Fast Tree Leaf Image Retrieval using a Probabilistic Multi-class Support Vector Machine Classifier

Ignazio Gallo, Alessandro Zamberletti, Simone Albertini,
Angelo Nodari, and Marco Vanetti

University of Insubria
Dipartimento di Scienze Teoriche ed Applicate
via Mazzini 5, Varese, Italy
ignazio.gallo@uninsubria.it

Abstract. Nowadays an increasing number of people own mobile phones with built-in camera, able to take pictures. Thus, having a fast and fully automatic algorithm of image retrieval is considered a promising way to identify plant leaves on a mobile device. Our solution proposes a Support Vector Machine that provides a multi-class probability estimation with radial basis function kernel based on two descriptors: PHOG and a variant of HAAR. With our method we placed seventh among all the fully automatic methods who participated in the *ImageCLEF Plant Identification 2012* task. As showed by the results, the proposed method is very fast and at the same time has a classification accuracy comparable with the state of the art in this domain, aspects which make this method feasible in practice.

Keywords: image classification, support vector machine, pyramid of histograms of orientation gradients

1 Introduction

This paper presents the participation of the *ArTe-Lab*[1] research laboratory (Applied Recognition Technology Laboratory) at University of Insubria in the *ImageCLEF Plant Identification 2012* task. The objective of this work is the retrieval of plant species, starting from the images belonging to the dataset proposed by the aformentioned task[2]; this problem can be solved with a good accuracy using one of the many algorithms proposed in literature, such as: LP-β [1], MKL [2], VLFeat [3], R.Forests [4], etc. However these algorithms require a lot of computional time and therefore they cannot be used to perform a real time classification of the images. We wanted to follow a different approach by designing and developing a fast algorithm that is able to obtain a good accuracy over all the types of images belonging to the dataset of the task subject of this study.

[1] http://artelab.dicom.uninsubria.it/
[2] http://www.imageclef.org/2012/plant/

In order to identify the species a leaf belongs to, we model the species as classes in a classification framework that is based on a set of simple visual features extracted only from the images; we do not consider any metadata annotated along with each image which belongs to the dataset. In particular, we decided to employ simple features based on edges and intensity values distribution, among all the most successful features proposed in literature for object classification. We address the problem of learning score functions, motivated by ranking tasks in information retrieval (IR). Given an image containing a leaf, the scoring function associates a score to each known class; these class labels are then presented to the user in a decreasing order of scores. The quality of this sorted list of class labels depends on the position (rank) of the labels that are relevant to the image. Since the user considers only the few first class labels, it is desirable to have an high precision on top scored ones. Learning to rank is equivalent to the problem of choosing an effective scoring function, using a training set of images for which relevant classes are known.

With our method we placed seventh among all the fully automatic methods who participated in the *ImageCLEF Plant Identification 2012* task.

2 The Proposed Method

The multi-class classification problem refers to assigning each of the observations into one of k classes. In this paper we focus on a technique that provides a multi-class probability estimation by combining all the pairwise comparisons [5], using a Support Vector Machine (SVM) [6] classifier. Pairwise coupling is a popular multi-class classification method that combines all comparisons for each pair of classes; this method can be reduced to a linear system which is easy to implement. In particular, we used the implementation found in LIBSVM library for support vector machines [7]. The predicted label is the one with the largest probability value but, using the same probability values we sort all the classes, from most to least likely.

The SVN receives as input two different descriptors: the first is the Pyramid of Histograms of Orientation Gradients (PHOG) [8] and the second is similar to the HAAR descriptor proposed by Viola and Jones [9]. We used 15 bins and 3 layers for the PHOG descriptor while the HAAR is composed by two different descriptors. Regarding the latest, the first HAAR descriptor is $D^1 = \{D^1_1, \ldots, D^1_{25}\}$ and it is computed as follows: each image is divided into a table of 25 rows and 3 columns; for each row j we calculate the sums S_i obtaining a descriptor's component as $D^1_j = S_1 + S_2 - S_3$. The second HAAR descriptor $D^2 = \{D^2_1, \ldots, D^2_{50}\}$ is computed similarly to the first one: each image is divided in 25 rows and 6 columns but, for each row j, we compute two components $D^2_j = S_2 - S_1 - S_3$ and $D^2_{j+1} = S_5 - S_4 - S_6$. Each descriptor is first normalized and then concatenated to the other in order to form the input pattern. Figure 1 shows a graphical representation of how the input images are transformed in patterns for the used SVM model. In the top row of the figure we can notice the three histograms extracted from the three levels of the PHOG descriptor, while

