Leaf image classification with exponential family Fisher vector

None

None

Abstract. In this work we present an experimental analysis of the use of exponential family Fisher vector to solve the problem of visual plant identification. We make a comparison of the encoding of different descriptors with this framework and we evaluate the performance on public datasets and compare these results with state of the art methods proposed in the literature. We show that eFV framework performs very well in the problem of plant classification.

Keywords: plant identificacion, exponential family Fisher vectors, image classification

1 Introduction

In recent years, there has been an increasing interest in the problem of plant species classification using visual information [3, 7, 13, 20]. Some reasons of this are the large number of endangered species and the high rates of deforestation due to the shift of the agricultural frontier and a poor urban planning. Plants have a crucial role in the life on earth and their carelessness cause irreversible problems to our society, such as global warming, loss of biodiversity and environmental damage [3, 22]. The problem presents a very interesting challenge, because it is almost impossible for common people and very difficult for trained ones such as farmers, wood exploiters or even botanists [5]. The reasons of this are many, among which we can name the large number of species, accounted for approximately 200000, the vast intra-class variability and a high visual similarities between classes [20].

In this work we address the problem using a recently proposed encoding called exponential family Fisher vectors (eFV) [19]. This encoding is a generalization of Gaussian based Fisher vector (FV) to the exponential family distribution, which allows to encode local descriptors in a large number of input domains such as real, integer or binary vectors and symmetric positive definite matrices (SPDM).

2 Related Work

A lot of preproccesing, feature extraction and classification algorithms to solve the problem have been proposed. These algorithms can be divided into two big groups, those using global descriptors and those based on local descriptors.

Within the first, the authors of [24, 25] propose the use of shape and texture global features obtained after a segmentation step for the classification of leaves images. In [13] a system using geometric descriptors, multi-scale distance matrix, invariant moments and a new set of global descriptors is proposed. The computation of these descriptors is based on the outline of the leaf, so a preprocessing step is necessary for proper countor extraction and according to the authors this step fails for some type of pinales. A semi-automatic algorithm that returns the most probable classes in descending confidence order is proposed in [20]. The used descriptors are globals and for their calculation the user has to mark the base and the apex of the leaf. In [3] is proposed a recognition method based on shape and texture global descriptors, which are sensitive to rotation, translation and scaling, so it is required an alignment algorithm before the extraction of descriptors.

With respect to the local descriptors based methods, in [9] is proposed a system based on sparse coding of SIFT descriptors and a similar method using a combination of descriptors including SIFT is presented in [16]. In [1], the authors propose the use of different local descriptors (SURF, Fourier, Rotation Invariant, LBP) encoded with FV to classify images of leaves taken with a natural background. In that work descriptors are calculated over Harris interest points and classified with an SVM in a OvA configuration. The authors of [15] use local descriptors (4 versions of SIFT and self-similarity) augmented with a polynomial method that takes into account neighbors descriptors and then encoded with FV. In [4] FV over SIFT and color moments are combined with CNN, including a preprocessing step to get the most representative bounding box of the image.

In this work we propose the use of local descriptores, encoded with a recent proposed framework, termed exponential family Fisher vectors (eFV) [19].

3 Method description

The proposed pipeline contains four stages, the first is dense extraction of visual descriptors, then these descriptors may or may not be reduced in dimensionality with PCA, after that, encoding is performed with eFV and finally these vectors are classified using SVM. A diagram of the pipeline can be viewed in the figure 1 and in the following we explain each of the parts.

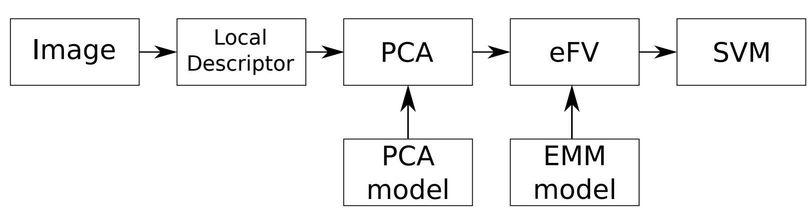

Fig. 1. Block diagram of the proposed system.

Table 1. Descriptors and corresponding distribution used for enconding.

Descriptor	Input Domain	Distribution
SIFT	$\mathbb{R}$	Gaussian
BRIEF, BinSIFT, LBP	$\{0, 1\}$	Bernoulli
COV	SPD Matrices	Wishart

3.1 Descriptors

Descriptors are extracted densely on a regular grid with the same step in both directions. Furthermore, these are calculated in the original image and in four scales, with a scaling factor of $\frac{1}{\sqrt{2}}$.

