Inria IMEDIA2's Participation at ImageCLEF 2012 Plant Identification Task

Vera Bakić, Itheri Yahiaoui, Sofiène Mouine, Saloua Litayem Ouertani, Wajih Ouertani, Anne Verroust-Blondet, Hervé Goëau, Alexis Joly

Inria IMEDIA2's participation at ImageCLEF 2012 plant identification task

Vera Bakić[1], Itheri Yahiaoui[1,3], Sofiene Mouine[1], Saloua Ouertani-Litayem[1], Wajih Ouertani[1], Anne Verroust-Blondet[1], Hervé Goëau[1], and Alexis Joly[2]

[1] Inria IMEDIA2 Team, Rocquencourt, France, name.surname@inria.fr, http://www-rocq.inria.fr/imedia/
[2] Inria ZENITH Team, Montpellier, France, alexis.joly@inria.fr, http://www-sop.inria.fr/teams/zenith/
[3] Laboratoire CReSTIC, Université de Reims, France, itheri.yahiaoui@univ-reims.fr

Abstract. This paper describes the participation of Inria IMEDIA2 Team, within the Pl@ntNet project[4], at the ImageCLEF2012 plant identification task. The runs used very distinct approaches, sometimes relying on similar extracted features. For *Scan* and *Scan-like* categories, the first two runs combine distinct local and contour approaches in two ways (late and early fusion), while the third run explores the learning capacity of a multi-class SVM technique on a contour based descriptor. For *Photograph* our runs used local features positioned towards the center of the image to reduce the impact of background features. In the second run, an automatic segmentation with a rejection criterion was attempted. In the third run, points were associated with interesting zones. In general, even if they were distinct, the methods used performed very well.

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

1 Introduction

The plant identification task of ImageCLEF2012 is a tree species identification based on leaf images. It was organized as a plant species retrieval task over 126 species with visual content being the main available information. Three types of image content were considered: *Scans* - scans with a white background, *Scan-like* - photographs with a white, uniform background and *Photographs* - unconstrained leaf's images acquired on trees with natural background.

A part of image dataset was provided as training data with full class labels at the beginning of the task, while the test dataset was provided several weeks later without labels. The training and test subsets were built so that the images from an individual plant are not present in both sets, which makes the task more similar to real external queries. The table below shows the composition of the data sets:

[4] http://www.plantnet-project.org/

Number of	Scan		Scan-like		Photograph		All	
	Train	Test	Train	Test	Train	Test	Train	Test
pictures	4870	1760	1819	907	1733	483	8422	3150
individual plants	310	–	118	–	253	–	673	–
authors	22	10	8	10	22	25	29	38

The identification score S was related to the rank of the correct species in the list of retrieved species as follows:

$$S = \frac{1}{U}\sum_{u=1}^{U}\frac{1}{P_u}\sum_{p=1}^{P_u}\frac{1}{N_{u,p}}\sum_{n=1}^{N_{u,p}} S_{u,p,n}, \text{ where}$$

U = number 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,
$S_{u,p,n}$ = score between 1 and 0 equal to the inverse of the rank of the right species (for the n^{th} picture taken from the p^{th} plant observed by the u^{th} user).

Inria IMEDIA2 team, within the Pl@ntNet project, submitted three runs, covering all three image categories. For *Scan* and *Scan-like* we tested early and late fusion of local and shape boundary features using KNN and SVM classifiers. For *Photograph* local features were selected using different geometric constraints and leaf detection and segmentation. In the following text, we will describe the *Scan* and *Scan-like* approaches in Section 2, while the *Photograph* ones will be presented in Section 3. The results are discussed in Section 4. Concluding remarks are in Section 5.

2 Methods used for *Scan* and *Scan-like*

Both contour and local interest points descriptions are useful for the leaf species identification: contour based descriptors capture the global shape of the leaves, while local descriptors associated to the extracted interest points retain their micro-texture. Our three runs test different approaches based on these two kinds of descriptors: the first two runs combine distinct local and contour approaches in two ways (late and early fusion in Sections 2.1 and 2.2), while the third run (Section 2.3) explores the learning capacity of a multi-class SVM technique on a contour based descriptor.

