Participation of INRIA & Pl@ntNet to ImageCLEF 2011 plant images classification task

Hervé Goëau[1], Alexis Joly[1], Itheri Yahiaoui[1], Pierre Bonnet[2], and Elise Mouysset[3]

[1] INRIA, IMEDIA team, France, name.surname@inria.fr,
http://www-rocq.inria.fr/imedia/
[2] CIRAD, UMR AMAP, France, pierre.bonnet@cirad.fr,
http://amap.cirad.fr/fr/index.php
[3] Tela Botanica, France, elise@tela-botanica.org, http://www.tela-botanica.org/

Abstract. This paper presents the participation of INRIA IMEDIA group and the Pl@ntNet project to ImageCLEF 2011 plant identification task. ImageCLEF's plant identification task provides a testbed for the system-oriented evaluation of tree species identification based on leaf images. The aim is to investigate image retrieval approaches in the context of crowdsourced images of leaves collected in a collaborative manner. IMEDIA submitted two runs to this task and obtained the best evaluation score for two of the three image categories addressed within the benchmark. The paper presents the two approaches employed, and provides an analysis of the obtained evaluation results.

Keywords: Pl@ntNet, IMEDIA, INRIA, ImageCLEF, plant, leaves, images, collection, identification, classification, evaluation, benchmark

1 Introduction

This paper presents the participation of INRIA IMEDIA group and the Pl@ntNet[4] project to the *plant identification task* that was organized within ImageCLEF 2011[5] for the system-oriented evaluation of visual based plant identification. This first year pilot task was more precisely focused on tree species identification based on leaf images. The task was organized as a classification task over 70 tree species with visual content being the main available information. Three types of image content were considered: leaf *scans*, leaf photographs with a white uniform background (referred as *scan-like* pictures) and unconstrained leaf's *photographs* acquired on trees with natural background. IMEDIA group, in collaboration with the Pl@ntNet project submitted two runs, one based on large-scale local features matching and rigid geometrical models, the other one based on segmentation and shape boundary features.

[4] http://www.plantnet-project.org/papyrus.php?langue=en
[5] http://www.imageclef.org/2011

2 Task description

The task was evaluated as a supervised classification problem with tree species used as class labels.

2.1 Training and Test data

A part of Pl@ntLeaves dataset was provided as training data whereas the remaining part was used later as test data. The training subset was built by randomly selecting 2/3 of the **individual plants** of each species (and not by randomly splitting the images themselves). So that pictures of leaves belonging to the same individual tree cannot be split across training and test data. This prevents identifying the species of a given tree thanks to its own leaves and that makes the task more realistic. In a real world application, it is indeed much unlikely that a user tries to identify a tree that is already present in the training data. Detailed statistics of the composition of the training and test data are provided in Table 1.

		Nb of pictures	**Nb of individual plants**	**Nb of contributors**
Scan	**Train**	2349	151	17
	Test	721	55	13
Scan-like	**Train**	717	51	2
	Test	180	13	1
Photograph	**Train**	930	72	2
	Test	539	33	3
All	**Train**	3996	269	17
	Test	1440	99	14

Table 1. Statistics of the composition of the training and test data

2.2 Task objective and evaluation metric

The goal of the task was to associate the correct tree species to each test image. Each participant was allowed to submit up to 3 runs built from different methods. As many species as possible can be associated to each test image, sorted by decreasing confidence score. Only the most confident species was however used in the primary evaluation metric described below. But providing an extended ranked list of species was encouraged in order to derive complementary statistics (e.g. recognition rate at other taxonomic levels, suggestion rate on top k species, etc.).

The primary metric used to evaluate the submitted runs was a *normalized classification rate* evaluated on the 1st species returned for each test image. Each test image is attributed with a score of 1 if the 1st returned species is correct and 0 if it is wrong. An average *normalized* score is then computed on all test images. A simple mean on all test images would indeed introduce some bias with regard to a real world identification system. Indeed, we remind that the Pl@ntLeaves dataset was built in a collaborative manner. So that few contributors might have provided much more pictures than many other contributors who provided few. Since we want to evaluate the ability of a system to provide correct answers to all users, we rather measure the mean of the average classification rate per author. Furthermore, some authors sometimes provided many pictures of the same individual plant (to enrich training data with less efforts). Since we want to evaluate the ability of a system to provide the correct answer based on a single plant observation, we also decided to average the classification rate on each individual plant. Finally, our primary metric was defined as the following average classification score S:

