

Inria's participation at ImageCLEF 2013 Plant Identification Task

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

Inria, France, `name.surname@inria.fr`

Abstract. This paper describes the participation of Inria within the Pl@ntNet project¹ at ImageCLEF2013 plant identification task. For the *SheetAsBackground* category (scans or photographs of leaves with a uniform background), the submitted runs used a multiscale triangle-based approaches, either alone or combined with other shape-based descriptors. For the *NaturalBackground* category (unconstrained photographs of leaves, flowers fruits, stems,...), the four submitted runs used local features extracted using different geometric constraints. Three of them were based on large scale matching of individual local feature, while the last one used a fisher vector representation. Metadata like the flowering date or/and plant identifier were successfully combined to the visual content. Overall the proposed methods performed very well for all categories and sub-categories.

Keywords: Pl@ntNet, Inria, ImageCLEF, plant, leaves, flowers, fruits, stem, multi-organ, image, collection, identification, classification, evaluation, benchmark

1 Introduction

The plant identification task of ImageCLEF2013 [3] [7] was organized as a plant species retrieval task over 250 species with visual content being the main available information. Two categories were considered: (i) a *SheetAsBackground* category containing exclusively leaves on a white background and (ii) a *NaturalBackground* category containing unconstrained photographs of leaves, flowers, fruits, stems and entire plant views. The identification score was related to the rank of the correct species in the list of retrieved species averaged over the authors of the pictures and the observed plants (the task organizer argues that this weighted metric reduce some bias induced by this particular context of botanical dataset as explained in last year plant task overview[8]).

Inria, within the Pl@ntNet project, submitted four runs, in both image categories. But the methods used different algorithms: for the *SheetAsBackground* queries, the submitted runs were based on shape boundary features (see Section 2), while large scale matching approaches or fisher vectors with SVM classifiers were used for the *NaturalBackground* category (see Section 3). Results are discussed in Section 4, followed by conclusions and perspectives in Section 5.

¹ <http://www.plantnet-project.org/>

2 Methods used for the *SheetAsBackground* category

We used the triangular representations presented in [16]: they are fast to compute and yielded promising results on scan-like images. Moreover, these efficient local approaches are robust to partial leaf occlusions.

Two declinations of multiscale triangle-based descriptors were used:

- **TOA**: a multiscale triangular shape descriptor, where the triangles are described by two successive oriented angles [16]. We used 400 sample contour points and 20 triangles associated to each contour point, with a distance $d = 2$ between the triangle points at two successive scales.

- **TSLA**: a multiscale triangular shape descriptor where the triangles are described by their lengths and an angle [16]. Here, the leaf contour is described by 400 sample points, each point is represented by 10 triangles, with a distance $d = 5$ between the triangle points at two successive scales.

We experimented also combinations of these representations with two complementary descriptions:

- **DFH - Shapes**: a 2D Directional Fragment Histogram (DFH) that computes directions and relative lengths on a succession of elementary fragments on the contour, associated with a set of geometric metrics of the shape (Shapes) [23].

- The **SC2** descriptor proposed within the advanced shape context [15], which computes the spatial correlation between the leaf salient points, computed by an Harris detector, and its margin.

All these descriptors require a preliminary leaf boundary extraction, performed here by using the Otsu thresholding method [18]. The species retrieval process involves a local matching process for TOA, TSLA and SC2, while global comparisons are performed for DFH - Shapes (see [23, 15, 16] for more details).

In summary, we used the following descriptors:

TOA in *Inria PlantNet Run 1*, TSLA in *Inria PlantNet Run 2*, TSLA + DFH - Shapes in *Inria PlantNet Run 3* and TOA + SC2 in *Inria PlantNet Run 4*.

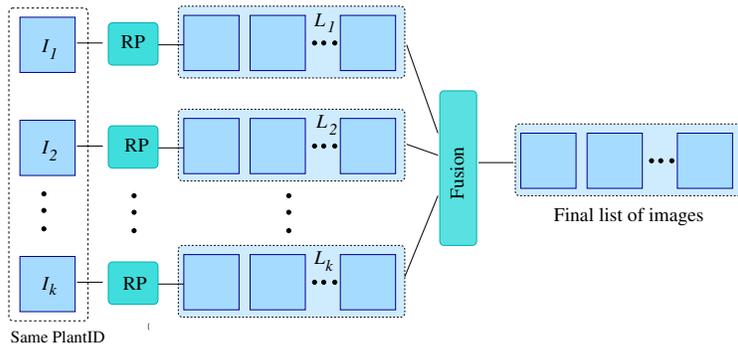


Fig. 1. Multiple image queries. I_1, \dots, I_k are leaf images associated to the same *PlantID* tag and RP is the retrieval process involving either TOA, TSLA, SC2 or DFH - Shapes.