2.1 Late fusion of a large-scale local features matching method and a shape boundary feature based method → *RUN1*

This method is inspired by two runs submitted by Inria IMEDIA Team at ImageCLEF 2011 [8]: the large-scale local features matching method was the best in the *Scan* category, while contour based method was the best in the *Scan-like* category. The two approaches are complementary because they retain both local and boundary shape information and we observed that they performed well last year on distinct test images. Thus, we decided to use a basic combination of these two methods with a fusion of their responses on the image level. This

allows us to keep the good results of each, while minimizing the impact of the erroneous response of one of the approaches. In addition, several major changes in both approaches were made to improve the performance of each or reduce the use of computing resources.

Large-scale local features matching. The basic algorithm applied for this method is: *(i)* Interest points detection, *(ii)* chosen local description for each interest point, *(iii)* local features matching.

(i) **Interest points detection** - 200 Harris points were used at four distinct resolutions with a scale factor equal to 0.8 between each resolution [9, 11]. We output up to 4 significant orientations for a patch around each point, where applicable. This method enabled us to boost the number of training samples and compensate for non-ideal single orientation detection, while keeping the number of points rather small. Figure 1 illustrates the process for sample patches. Finally, the number of points rose to an average of 343 per image[5].

Fig. 1. Multiple orientations: from an original patch (left), in addition to the major orientation (2^{nd} column), other significant orientations are output (3^{rd}, 4^{th} column)).

(ii) **Local features** are extracted around each Harris point from an image patch oriented and scaled according to the given orientation and scale: SURF [2] is based on sums of 2D Haar wavelet responses, we used OpenSURF implementation [5]; a 16-dim. histogram based on the Hough transform [6]; a 20-dim. Fourier histogram [6] and an 8-dim. Edge Orientation Histogram - were concatenated, resulting in a 108-dimensional vector.

(iii) **Matching** - Signatures in the training dataset are compressed and indexed using RMMH method [10, 8]. Local features of a query image are compressed through a 256-bit hash code and its approximate 30-nearest neighbors are searched to obtain a set of candidate matches. To an image, a score equal to the number of its local features matched is assigned. Then, images are re-ranked according to their score.

Contour based descriptor. It consists in a leaf boundary based descriptor that combines two complementary information: (i) The Directional Fragment Histogram descriptor with arbitrary parameters, introduced in [17]. This descriptor has the advantage that it outlines local orientation variations of the leaf margin which is a discriminant key indicator of leaf species. It encodes the relative distribution density of groups of contour points with uniform orientation. The DFH descriptor succeed to detail local properties of the leaf boundary. (ii) For the global geometric properties of the leaf form, we used six shape parameters as in [8]: circularity, convexity, solidity, rectangularity, sphericity and ellipse

[5] Same results were achieved on the last year's task with 500 points per image.

variance. Segmentation was done automatically using Otsu algorithm [15], with addition of automatic selection of a channel that gives the best separation.

Late fusion. The two methods described above will each return a list of images belonging to the same training database but ordered differently: a same image may be present in both lists, but at a different rank. Then, the two lists are merged by setting the rank of an image to a minimum of the ranks in each list. In this way, we preserve a good position of an image returned by one method and ignore the other, presumably, incorrect rank. After the fusion, the unified image list is re-ranked according to the new scores.

Classification with a top-knn decision rule. To obtain the species list from the images list, we counted the number of occurrences of each species in the top 15 images. The list of species was then re-ranked according to their score, to obtain the final ordering of the proposed species.

Training data and descriptor choices. In order to determine the efficiency of a descriptor and which image category to use for training, we performed a series of "leave-individual-out" tests on the training data itself. That is, in the score calculation for an image from the training database, we excluded from the returned response all the images belonging to the same individual as the query image. In addition, we averaged the score as for the official one. We concluded that the combination of the presented descriptors gives the best results on the train data sets and that the *Scan* images should be searched in *Scan* dataset, while the *Scan-like* should be searched in the union of *Scan* and *Scan-like*. Assuming that the test images will perform similarly to the train images, we decided to keep the above parameters for the submitted run.

2.2 Advanced shape context → *RUN2*

The methods used in this run are based on an advanced shape context approach [14], which extends the standard shape context [3]. Here, two different sets of points are distinguished when computing the shape contexts: the *voting set*, i.e. the points used to describe the coarse arrangement of the shape and *the computing set* containing the points where the shape contexts are computed. Two scenarios are proposed by varying the computing set $\mathcal{C}$ and the voting set $\mathcal{V}$ of points in the image.