The selected descriptors for this work are SIFT, binarized SIFT (BinSIFT), BRIEF and a variation of COV [21]. PCA was only applied to SIFT and BinSIFT descritors.

3.2 Exponential family Fisher vector

The FV [18] representation is actually one of the most robust for image classification [14] and fine-grained classification [8]. This representation encodes an image as a gradient vector that characterize the distribution of the samples with respect to the parameters of a probabilistic model. The eFV enconding used in this work is a generalization of the FV which extends the descriptor input domains, to real, integer or binary vectors, and SPD matrices.

Suppose an image I on which we extract N low-level descriptors $\mathbf{x}_i \in \mathbb{R}^D$, $\mathbf{X} = \{\mathbf{x}_1, ..., \mathbf{x}_N\}$. Let $\lambda = \{\alpha_k, \eta_k : k = 1, ..., K\}$ be the parameters of a pdf $p(\mathbf{x}_i|\lambda)$ modelling the generation process of descriptors. The eFV of I is defined as:

$$g(\mathbf{X}) \triangleq \frac{1}{\sqrt{N}} \sum_{n=1}^{N} L_\lambda \nabla_\lambda \ \mathsf{log} \ p(\mathbf{x}_n|\lambda)$$

where ∇_λ denotes the gradient with respect to λ and L_λ is a normalizer.

We model $p(\mathbf{x}_n|\lambda)$ as a K mixture model with param λ:

$$p(\mathbf{x}_i|\lambda) = \sum_{k=1}^{K} w_k p_k(\mathbf{x}_i), \quad w_k > 0 \ \ \forall k, \quad \sum_{k=1}^{K} w_k = 1$$

w_k are the mixture weights and $p_k(\mathbf{x})$ as a member of the q-parameter exponential family:

$$p_k(\mathbf{x}|\eta_k) = h(\mathbf{x}) exp(\langle \eta_k, T_k(\mathbf{x}) \rangle - \psi(\eta_k))$$

For a more detailed derivation of the eFV encoding we suggest see [19].

In table 1 is shown the descriptor and its corresponding exponential family distribution used for encoding.

3.3 Classifier

For eFV classification we used SVM with a linear kernel trained with SGD, because it is the normal selection for this type of codifications [18, 19]. The use of non-linear kernels is problematic due to the high dimensionality of the vectors.

4 Experiments

To evaluate the eFV encoding, we perform experiments on different public datasets, commonly used for this task and we compare our results with different state of the art algorithms.

4.1 Datasets

The first dataset is the presented in [23], known as Flavia, which contains 1907 images of leaves from 32 classes of trees, with a minimun of 50 samples per class and a maximun of 72. The normal procedure of evaluation is to leave 10 samples of each class for test and train on the rest.

The second dataset is known as Foliage [11], which contains 120 samples for each of 60 species of trees. The recommended procedure of evaluation is to take 100 samples for training and 20 for testing for each class.

The last two, are the used on the plant identification challenge organized in the ImageCLEF 2012 and 2013. The first of these datasets, PlantCLEF2012 [6], contains 11572 images of 126 species of trees divided in three types, scan, scan-like and photograph. The second, PlantCLEF2013 [5], contains 26077 images of 250 tree species of two types, sheet as background and natural background. Also, the NaturalBackground images are divided into 5 types, images of entire plant, flower, fruit, leaf and stem.

4.2 Experimental configuration

As already mentioned, local descriptors were calculated on a regular grid and in four image scales with a factor of $\frac{1}{\sqrt{2}}$. In the case of SIFT and BinSIFT descriptors, a step of dimensionality reduction using PCA was applied and the resulting dimensionality was 78. On these descriptors, a 256 component family exponential mixture model was adjusted, which was then used to calculate the eFV encoding, according to the configuration shown in the table 1. Table 2 shows a resume of the different eFV configurations and its short name for further reference. eFV computing was done with the library mentioned in [19].

Furthermore, we propose the use of the results obtained using the descriptores based on CNN proposed in [17] as a baseline for comparison. These descriptors were computed such as the output of the 7th layer (fc7) of the convolutional neural network available in [10] and then classified with an SVM. This baseline is referred in the following as CNN+SVM.

Table 2. eFV configurations and short names.

Short Name	Descriptor	PCA	Exponential Mixture Model
BRIEF-BMM-eFV	BRIEF	No	Bernoulli
SIFT-PCA-GMM-eFV	SIFT	Yes	Gaussian
COV-WMM-eFV	Covariance	No	Wishart
LBP-BMM-eFV	LBP	No	Bernoulli
BinSIFT-BMM-eFV	BinSIFT	Yes	Bernoulli

4.3 Results

Table 3 shows the accuracy of different configurations of the proposed method on the Flavia and Foliage datasets together with recent results available in the literature. The accuracy is obtained as the percent of well classified samples.