Multiple image queries. For each run, we used the fact that images in the test dataset are associated with plant observations in order to perform multiple image queries for leaf images having the same *PlantID* value (cf. Figure 1). More precisely, for each descriptor:

- We first grouped all the images I_1, \dots, I_k coming from the same plant observation using the *PlantID* in metadata.
- Then, we computed the retrieved images similarity ranking lists L_1, \dots, L_k corresponding to the query images I_1, \dots, I_k .
- Finally, the 100 first image results were kept for each list and were merged into a final list L using a late fusion with the Leave Out algorithm (LO) [14] (lists are merged by setting the rank of an image to the minimum of the ranks in each list). Thus, the best position of an image among the returned lists is kept.

Descriptors combination. For the two last runs, the descriptors combination has been performed by a late fusion on the feature similarity ranking lists resulting from the multiple image queries using the same LO fusion process.

3 Methods used for the *NaturalBackground* category

For the *NaturalBackground* category, we explored the following directions: the fact that a plant view organ is generally centred (Section 3.1 and 3.4), the fact that there are sometimes several images from the same plant observation, the fact that species have not the same flowering periods (Section 3.2), and finally whether the automatic segmentation improves performances (Section 3.3).

The 4 runs are based on the same local features:

Interest points detection. Harris corners were used at four distinct resolutions with multiple orientations ([6]). In addition, as in [1], to minimize the effect of the cluttered background, a rhomboid-shaped mask was applied to the input image and more weight was given to the points closer to the center of the image. Fig. 2(a) illustrates the detected points. About 400 points per image were output.

Local features are extracted around each interest point from an oriented and scaled patch: rotation invariant Local Binary Pattern (ri-LBP) [17]; SURF [2] (sums of 2D Haar wavelet responses), we used OpenSURF [4]; a 20-dim. Fourier histogram [5]; an 8-dim. Edge Orientation Histogram (EOH); a 27-bin weighted RGB histogram (wght-RGB) [5] and a 30-bin HSV histogram;

3.1 Local features and weighted descriptor fusion → *RUN1*

After a series of tests with the training data, we concluded that not all type of local features should be used for all views. As might be expected, color is dominant for the flower but may lead to confusion for leaves, while texture plays an important role for stem and fruits. The most discriminant features retained for the methods are ri-LBP, SURF, Fourier, wght-RGB, histo-HSV for *Flowers*, *Fruit*, *Stem* and *Entire*, while only ri-LBP, SURF, Fourier, EOH are computed for *Leaves*. For each combination of one view and one type of feature, local

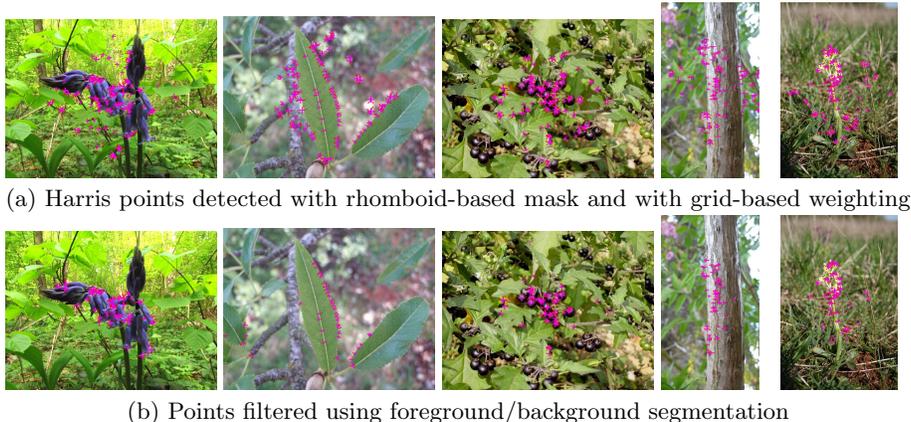


Fig. 2. Examples of filtering of the interest points with the background segmentation

features are computed, compressed and indexed using RMMH method [12, 9] using a 256-bit hash code, which led to a total of 24 unique hash tables.

For a query image Q , according to its view, the basic algorithm applied is: (i) to retrieve lists of similar images for each type of feature, (ii) transform each list of images into a probability distribution by species, (iii) merge probabilities for all types of features.