SC0: Spatial relations between margin points. Here the computing set $\mathcal{C}$ and the voting set $\mathcal{V}$ are identical. They involve only margin points i.e. n points extracted from the margin by a uniform quantization: $\mathcal{C} = \mathcal{V} = \{\text{margin points}\}$ (Figure 2 (a)). This description corresponds to the shape context proposed by Belongie et al. [3] with a different matching method. Note that the venation network is not introduced here. Segmentation algorithm was the same as in Section 2.1.

SC2: Spatial relations between salient and margin points. Here we want to measure the spatial relationships between the salient points described in the

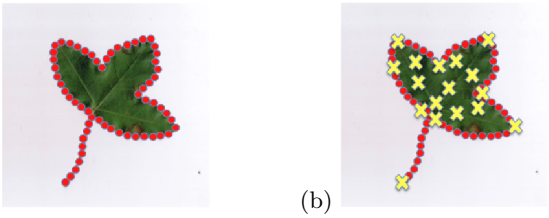


Fig. 2. Points used in scenario SC0 (a) and SC2 (b). The small circles represent the sample points on the leaf margin. The crosses represent Harris salient points.

context defined by the leaf margin (Figure 2 (b)). The voting set of points $\mathcal{V}$ is composed of all the margin points. The Harris points form the computing set $\mathcal{C}$: $\mathcal{C} \neq \mathcal{V}$, $\mathcal{C} = \{\text{salient points}\}$ and $\mathcal{V} = \{\text{margin points}\}$. As mentioned above, the salient points may lay inside the leaf or may belong to the leaf margin. Our aim is to study the correlation between the venation network and the margin of the leaves belonging to the same species.

Local features. The advanced shape context captures a spatial configuration of points without taking into account local properties of the image around the set $\mathcal{C}$ of computing points. Thus, to enrich the description, a set of local features computed on the neighborhood of each point of $\mathcal{C}$ is introduced. As the color is not a discriminant feature for leaves, we focus on texture and shape. Three local features are extracted from the gray-level of an image patch located around each point: a 16-dim. Hough histogram; a 40-dim. Fourier histogram [6] and an 8-dim. classical Edge Orientation Histogram, which is known to be suitable for non-uniform textures. These three features have given promising results when associated with Harris points on scans of leaves in [8].

Matching Method. The features matching, is done by a Multi Probe Locality Sensitive Hashing technique [12] and the distance L_2 is used to compute the similarity between two feature vectors. The principle of this algorithm is to project all the features in an L dimensional space and to use hash functions to reduce the search and the time cost.

Classification with a top-knn decision rule is as in Section 2.1.

Descriptor choices. After an evaluation similar to the one described in Section 2.1 on the training data, *RUN2* was constructed as follows:
- A combination of SC2 and local features with 200 Harris points for *Scan*;
- SC0 with 200 sample points on the leaf margin for *Scan-like*.

2.3 SVM multi-class classification → *RUN3*

This run is performed in order to explore the enhancement of shape feature behaviour through a learning schema. In our experiments we have tested 6 strategies to do classification (see Table 2). To do so, we split the training dataset into train and validation sets with different proportions which also respects the

non-split of images coming from individual plants (the relative images of a given individual plant are either used in train or in validation and not in both). We performed a multi-class SVM technique on a contour based descriptor [17]. Indeed, we adopted a one-vs-one schema [13], which offers more balanced elementary binary classification compared to one-vs-all. It decomposes the classification problem into several binary classification tasks. They are built to discriminate between each pair of classes, while discarding the rest of the classes. If K is the number of classes, one will need to train one binary classifier for each of the possible two classes combinations. This procedure will generate $K(K-1)/2$ binary classifiers. When applied to a test data, a voting is performed among the classifiers and the class is predicted according to the maximum voting strategy [7]. As a kernel choice we conventionally adopted a linear kernel on the 630-dimensional shape features.

We notice the surprising effect of performance's increase between the 70%-30% splitting and the 50%-50% one. That might come from less overfitting on the training subset. For the run, we have chosen to keep learning on both *Scan* and *Scan-like* images together without a prior distinction. That sounds more stable according to the preliminary tests (see rows 5 and 6 in the table 2). For more classes recall and since the final run evaluation takes that into consideration, we have also kept the whole classes ranked according to the voting strategy [7].

