

ZhaoHFUT at ImageCLEF 2012 Plant Identification Task

Peng Zheng¹, Zhong-Qiu Zhao¹, and Herve Glotin^{2,3}

¹ The College of Computer Science and Information Engineering,
Hefei University of Technology, China,

² Systems and Information Sciences lab, LSIS CNRS & Univ. of Sud-Toulon Var - La
Garde, 83957, France,

³ Institut Iniversitaire de France, Paris, France,
nicknina2006@163.com, zhongqiuzhao@gmail.com, glotin@univ-tln.fr

Abstract. This paper presents the contribution of ZhaoHFUT group to the ImageCLEF 2012 Plant identification task. The task involves identifying various species of trees based on images of their leaves. In this task, we adopted the main structure of the ScSPM model and another extension of Scale Invariant Feature Transform (SIFT) descriptors, namely flip SIFT descriptors, to investigate the performance of them considering their good performance in object classification. Although our results are not quite promising as compared to other participant groups, they can still guide our work in this field for some conclusions reached.

Keywords: Plant identification, feature extraction, ScSPM, flip SIFT descriptors

1 Introduction

There are about 400,000 species in the world, of which only 270,000 have been named and identified by botanists. Accurate knowledge of the identity, geographic distribution and uses of plants is essential for agriculture development and biodiversity. Although there are many researches in this field, the taxonomy of species is still a hard task even for professionals (such as farmers or wood exploiters) or for the botanists themselves because such basic information is usually partially available for professional researchers and the flowers and fruits are not always available for studies throughout the year while the leaves which can be gotten for most of the time don't have sufficient visible characteristics to differentiate between many species.

Methods of computer vision such as image retrieval technologies can be helpful in reducing this taxonomic gap. Also evaluating recent advances of the IR community on this challenging task might therefore have a strong impact. The ImageCLEF series use this idea and propose an ongoing campaign on plant identification task. The goal of the task is to retrieve the correct species among the top k species of a ranked list of retrieved species for each test image and

provide a forum for researchers working on image analysis and artificial intelligence methods to share their ideas and compare their systems in order to help the taxonomic process with leaves information [1] [2].

In the continuity of last year pilot, the task will be focused on tree species identification based on leaf images. The main novelties compared to last year are the following:

- (1) More species: the number of species will be this year 126, which is an important step towards covering the entire flora of a given region.
- (2) Plant species retrieval vs. pure classification: last year's task was organized as a pure classification task over 70 species, whereas this year evaluation metric will be slightly modified in order to consider a ranked list of retrieved species rather than a single brute determination [1].

So the task this year is much more challenge than that of last year and the evaluation is also quite reasonable compared with that of last year.

The following Section 2 describes the materials and methods used. In Section 3, we describe the results of our experiments. Finally, some conclusions reached are presented in Section 4.

2 Material and Methods

2.1 Database

The task will be based on Pl@ntLeaves II dataset which focuses on 126 tree species from French Mediterranean area. This database is maintained by the French project Pl@ntNet (INRIA, CIRAD, Telabotanica). It contains 11572 pictures subdivided into three different kinds of pictures: scans (57%), scan-like photos (24%) and free natural photos (19%)[1].

1. Scan: These images are oriented vertically along the main natural axis and with the petiole visible collected using flatbed scanners.
2. Scan-like photos: Those images have uniform background but with some luminance variations, optical distortions, shadows and color derivations which look similar to the scans images.
3. Free natural photos: No acquisition protocol is used, which results in a non-uniform background, rotated and bad-scaled images, among others problems. And the images are taken directly on the trees.

All data are published under a creative commons license. Each image has an xml file associated that contain the date, type (single leaf, single dead leaf or foliage), name of the author and GPS coordinates of the observation etc. But we don't make use of this kind of information and all the data are extracted from the images directly without any segmentation operations. So the experiments are fully automatic without any human assistance.

The training data finally results in 8422 images (4870 scans, 1819 scan-like photos, 1733 natural photos) with full xml files associated to them. The test data results in 3150 images (1760 scans, 907 scan-like photos, 483 natural photos) with purged xml files (i.e. without the taxon information that has to be predicted)[1].

2.2 Methods

As the leaf is one kind of object, the methods applied to object classification can be adopted to perform plant identification. Sparse representation is a new method used to solve this problem and achieve promising results [3]. Here we utilize the ScSPM model proposed by Yang to finish this task. ScSPM is an algorithm which combined spatial pyramid matching and sparse representation getting good results in object classification task [4]. This algorithm has 4 stages as follows.