(i) **Image retrieval:** A response list R is composed of similar images I_i ranked by a score S_i . The score S_i is the number of matches between an image I_i and the query image Q , more precisely the number of instances of image I_i retrieved through the nearest neighbors lists of each local feature of the query image Q : the description of a local feature is compressed with RMMH and its approximate 30-nearest neighbors are searched by probing multiple neighboring buckets in the consulted hash table (according to the a posteriori multi-probe algorithm described in [13]). We define i as the rank of the image in R (we limit i to 300), M_i the plant observation id (referenced as *IndividualPlantId* in the metadata), and C_i the label identifying a species.

(ii) **Probability distribution:** For converting an image response list to a species probability distribution, we use an adaptive rule focusing on plant observations (rather than images). Indeed, the more a species will be represented in a response through various plants observed by distinct users at different dates and locations, the more the associated images will be informative for predicting a species. In contrast, numerous redundant near duplicate images from the same plant observation will not be really informative for predicting a species.

Instead of using the top- K images for decision (as in the last year's runs [8]), we search for top- K'' classes (species) represented with at least K' different plant observations. The values of K' and K'' are determined empirically based on the given training database and are constant for a database: K' is a per-

		Query image from class <i>Cercis siliquastrum</i> :								
(a)	Sp.	Pistacia terebinthus	Robinia pseudoacacia			Cercis siliquastrum				
(b)	Obs.	1	2 (3 images / obser.)			3	4	5	6	
(c)	Images									
(d)	$ S_i $	21	12	11	10	11	10	9	6	
(e)	$ S^c $	21	(12 + 11)/2 = 11.5, ignore 3 rd img.			11	10	9	6	
(f)	S^c	1-im., 1-obs., ignore	11.5			11 + 10 + 9 + 6 = 36				
(g)	$P(C^k)$	0	0.24			0.76				

Fig. 3. Example of the probabilities computation from a ranked list of images.

centage of the average size of the training class, while K'' is a percentage of the number of training classes. The response R is scanned from the most to the least similar image, the counter of the number of classes with at least K' images is incremented accordingly, and when we find K'' such classes, we stop the scanning of the response. The adaptive criterion is important in order to avoid the noise in the final response: (a) had we searched for a fixed (and not dependent on the training data) number of classes with at least K' plant observations, we would often output classes that are not relevant to fill-in the pre-defined requirement, or (b) had we output a fixed number of most similar classes, we would give more weight to the classes with small number of plant observations and would not reward the fact that some classes are well represented in terms of plant observations. Finally, our K'' per class resulted in the K per image of 85, ranging from 15 to 205, for the most to the least different-plant-observation-containing response; organ-wise, the values of K ranged from 66 for Stem to 101 for Fruit.

Moreover, to eliminate the redundant images, we consider only two most similar images per one plant observation: the score per plant observation S^m is a simple average of the image scores S_i . The score for a class is a simple sum of the scores of its plant observations S^m : this step actually favors the well-represented classes, and penalizes the classes with small number of plant observations. Finally, the classes with only one image coming from one plant observation are removed from the list as outliers.

Figure 3 shows an example of the probability computation from a ranked list of 8 images: the first returned image has high score S_i , however, it is the only image of the only plant observation for its class and this image is ignored in

Organ	W_{unif}	SURF	Fourier	EOH	ri-LBP	wght-RGB	histo-HSV	$\Delta(W^{avg})$	$\Delta^{avg}(W)$
Flower	0.2	0.169	0.171	-	0.209	0.230	0.221	0.061	0.178
Entire	0.2	0.192	0.175	-	0.196	0.221	0.216	0.046	0.158
Fruit	0.2	0.188	0.182	-	0.186	0.223	0.221	0.041	0.158
Leaf	0.25	0.271	0.233	0.237	0.259	-	-	0.038	0.176
Stem	0.2	0.197	0.216	-	0.210	0.192	0.184	0.032	0.173

Table 1. Distribution of average weights attributed to each local feature for test images (per organ) and average maximum weight delta over all test images.

the final classes list; next three images belong to the same plant observation, so we will keep only the first two scores S_i ; finally, the last four images belong to the same class as the query image, however, unlike for the previous group, they are all from different plant observations. Thus, we will use all scores in the calculation of the class probability.