Strategy	Train set	Validation set	Decision, 2 splits	
1^{st}	Scan-like	Scan-like	0.4578	0.4533
2^{nd}	Scan	Scan-like	0.1526	0.4464
3^{rd}	Scan-like	Scan	0.3180	0.3106
4^{th}	Scan	Scan	0.3485	0.6151
5^{th}	Scan + Scan-like	Scan	**0.4857**	**0.5056**
6^{th}	Scan + Scan-like	Scan-like	**0.4145**	**0.4236**

Table 1. Strategies tested and their performance on two splits of scan and scan like datasets. From left to right: (70% train, 30% test) and (50% train, 50% test)

3 Methods used for *Photograph*

For the *Photograph* images, the background around the leaf is not uniform (sand, stones, other leaves), and the leaves may be deformed or mutually occluded. The shape boundary features used for *Scan* and *Scan-like* in Section 2.1 are unsuitable in this case: the automatic segmentation of the leaf and background is far from perfect and the detected shape is not good. In addition, the Harris points detector may detect mainly the points in the background. In our runs we explored the following directions: the fact that the leaf is, in general, centered (Section 3.1); whether the automatic segmentation improves performance (Section 3.2) and finally, using multi-class SVM on embedded local features (Section 3.3).

3.1 Rhomboid masking and local features matching → *RUN2*

The basic algorithm applied for this method is the same as described in Section 2.1 for local features. The major difference is in the selection of Harris points, where masking and points weighting were applied.

Rhomboid filtering. In order to minimize the effect of the cluttered background, we modified the input image for the Harris points detector. The assumption is that the leaf is centered, so we masked with an adaptable rhomboid-shape the corners of the image; the transition from foreground to masked out region was smooth to avoid that the points get detected on the mask boundary. Figure 3 illustrates (a) the points detected in the original image, (b) the masked image used as the new input and (c) the points detected in the masked image.

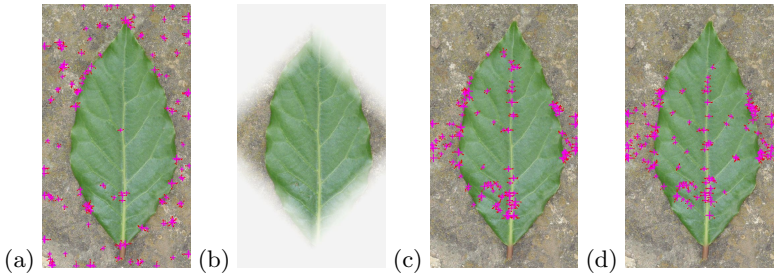


Fig. 3. Rhomboid filtering and grid weighting: (a) original points, (b) masked image, (c) points detected in image (b), (d) points detected with grid weighting

Grid-based points weighting. In the above detection, we noticed that even if the small amount of the cluttered background appears, most of the points are still located in the background. Thus, before the selection of the best 200 points, we applied a weighting scheme based on a grid (of size 7x7): the number of points allocated for the current scale is distributed in the grid cells using Gaussian-like distribution - the closer to the center, the more points allocated and selected. Figure 3 (d) illustrates the final selection of the points using grid- and centered-based weighting.

Training data and descriptors choice. For the selection of descriptors and training dataset we used the same procedure as in Section 2.1. Only the *photograph* data were used as the training database.

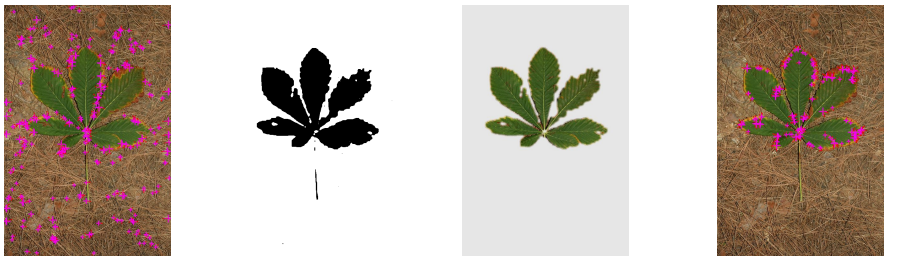
3.2 Segmentation, filtering and local features matching → *RUN1*

Another approach for the *Photograph* images was to attempt the segmentation, where it was possible, and to reject the points that do not belong to the leaf region. In this case, we use the complete training database, with *Scan*, *Scan-like* and *Photograph* images. The advantage of this method is that we could use it for all types of test images. We used the provided "content" annotations: *Leaf*, *Picked leaf* and *Leafage* images and applied different processing for each type. The algorithm and the decision rule were the same as in Section 3.1.