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

U : number of users (who have at least one image in the test data)
P_u : number of individual plants observed by the u-th user
$N_{u,p}$: number of pictures taken from the p-th plant observed by the u-th user
$s_{u,p,n}$: classification score (1 or 0) for the n-th picture taken from the p-th plant observed by the u-th user

It is important to notice that while making the task more realistic, the normalized classification score also makes it more difficult. Indeed, it works as if a bias was introduced between the statistics of the training data and the one of the test data. It highlights the fact that bias-robust machine learning and computer vision methods should be preferred to train such real-world collaborative data. Finally, to isolate and evaluate the impact of the image acquisition type (scan, scan-like, photograph), a normalized classification score S was computed for each type separately. Participants were therefore allowed to train distinct classifiers, use different training subsets or use distinct methods for each data type.

3 Description of used methods

3.1 Large-scale local features matching and rigid geometrical models → *inria_imedia_plantnet_run1*

State-of-the-art methods addressing leaf-based identification of leaves are mostly based on leaf segmentation and shape boundary features [2, 12, 3, 15, 1]. Segmentation-based approaches have however several strong limitations including the presence of clutter and background information as well as other acquisition shortcomings (shadows, leaflets occlusion, holes, cropping, etc.). These issues are particularly critical in a crowdsourcing environment where we do not control accurately the acquisition protocol. Alternatively, our first run is based on local features and large-scale matching. Indeed, we realized that large-scale object retrieval methods [14, 10], usually aimed at retrieving rigid objects (buildings, logos, etc.), do work surprisingly well on leaves. This can be explained by the fact that even if only a small fraction of the leaf remains affine invariant, this is sufficient to discriminate it from other species. Concretely, our system is based on the following steps: (i) Local features extraction (mixed texture & shape features computed around Harris points) (ii) Local features matching with an efficient hashing-based indexing scheme (iii) Spatially consistent matches filtering with a RANSAC algorithm using a rigid transform model (iv) Basic top-K decision rule as classifier: for each species, the number of occurrences in the top-K images returned is used as its score.

Besides clutter robustness, the method has several advantages: it does not require any complex training phase allowing fast dynamic insertion of new crowdsourced training data, and it is weakly affected by unbalanced class distribution thanks to the selectivity of the spatial consistency filtering.

Mixed texture & shape local features Rather than using classical SIFT features computed around DoG points, we employed multi-resolution color Harris points ([5] and [8]). Indeed, we remarked that Harris corners were much more representative of relevant patterns of the leaves than the DoG points. Leaf boundary corners detected by Harris detector are notably much more stable than the blobs detected by DoG (which are visually mainly noise). We used the color version of Harris detector [5] that has usually a better repeatability. Finally we extracted the points at four distinct resolutions (as in [8]) to deal with scaling and blurring (with a scale factor equal to 0.8 between each resolution). The number of Harris points extracted per image was limited to 500 (with a log-scale maximum number of points per resolution).

Local features: hough_4_4, eoh_8, fourier_8_32 are extracted around each Harris point from an image patch oriented according to the principal orientation and scaled according to the resolution at which the Harris corner was detected.

hough_4_4 is a 16 dimensional histogram based on ideas inspired from the Hough transform and is used to represent simple shapes in an image [4].

fourier_8_32 is a Fourier histogram used as a texture descriptor describing the distribution of the spectral power density within the complex frequency plane. It can differentiate between the low, middle and high frequencies and between different angles the salient features have in a patch [4].

eoh_8 is a 8 dimensional classical Edge Orientation Histogram used for describing shapes in images and gives here the distribution of gradients on 8 directions in a patch.

Finally, we use as local features the concatenation of these 3 local features, resulting in a 280-dimensional feature vector extracted around each Harris point.

Local features compression with RMMH Random Maximum Margin Hashing (RMMH) [7] is a new data dependent hashing method that we recently introduced for the efficient embedding of high-dimensional feature vectors. The main idea of RMMH is to train balanced and independent binary partitions of the high-dimensional space by training svm's on purely random splits of the data, regardless the closeness of the training samples and without any form of supervision. It allows to generate consistently more independent hash functions than previous data dependent hashing methods while keeping a better embedding than classical data independent random projections such as LSH [7]. In this work, each local feature vector was embedded into a 256-bits hash code using RMMH with a linear kernel (inner product) and M=32 training samples per hash function (i.e. per bit). The distance between two local features is finally computed as a Hamming distance between their two hash codes.