Feature Extraction. Feature extraction has great effects on the stages afterwards. As SIFT descriptors are widely used in pattern recognition, signal processing and computer vision and suitable for classification, in this paper we extracted SIFT descriptors densely in images and normalized them to form the feature vectors [5]. Also as symmetry can be found in leaves, if we consider this kind of symmetry in feature extraction, a better result may be able to gotten. So here we take advantage of an extension version of SIFT descriptors, namely flip SIFT descriptors, to do another experiment [6]. The descriptors can be generated by the software package provided by the author. We change it by extracting them densely in the images and omit some descriptors which don't meet the requirements.

Sparse Coding. Since the introduction of sparse coding, many algorithms have been proposed to solve this problem such as MP, OMP which solve the l_0 -regularization problem:

$$\min_a \|y - Xa\|_2^2 + \lambda \|a\|_0 \quad \text{subject to } \|x_i\| \leq 1 . \quad (1)$$

and LASSO that solves the l_1 -regularization problem:

$$\min_a \|y - Xa\|_2^2 + \lambda \|a\|_1 \quad \text{subject to } \|x_i\| \leq 1 . \quad (2)$$

Feature search sign is a method solving this problem with high efficiency, so we use it to do sparse coding for saving the time for experiments [7].

Spatial Pooling Scheme. Instead of the histogram representation in traditional SPM method [8], this scheme divided the image into several scales and on each scale it performed pooling scheme in the patch formed.

Usually there are three different pooling mechanisms which can be used in this model:

- (1) Average pooling which calculates the mean pooling of absolute value.
- (2) Maximum value pooling which gets the maximum absolute value in a local patch.
- (3) Energy pooling which calculates the square root of mean of energy values in a patch formed.

The operation of maximum pooling can get the best result usually, so we just report the result of it.

Classification. The last step of this framework is the training of a classifier to classify the unitary normalized test vectors obtained from the previous stages. As linear classifiers can earn a good result, we use linear Supported Vector Machine (SVM) classifiers to finish this work and the parameters for SVMs are tuned by cross-validation. Also the results for the runs are formed by normalizing the degrees of attribution in different SVMs.

3 Experiments and Results

3.1 Evaluation Metric

The only thing that matters in the primary metric evaluating the submitted runs is the rank of the correct species in the list of retrieved species. Each test image will be assigned a score between 0 and 1 which decreases quickly with the increment of the rank of the correct species. At last, a simple mean on all test images will be computed. Since we want to evaluate the ability of a system to provide correct answers to all users and based on a single plant observation, we rather measure the mean of the average classification rate per author and average the classification rate on each individual plant respectively. Finally, the final metric is defined as the following average classification score S :

$$S = \frac{1}{U} \sum_{u=1}^U \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{n=1}^{N_{u,p}} s_{u,p,n} . \quad (3)$$

where U is the number of users (who have at least one image in the test data), P_u is the number of individual plants observed by the u -th user, $N_{u,p}$ is the number of pictures taken from the p -th plant observed by the u -th user and $s_{u,p,n}$ is classification score for the n -th picture taken from the p -th plant observed by the u -th user. Considering different image acquisition type may have different effects, an average classification score S will be computed separately for each type [1].

3.2 The Results

According to the rule of the campaign, we submitted three different runs to the plant identification task. And the details are as follows.

In the first run, we just extracted the dense-SIFT descriptors from the images, used ScSPM model for generating the feature vectors and took the ls-svm package to form classifiers to classify the test images in the database [9].

In the second run, we changed the descriptors into flip SIFT ones in hope of taking symmetry into account and the others kept the same as the first run.

In the third run, also the last run, we fused the classifiers of the two runs above to form a new kind of degree of attribution. Then the normalization of the terms formed the rank list we submitted.

The results are showed in Table 1. From it we can find that the third run achieves the best results in all kinds of pictures. That's to say, the fusion scheme

usually can reach a better result for combination of different methods advantages. The run for the SIFT descriptors is in the middle place of the three runs and the one for flip SIFT descriptors gets the worst results which is in contradiction with our expectation. In fact, we have done comparison experiments on these two kinds of descriptors by using the species which have more than 30 images in training database with MAP as the evaluation term. Then the result of flip SIFT descriptors is better than that of original SIFT descriptors. So we can speculate that the reasons for this phenomenon may have a relation to the number of training images in certain species which have very few images and the parameters of the SVMs which are the same as those of the first run for convention and time limit.

