

Sabancı-Okan System at LifeCLEF 2014 Plant Identification Competition

Berin Yanikoglu¹, S. Tolga Yildiran¹, Caglar Tirkaz¹, and Erchan Aptoula²

¹ Sabanci University, Istanbul, Turkey

² Okan University, Istanbul, Turkey

{berrin, stolgay, caglart}@sabanciuniv.edu

erchan.aptoula@okan.edu.tr

Abstract. We describe our system in 2014 LifeCLEF [1] Plant Identification Competition. The sub-system for isolated leaf category (LeafScans) was basically the same as last year [2], while plant photographs in all the remaining categories were classified using either local descriptors or deep learning techniques. However, due to large amount of data, large number of classes and shortage of time, our system was not very successful in the plant photograph sub-categories; but we obtained better results in isolated leaf images. As announced by the organizers, we obtained an inverse rank score of 0.127 overall and 0.449 for isolated leaves.

1 Overview

Plant identification campaign within LifeCLEF 2014 was similar to previous years, but it was in larger scale with twice the number of classes and images as of last year [3]. The dataset consists of isolated leaf images called LeafScans, consisting of scanned or scan-like leaf photographs and plant photographs in different categories (e.g. Flower, Fruit, Stem). In summary, the dataset contained a total of 47,815 images (11,335 pictures of isolated leaves and 36,480 plant photographs).

Our submission was very similar to that submitted in 2013 for the case of isolated leaves, while we started to build a new system for plant photographs using local descriptors and deep learning techniques. We split the task to share the work load: for Flower, Fruit and Entire categories, we used dense-SIFT descriptors, while for Branch and Leaf categories we used convolutional neural networks (CNN). The stem category was also recognized using globally extracted texture and color features, similar to LeafScans.

In both of these two approaches, the main problem was the large number of classes, which resulted in long training times. As a result, we exploited the meta-data wherever applicable: namely to split the flower/fruit categories according to flowering/fruit bearing periods and trained a separate classifier for each time period. In this way, we aimed to reduce the number of classes dealt by each classifier. In the other categories (Stem, Branch, Entire and Leaf) time information did not seem very useful and was not used; in fact we did not use meta-data anywhere else in the system.

2 Preprocessing

Preprocessing stages were present only for the isolated leaf images. Specifically, we align the leaf's major axis with the vertical through principal component analysis, with additional correction coming from the leaf petiole's location. Then size normalization is achieved by normalizing the leaf height to 600 pixels, preserving the aspect ratio. Orientation normalization done this way is quite successful, however is not error-proof. Furthermore, errors in orientation normalization typically cause overall errors because most of the features are sensitive to orientation. There was no preprocessing for plant photographs.

3 Features

3.1 Features for LeafScans

The descriptors used for characterizing the samples of the LeafScan category are identical to those used during the plant identification track of ImageCLEF 2013 [2]. In detail they are as follows:

Basic Shape Statistics (BSS): provides contour related information by computing basic statistical measures from the distance to centroid curve. In particular, once the image centroid is located, we compute the contour pixels' Euclidean distance to it, thus resulting in a numerical sequence. After sorting the said sequence, we extract the following basic measures from it:

$$BSS = \{maximum, minimum, median, variance\} \quad (1)$$

Area Width Factor (AWF) is computed on grayscale and constitutes a slight variation of the *leaf width factor* introduced in Ref. [4]. Specifically, given an isolated leaf image, it is first divided into n strips, perpendicular to its major axis. For the final n -dimensional feature, we compute the volume (Vol , i.e. sum of pixel values) of each strip (Vol_i) normalized by the global volume (Vol):

$$AWF = \{Vol_i/Vol\}_{1 \leq i \leq n} \quad (2)$$

Regional Moments of Inertia (RMI) is relatively similar to AWF. It requires an identical image subdivision system, differing only in the characterization of each strip. To explain, instead of using the sum of pixel values, each strip is described by means of the mean Euclidean distance between its centroid and contour pixels [5].

Angle Code Histogram (ACH) has been used in Ref. [6] for tree leaf classification. Given the binary segmentation mask, it consists in first subsampling the contour points, followed by computing the angles of successive point triplets. The final feature is formed by the normalized histogram of the computed angles.

Edge Background/Foreground Ratio Histogram is computed on the binary mask of its input and it consists in calculating the ratio of background to foreground pixels in a subwindow centered on each edge pixel. The normalized histogram of the said ratios constitutes the end feature vector.

Orientation Histogram (OH) is computed on grayscale data. After computing the orientation map using a 11x11 edge detection operator for determining the dominant orientation at each pixel, the feature vector is computed as the normalized histogram of n bins of dominant orientations.

Circular Covariance Histogram (CCH) and Rotation Invariant Point Triplets (RIT) are both texture descriptors [7] based on the morphological covariance operator. They operate on grayscale images, and focus on extracting periodicity patterns by means of morphological openings and closings using circular structuring elements.