Table 1: Comparison, in terms of overall accuracy, of different images classification algorithms on the Caltech-101 and the Drezzy-46 dataset. It is also reported the average time required to perform the classification of a single image.

	Overall Accuracy %		
Algorithm	Caltech-101	Drezzy-46	Avg. Time
LP-β	82.10	80.12	83.0s
MKL	73.70	88.89	88.0s
VLFeat	65.00	71.30	5.27s
R.Forests*	80.00	-	45.0s
HOG-SVM	28.26	32.90	0.014s
PHOG-SVM	54.00	64.87	0.047s
HAAR**	-	-	0.001s

*the source code is not available
**taken individually this feature is not relevant in the classification process

each histogram is constructed by concatenating the histograms extracted from each cell. In the bottom row we can notice the two HAAR descriptors, each of which is transformed into a new histogram to be concatenated to the previous.

The SVM is trained for probability estimation and it uses a radial basis function as kernel and $C = 8$, $\gamma = 2$ as main parameters.

We compared PHOG and Haar descriptors along with many others in order to determine the most stable and robust ones.

3 Experiments

In order to choose the best features to manage the problem addressed in this study, we started our experiments comparing different classification algorithms found in the literature: a multiclass method called LP-β [1], a multi-kernel method called MKL [2], an algorithm based on Bag of Words called VLFeat [3] and an image classification approach based on R.Forests named Random Forests (R.Forests) [4]. We also considered simpler approaches based on HOG [10] and PHOG [8] features, using a SVM classifier [6]. We also evaluated an interesting feature: the HAAR descriptor proposed by Viola and Jones [9]; when taken individually this feature has not been shown to be relevant in classification but, in the experimental phase, we have verified that by combining it with other features we increase the object classification accuracy.

These features are compared using two standard datasets: the Caltech-101 [11] which contains 9,146 images of generic objects belonging to 101 classes, and the Drezzy-46 [12] which is composed by 46 classes of different commercial products crawled from the web and, for each class, there are approximately 100 images, leading to a total amount of about 4600 images. The results are reported in Table 1, we also report the computational time, evaluated using a single thread C# code, on a Intel®Core^{TM}i5 CPU at 2.30GHz.

Because our goal is focused on the development of an application able to properly work on a mobile device, we were looking for good features that can

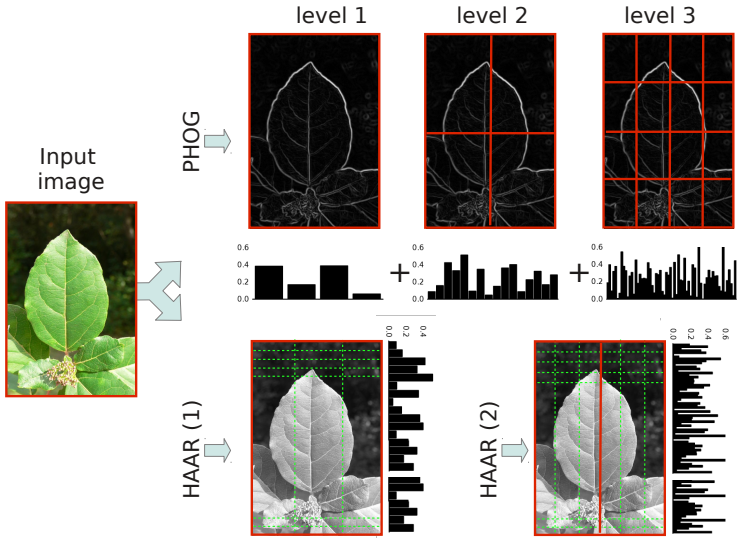

Fig. 1: A representative example of the PHOG and HAAR features extracted from an image belonging to the dataset used. Each features is normalized and linked to the others, in order to construct the representative pattern of the input image.