Local features indexing and matching with AMP-RMMH We also used RMMH for indexing purposes using the multi-probe hashing method described in [9]. The 20 first bits of the hash codes were used to create a hash table and all binary hash codes of the full training set were mapped into it (resulting in about 2 millions 256-bits hash codes mapped in a 2^{20} size hash table). At query time, each local feature of the query image is compressed with RMMH through a 256-bit hash code and its approximate 600-nearest neighbors are searched by probing multiple neighboring buckets in the hash table (according to the a posteriori multi-probe algorithm described in [9]). This step returns a large set of candidate local feature matches than can be reorganized image by image to finally obtain a set of candidate images each with a set of candidate matches.

Reranking with rigid geometrical models A last step is finally applied to re-rank the candidate images (retrieved from the training set) according to their geometrical consistency with the query local features (as in [14] or [10]). We therefore estimate a translation-scale-rotation geometric model between the query image and each retrieved image. This is done using a RANSAC-like algorithm working only on points positions, so that it uses random pairs of matches to build candidate transform parameters. The final score for each image is computed as the number of inlier matches (i.e. the ones that respect the estimated translation-scale-rotation geometric model). All images that were returned by the former step are finally re-ranked according to this geometrical consistency score.

Classification with a top-k decision rule Best species label is finally computed by voting on the top-10 returned training images (ranked by geometrical consistency score).

Training data strategy Since training and test leaf images are categorized in three distinct image types (scans, scan-like photos and unconstrained photos), an important question is which training images types should be used for which test image type. Few leave-one-out experiments performed on the training set itself did show us that using only scans as training images for all test images might be more effective than other strategies (e.g. using all training images for all test image types or using only the same image type for training and testing). This can be explained by the fact that scan images do not contain any noisy background so that all local features included in the trained index are actually parts of the leave and not distractors as in unconstrained photographs.

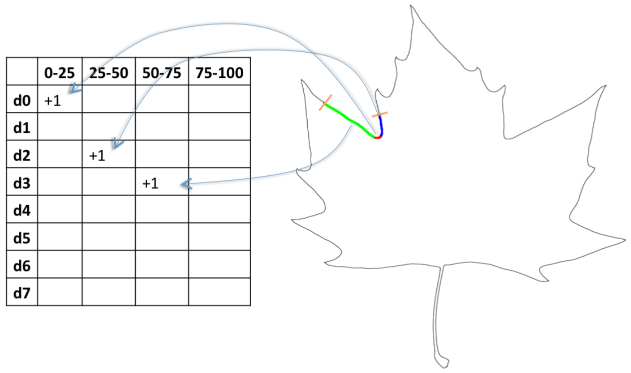
3.2 Directional Fragment Histogram and geometric parameters on shape boundary→ *inria_imedia_plantnet_run2*

The method used in the second run is very distinct from the first one and is closer from state-of-the-art methods based on leaf segmentation and shape boundary features. We use a shape boundary descriptor called Directional Fragment Histogram introduced in a previous work on botanical images database [15], and to combine it with usual geometric parameters on shapes. The method described below focuses on scans and scan-like images. For photographs, results were produced by using classical global descriptors (Fourier histogram, Hough histogram, HSV color histogram, Global and Local Edge Orientation Histograms) implemented in the framework developed in IMEDIA team (more details of these global descriptors can be found in [4] and [6]).

As almost boundary-based shape description methods, the first step deals with image segmentation. We use the classical Otsu adaptive thresholding method [13], applied widely in the literature due to its content-independent characteristic. Then two distinct feature extractions are applied in order to obtain a set of two vector descriptors, one containing the boundary description with a Directional Fragment Histogram, and the other containing 8 distinct geometric parameters.

Boundary description with Directional Fragment Histogram This method was introduced and applied successfully in a previous work on botanical data in 2006 [15]. The main idea is to consider that each element of a contour has a relative orientation with respect to its neighbors. The method consists in to slide a segment over the contour of the shape and to identify groups of elements having the same direction within the segment. Such groups are called Directional Fragments, and then the DFH codes the frequency distribution and relative length of groups of elements. Figure 1 gives an example of the extraction of the Directional Fragment Histogram during one position of the sliding segment (colored in three fragments green, red and blue) along the contour. In this example the DFH is a 32-dimensional histogram given by 8 orientations d_0 to d_7 combined with 4 balanced ranges of relative lengths, (the lengths of the fragments seen as a percentage of the segment length). In this position the sliding segment is counted 3 times at 3 distinct orientations and lengths. At the end of the procedure DFH is finally normalized by the number of all the possible segments.

