Computer Vision IBIO4490 - PHOW Lab Report

Paula Florez Herrera (pp.florez2215@uniandes.edu.co)

***Abstract*—One of the main problems being treated on the field of Computer Vision is the recognition and classification of images according to categories determined by the human experience. For this laboratory, a dataset of images and code by Caltech and an image dataset by ImageNet will be used. The objective is to understand the functioning of the Caltech code and then apply it to classify the ImageNet dataset to further compare its accuracy and performance.**

I. Introduction and Database

PHOW, or Pyramid Histogram of visual Words, is a representation that describes images according to their appearance, based on the implementation of multi-level Scale-Invariant Feature Transform or SIFT. The PHOW code used for this laboratory comes from the Caltech 101 database but will also be used to classify the images from the ImageNet dataset.

The databases used for this laboratory come from Caltech [1] and ImageNet, the Caltech Dataset contains around 9145 images of 102 different categories. Examples of the images contained in this dataset are shown on Figure 1 and 2.

Fig. 1. Example of Caltech Database: Category Plane

Fig. 2. Example of Caltech Database: Category Plane

As it is shown on these images, the main orientation of the object to identify remains the same, which makes it a rather easy recognition problem. The images seen on the ImageNet database are more complicated to recognize and this might cause a slightly different result when running the Caltech code with these images. Examples of this dataset for the one of its categories are shown on Figure 3 and 4.

Fig. 3. Example of ImageNet Database: Category Nail

II. Methods and Adjustments

A. PHOW Caltech

The acquired code for the Caltech Database starts by configuring the network with the sizes of the training and testing sets, and also several variables that permit the continuity of the code.

The next section allows the checking and downloading of the Caltech database, in order to run the following code with the images found on the specified directory. After that, the data setup section arranges the variables from the following code to the previous dataset to be evaluated.

After all the setup, the classifier is trained using PHOW and the spatial histograms and texton maps are computed with the functions downloaded from the guide.

The last sections of the code are related to the training, testing and final evaluation of the Support Vector Machine, in which the images are evaluated for the categories available and

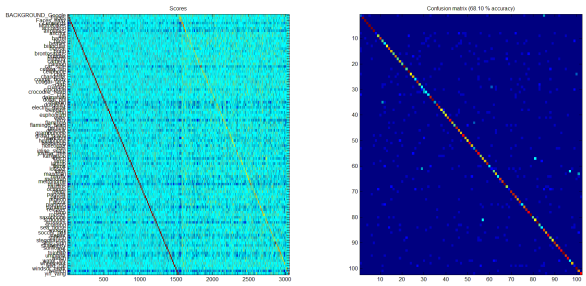

Fig. 5. Confusion Matrix for the Caltech Database

Fig. 4. Example of ImageNet Database: Category Nail

then compared to be in the appropriate one with a confusion matrix. This result is shown on figure 5.

B. PHOW for ImageNet Database

In order to adapt the code for the ImageNet Database, only two modifications were made, however the final modified code presented an issue that made it impossible to obtain results. The modifications made where the resetting of the path for the image database and the switching of the extension of the images from .jpg to .JPEG. However, the modifications seemed not to be enough since the code stopped before completing the testing and this made it impossible to evaluate the results of this method on the database.

III. Conclusions

The implementation of the provided method on the new database proved to need more adaptation to obtain results. What was expected was that the 68.1 percent accuracy from the method on the Caltech database was reduced when used with ImageNet because of the nature of the images in both databases. The images from Caltech were less variable in orientation and form, which allowed the method to obtain a higher accuracy, something that would not occur with the ImageNet database.

References

[1] Vision.caltech.edu,. 'Caltech 101'. N.p., 2015. Web. 01 Apr. 2015.