# The ImageCLEF 2012 Plant identification Task

Hervé Goëau[1], Pierre Bonnet[2], Alexis Joly[1], Itheri Yahiaoui[1,5], Daniel
Barthelemy[3], Nozha Boujemaa[1], and Jean-François Molino[4]

[1] INRIA, IMEDIA & ZENITH teams, France, name.surname@inria.fr,
http://www-rocq.inria.fr/imedia/, http://www-sop.inria.fr/teams/zenith/
[2] CIRAD, UMR AMAP, France, pierre.bonnet@cirad.fr, http://amap.cirad.fr
[3] CIRAD, BIOS Direction and INRA, UMR AMAP, F-34398, France,
daniel.barthelemy@cirad.fr, http://amap.cirad.fr/fr/index.php
[4] IRD, UMR AMAP, France, jean-francois.molino@ird.fr, http://amap.cirad.fr
[5] Laboratoire CReSTIC, Université de Reims, France, itheri.yahiaoui@univ-reims.fr

**Abstract.** The ImageCLEF's plant identification task provides a testbed
for the system-oriented evaluation of plant identification, more precisely
on the 126 tree species identification based on leaf images. Three types
of image content are considered: *Scan*, *Scan-like* (leaf photographs with
a white uniform background), and *Photograph* (unconstrained leaf with
natural background). The main originality of this data is that it was
specifically built through a citizen sciences initiative conducted by Tela
Botanica, a French social network of amateur and expert botanists.
This makes the task closer to the conditions of a real-world application.
This overview presents more precisely the resources and assessments of
task, summarizes the retrieval approaches employed by the participating
groups, and provides an analysis of the main evaluation results. With a
total of eleven groups from eight countries and with a total of 30 runs
submitted, involving distinct and original methods, this second year pilot
task confirms Image Retrieval community interest for biodiversity and
botany, and highlights further challenging studies in plant identification.

**Keywords:** ImageCLEF, plant, leaves, images, collection, identifica-
tion, classification, evaluation, benchmark

## 1   Introduction

Convergence of multidisciplinary research is more and more considered as the
*next big thing* to answer profound challenges of humanity related to health, bio-
diversity or sustainable energy. The integration of life sciences and computer sci-
ences has a major role to play towards managing and analyzing cross-disciplinary
scientific data at a global scale. More specifically, building accurate knowledge
of the identity, geographic distribution and uses of plants is essential if agri-
cultural development is to be successful and biodiversity is to be conserved.
Unfortunately, such basic information is often only partially available for pro-
fessional stakeholders, teachers, scientists and citizens, and often incomplete for
ecosystems that possess the highest plant diversity. A noticeable consequence,

expressed as the *taxonomic gap*, is that identifying plant species is usually impossible for the general public, and often a difficult task for professionals, such as farmers or wood exploiters and even for the botanists themselves. The only way to overcome this problem is to speed up the collection and integration of raw observation data, while simultaneously providing to potential users an easy and efficient access to this botanical knowledge. In this context, content-based visual identification of plant's images is considered as one of the most promising solution to help bridging the taxonomic gap. Evaluating recent advances of the IR community on this challenging task is therefore an important issue.

This paper presents the *plant identification task* that was organized for the second year running within ImageCLEF 2012[6] dedicated to the system-oriented evaluation of visual based plant identification. The task was again focused on tree species identification based on leaf images, but with more species (126 instead of 70) which is an important step towards covering the entire flora of a given region. The task was more related to to a retrieval task instead of a pure classification task in order to consider a ranked list of retrieved species rather than a single brute determination. Visual content was being the main available information but with additional information including contextual meta-data (author, date, locality name and geotag, names at different taxonomic ranks) and some EXIF data. 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. The main originality of this data is that it was specifically built through a citizen sciences initiative conducted by Telabotanica[7], a French social network of amateur and expert botanists. This makes the task closer to the conditions of a real-world application: (i) leaves of the same species are coming from distinct trees living in distinct areas (ii) pictures and scans are taken by different users that might not used the same protocol to collect the leaves and/or acquire the images (iii) pictures and scans are taken at different periods in the year.