Table 3. Accuracy of different configurations of eFV and results in the literature on datasets Flavia and Foliage

Method	Acc. Flavia	Acc. Foliage
CNN+SVM	99.06	**99.33**
SIFT-PCA-GMM-eFV	99.06	98.75
COV-WMM-eFV	**99.38**	98.25
LBP-BMM-eFV	95.62	93.25
BinSIFT-BMM-eFV	89.06	94.33
BRIEF-BMM-eFV	74.06	67.83
GLC [13]	93.00	-
SC [9]	95.47	-
CS [20]	97.00	-
GLS [12]	97.19	95.00
ICM [22]	97.82	-

As we can see in table 3 the best performant descriptors encoded with eFV are SIFT and COV, and their accuracy on Flavia and Foliage is above the recents method proposed in the literature. We have the say that in Foliage the baseline CNN+SVM has the best accuracy. In tables 4 and 5, we compare the results of our algorithm with the best results on the PlantCLEF2012 and PlantCLEF2013 challenges. The score is computed using the scripts provided with the datasets. In bold letters we highlight the best accuracy for each type of image. For these two datasets we only show the accuracy for SIFT and COV descriptors.

In PlantCLEF2012 dataset (table 4) the enconding of SIFT descriptors with eFV shows the best performance for Scan-like, Photos and Average, and the baseline system CNN+SVM, shows the best performance for Scan type of images.

For the dataset PlantCLEF2013, the best performance for SheetAsBackground images is obtained with the method proposed in [24] but this method

Table 4. Classification results on PlantCLEF2012 for the 3 types of images and on average.

Method	Scan	Scan-like	Photos	Average
CNN+SVM	**0.65**	0.51	0.40	0.520
SIFT-PCA-GMM-eFV	0.62	**0.74**	**0.44**	**0.60**
COV-WMM-eFV	0.481	0.432	0.240	0.384
SABANCI OKAN [25]	0.58	0.55	0.22	0.16
INRIA [2]	0.39	0.59	0.21	0.40
LSIS DYNI [16]	0.41	0.42	0.32	0.42

fails for the NaturalBackground images as we can see in the table 5. Also this method has the drawbacks of a preprocessing segmentation step which is inapplicable for NaturalBackground images. For NaturalBackground images, the best performance is achieved with the method presented in [15] based in a complex scheme of late-fusion of 4 versions of SIFT and self-similarity encoded with a polynomial embedding of descriptors encoded with FV. Also, the last method uses metadata information of the test set, in particular the type of NaturalBackground image.

Table 5. Classification results on PlantCLEF2013 for the 2 types of images.

Method	SheetAsBackground	NaturalBackground
CNN+SVM	0.557	0.403
SIFT-PCA-GMM-eFV	0.594	0.365
COV-WMM-eFV	0.363	0.181
SABANCI OKAN [24]	**0.607**	0.181
NlabUTokio [15]	0.502	**0.393**

5 Conclusions

We present a detailed empirical evaluation of different eFV configurations applied to the problem of plant identification. We perform experiments on different public datasets and compare our results with state of the art algorithms and the obtained results in some experiments are better than the state of the art. In most of the cases the best configuration is SIFT descriptors encoded with eFV, but the baseline using CNN and SVM also performs very well.

The advantages of the proposed method are that it does not need a preprocessing step for the leaf countor extraction because it is based on local descriptors, it permits the use of different descriptors in an unified framework, it is not based in handcrafted or ad-hoc descriptors and it is simpler than some of the existing algorithms. Furthermore, unlike other methods it can be applied on images of leaves with a simple background or with complex background as we demonstrated on the experiments.