***Scan* and *Scan-like* processing** All training images of these types were processed using the same algorithm as in Section 2.1, local features part.

***Leafage* processing** All images of this type were processed using the same algorithm as in Section 3.1. The reason for this choice is that *Leafage* images have multiple leaves and are extremely cluttered and hard for segmentation.

***Leaf* and *Picked leaf* processing.** For each image we attempted segmentation using Otsu algorithm [15], with addition of automatic selection of a channel that gives the best separation and LUV channels were used as they give better separation for cluttered backgrounds. Then, we automatically verified if the region was well-formed or if the foreground and background classes were too mixed: under the assumption that we have two classes (Figure 4 (b)), we calculated the average distance of each point from the region centers. If the regions were mainly centered and the difference of the distances was more than 20%, segmentation was accepted as good, otherwise, we rejected it and the image was processed as if it was labeled *Leafage*. For the correct segmentation, the biggest found region that does not touch the image boundary was considered as the leaf. Figure 4 illustrates the process of correct (top) and failed (bottom) segmentation. We output 400 points from the Harris detector, however, for the final points list, we keep up to 200 points that do belong to the leaf region (Figure 4 (c)). With the addition of the multi-orientation points, the number of points per image rose to an average of 395 for the *Photograph*. Figures 4 (a) and (d) show the points distribution for the original detection and after filtering or rejecting.



(a) orig. points (b) segm. mask (c) masked img (d) filtered points

(a) orig. points (b) segm. mask (c) masked img (d) points with rhomboid

Fig. 4. Segmentation of photographs: correct (top) and failed and rejected (bottom)

3.3 Multi-Class SVM on embedded local features → *RUN3*

In this run, we explore a data-independent bag of word schema for description and automated partial images zones definition. On this representation, we perform multi-class SVM for learning and predicting plant leaves' classes. The run is motivated by the joint retrieval and learning schema presented in [16], however, we do not have image's interesting zones annotations provided by a user. Instead, we were interested on bounding boxes sampling approaches defining the zones.

Bounding box definition. We used the objectness measure introduced in [1], which is a class generic object detector, quantifying how likely a part (e.g. win-

dow) of an image contains an object of any class. During learning objectness cues' parameters, we did not use plants' images with ground truth relevant zones. Instead, we learned using generic VOC Pascal classes. We expect a broad learning transfer from generic objects to leaves' world. We arguably consider few number of windows per image (10), since images mostly have a single or few interest zones. Figure 5 shows an automatic selection of windows expected as object's delimiter.

Fig. 5. Automatically defined bounding boxes (on sample test images)

Representations. We used the local features from Section 3.1 projected through an efficient approach for feature set representation based on random histograms [4]. Two representations were generated for each image: (i) the features falling within bounding boxes, and (ii) the features in the whole image. Embedding parameters are chosen such that the final representation's histograms have a considerably high dimension of 20480 bins. To evaluate which representation is more suitable, we tested the following combinations using a train/validation dataset splits and a linear multi-class SVM (as in Section 2.3): bounding box or whole image features used for train or validation sets (Table 2, left 3 columns). For both 1^{st} and 4^{th} strategies, we adopt a ranking based on a voting of classes. Each window votes one time for the predicted class, the final vote was the number of times a class is voted by the bounding boxes of a given image. Table 2 shows the decision results for the three splits. It is clear that predicting on sampled bounding boxes, while learning on the whole image noticeably outperforms other strategies. Thus, we used the 4^{th} strategy. Since the final score explores a ranking of classes, to prevent loosing the relevant class, we keep the best 4 classes voted within each image.

	Train set	Test set	Decision, 3 splits	Avg.
1^{st}	bounding box	bounding box	0.001 0.001 0.000	0.001
2^{nd}	bounding box	whole image	0.250 0.254 0.269	0.257
3^{rd}	whole image	whole image	0.251 0.265 0.255	0.257
4^{th}	whole image	bounding box	0.325 0.291 0.345	**0.320**

Table 2. Strategies tested and their performance on three splits of dataset

4 Results

Table 3 resumes the official scores for the first 12 submitted runs out of 30 total. Our three runs give generally good results and are placed in the 10 top runs for different categories. For the *Scan* category, *RUN1* and *RUN3* are in the top 3 while *RUN2* is placed 8^{th}. For the *Scan-like*, the three runs were in the top 5 and

RUN2 gave the best result of the task while *RUN1* is 3^{rd} and *RUN3* 5^{th}. For the more difficult category of *Photograph* *RUN1*, *RUN2* and *RUN3* were respectively 4^{th}, 5^{th} and 10^{th} among 21 runs that used a full automatic approach.