be quickly computed and do not require too much memory. Therefore we have selected the features considering the best balance between accuracy, computational time and memory requirements: PHOG and HAAR descriptors. In particular, the parameters of the PHOG feature were tuned in order to obtain the best compromise between speed and accuracy. Regarding the HAAR feature, we computed the two descriptors by setting the width of the rectangle to fit the image width of each image belonging to the *ImageCLEF Plant Identification 2012* task dataset, on the basis of the results obtained during the experimental phase. Even if the predicted output label of our solution is the one having the largest probability value, using the same probability we can provide as output a list of class labels ordered by likelihood.

After selecting the features to adopt in our model, we evaluated the performance of classification on the Pl@ntLeaves dataset which has been built for the *ImageCLEF Plant Identification 2012* competition, showed in Figure 2. This dataset contains around 11572 pictures subdivided into 3 different kinds of pictures: scans (Scan), scan-like photos (Pseudoscan) and free natural photos (Photograph). Each picture represents a plant leaf belonging to one of the 126 species present in the dataset. In the Scan category, each scan shows the upper-side of one leaf on a uniform background, centered and oriented vertically along the main natural axis; in the scan-like category the images are similar to the previous category but there are some luminance variations, optical distortions and shadows due to the not flattened acquisition method; in the free natural photos category

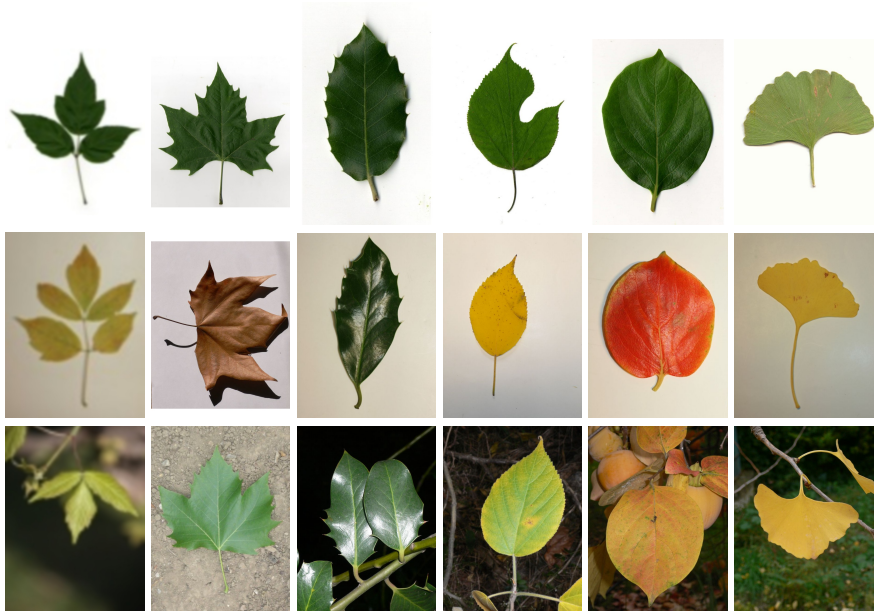

Fig. 2: Examples of images belonging to the Pl@ntLeaves dataset. The dataset is divided in three categories: scans, scan-like photos and free natural photos.

the images are taken directly on the trees, therefore they may contain one or more leaves with different and complex backgrounds (such as branches, leafage, a trunk, the ground, the sky etc.) and various orientations. We trained a single SVM model for all three categories of images, even if the category Photograph deserves at least a preprocessing in order to improve the overall performance.

We used the overall-accuracy to evaluate the results on the Pl@ntLeaves dataset and we obtained an average score of 0.30; considering individual types of images we obtained: 0.40 for annotations of type Scan, 0.37 for Pseudoscan and 0.14 for Photograph. A comparison with the results obtained by the other participants can be found in the figures 3,4 and 5. Our solution was implemented in C++ and tests were conducted on a Linux machine with CPU Intel Core2Duo 4500 2.02GHz. The average time required to transform a single image into a pattern for the SVM model is $90ms$ (that approximately corresponds to the time required to compute the single feature PHOG), while to predict the ranking for a single image the trained SVM needs approximately $20ms$.

