

SIFT, BoW architecture and one-against-all Support vector machine

Mohamed Issolah, Diane Lingrand, Frederic Precioso

I3S Lab - UMR 7271 UNS-CNRS
2000, route des Lucioles - Les Algorithmes - bt. Euclide B
06900 Sophia Antipolis - France

Abstract. For this first participation to ImageClef Plant Identification, we build on the reference Bag-of-Word framework (BoW). We extract Points-of-Interest (PoI) using the SIFT detector in every image and describe each local feature with the SIFT descriptor. The visual dictionary is built with a K-means algorithm of 100 clusters on the local features. Each image is then represented by its histogram onto the dictionary using hard-assignment strategy. We classify the images with as many binary one-against-all Support Vector Machines as the number of plant classes per organ types. Our aim is to evaluate for the plant identification task a classic baseline of multi-class image categorization. Our first results illustrate how difficult this task is and that a framework which has become a standard baseline for classifying general image datasets is not immediately relevant on Plant Identification data.

1 Our system

Our system for ImageClef Plant Identification task [1] is built on the reference Bag-of-Word framework (BoW) and a set of binary Support Vector Machines. We use the XML metadata file provided with the images to extract information as Type of content or Background type. Let us now detail our settings.

1.1 Feature extraction and Image description

We extract Points-of-Interest (PoI) using both the SIFT detector and the SIFT descriptor in each image. We extract about 1000 points in each image, with standard settings of Opencv C++ library:

- Number of layers per Octave: 3
- The minimum threshold to consider a point as PoI: 0.04
- σ of Gaussian: 1.6

Then we build a visual dictionary using a K-means where K is set to 100. Finally we represent each image by its histogram obtained by the hard assignment of each local feature to BoW clusters. The histogram is normalized by the size of the BoW.

1.2 Learning

We build one-against-all SVMs for each plant class and for each organ type and we exploit the XML metadata provided by the Challenge during the training phase. We thereby identify type of content information to train separately organ-oriented SVMs. The same SVM configuration is used for any organ:

- $C = 100$
- The kernel type is LINEAR
- Number of iteration = 10000

In the end, we obtain a trained SVM per class of plant and per organ type among:

- Entire
- Stem
- Fruit
- Flower
- Leaf NaturalBackground
- Leaf SheetAsBackground

SVM outputs are organized into vectors, one vector per organ type, as depicted in figure 1

$$\begin{pmatrix} class_{c_11} \\ class_{c_12} \\ class_{c_13} \\ \vdots \\ class_{c_1n} \end{pmatrix} \begin{pmatrix} class_{c_21} \\ class_{c_22} \\ class_{c_23} \\ \vdots \\ class_{c_2n} \end{pmatrix} \dots \begin{pmatrix} class_{c_61} \\ class_{c_62} \\ class_{c_63} \\ \vdots \\ class_{c_6n} \end{pmatrix}$$

Fig. 1. Each vector contains the classification scores obtained with n one-against-all SVMs (one SVM score per plant class) with respect to one of the 6 organs .

In 2013 dataset, the max number of plant classes is $n = 250$.

1.3 Confidence computation

For one test image, our system performs keypoint extraction and description, then from the XML file associated to the test image, we extract organ and type information. We then classify the test image histogram using all the SVMs corresponding to the associated organ. For example, if the organ of the image is Leaf and the type is SheetAsBackground, our system executes the set of n SVMs corresponding to Leaf vector. The final class corresponds to the class associated with the SVM providing the highest confidence score.

Figure 2 summarizes all the steps of our system.