(iii) **Probability fusion:** At this step we have several species probability distributions, one for each kind of feature and we use a weighted fusion in order to obtain a final probability distribution. Let us define F as a set of local features and $P(C_f^k)$ as probability of class C^k for feature $f \in F$. In order to reflect the discriminating power of each local feature, we define the final probability:

$$P(C^k) = \sum_{f \in F} w(f) * P(C_f^k), \text{ where } w(f) = \frac{\max_{\forall k} P(C_f^k)}{\sum_{f \in F} \max_{\forall k} P(C_f^k)}$$

Table 1 shows the average weights attributed to each local feature for all test images, displayed per organ. The last two columns show the difference between the minimal and maximal average weight per feature and the average difference between the minimal and maximal weight per image. Color-based features have higher weights for *Flower*, *Entire* and *Fruit*, than gray-based ones. For gray-based features, ri-LBP contributes more than SURF and Fourier for *Flower*, while it is Fourier which contributes the most for *Stem*. For *Leaf*, SURF and ri-LBP contributed more than Fourier and EOH. We can note that all the average weights are actually rather close to the W_{unif} , and that all the chosen local features play an important role in the overall decision. The values $\Delta^{avg}(W) \approx W_{unif}$ show that on the image level, there was always a local feature that had the weight significantly lower than the others, thus the influence of that, presumably, confusing response was minimized in the overall response.

3.2 Multi-image queries and filtering by flowering period \rightarrow *RUN2*

This run is an extension of the *RUN1*, with two modifications making use of metadata: (i) the responses from all images belonging to the same plant observation are merged and (ii) only for the *Flower* images, flowering dates for filtering irrelevant species in responses.

Multi-image queries: As in the training dataset, images from the test dataset are sometimes related to a same plant observation which was explicitly men-

(a)	Entire 1 image	Fruit 2 images	Leaf 5 images	Stem 2 images
Indiv.'s images				
(b)	1 Cra.m. 0.06 2 Eup.p. 0.05 3 Mel.a. 0.04 4 Pop.a. 0.04 5 Hed.h. 0.04	Vib.t. 0.08 Vib.t. 0.14 Bet.u. 0.05 Que.c. 0.06 Arb.u. 0.04 Til.p. 0.06 Lig.v. 0.04 Mel.a. 0.05 Que.p. 0.04 Rob.p. 0.04	Cra.m. 0.1 Syr.v. 0.08 Que.i. 0.07 Que.i. 0.08 Gin.b. 0.07 Pit.t. 0.07 Syr.v. 0.06 Ner.o. 0.06 Lab.a. 0.04 Dio.k. 0.06	Pla.x. 0.33 Pla.x. 0.37 Cel.a. 0.1 Pis.l. 0.06 Lau.n. 0.06 Pop.a. 0.05 Fra.o. 0.04 Lau.n. 0.05 Pit.t. 0.04 Vib.t. 0.05
(c)	1 Cra.m. 0.06 2 Eup.p. 0.05 3 Mel.a. 0.04 4 Pop.a. 0.04 5 Hed.h. 0.04 6 Fra.e. 0.03	Vib.t. 0.10 Que.c. 0.04 Til.p. 0.04 Bet.u. 0.04 Mel.a. 0.04 Rob.p. 0.03	Syr.v. 0.05 Cra.m. 0.05 Ace.o. 0.04 Que.i. 0.04 Cer.s. 0.03 Fag.s. 0.03 ...22 Vib.t. 0.02	Pla.x. 0.30 Cel.a. 0.08 Lau.n. 0.05 Pis.l. 0.04 Pop.a. 0.04 Vib.t. 0.04
(d)	0.112	0.199	0.098	0.591
(e)		1 Platanus x hispanica 0.18 2 Celtis australis 0.05 3 Viburnum tinus 0.05 4 Laurus nobilis 0.03 5 Populus alba 0.03 ...10 Crataegus monogyna 0.02 ...32 Syringa vulgaris 0.01		1 st in Stem 2 nd in Stem 1st in Fruit 3 rd in Stem 4 th in Entire 1 st in Entire 1 st in Leaf

Fig. 4. Multi-image multi-organ query simulation for plant id 4165 in the training dataset. The correct species is highlighted in bold in all responses where it was present.

tioned in metadata with the tag *IndividualPlantId*. Thus, we have some query images showing the same plant on distinct organs with different angles (Figure 4). In such cases, we try to take advantage from the complementarity of the views in order to compute a unique ranking list of species for one plant id. Then, the ranking list was repeated in the run file for each image of the associated plant. For each plant associated to at least two image queries, the fusion is applied in two-stages:

1. single-view level: if several query images (I_o) are related to the same organ o , we kept for each class C^k the highest probability: $P_o(C^k) = \max_{\forall i \in I_o} P_i(C^k)$

2. multi-organ level: for responses of different organs (O), we apply the same the weighted fusion of probabilities used as in Section 3.1 where local features F are replaced with the organs set O . As discussed, this approach underlines the discriminant power of an organ.