References

1. Bakic, V., Mouine, S., Ouertani-Litayem, S., Verroust-Blondet, A., Yahiaoui, I., Goëau, H., Joly, A.: Inria's participation at imageclef 2013 plant identification task. In: CLEF (Online Working Notes/Labs/Workshop) 2013 (2013)
2. Bakic, V., Yahiaoui, I., Mouine, S., Ouertani, S.L., Ouertani, W., Verroust-Blondet, A., Goëau, H., Joly, A.: Inria imedia2's participation at imageclef 2012 plant identification task. In: CLEF (Online Working Notes/Labs/Workshop) 2012 (2012)
3. Chaki, J., Parekh, R., Bhattacharya, S.: Plant leaf recognition using texture and shape features with neural classifiers. Pattern Recognition Letters 58, 61–68 (2015)
4. Chen, Q., Abedini, M., Garnavi, R., Liang, X.: Ibm research australia at lifeclef2014: Plant identification task. In: Working notes of CLEF 2014 conference (2014)
5. Goëau, H., Bonnet, P., Joly, A., Bakic, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The imageclef 2013 plant identification task. In: CLEF (2013)
6. Goëau, H., Bonnet, P., Joly, A., Yahiaoui, I., Bakic, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The imageclef 2012 plant identification task. In: CLEF (2012)
7. Goëau, H., Joly, A., Bonnet, P., Selmi, S., Molino, J.F., Barthélémy, D., Boujemaa, N.: Lifeclef plant identification task 2014. CLEF2014 Working Notes. Working Notes for CLEF 2014 Conference, Sheffield, UK, September 15-18, 2014 pp. 598–615 (2014)
8. Gosselin, P.H., Murray, N., Jégou, H., Perronnin, F.: Revisiting the fisher vector for fine-grained classification. Pattern Recognition Letters 49, 92–98 (2014)
9. Hsiao, J.K., Kang, L.W., Chang, C.L., Lin, C.Y.: Comparative study of leaf image recognition with a novel learning-based approach. In: Science and Information Conference (SAI), 2014. pp. 389–393. IEEE (2014)
10. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R., Guadarrama, S., Darrell, T.: Caffe: Convolutional architecture for fast feature embedding. In: Proceedings of the ACM International Conference on Multimedia. pp. 675–678. ACM (2014)
11. Kadir, A., Nugroho, L., Susanto, A., Santosa, P.: Neural network application on foliage plant identification. International Journal of Computer Applications 29(9), 15–22 (2011)
12. Kadir, A.: A model of plant identification system using glcm, lacunarity and shen features. arXiv preprint arXiv:1410.0969 (2014)
13. Kalyoncu, C., Toygar, Ö.: Geometric leaf classification. Computer Vision and Image Understanding (2014)
14. Ken Chatfield, Victor Lempitsky, A.V., Zisserman, A.: The devil is in the details: an evaluation of recent feature encoding methods. In: Proc. BMVC. pp. 76.1–76.12 (2011), http://dx.doi.org/10.5244/C.25.76
15. Nakayama, H.: Nlab-utokyo at imageclef 2013 plant identification task. In: Working notes of CLEF 2013 conference (2013)
16. Paris, S., Halkias, X., Glotin, H.: Participation of lsis/dyni to imageclef 2012 plant images classification task. In: CLEF (Online Working Notes/Labs/Workshop) (2012)

17. Razavian, A.S., Azizpour, H., Sullivan, J., Carlsson, S.: Cnn features off-the-shelf: an astounding baseline for recognition. In: Computer Vision and Pattern Recognition Workshops (CVPRW), 2014 IEEE Conference on. pp. 512–519. IEEE (2014)
18. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.J.: Image classification with the fisher vector: Theory and practice. International Journal of Computer Vision 105(3), 222–245 (2013)
19. Sánchez, J., Redolfi, J.: Exponential family fisher vector for image classification. Pattern Recognition Letters 59, 26–32 (2015)
20. Sfar, A.R., Boujemaa, N., Geman, D.: Confidence sets for fine-grained categorization and plant species identification. International Journal of Computer Vision pp. 1–21 (2014)
21. Tuzel, O., Porikli, F., Meer, P.: Region covariance: A fast descriptor for detection and classification. In: Computer Vision–ECCV 2006, pp. 589–600. Springer (2006)
22. Wang, Z., Sun, X., Ma, Y., Zhang, H., Ma, Y., Xie, W., Zhang, Y.: Plant recognition based on intersecting cortical model. In: Neural Networks (IJCNN), 2014 International Joint Conference on. pp. 975–980. IEEE (2014)
23. Wu, S.G., Bao, F.S., Xu, E.Y., Wang, Y.X., Chang, Y.F., Xiang, Q.L.: A leaf recognition algorithm for plant classification using probabilistic neural network. In: Signal Processing and Information Technology, 2007 IEEE International Symposium on. pp. 11–16. IEEE (2007)
24. Yanikoglu, B., Aptoula, E., Yildiran, S.T.: Sabanci-okan system at imageclef 2013 plant identification competition. In: Working notes of CLEF 2013 conference (2013)
25. Yanikoglu, B.A., Aptoula, E., Tirkaz, C.: Sabanci-okan system at imageclef 2012: Combining features and classifiers for plant identification. In: CLEF (Online Working Notes/Labs/Workshop) (2012)