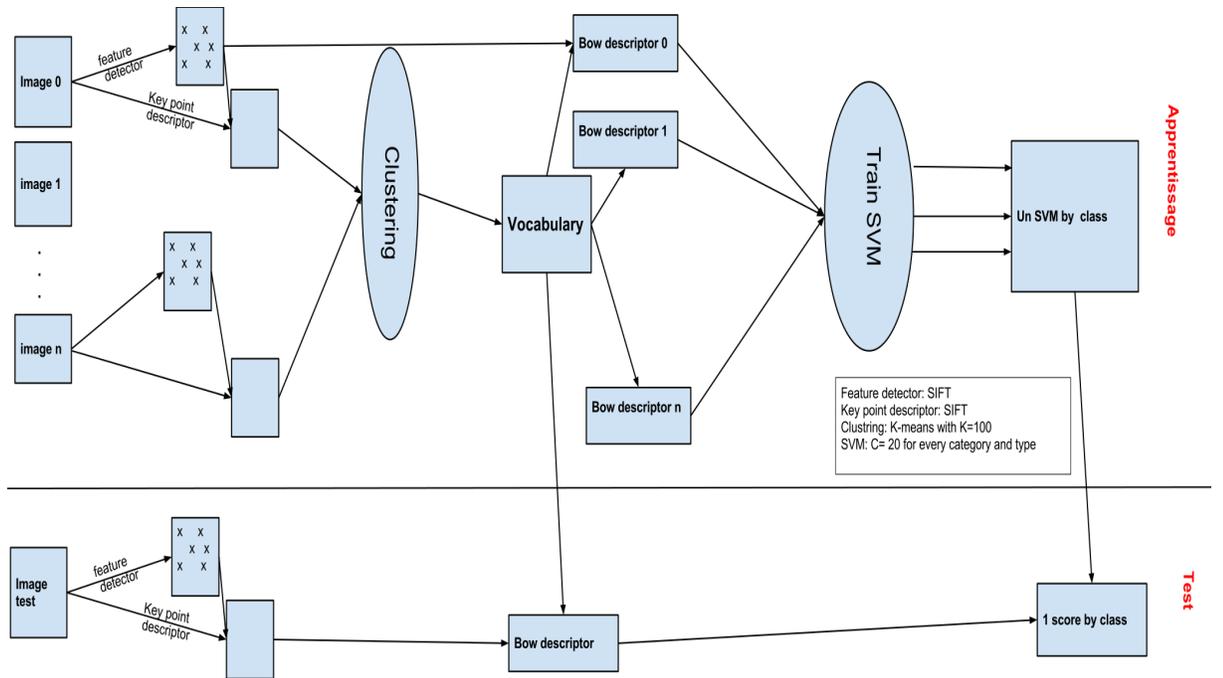


Fig. 2. Global schema of system

2 Experiments

Experimentation focus on three steps of our standard framework:

1. Configuration of K-means
2. Configuration of SVM
3. Decreasing amount of data to process

For the first configuration, different parameters are to be determined but the main one to choose is K, the number of clusters. The higher the number of clusters, the more discriminant the Bag-of-Words histogram representation. The number of clusters is fixed to 100 because the K-means configuration is common to all different organs and plant classes, it must preserve the generalization capability of the BoW representation and must require a reasonable computation time.

The same holds for the SVMs which have common settings for all organs and all classes in our baseline implementation. The bigger the parameter C , the lower the error rate. C is set to 100.

The Clustering is the most demanding step in terms of computational intensity. In order to reduce its impact, only 100 points among the 1000 extracted in each image are considered during the clustering. The keypoints to be discarded are chosen randomly.

2.1 Resources

For implementation we use the OpenCV libraries which offers different type of detection and different methods for learning. The implementation language is C++ for efficiency and speed. The LibXML libraries are used for XML parsing. The program has been launched on a server made of 2 Processors Intel Xeon X5675 at 3,06GHz, 6 Cores and 24GB RAM DDR3-1333MHz.

2.2 Results

The goal of our proposition is to evaluate a standard framework for image categorization in multimedia databases, using the classic local feature, SIFT, (both for the detector and the descriptor), a BoW architecture and as many one-against-all SVMs as binary classification required.

The results are presented in figure 3, 4 and 5

Run name	runfilename	Entire	Flower	Fruit	Leaf	Stem	NaturalBackground
I3S Run 1	1368034466828_new_100	0.017	0.023	0.041	0.038	0.025	0.026
I3S Run 2	1368165605197_new2_100	0.017	0.023	0.041	0.038	0.025	0.026

Fig. 3. Results on ImageClef 2013 with respect to the organs

3 Conclusion

As a first participation to ImageClef Plant Identification challenge, we have implemented a standard framework which proved to be powerful for image categorization in multimedia database. To do so, we have considered SIFT algorithm for both local feature detection and description, then represented each image with its histogram on the visual dictionary. The resulting histograms for each image are classified by one-against-all SVMs, one SVM per plant class and per organ. Despite the efficiency of such architecture for image categorization, the results are somewhat disappointing on ImageClef Plant Identification task. We are currently working on how to optimize all the parameters of our method to achieve better results.