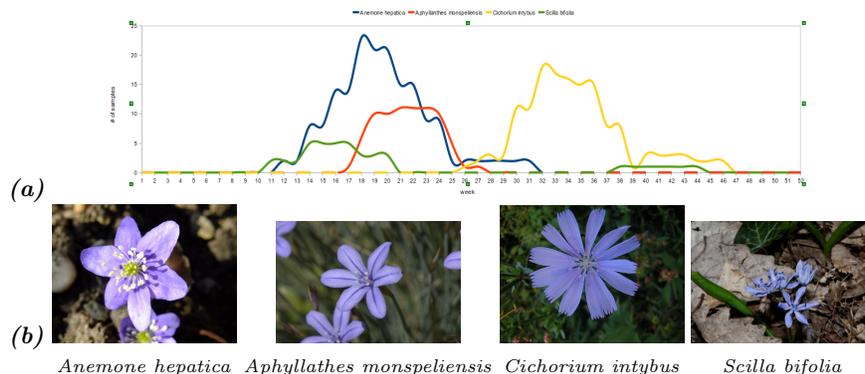


Fig. 5. Flowering periods for sample images with similar color and/or texture

Figure 4 shows an example of the fusion on one set of query images from one plant observation ($IndividualPlantID = 4165$) from the training dataset in a leave one out procedure during preliminary tests. The correct species is *Viburnum tinus* and 4 views are represented by 10 images (4(a) shows sample images for each view). 4(b) shows the first proposed species per image and we can see that *Viburnum tinus* does not appear in all responses. 4(c) shows the responses combined per view. We can note here that for *Entire* and *Leaf*, the proposed species have lower probability with respect to the species proposed for other views, which results to higher weights assigned for *Stem* and *Fruit* (line 4(d)). Finally, the ranking list of proposed species 4(e) is lead by the first species from the, presumably, least confusing view response (*Stem*) while the correct species is at the 3rd position. Finally, it is not only 2 query images of *Fruit*, but all the 10 query images which are associated with a relevant species response.

Filtering by flowering period: Unlike other views, *Flower* has an important feature: in many species these plant organs are present for just a quite short period of time, and each species has its own flowering periods. The metadata contains the date when the photograph was taken, for the training, as well as for the test datasets. A post-processing treatment is applied to the list of species obtained through pure visual search, and only the species for which the date of query image was observed in the training data were retained.

The flowering period histogram (Figure 5 (a)) for a species is constructed by week, with ± 3 additional weeks to account for geographical and year-to-year differences. Given a training image of class C^k , taken in week w , histogram bins $H^{C^k}(h)$, $h = w - 3, \dots, w + 3$ are incremented. For a query image Q taken in week w^Q , an histogram H^Q is constructed in the same manner, and finally a species C^k is retained if $\exists w | H^Q(w) > 0 \wedge H^{C^k}(w) > 0$.

Figure 5 (a) shows the flowering periods for four species that have similar color and texture. *Cichorium intybus* flowers appear clearly later than the other

species over a year. Thus, any query images in this period will exclude the three other species, even if the visual content are very similar to these species.

3.3 Automatic segmentation \rightarrow *RUN3*

In this run we tested the automatic fore-/back-ground segmentation in an attempt to reduce the number of the Harris points in the (cluttered) background. The segmentation algorithm used was Otsu [18], with an addition of automatic selection of one of LUV colorspace channels that gives the best separation. Then, we automatically checked if the region was well-formed or if the fore- and back-ground classes were too mixed as in [1]. For the correct segmentation, all the regions that do not touch the boundary significantly are considered as foreground object; for the rejected segmentation, we used rhomboid mask. Only the points that fall in the foreground regions were kept. Overall, the number of points was reduced by 30%, varying from 16% for Flower to 40% for Stem.

Figure 2 shows (a) initial detected points and (b) the filtered set. The distribution in (a) is clearly in a rhomboid shape, while in the case (b) the points are all in the foreground object or on its edges.

3.4 Multi-modal image representation through embedded local features \rightarrow *RUN4*

This run is performed in order to explore combining local features through an embedding schema for images representation and multi-class learning approach. We have used fisher vectors representation [20] for the embedding of the local features as it has been proved as a successful extension of the popular bag-of-visual word (BOV) representation [22].