Table 1. Results for all runs in plant database.

Run name	Retrieval type	run-type	scan	pseudo-scan	photograph	avg score
ZhaoHFUT_run1	Visual	Automatic	0,30	0,24	0,09	0,21
ZhaoHFUT_run2	Visual	Automatic	0,24	0,17	0,09	0,17
ZhaoHFUT_run3	Visual	Automatic	0,32	0,26	0,11	0,23

Also we can find that of the three kinds of images, the results for the scan images are the best, then those of the pseudo-scan images and the ones of the photograph images come at last. The differences between the results of the first two kinds of images may due to optical distortions and shadows in the images which can not be recognized by dense SIFT descriptors. The reason for the worst performance in classifying the photograph images is that the model used considers the natural scenes as parts of leaves which does harm to the identification task afterward. Maybe taking segmentation operation at first for this kind of images is a good choice.

For the limitation of pages, we don't present it here and you can find it on the homepage of this task [1]. The results can be divided into two parts, namely fully automatic and others, at first. In the first two kinds of images, the best of ours gets the sixth place in automatic ones. While for the last kind of images, the results of ours fall behind the others and get the last several positions. For the average classification score, we can earn the sixth place in automatic ones. That's to say, the results are not too bad for our first participation in this task and the results can be improved a lot if we improve the performance of classifying the photograph images.

4 Conclusion

This is our first try in the plant identification task. Although the results are not quite promising, we still reach some conclusions and gain some experience.

(1) At first, as there are some species which have only a few images, the bias for these species may be emphasized and combining some retrieval methods is a good choice.

(2) Then the flip SIFT descriptors may not be suitable for this task if we can't tune the parameters individually and change it into dense ones in a proper way.

(3) After that, building different classifiers for different types of images may be helpful and the ScSPM model probably doesn't meet the requirements of classifying the photograph images and it suggests us selecting some other algorithms for this type of images.

(4) Also it may not be proper for the prediction lists of SVMs to represent the results directly and there are some problems in the normalization scheme. And nonlinear classifiers may improve the results.

(5) At last, the provided xml files are not used in our experiments and maybe they will provide complementary information for the visual features.

As this kind of system is quite helpful for researches and agricultural development, we will continue to do researches in this field. And by caring about the details and altering the methods used, we hope to further improve our experiment in future tasks.

Acknowledges. This research has been supported by the National Natural Science Foundation of China (61005007), the National 863 Program of China (2012AA011005), the China Post-doctoral Science Foundation (20100470825, 201104324) and Cognilego ANR 2010-CORD-013. The authors would like to thank Jianchao Yang, who provided the SIFT and ScSPM codes for sharing.

References

1. ImageCLEF 2012 Plant identification task. <http://www.imageclef.org/2012/plant>
2. Hervé Goeau, Pierre Bonnet, Alexis Joly, Itheri Yahiaoui, Daniel Barthelemy, Nozha Boujemaa, Jean-François Molino, The ImageCLEF 2012 plant identification task, CLEF 2012 working notes, Rome, Italy, 2012.
3. Olshausen B A, Field D J.: Sparse coding with an overcomplete basis set: a strategy employed by V1? *Vision research*. vol.37(23), pp.3311-3325, DOI: 10.1016/S0042-6989(97)00169-7 (1997)
4. Yang Jianchao, Yu Kai, Gong Yihong, etc: Linear Spatial Pyramid Matching Using Sparse Coding for Image Classification. *IEEE Conference on Computer Vision and Pattern Recognition*, 1794-1801(2009)
5. Lowe, David G.: Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision*, vol 2, pp.1150-1157(1999)
6. lip-vireo package for flip SIFT descriptors. <http://www.cs.cityu.edu.hk/~wzhao2/>
7. H. Lee, A. Battle, R. Raina, and A. Y. Ng.: Efficient sparse coding algorithms. *Advances in Neural Information Processing Systems (NIPS)*, pp. 801-808(2007).
8. S. Lazebnik, C. Schmid, and J. Ponce: Beyond bags of features: Spatial pyramid matching for recognizing natural scene categories. In *CVPR(2006)*
9. lssvm matlab package. <http://www.esat.kuleuven.be/sista/lssvmlab/>