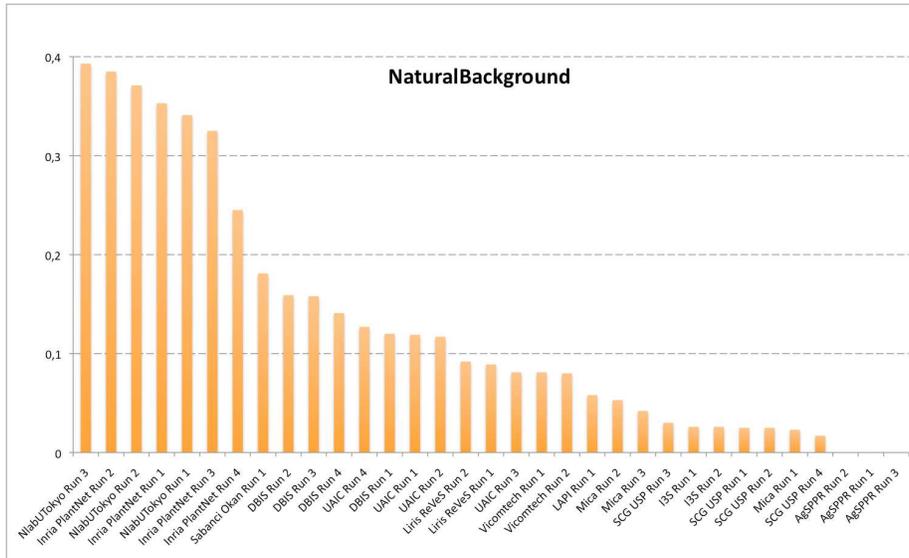


Fig. 4. NaturalBackgroundScores

References

1. Caputo, B., Muller, H., Thomee, B., Villegas, M., Paredes, R., Zellhofer, D., Goeau, H., Joly, A., Bonnet, P., Gomez, J.M., Varea, I.G., Cazorla, M.: Imageclef 2013: the vision, the data and the open challenges. Proceedings CLEF 2013, LNCS (2013)

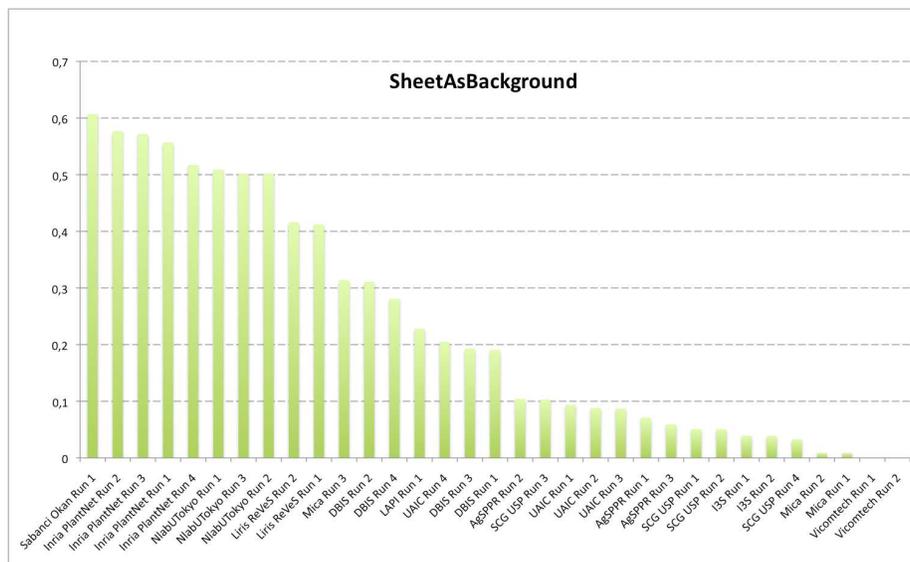


Fig. 5. SheetAsBackgroundScores