Image representation: Fisher vectors representation (FV) extends the BOV representation where the local patches are described by their deviation from a generative Gaussian mixture model (GMM). Let $X = \{x_t, x_t \in \mathfrak{R}^D, t = 1..T\}$ be the set of T D-dimensional local descriptors of an image I. Assuming that the generation process of X can be modelled by the GMM model with parameters λ , we can characterize an image I by the following gradient vector [10]:

$$G_\lambda^X = \frac{1}{T} \nabla_\lambda \sum_{t=1}^T \log \mu_\lambda(x_t)$$

where $\lambda = \{\omega_i, \mu_i, \sigma_i, i = 1..K\}$ and ω_i , μ_i and σ_i are respectively the mixture weight, mean vector and variance matrix (assumed diagonal) of Gaussian μ_i . We used the gradient vector wrt. the mean only on GMM with K Gaussians as it was presented in [11]: Jegou et al. stated that in a set of experiments, for a fixed FV size, the option led to a very similar result to that of gradient with respect to the variance and mean. In that case, the FV is the concatenation of the gradient vectors G_i^X with respect to the mean μ_i of Gaussian i . It is therefore K d-dimensional, K being the number of Gaussians and d the local descriptor dimensionality.

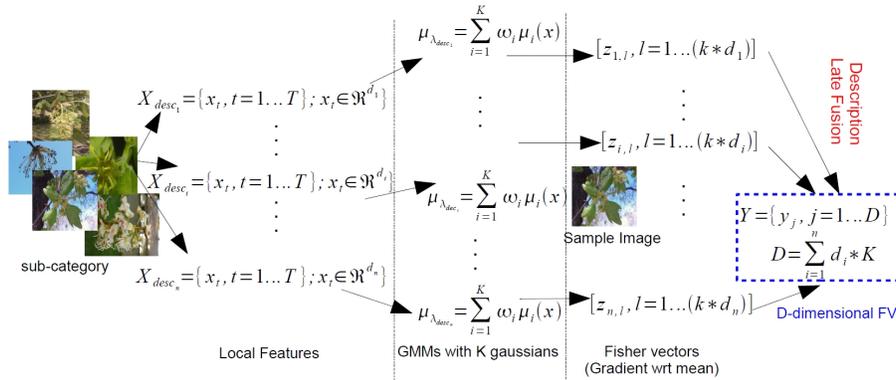


Fig. 6. Multimodal image representation through embedded local features

Vocabulary learning: For each image I we used the same local features extracted for the three previous runs (Section 3.1). Let $X_{desc_i} = \{x_t, x_t \in \mathfrak{R}^{d_i}, t = 1 \dots T\}$ be the set of T d_i -dimensional local descriptors of an image I according to a descriptor $desc_i$. Thus each image is represented by a set of descriptors $\{X_{desc_1}, \dots, X_{desc_n}\}$, and for each descriptor $desc_i$, we estimated a d_i -dimensional GMM with $K = 256$. As the GMM estimation is an unsupervised step, we used both the train and test datasets in order to enhance the representation. One GMM was estimated for each sub-category (*Flower, Fruit, Stem, Entire, Leaf*) and description space. We computed separately a FV for each local description using implementation [11]. The generated FVs are power and L2-normalized [21]. Finally, each image I is characterized by a concatenation of the FVs. Assuming that Y is the fusion of the FV representations of the image I , Y is a D -dimensional vector:

$$Y = \{y_j, j = 1 \dots D\} \text{ with } D = \sum_{i=1}^n d_i * K$$

Classifier training: For each sub-category we learned a standard approach with a linear one-versus-all multi-class SVM classifier. For efficiency, we trained our classifiers using Stochastic Gradient Descent (SGD) algorithm as in [19].

4 Results

SheetAsBackground. The four methods obtained very good results, all of them being in the top 5 with the run *Inria PlantNet Run 2* reaching the second position. We discuss the fact that the ranking of the 4 runs differs from what we observed during preliminary tests with a Leave One Out procedure on the training dataset (see Table 2). For instance TSLA method, which had the lowest preliminary results with a 4th position on the training dataset, is finally the 1st on the testing dataset. As noticed in [16], the respective orders of the TSLA and TOA scores may depend on the dataset. Here, the train and the test datasets

may lead to different results as they do not involve the same number of authors (36 for the train and 14 for the test) and the same number of individual plants (732 for the train and 150 for the test) and these two parameters are taken into account in the ImageCLEF scores. Nevertheless, the two scores are close, thus we can consider that this two descriptors have similar performances.

Method	ImageCLEF run	Training dataset score	ImageCLEF score
TOA	Inria PlantNet Run 1	0.763	0.557
TSLA	Inria PlantNet Run 2	0.726	0.577
TSLA+DFH-Shapes	Inria PlantNet Run 3	0.766	0.572
TOA+SC2	Inria PlantNet Run 4	0.753	0.517

Table 2. *SheetAsBackground* scores of the 4 runs on the training dataset during preliminary experiments compared with the official ImageCLEF scores on test dataset. Best scores are in bold for each column.