Boundary description with geometric parameters In order to improve performances in plant identification, we chose to combine the DHF descriptor



	0-25	25-50	50-75	75-100
d0	+1			
d1				
d2		+1		
d3			+1	
d4				
d5				
d6				
d7				

Fig. 1. Extraction of the Directional Fragment Histogram during one position of the sliding segment (colored in three fragments green, red and blue).

with 8 morphological features used in plant identification literature, like Aspect Ratio, Rectangularity, Convex Area Ratio, Convex Perimeter Ratio, Sphericity, Eccentricity and Form Factor. The table 2 gives the 8 geometric parameters used for the task. Most of these parameters were succesfuly experimented in [11], but on a limited numbers of 6 species related in fact to 6 very distinct morphological categories of simple leaf shapes. The Plant Identification task was the opportunity to experiment these shape parameters on much more species, on much more morphological categories of leaf shapes, with simple and compound leaves, and for certain with more visual ambiguities between species.

Classification with a top-k decision rule Finally, the boundary is described by two vectors, one 8-dimensional vector containing the shape parameters, and a DFH histogram. A balanced weighted sum of L1 distances on these two vectors is used as similarity measure between an image test and a training image. Best species label is finally computed by voting on the top-10 returned training images, as in the previous first run.

4 Results

Figures 2, 3 and 4 present the normalized classification scores of the 20 submitted runs for each of the three image types. Figure 5 presents the mean performances averaged over the 3 image types. Table 3 finally presents the same results but with detailed numerical values.

The two runs submitted by IMEDIA, in spite of theirs theoretical differences, gave both good results, and obtained the best evaluation scores for two

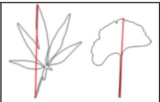
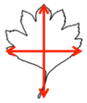
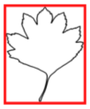
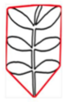
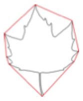
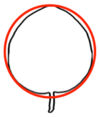
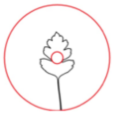
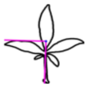
Diameter	Maximum length contained in the shape.		
Aspect Ratio	Ratio between the maximum length D_{max} and the minimum lenght D_{min} of the minimum bounding box of the shape.		$\frac{D_{max}}{D_{min}}$
Rectangularity	Ratio between the area A_s of the shape and the area A_b of the minimal bounding box.		$\frac{A_s}{A_b}$
Convex Area Ratio	Ratio between the shape area A_s and the convex hull area A_h.		$\frac{A_s}{A_h}$
Convex Perimeter Ratio	Ratio between the shape perimeter P_s and the convex hull perimeter P_h.		$\frac{P_s}{P_h}$
Form Factor	The form factor can be interpreted as the "roundness" of the shape and is a ratio between the area A_s and the (squared) perimeter P_s of the shape.		$\frac{4\pi A_s}{P_s^2}$
Sphericity	Ratio between the radius r_i of the incircle of the shape and the radius r_e of the excircle of the shape.		$\frac{r_i}{r_e}$
Eccentricity	Ratio of the major principal axis λ_1 over the minor principal axis λ_2		$\frac{\lambda_1}{\lambda_2}$

Table 2. The eight geometric parameters used in the second run *inria_imedia_plantnet_run2*.

of the three image categories addressed within the benchmark, on scans for the run *inria_imedia_plantnet_run1*, and on scan-like images for the second run *inria_imedia_plantnet_run2*.

Considering the first run, the approach based on large-scale local features matching and rigid geometrical models gives surprisingly better results on scans than state of the arts methods based on shape boundary features.

Considering the image types and the results for all teams, performances are degrading with the complexity of the acquisition image type. Indeed, scans are more easy to identify than scan-like photos and unconstrained photos are much more difficult. This is can be seen in figure 5 where the relative scores of each image type are highlighted by distinct colors. However, if this "rule" is true for the first run *inria_imedia_plantnet_run1* , it is not for the second run *inria_imedia_plantnet_run2* (and also for 5 other runs). It is difficult to give a precise reason of these results, but numerous unsuccessful scan tests have a relatively poor quality, coming from a low resolution original scan, noisy with a non uniform and gradually yellow colored background with blurred content. These unsuccessful scans maybe indicate a weakness at the very first step of automatic segmentation.