We want to emphasize that the proposed method does not use any a priori or metadata information but is based only on the information extracted from the content of each image and for this reason we believe that it is a very promising result.

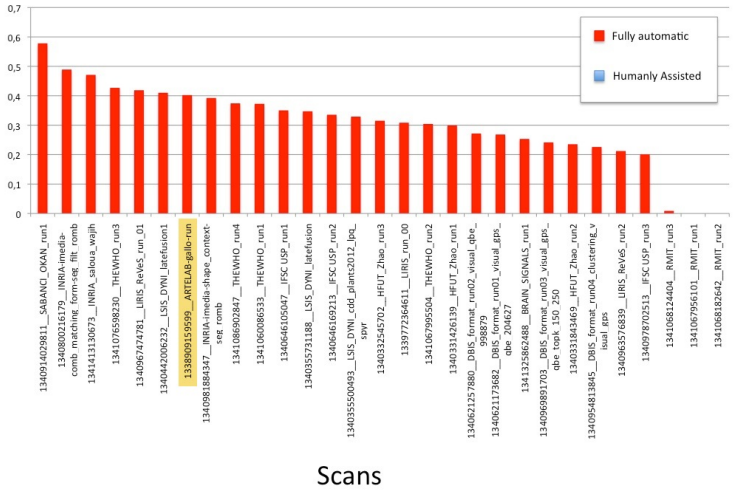

Fig. 3: Comparative results showing the performances of the proposed solution (*ARTELAB-gallo*) applied on the Scan category of the Pl@ntLeaves dataset.

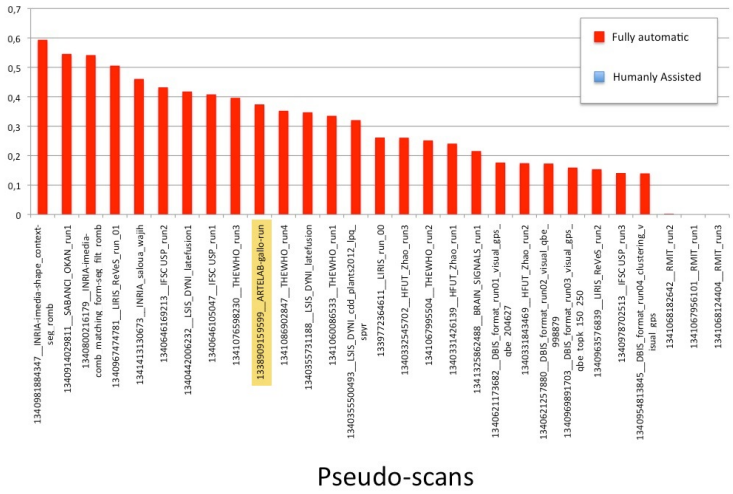

Fig. 4: Comparative results showing the performances of the proposed solution (*ARTELAB-gallo*) applied on the Pseudoscan category of the Pl@ntLeaves dataset.

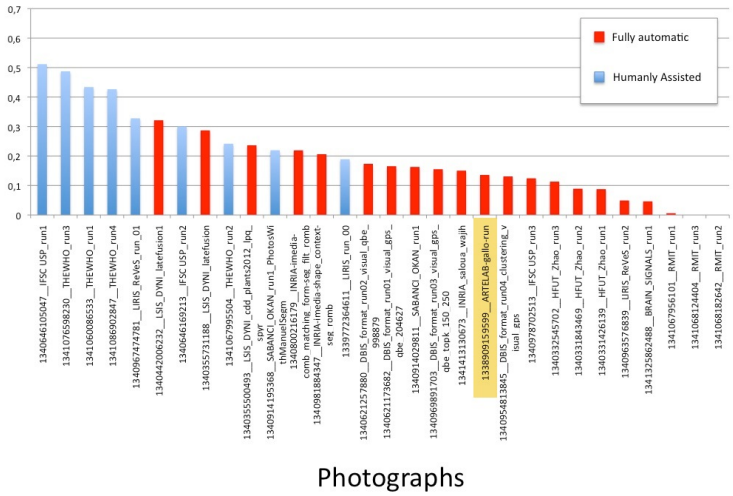

Fig. 5: Comparative results showing the performances of the proposed solution (*ARTELAB-gallo*) applied on the Photograph category of the Pl@ntLeaves dataset.