Moreover, the combinations TSLA+DFH-Shapes and TOA+SC2 increase the scores of TSLA or TOA on the training dataset while it is not the case on the test images. This may be due to some unsuccessful automatic extractions of the leaf contour on a part of the test dataset as illustrated in Figure 7. The quality of the automatic segmentation is closely dependent on the noise that may be present in the image. Multiscale triangular methods are robust to partial occlusion and to local contour deformations, since the local descriptors associated to the contour points describe only a portion of the leaf boundary (cf. [16] for a discussion on this subject). But the local descriptors associated to the salient points in SC2 represent the relative positions of the points with respect to the margin and are sensitive to contour deformations. It is also the case for DFH-Shapes, which provides a description of the whole contour. Thus combining any multiscale triangular descriptor with either SC2 or DFH-Shapes decreases the score when the contour is incorrectly extracted.



Fig. 7. Examples of unsuccessful segmentation on test dataset images.

NaturalBackground. Table 3 summarizes the official scores for the first 7 submitted runs together with our own runs. *Inria PlantNet Run 2* is 2nd overall and the 1st for the sub-category *Flower*. *Inria PlantNet Run 1, 3, 4* are respectively

Run	Overall score	Score par organ				
		Entire	Flower	Fruit	Leaf	Stem
NlabUTokyo Run 3	0.393	0.297	0.472	0.311	0.275	0.253
Inria PlantNet Run 2	→ 0.385	→ 0.274	0.494	0.260	0.272	0.240
NlabUTokyo Run 2	0.371	0.273	0.484	0.259	0.273	0.285
Inria PlantNet Run 1	→ 0.353	→ 0.254	0.437	0.249	0.240	0.211
NlabUTokyo Run 1	0.341	0.236	0.423	0.209	0.269	0.276
Inria PlantNet Run 3	→ 0.325	→ 0.216	0.421	0.238	0.195	0.176
Inria PlantNet Run 4	→ 0.245	→ 0.150	0.327	0.137	0.165	0.171

Table 3. Official scores for the first two teams for *NaturalBackground*. The best result per image type is highlighted in bold, Inria PlantNet runs are marked with →.

4th, 5th and 7th. Basically, our scores are closely interlaced with the ones from *NlabUTokyo*, whereas the runs of all other groups are significantly lower.

Inria PlantNet Run 2 had the best scores within our runs as well in details for the 5 sub-categories. We can note also that overall gain for *Inria PlantNet Run 2* with respect to *Inria PlantNet Run 1* was 0.032, ranging from 0.011 for *Fruit* to 0.057 for *Flower*, which shows the benefits that can be done when using multi-image queries and flowering period filtering.

Inria PlantNet Run 3 achieved only 0.028 (8%) smaller score than *Inria PlantNet Run 1*, while using 30% less interest points, which is an interesting result if we consider memory resources for implementation in a final system. But we must note that for *Flower* and *Fruit*, the losses were much smaller (3-4%) comparing to *Entire*, *Leaf* and *Stem* (15-19%). We suspect that this is due to the segmentation problems with the later three views and that in those cases, alternative segmentation methods could be applied.

Inria PlantNet Run 4 reached a quite good results if we compare with all submitted runs. However, the method experimented obtained lower score than the method used in *Inria PlantNet Run 1* which was using the same local features. One explanation can be that the choice of a fusion by concatenation of Fisher Vectors computed in separated features space was not the most relevant fusion scheme because the discriminating powers of each local feature was not visible, or the concatenation led to normalisation problems. Also, the fisher embedding itself might degrade performances compared to the independent knn search of each local feature.

5 Conclusions

Inria submitted four runs that used different processing methods for *SheetAs-Background* and *NaturalBackground* categories. The task was more challenging this year, with twice of species and images, and more visual diversity with the 5 different plant views. Our methods performed very well and our best runs were 2nd in both categories.

For *SheetAsBackground*, the results obtained by our four runs (in the top 5 positions) confirmed the effectiveness of our multiscale triangle based approaches [16] for the identification of plant species from scan and scan-like leaf images.

For *NaturalBackground*, we tested local features matching with several declinations as the use of metadata or automatic segmentation. For the first three runs, we proposed to use a weighted fusion scheme for combining the responses from multiple indexes build for each combination of type of views and type of local features. It underlines the discriminant power of a non-confusing response. With the use of metadata, such as the plant ids and the date of photograph, we showed that multi-image queries and filtering by flowering period can clearly increase the baseline performance. On the other hand, with the use of automatic segmentation, we could reduce the size of the training features, while having a rather small performance decrease. We also note that, as for other participants, the score for *Flower* category is double the score for other categories and that the best *Flower* scores are approaching the scores obtained in *SheetAsBackground* category.