Run id	Participant	Scans	Scan-like	Photographs	Mean
IFSC USP_run2	IFSC	**0,562**	0,402	**0,523**	**0,496**
inria_imedia_plantnet_run1	INRIA	**0,685**	0,464	0,197	**0,449**
IFSC USP_run1	IFSC	0,411	0,430	**0,503**	**0,448**
LIRIS_run3	LIRIS	0,546	0,513	**0,251**	0,437
LIRIS_run1	LIRIS	0,539	**0,543**	0,208	0,430
Sabanci-Okan-run1	SABANCI-OKAN	**0,682**	0,476	0,053	0,404
LIRIS_run2	LIRIS	0,530	0,508	0,169	0,403
LIRIS_run4	LIRIS	0,537	**0,538**	0,121	0,399
inria_imedia_plantnet_run2	INRIA	0,477	**0,554**	0,090	0,374
IFSC USP_run3	IFSC	0,356	0,187	0,116	0,220
kmimmis_run4	KMIMMIS	0,384	0,066	0,101	0,184
kmimmis_run1	KMIMMIS	0,384	0,066	0,040	0,163
UAIC2011_Run01	UAIC	0,199	0,059	0,209	0,156
kmimmis_run3	KMIMMIS	0,284	0,011	0,060	0,118
UAIC2011_Run03	UAIC	0,092	0,163	0,046	0,100
kmimmis_run2	KMIMMIS	0,098	0,028	0,102	0,076
RMIT_run1	RMIT	0,071	0,000	0,098	0,056
RMIT_run2	RMIT	0,061	0,032	0,043	0,045
daedalus_run1	DAEDALUS	0,043	0,025	0,055	0,041
UAIC2011_Run02	UAIC	0,000	0,000	0,042	0,014

Table 3. Normalized classification scores for each run and each image type. Top 3 results per image type are highlighted in bold

5 Conclusions

For IMEDIA these results are very promising considering the complementarity of the two very distinct methods. Surprisingly, the matching approach gives the best evaluation score of the task on scans than state of the arts methods

