In the method of PHOW; first, the image is divided in a regular grid or divide in a fine sub regions, which are called pyramids. Next, extract the Dense Scale Invariant Feature Transform (SIFT) descriptors, the visual features of image. However, the SIFT desciptors is necesary asign in

a dictionary of characteristics, for this reason, it is necesary a clustering method that groups the data, in this case K - Means. [3]

The second stage, the visual word histograms (HOW) are generated. This histogram is constructed by the dictionary of the cisual words, where it is sought to assign each characteristic to the closest visual word. In this way, a classifier is generated by Support Vector Machine (SVM). The training is done with the histories already made, determining the one that maximizes the margin of separation between the two classes, since several categories apply multi SVM. [3]

For this practice, the different categories of the database were classified using PHOW. In this case, the vlfeat library of Matlab was used, which allowed to classify the different categories 101 in Caltech 101 and 200 in ImageNet200. Making use of the function of phow caltech 101 different parameters were modified with which the changes in the Acurracy could be observed, as it is the case of the number of images, classes, among others.

3. Results

The results obtained by using the function phow_caltech101 () are presented, which downloads the Caltech 101 database and allows its classification by means of the variation of parameters such as the number of images in test and training, the number of categories , the number of SVM that are trained among others. First, the results obtained are shown for the default values in the function.

Parameters	Values
# Train images	15
# Test images	15
# Classes	5
# Spatial X	2
# Spatial Y	2
# Words	300
# SVM	300

Table 1. Default parameters for Tiny problem

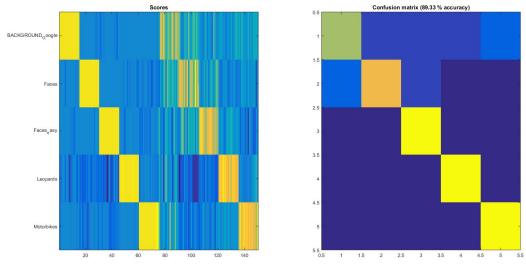

Figure 3. Confusion matrix for the default parameters

With the default values, we have an Acurracy of 89.33%. In this way, variations of the parameters are made and the condition of Tiny problem is deactivated.

Changed parameter	Values	Acurracy
# Train images	45	93.33%
# Classes	15	69.11%
# Classes, # Spatial X/Y	51, [2 4]	59.08%
# Train images, # Classes, # Spatial X/Y,	30, 102, [3 5]	57.52%

Table 2. Results with the change of parameters for Caltech 101

Now, the results obtained using the ImageNet200 database are presented, with which the following results were obtained with the variation of parameters. The initial parameters are 100 image for train, 100 image for test, 200 number of classes and [2 4] for Spatial X/Y. With this parametres, the acurracy was 0.00

Changed parameter	Values	Acurracy
# Train images	200	0.00%
# Classes	5	24.00%
# Classes, # Spatial X/Y	5, [3 5]	11.40%
# Words, # Spatial X/Y,	800, [3 5]	5.52%
# Classes, SVM, # Words	5, 15, 600	22.40%

Table 3. Results with the change of parameters for Image Net 200

Within the classification process, the classes that seem easiest to classify are:

- Wood Spider
- White Wolf
- Zebra
- Web Site
- Bookcase

Within the classification process, the classes that seem most difficult to classify are:

- Bedlington terrier
- Chihuahua
- Scottish deerhound
- Bull Frog
- Carbonara

4. Discussion

from the variation of parameters for the function of clatech 101 you can notice changes in the Acurracy of the function, the most relevant is the change of the number of categories, in front of a smaller amount you get better results of classification, by increasing the number of classes a reduction was obtained; however in the case of caltech 101 it is lowered to 57.2 % for all categories, being a good result for what is expected or chance.

On the other hand, by increasing the number of train images, you can also see a better result since the amount of information is increased where you can present variation or information that was not previously considered. However, by increasing the number of test images that are equal to those of training reduces the Acurracy, evidencing deficiency or variation in other test data. These changes modify the results between 5 to 10 %. Within the variation for the SVM number and the spatial number X / Y, no noticeable or significant change was noticed; additionally, the function has internal paramenters as the number of bins that can not be modified.

In the case of the databases, the difference between the two results could be evidenced, if the complete database is used in the case of Imagenet, it has a result of 0 and in the case of Caltech 101 it has 57.52 %, this makes notice the difference between the two results. You can start by the number of categories that is almost double reducing the probability of hitting. However, when comparing the number of classes the number in ImageNet is better 24 % for 5 categories and 93.33 % in caltech 101; in this way, it is evident that the code is designed for the Caltech database

On the other hand, the number of words parameter has an increase effect in the ACA, as it increases. This is because the descriptor of the images depends on the number of words, therefore, a greater number of words will obtain a more accurate descriptor of the images, which allows a more discriminative classification.

In the case of the most easy to classify categories and the most difficult, it is because the SIFT form descriptors obtain comparators that allow them to be discriminated against the other classes, it can be said that they have the most defined patterns to define the object . On the other hand, in the case of the most difficult due to its low performance and because the images have a high resemblance as is the case of breed of dogs. Difficulty training the classifier and generating similar values for the SVM

Finally, the function can improve if the paramentros optimos that allow to improve the results; however, it would be an iterative process. Thus, the code could be improved by applying a previous classifier or allowing to obtain an amount of information such as a bank of filters and textones. Additionally, the SVM could be changed by Random Forest, this allows a better performance for the case of multi category.

5. Conclusions

- The performance of PHOW depends on the selection of the hyperparameters, in order to obtain better results, an analysis of the parameters of the function must be performed, as is the case of the number of categories, images used for training and one of the most relevant is the number of words, it should be noted that the model changes according to the databases.
- It was determined that the most influential parametors are the number of training images, and the number of visual words, other parameters such as the number of test images or the number of classes are evident but in the time of the practice or research can not be changed just for better results.
- It was possible to observe the difference between both databases when evaluating the classification function, in the case of the images for ImageNet the best results were not obtained, because the code used was designed for problem of classification and recognition in the Caltech101 database. Additionally, Caltech 101 has categories that can be more easily discriminated against compared to ImageNet200.

References

[1] W. S. R. L. L.-J. L. K. F.-F. L. Deng, Jia; Dong. Imagenet: A large-scale hierarchical image database. Conference on Computer Vision and Pattern Recognition, 2009.

[2] R. F. L. Fei-Fei and P. Perona. Learning generative visual models from few training examples: an incremental bayesian approach tested on 101 object categories. 2004.

[3] W. J. Y. H. L. J. X. Ping. Phow based feature detection for head pose estimation. Communication Technology (ICCT), 2015 IEEE 16th International Conference, 2015.