## 2    Task resources

### 2.1    The Pl@ntLeaves dataset

Building effective computer vision and machine learning techniques is not the only side of the *taxonomic gap* problem. Speeding-up the collection of raw observation data is clearly another crucial one. The most promising approach in that way is to build real-world collaborative systems allowing any user to enrich the global visual botanical knowledge [15]. To build the evaluation data of ImageCLEF plant identification task, we therefore set up a *citizen science* project around the identification of common woody species covering the Metropolitan French territory. This was done in collaboration with TelaBotanica social network and with researchers specialized in computational botany.

---

[6] http://www.imageclef.org/2012
[7] http://www.tela-botanica.org/

Technically, images and associated tags were collected through a crowd-sourcing web application [15] and were all validated by expert botanists. Several cycles of such collaborative data collection and taxonomical validation occurred.

Scans of leaves were first collected over the two summers 2009 and 2010 thanks to the work of active contributors from TelaBotanica social network. The idea of collecting only scans during these two seasons periods was to initialize training data with limited noisy background and to focus on plant variability rather than mixed plant and view conditions variability. This allowed to collect 2228 scans over 55 species. A public version of the web application[8] was then opened in October 2010 and additional data were collected up to March 2011. The new collected images were either *scans*, or photographs with uniform background (referred as *scan-like* photos), or unconstrained *photographs* with natural background. They involved besides 15 new species from the previous set of 55 species. Since April 2011 botanists from TelaBotanica contribute regularly every month on more and more tree species (174 on leaves at the time of writing), on more localities all over France and neighbouring countries, with more growing stages with spring leaves, mature summer leaves, dry autumn leaves and evergreen leaves. This nonstop data collecting over the months introduce slowly visual and morphological variabilities.

The Pl@ntLeaves dataset used within ImageCLEF2012 finally contained 11572 images: 6630 scans, 2726 scan-like photos and 2216 photographs (see Figure 1).
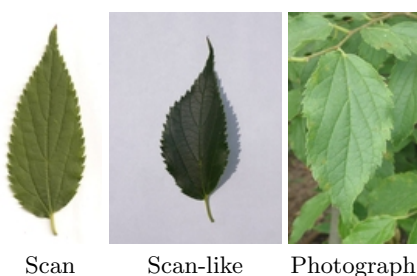


Scan      Scan-like      Photograph

**Fig. 1.** The 3 image types illustrated with the same species *Celtis australis L.*

### 2.2 Pl@ntLeaves metadata

Each image of Pl@ntLeaves dataset is associated with the following meta-data:

- *Date* date and time of shot or scan
- *Type* (acquisition type: *Scan*, *Scan-like* or *Photograph*)
- *Content* type (see figure3):
  - *Leaf* (tagged by default for all images, supposing that a picture is focused one single leaf),

---

[8] it is closed now, but a newer a application can be found at http://identify.plantnet-project.org/en/base/plantscan

- • *Picked leaf* most of the time mature leaf - not dry - on the floor,
- • *Upper side* and *Lower side* involving generally 2 images from one leaf,
- • *Branch* for some pictures containing enumerable leaves,
- • *Leafage.*
- – *Taxon* full taxon name according the botanical database[2](Regnum, Class, Subclass, Superorder, Order, Family, Genus, Species)
- – *VernacularNames* English commun name
- – *Author* name of the author of the picture
- – *Organization* name of the organization of the author
- – *Locality* locality name (a district or a country division or a region)
- – *GPSLocality* GPS coordinates of the observation

These meta-data are stored in independent xml files, one for each image. Figure 2 displays an example image with its associated xml data.

Additional but partial meta-data information can be found in the image's EXIF, and might include the camera or the scanner model, the image resolution and dimension, the optical parameters, the white balance, the light measures, etc.