Run	Run type	Official scores			
		Scan	*Scan-like*	*photo*	**Avg**
SABANCI OKAN run 2	Humanly Assisted	**0.58**	0.55	0.22	**0.45**
THEWHO run 3	Humanly Assisted	0.43	0.40	0.49	0.44
SABANCI OKAN run 1	Automatic	**0.58**	0.55	0.16	0.43
IFSC USP run 1	Humanly Assisted	0.35	0.41	**0.51**	0.42
LIRIS reves run 1	Humanly Assisted	0.42	0.51	0.33	0.42
RUN1: INRIA Imedia PlantNet run 1	Automatic	0.49	0.54	0.22	0.42
RUN2: INRIA Imedia PlantNet run 2	Automatic	0.39	**0.59**	0.21	0.40
THEWHO run 4	Humanly Assisted	0.37	0.53	0.43	0.38
LSIS DYNI run 3	Automatic	0.41	0.42	**0.32**	0.38
THEWHO run 1	Humanly Assisted	0.37	0.34	0.43	0.38
RUN3: INRIA Imedia PlantNet run 3	Automatic	0.47	0.46	0.15	0.36
IFSC USP run 2	Humanly Assisted	0.34	0.43	0.30	0.36

Table 3. Normalized classification scores and estimated species rank for the first 12 runs. The best result per image type is highlighted in bold.

***Scan* and *Scan-like*.** With respect to last year's absolute score values, we expected that this year's scores achieve at least the same value if not higher. Indeed, this year the score reflects the correct species rank (which gives always a value distinct from zero), while last year the score was a pure classification score (0 or 1). However, this was neither the case for last year's participants. The reasons are probably that this year the task was more challenging, with the number of species almost doubled and higher number of contributors, which increased the overall visual diversity of images and species. For instance we observed that train and test *Scan* images are rather different: most of the train dataset leaves are mature and green, while a good number of those in the test dataset are young or dead. For *Scan-like*, the image dataset has varying quality like illumination changes, more or less pronounced shadows and variations in the background.

In spite of these difficulties *RUN1* performs well and confirms the good results obtained in the last year's task. The late fusion of two complementary approaches keeps the good performances for both *Scan* and *Scan-like* categories, while last year each method had the best performances in only one category.

For *Scan-like*, *RUN2* achieved the best performances in this year's task. We suppose that the delayed use of the boundary points for the matching phase brought as more similar the partial shape boundary information. This may have compensated for the variations in the leaf boundary (leaf poses, missing leaflets) and imperfect segmentations (shadows), as they were not taken as global information. For *Scan* results are lower than expected. As we noted a significant intra-class variations in the micro-texture, we suppose that the early fusion between the shape and local descriptors was less robust to them. However, this

problem could be compensated by more suitable local features to this visual content.

The *RUN3* gave also very good results on both *Scan* and *Scan-like* categories, with more or less the same score, which seems coherent with the fact that the two types of images were considered as one in order to compensate a lack of examples for some species. Moreover, this method tends toward the score of *RUN1*, notably for *Scan*, while it used only a subset of contour shape descriptors.

Photographs. *RUN1* and *RUN2* gave good results, if we consider only fully automatic runs. Our approaches gave even more or less the same results as some methods with human interactions.

RUN1 had slightly higher score than *RUN2*, which shows us that the automatic segmentation and the use of the complete database for training proved useful. It is important to note here that the segmentation was performed over 35% of all *Photograph* images following the "leaf" and "picked leaf" tags and the rejection criterion. Within the segmented images, only 2% were complete misses (e.g. no leaf part included), and 5% with partial leaf information. This illustrates the idea that it is better to focus on well framed (and segmentable) pictures and reject the ones that are too cluttered.

For its part, *RUN3* gave intermediate performances within a group of runs with automatic approaches, which had more or less similar scores around 0.15. These results are lower than expected, which could be due to the fact that the bounding boxes did not quite correspond to the central image part with the object of interest (the leaf). We are convinced that this method could be improved for instance by using points detection for each interest box separately, and also by using more specialized models of object detection learned on plant images rather than on the general objects.