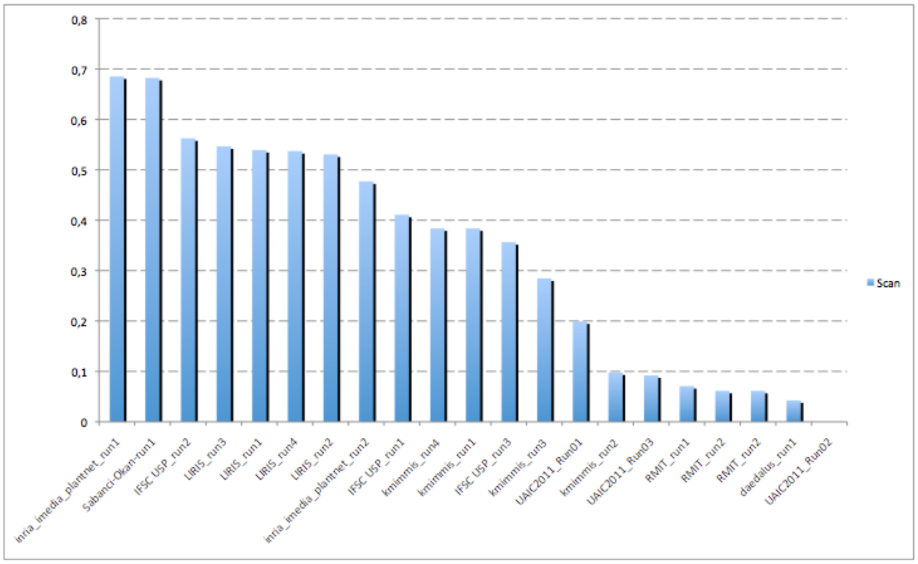


Fig. 2. Normalized classification scores for scan images

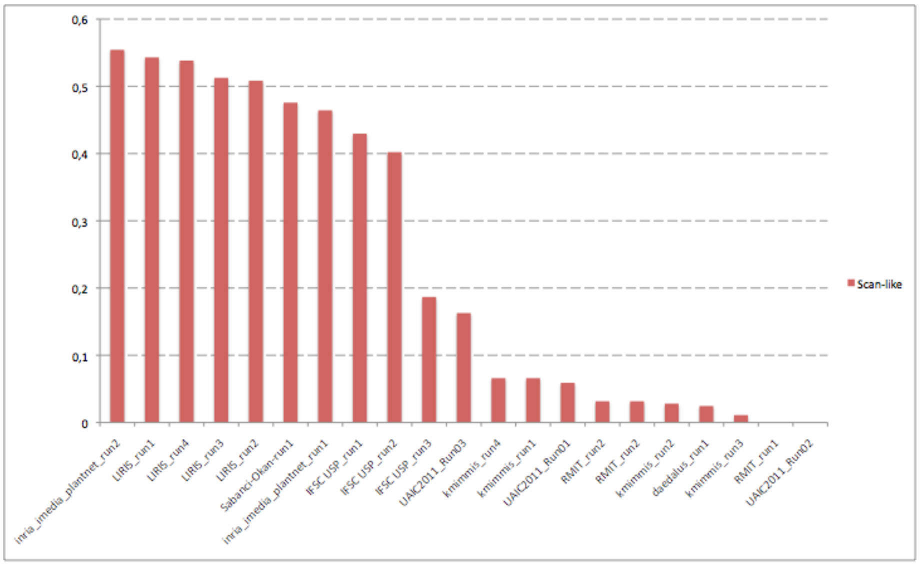


Fig. 3. Normalized classification scores for scan-like photos

based on shape boundary features. This is an important result that opens further investigations in matching based approaches applied to plant identification.