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

References

1. Bakić, V., Yahiaoui, I., Mouine, S., Ouertani-Litayem, S., Ouertani, W., Verroust-Blondet, A., Goëau, H., Joly, A.: Inria imedia2's participation at ImageCLEF 2012 plant identification task. In: CLEF (Notebook Papers/Labs/Workshop) (2012)
2. Bay, H., Ess, A., Tuytelaars, T., Van Gool, L.: Surf: Speeded up robust features. In: Computer Vision and Image Understanding (CVIU) (2008)
3. Caputo, B., Muller, H., Thomee, B., Villegas, M., Paredes, R., Zellhofer, D., Goëau, H., Joly, A., Bonnet, P., Gomez, J.M., Varea, I.G., Cazorla, M.: ImageCLEF 2013: the vision, the data and the open challenges. In: Proc CLEF 2013. LNCS (2013)
4. Evans, C.: Notes on the opensurf library. Tech. rep., University of Bristol (2009)
5. Ferecatu, M.: Image retrieval with active relevance feedback using both visual and keyword-based descriptors. Ph.D. thesis, Univ. Versailles St-Quentin (2005)
6. Goëau, H., Bonnet, P., Barbe, J., Bakić, V., Joly, A., Molino, J.F., Barthélémy, D., Boujemaa, N.: Multi-organ plant identification. In: Workshop on Multimedia Analysis for Ecological Data Proceedings. MAED '12 (2012)
7. Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The ImageCLEF 2013 plant identification task. In: CLEF 2013 Working Notes
8. Goëau, H., Bonnet, P., Joly, A., Bakić, V., Barthélémy, D., Boujemaa, N., Molino, J.F.: The ImageCLEF 2012 plant identification task. In: CLEF Working Notes (2012)
9. 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)
10. Jaakkola, T., Haussler, D.: Exploiting generative models in discriminative classifiers. In: In Advances in Neural Information Processing Systems 11 (1998)
11. Jégou, H., Perronnin, F., Douze, M., Sánchez, J., Pérez, P., Schmid, C.: Aggregating local image descriptors into compact codes. IEEE Transactions on Pattern Analysis and Machine Intelligence (2011), <http://hal.inria.fr/inria-00633013>

12. Joly, A., Buisson, O.: Random maximum margin hashing. In: CVPR 2011 (2011)
13. Joly, A., Buisson, O.: A posteriori multi-probe locality sensitive hashing. In: Proceedings of the 16th ACM international conference on Multimedia. MM '08 (2008)
14. Jović, M., Hatakeyama, Y., Dong, F., Hirota, K.: Image retrieval based on similarity score fusion from feature similarity ranking lists. In: International conference on Fuzzy Systems and Knowledge Discovery (2006)
15. Mouine, S., Yahiaoui, I., Verroust-Blondet, A.: Advanced shape context for plant species identification using leaf image retrieval. In: ACM International Conference on Multimedia Retrieval, ICMR'12. pp. 49–56 (Jun 2012)
16. Mouine, S., Yahiaoui, I., Verroust-Blondet, A.: A shape-based approach for leaf classification using a multiscale triangular representation. In: ACM International Conference on Multimedia Retrieval, ICMR'13. pp. 309–310 (Apr 2013)
17. Ojala, T., Pietikäinen, M., Mäenpää, T.: Gray scale and rotation invariant texture classification with local binary patterns. In: Computer Vision - ECCV 2000, Lecture Notes in Computer Science, vol. 1842. Springer Berlin Heidelberg (2000)
18. Otsu, N.: A Threshold Selection Method From Gray-Level Histogram. IEEE Trans. Syst., Man, Cybern. (1979)
19. Perronnin, F., Akata, Z., Harchaoui, Z., Schmid, C.: In: IEEE Computer Vision and Pattern Recognition (CVPR). Providence (RI), United States (2012)
20. Perronnin, F., Dance, C.R.: Fisher kernels on visual vocabularies for image categorization. In: CVPR (2007)
21. Perronnin, F., Sánchez, J., Mensink, T.: Improving the fisher kernel for large-scale image classification. In: ECCV. Berlin, Heidelberg (2010)
22. Sánchez, J., Perronnin, F., Mensink, T., Verbeek, J.: Image classification with the Fisher vector: Theory and practice. Tech. rep., Inria, project team LEAR (2013)
23. Yahiaoui, I., Mzoughi, O., Boujemaa, N.: Leaf shape descriptor for tree species identification. In: ICME (2012)