Color Auto-correlogram (AC) was used for color description [8]. It was computed in the LSH color space after a non-uniform subquantization to 63 colors (7 levels for hue, 3 for saturation and 3 for luminance). The color autocorrelogram describes the spatial correlation of colors. It consists of a table where the entry (i, j) denotes the probability of encountering two pixels of color i at a distance of j pixels.

Saturation-weighted Hue Histogram (SWHH) was also used as a color descriptor, where the total value of each bin $W_\theta, \theta \in [0, 360]$ is calculated as:

$$W_\theta = \sum_x S_x \delta_{\theta H_x} \quad (3)$$

where H_x and S_x are the hue and saturation values at position x and δ_{ij} the Kronecker delta function [9]. As far as the color space is concerned, we used LSH [10] since it provides a saturation representation independent of luminance.

We used a single Support Vector Machine (SVM) classifier, as described in [2], to classify LeafScan images based on these features.

3.2 Features for Plant Photos

We used local descriptors and deep learning techniques for the task of recognizing photographs of plants, where shape descriptors are not useful and global color and texture information is of limited use. In Stem category only, we used texture and color information globally, as the task seemed somewhat easier than others and due to lack of time.

Local Descriptors - Dense SIFT This approach is applied to Flower, Fruit and Entire categories. Due to the large number of classes that increase training times and decrease accuracy, we first divided each category into sub-categories,

using the date information from meta-data. Then we trained a separate classifier for each sub-category (15 of them in total), in order to reduce the problems and increase accuracy.

In the dense-SIFT approach, 16-by-16 patches are collected from images and SIFT descriptors are calculated for each patch [11, 12]. Then, these descriptors are clustered using k-means clustering [13] to produce a visual word dictionary. We experimented with the size of this dictionary and selected 1,200 words, as it gave the best result with validation data. Once the visual word dictionary is selected, each image is described as a histogram of these visual words, using the Bag of Words model [14]. Finally, a linear Support Vector Machine (SVM) is trained to differentiate between classes in each sub-category, using these histograms as features [15].

We used the Weka toolbox [16] for the implementation where Stochastic Dual Coordinate Ascent [17] solver method is used as the optimization method. For parameter optimization, grid search is used, where $C = 10$ was selected as the cost value.

Deep Learning - Convolutional Neural Network: We trained a convolutional neural network (CNN) for the Leaf and Branch categories, that were seen as the most challenging sub-categories. The CNN we employed contains 8 layers where the output of the last fully connected softmax layer produces a distribution over the class labels. The first and the fourth layers are convolutional layers with 5x5 kernels; the second and the fifth layers are local contrast normalization layers; the third and the sixth layers are max-pooling layers; the seventh layer is a locally connected convolutional layer. The Rectified Linear Units (ReLU) non-linearity is applied to the output of every convolutional and fully-connected layer [18–20]

In order to process each color image, the largest square region of the image is cropped and down-scaled to 32x32x3 for further processing. The first convolutional layer filters the 32x32x3 input image with 32 kernels of size 5x5x3; the second convolutional layer takes as input the (response-normalized and pooled) output of the first convolutional layer and filters it with 64 kernels of size 5x5x32; the third convolutional layer, a locally-connected layer with unshared weights, has 64 kernels of size 3x3x64; the final layer is a fully connected soft-max layer.

In order to reduce over-fitting we used two techniques: data augmentation and drop-out. For data augmentation we applied label preserving transformations on the image data such as reflections, small rotations and translations whereas dropout is employed in the third convolutional layer, just before the soft-max layer.

Global Features - Stem Category: For the characterization of the samples of the Stem category, we employed a combination of texture and color descriptors obtained from the whole image. The reason was two-fold: lack of time and the impression that this sub-category was relatively simpler compared to other sub-categories.

We used as descriptors those used with isolated leaf images (i.e. OH, CCH, RIT), as well as the *Morphological Covariance* (MC). *MC* is the morphological equivalent of the usual covariance operator. It is based on successive erosions by means of point couples at various distances and orientations:

$$MC(f) = Vol(\varepsilon_{P_{2,v}}(f)) / Vol(f) \quad (4)$$

where ε denotes the erosion operator, $P_{2,v}$ a point couple separated by a vector \mathbf{v} and Vol the image volume, i.e. the sum of pixel values. As such it can capture the directionality as well as periodicity of its input [7].

4 Results

For training classifiers and validating our methods, we split the provided training set into two (*Train* and *Validation* sets), separately for each sub-category. In doing the split of the available development data, we kept all images of an individual plant in the *Train* or *Validation* splits, in order to reduce overfitting.

We report the top-1 recognition accuracy of our system on this *Validation* data, in the fourth column of Table 1, since the labels for the Test set are not yet available. We also report the inverse-rank scores obtained from *Test* data, announced by the competition organizers, in the fifth column. Note that the top-1 accuracies and the inverse-rank scores are not directly comparable, due to data set change, different metrics, and also because the official score uses a user-based averaging, while the accuracy results use an average over images.

Table 1. Accuracy in each subcategory, as measured during development over validation split of the development data.