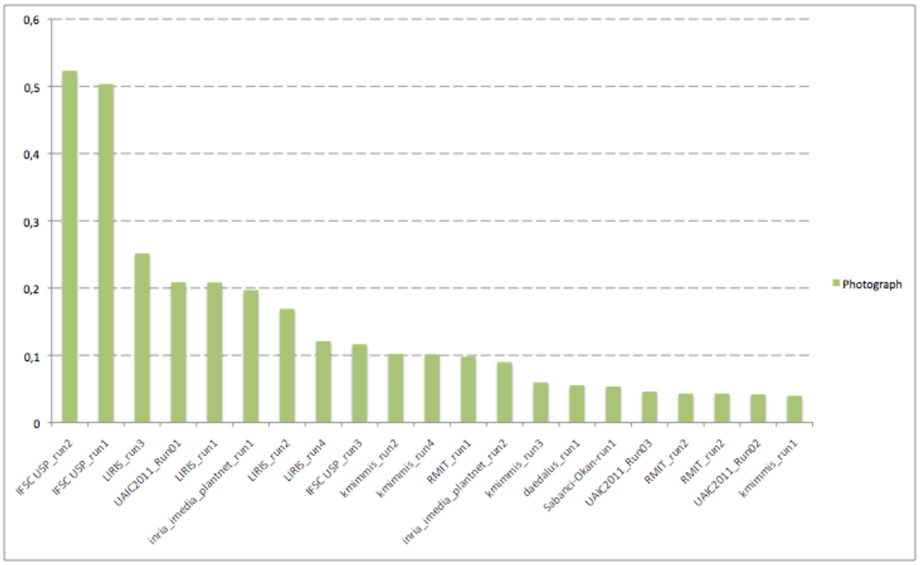


Fig. 4. Normalized classification scores for photographs

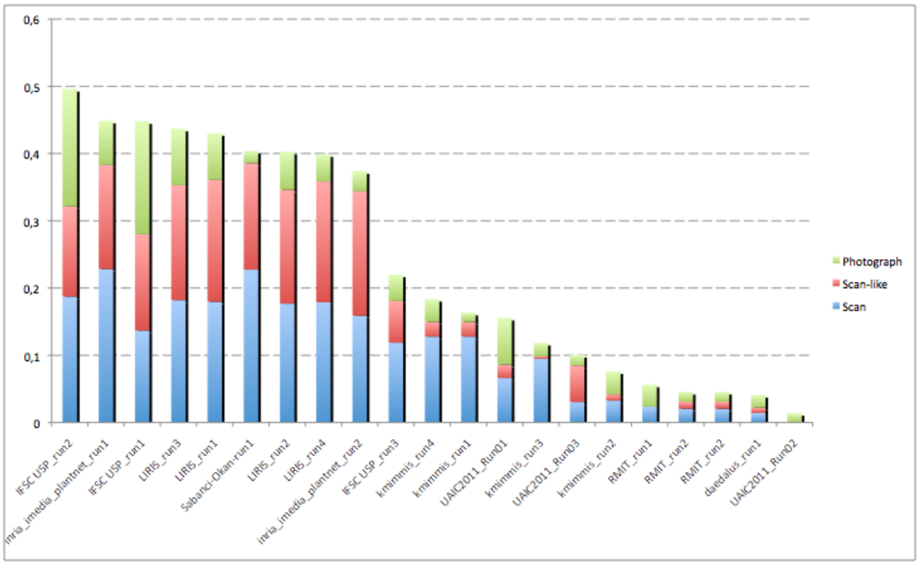


Fig. 5. Normalized classification scores averaged over all image types

Initially aimed at retrieving rigid objects, this original approach for plant leaf identification can be certainly improved in order to be more robust to other kind of images as the scan-like pictures and photographs, maybe by considering a part-based model approach. The very good results with the second method based on shape boundary description, let us to plan improvement by combining it with the matching approach. Indeed, by considering all test images, only 22% of the images are successful at the same time for the two methods, which let us to aim a significant room for improvement.

Acknowledgement

This work was funded by the Agropolis fundation through the project Pl@ntNet (http://www.plantnet-project.org/)

References

1. Backes, A.R., Casanova, D., Bruno, O.M.: A complex network-based approach for boundary shape analysis. Pattern Recognition 42(1), 54 – 67 (2009)
2. Belhumeur, P., Chen, D., Feiner, S., Jacobs, D., Kress, W., Ling, H., Lopez, I., Ramamoorthi, R., Sheorey, S., White, S., Zhang, L.: Searching the world's herbaria: A system for visual identification of plant species. In: ECCV, pp. 116–129 (2008)
3. Bruno, O.M., de Oliveira Plotze, R., Falvo, M., de Castro, M.: Fractal dimension applied to plant identification. Information Sciences 178(12), 2722 – 2733 (2008)
4. Ferecatu, M.: Image retrieval with active relevance feedback using both visual and keyword-based descriptors. Ph.D. thesis, Université de Versailles Saint-Quentin-en-Yvelines (jul 2005)
5. Gouet, V., Boujemaa, N.: Object-based queries using color points of interest. Content-Based Access of Image and Video Libraries, IEEE Workshop on 0, 30 (2001)
6. Hervé, N., Boujemaa, N.: Image annotation: which approach for realistic databases? In: Proceedings of the 6th ACM international conference on Image and video retrieval. pp. 170–177. CIVR '07, ACM, New York, NY, USA (2007)
7. Joly, A., Buisson, O.: Random maximum margin hashing. In: CVPR 2011. pp. 573–580 (2011)
8. Joly, A.: New local descriptors based on dissociated dipoles. In: Proceedings of the 6th ACM international conference on Image and video retrieval. pp. 573–580 (2007)
9. Joly, A., Buisson, O.: A posteriori multi-probe locality sensitive hashing. In: Proceeding of the 16th ACM international conference on Multimedia. pp. 209–218 (2008)
10. Joly, A., Buisson, O.: Logo retrieval with a contrario visual query expansion. In: Proceedings of the seventeen ACM international conference on Multimedia. pp. 581–584 (2009)
11. Knight, D., Painter, H., Potter, M.: Automatic plant leaf classification for a mobile field guide. Tech. rep.
12. Neto, J.C., Meyer, G.E., Jones, D.D., Samal, A.K.: Plant species identification using elliptic fourier leaf shape analysis. Computers and Electronics in Agriculture 50(2), 121 – 134 (2006)
13. Otsu, N.: A Threshold Selection Method From Gray-Level Histogram. IEEE Trans. Syst., Man, Cybern. 9, 62–66 (1979)

14. Philbin, J., Chum, O., Isard, M., Sivic, J., Zisserman, A.: Object retrieval with large vocabularies and fast spatial matching. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (2007)
15. Yahiaoui, I., Herve, N., Boujemaa, N.: Shape-based image retrieval in botanical collections. In: Advances in Multimedia Information Processing - PCM 2006, vol. 4261, pp. 357–364 (2006)