4 Conclusions

Our group submitted just one run in our first participation in the *ImageCLEF Plant Identification 2012* task; in this paper we described a simple plant species retrieval model based on a probabilistic SVN using PHOG and HAAR features.

With our proposed solution we placed seventh among all the fully automatic methods who participated in the aformentioned task and this is a good result because our approach can perform the classification of a single image, belonging to the Pl@ntLeaves dataset, in real time on a low computational power machine. Looking at the results, we can see that the proposed method obtains good results in the classification of the images belonging to the first two classes of images (Scan and Pseudoscan), while it needs much more preprocessing and segmentation work to be effective in the classification of the images belonging to the Photograph class.

Concerning the computational time, our algorithm turned out to be very fast due to the fact that we employed simple features and also because the Support Vector Machine using a radial basis function kernel requires linear time in the feature pattern size for the predictions phase.

This work opens a possible deep study of the features adopted to enhance the classification accuracy. Our simple approach allows the proposed algorithm to be very fast, but we could evaluate the possibility of adopting more complex features or even exploiting the pre-processing phases, such as image segmentation

or partitioning, in order to improve the number of correct classifications, in spite of the computational performances.

Another possible future improvement of our model lies in the possibility of exploiting the metadata informations associated to the images belonging to the Pl@ntLeaves dataset, in order to enhance the classification accuracy; for example, we could identify the features that characterize the plants from each different region, as we can obtain the geographical place where the leaf had been taken from the metadata.

References

1. Gehler, P.V., Nowozin, S.: On feature combination for multiclass object classification. In: ICCV, IEEE (2009) 221–228
2. Vedaldi, A., Gulshan, V., Varma, M., Zisserman, A.: Multiple kernels for object detection. In: Proceedings of the International Conference on Computer Vision (ICCV). (2009)
3. Vedaldi, A., Fulkerson, B.: Vlfeat: an open and portable library of computer vision algorithms. In Bimbo, A.D., Chang, S.F., Smeulders, A.W.M., eds.: ACM Multimedia, ACM (2010) 1469–1472
4. Bosch, A., Zisserman, A., Munoz, X.: Image classification using random forests and ferns. In: IEEE International Conference on Computer Vision. (2007)
5. Wu, T.F., Lin, C.J., Weng, R.C.: Probability estimates for multi-class classification by pairwise coupling. J. Mach. Learn. Res. **5** (December 2004) 975–1005
6. Cortes, C., Vapnik, V.: Support vector networks. Machine Learning **20** (1995) 273–297
7. Chang, C.C., Lin, C.J.: LIBSVM: A library for support vector machines. ACM Transactions on Intelligent Systems and Technology **2** (2011) 27:1–27:27 Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.
8. Bosch, A., Zisserman, A., Munoz, X.: Representing shape with a spatial pyramid kernel. In: Proceedings of the 6th ACM international conference on Image and video retrieval. CIVR '07, ACM (2007) 401–408
9. Viola, P.A., Jones, M.J.: Robust real-time face detection. International Journal of Computer Vision **57**(2) (2004) 137–154
10. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: Proc. CVPR. (2005) 886–893
11. L. Fei-Fei; Fergus, R.P.: One-shot learning of object categories. IEEE Transactions on Pattern Analysis Machine Intelligence **28** (April 2006) 594–611
12. Nodari, A., Ghiringhelli, M., Albertini, S., Vanetti, M., Gallo, I.: A mobile visual search application for content based image retrieval in the fashion domain. In: Workshop on Content-Based Multimedia Indexing (CBMI2012). (2012)