Content	# Train images/ /#Classes	# Validation images/ /#Classes	Accuracy (%)	Inverse-Rank
Global				
LeafScan	9,532/212	1,802/133	69.70	0.449
Stem	2,526/368	940/282	25.15	0.089
Dense-SIFT				
Flower	11,169/484	1,993/464	21.74	0.149
Fruit	3,033/374	721/269	19.12	0.118
Entire	5,348/490	1,008/476	9.55	0.077
CNN				
Leaf	5,660/470	2,094/439	20.68	0.066
Branch	1,414/356	573/270	12.64	0.007

The way we split the training data resulted in having fewer number of classes in the *Validation* set, as seen in Table 1, but we did not want to split images of the same individual plant across the two sets (as some of them could be very similar). As a result, the validation performance was not fully informative, as some classes were not in the set; but this was considered preferable to alternatives.

Our results are not very good except for the LeafScan category where the task is easier and we have obtained very good results in the last years [21, 2]. In fact, all of our inverse-rank scores are significantly lower compared to last year's results. This may be attributed to the more complex problem (with almost twice the number of classes) and also no time to quite polish up our new methods.

Acknowledgements This project is supported by Turkish Scientific and Research Council of Turkey (TUBITAK), under project number 113E499.

References

1. Joly, A., Müller, H., Goëau, H., Glotin, H., Spampinato, C., Rauber, A., Bonnet, P., Vellinga, W.P., Fisher, B.: Lifeclef 2014: multimedia life species identification challenges. In: Proceedings of CLEF 2014. (2014)
2. Yanikoglu, B., Aptoula, E., Yildiran, S.T.: Sabanci-Okan system at ImageClef 2013. In: CLEF (Notebook Papers/Labs/Workshop). (2013)
3. Goëau, H., Joly, A., Bonnet, P., Molino, J.F., Barthélémy, D., Boujemaa, N.: Lifeclef plant identification task 2014. In: CLEF working notes 2014. (2014)
4. Hossain, J., Amin, M.A.: Leaf shape identification based plant biometrics. In: International Conference on Computer and Information Technology, Dhaka, Bangladesh (2010) 458–463
5. Knight, D., Painter, J., Potter, M.: Automatic plant leaf classification for a mobile field guide (2010)
6. Wang, Z., Chi, Z., Feng, D.: Shape based leaf image retrieval. *IEE Proceedings in Vision, Image and Signal Processing* **150** (2003) 34–43
7. Aptoula, E.: Extending morphological covariance. *Pattern Recognition* **45** (2012) 4524–4535
8. Huang, J., Kumar, S.R., Mitra, M., Zhu, W.J., Zabih, R.: Image indexing using color correlogram. In: CVPR, San Juan, Puerto Rico (1997) 762–768
9. Hanbury, A.: Circular statistics applied to colour images. In: Computer Vision Winter Workshop, Valtice, Czech Republic (2003) 55–60
10. Aptoula, E., Lefèvre, S.: On the morphological processing of hue. *Image and Vision Computing* **27** (2009) 1394–1401
11. Lowe, D.G.: Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision* **60** (2004) 91–110
12. Liu, C., Yuen, J., Torralba, A., Sivic, J., Freeman, W.T.: Sift flow: Dense correspondence across different scenes. In: Proceedings of the 10th European Conference on Computer Vision: Part III. ECCV '08, Berlin, Heidelberg, Springer-Verlag (2008) 28–42
13. Kanungo, T., Mount, D.M., Netanyahu, N.S., Piatko, C.D., Silverman, R., Wu, A.Y.: An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Trans. Pattern Anal. Mach. Intell.* **24** (2002) 881–892
14. Wallach, H.M.: Topic modeling: Beyond bag-of-words. In: Proceedings of the 23rd International Conference on Machine Learning. ICML '06, New York, NY, USA, ACM (2006) 977–984
15. Joachims, T.: Training linear svms in linear time. In: Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '06, New York, NY, USA, ACM (2006) 217–226

16. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H.: The weka data mining software: an update. *SIGKDD Explor. Newsl.* **11** (2009) 10–18
17. Hsieh, C.J., Chang, K.W., Lin, C.J., Keerthi, S.S., Sundararajan, S.: A dual coordinate descent method for large-scale linear svm. In: *Proceedings of the 25th International Conference on Machine Learning. ICML '08*, New York, NY, USA, ACM (2008) 408–415
18. Nair, V., Hinton, G.E.: Rectified linear units improve restricted boltzmann machines. In *Frnkranz, J., Joachims, T., eds.: ICML, Omnipress (2010)* 807–814
19. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In *Bartlett, P.L., Pereira, F.C.N., Burges, C.J.C., Bottou, L., Weinberger, K.Q., eds.: NIPS. (2012)* 1106–1114
20. Hinton, G.E., Srivastava, N., Krizhevsky, A., Sutskever, I., Salakhutdinov, R.: Improving neural networks by preventing co-adaptation of feature detectors. *CoRR abs/1207.0580* (2012)
21. Yanikoglu, B., Aptoula, E., Tirkaz, C.: Sabanci-Okan system at ImageClef 2012: Combining features and classifiers for plant identification. In: *CLEF (Notebook Papers/Labs/Workshop)*. (2012)