5 Conclusions

Inria IMEDIA2 Team submitted runs that used distinct approaches, sometimes relying on similar extracted features. Despite these differences, the methods used performed well and the three runs are placed in the 10 top runs for each category.

Again for the second year we obtained very good results on *Scan* and *Scan-like*. However, with respect to last year's absolute score value, we achieved lower scores, as other previous participants, in spite of the improvement of our methods, the proposition of new approaches and a less strict metric. This highlight the fact that plant identification from leaf *Scans* and *Scan-like* images is far from being resolved, especially when will consider more than 500 tree species as it can be observed in France for instance. We will try to improve our methods with the study of new features even more suitable to the visual diversity introduced by new image contributors, by exploring new classification and combination approaches with the metadatas (gps, dates, hierarchical taxonomy information).

For *Photograph*, it is difficult to compare results obtained with full automatic approaches to semi-automatic approaches with human assistance. If we consider only full-automatic approaches, we obtained promising results that we hope to reproduce on plant organs like flower or fruit, for which it is far more difficult to

have *Scan* or *Scan-like* images. Considering fully automatic or humanly assisted approaches, on one hand, we notice that with human assistance runs from other teams tend to have quite similar absolute scores as best *Scan* and *Scan-like* runs. On the other hand, we notice also that several automatic approaches give better performances than assisted ones. Maybe we will have to consider, alongside improving automatic approaches, several human interactions, like semi-supervised segmentation for test images only.

For all three categories, we have to pursuit that the correct species is returned within the top 5 proposed species. This would make our methods suitable for a mobile based recognition application.

Acknowledgments. Part of this work was funded by the Agropolis foundation through the project Pl@ntNet (http://www.plantnet-project.org/)

References

1. Alexe, B., Deselaers, T., Ferrari, V.: What is an object? In: CVPR (2010)
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Computer Vision and Image Understanding (CVIU) (2008)
3. Belongie, S., Malik, J., Puzicha, J.: Shape matching and object recognition using shape contexts. IEEE Trans. Pattern Anal. Mach. Intell. (2002)
4. Dong, W., Wang, Z., Charikar, M., Li, K.: Efficiently matching sets of features with random histograms. In: ACM Multimedia Conference (2008)
5. Evans, C.: Notes on the opensurf library. Tech. rep., University of Bristol (2009)
6. Ferecatu, M.: Image retrieval with active relevance feedback using both visual and keyword-based descriptors. Ph.D. thesis, University of Versailles St-Quentin-en-Yvelines (2005)
7. Friedman, J.H.: Another approach to polychotomous classification. Tech. rep. (96)
8. Goëau, H., Joly, A., Yahiaoui, I., Bonnet, P., Mouysset, E.: Participation of INRIA& Pl@ntNet to ImageCLEF 2011 plant images classification task. In: CLEF (Notebook Papers/Labs/Workshop) (2011)
9. Gouet, V., Boujemaa, N.: Object-based queries using color points of interest. Content-Based Access of Image and Video Libraries, IEEE Workshop on (2001)
10. Joly, A., Buisson, O.: Random maximum margin hashing. In: CVPR 2011 (2011)
11. Joly, A.: New local descriptors based on dissociated dipoles. In: Proceedings of the 6th ACM international conference on Image and video retrieval (2007)
12. Joly, A., Buisson, O.: A posteriori multi-probe locality sensitive hashing. In: 16th ACM international conference on Multimedia (2008)
13. Knerr, S., Personnaz, L., Dreyfus, G.: Single-layer learning revisited: A stepwise procedure for building and training a neural network. In: Neurocomputing: Algorithms, Architectures and Applications (1990)
14. Mouine, S., Yahiaoui, I., Verroust-Blondet, A.: Advanced shape context for plant species identification using leaf image retrieval. In: Proceedings of the 2nd ACM International Conference on Multimedia Retrieval (2012)
15. Otsu, N.: A Threshold Selection Method From Gray-Level Histogram. IEEE Trans. Syst., Man, Cybern. (1979)
16. Ouertani, W., Crucianu, M., Boujemaa, N.: Interactive learning of heterogeneous visual concepts with local features. In: MM '10: Proceedings of the seventeen ACM international conference on Multimedia (2010)
17. Yahiaoui, I., Herve, N., Boujemaa, N.: Shape-based image retrieval in botanical collections. In: Advances in Multimedia Information Processing - PCM 2006 (2006)