**Fig. 2.** An image of Pl@ntLeaves dataset and its associated metadata

### 2.3 Pl@ntLeaves variability

The main originality of Pl@ntLeaves compared to previous leaf datasets, such as the Swedish dataset [22], the ICL dataset [1] or the Smithsonian one [3], is that it was built in a collaborative manner through a citizen sciences initiative. This makes it closer to the conditions of a real-world application: (i) leaves of the same species are coming from distinct trees living in distinct areas (ii) pictures and scans are taken by different users that might not used the same protocol to collect the leaves and/or acquire the images (iii) pictures and scans are taken at different periods in the year. Intra-species visual variability and view conditions variability are therefore more stressed-out which makes the identification more

| Scan | | Scan-like | Photograph | | |
|---|---|---|---|---|---|
| Upper side | Lower side | Branch | Picked leaf | Branch | Leafage |
|  | | | | | |

**Fig. 3.** Available sub-categories.

realistic but more complex. Figures 4 to 11 provide illustrations of the intra-species visual variability over several criteria including growing stages, colors, global shape, margin appearance, number and relative positions of leaflets, compound leaf structures and lobe variations. On the other side, Figure 12 illustrates the light reflection and shadows variations of *Scan-like* photos. It shows that this acquisition protocol is actually very different than pure scans. Both share the property of a limited noisy background but *Scan-like* photos are more complex due to the lighting conditions variability (flash, sunny weather, etc.) and the *un-flatness* of leaves. Finally, the variability of unconstrained *photographs* acquired on the tree and with natural background is definitely a much more challenging issue as illustrated in Figure 13.



Spring 2012   Summer 2010   Autumn 2011   Winter 2011

**Fig. 4.** Four leaves from distinct plants of one same species (Platanus x hispanica Wild) collected at different growing stages.

## 3 Task description

### 3.1 Training and Test data

The task was evaluated as a supervised classification problem with tree species used as class labels. A part of Pl@ntLeaves dataset was provided as training data whereas the remaining part was used later as test data. The two subsets

**Fig. 5.** Color variation of *Cotinus coggygria Scop.* (Eurasian smoketree)



**Fig. 6.** Global shape variation of *Corylus avellana L.* (European Hazel)



**Fig. 7.** Leaf's margin variation of *Quercus ilex L.* (Holm oak)



**Fig. 8.** Number of leaflets variation of *Fraxinus angustifolia Vahl* (Narrow-leafed Ash)



**Fig. 9.** Leaflets relative position variation of *Vitex agnus-castus L.* (Judas Tree)

were created with the respect to the individual plants: 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

**Fig. 10.** Two compound leaves organisations of one same plant of *Gleditsia triacanthos L.* (Honey locust): *pinnate* and *bipinnate*.



**Fig. 11.** Lobes variation of *Broussonetia papyrifera (L.) L'Hér. ex Vent.* (Paper mulberry) on one same plant.



**Fig. 12.** Light reflection and shadows variation of scan-like photos of *Magnolia Grandiflora L.* (Southern Magnolia)



**Fig. 13.** Variability of unconstrained photographs of *Acer platanoides* (Norway Maple)

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. The training subset was built first by integrating the whole ImageCLEF2011 plant task dataset. New pictures were then added to the training and test data, by selecting randomly new **individual plants**, in a such way to reach so far as possible a ratio around 2 individual plants in the training data for 1 individual plant in the test data. Detailed statistics of the composition of

the training and test data are provided in 1.

| | | Pictures | Individual plants | Contributors |
|---|---|---|---|---|
| **Scan** | **Train** | 4870 | 310 | 22 |
| | **Test** | 1760 | 157 | 10 |
| **Scan-like** | **Train** | 1819 | 118 | 8 |
| | **Test** | 907 | 85 | 10 |
| **Photograph** | **Train** | 1733 | 253 | 22 |
| | **Test** | 483 | 213 | 25 |
| **All** | **Train** | 8422 | 681 | 29 |
| | **Test** | 3150 | 455 | 38 |

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

### 3.2 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. Compared to last year the task was more related to plant species retrieval than a pure classification problem. This year evaluation metric was slightly modified in order to consider a ranked list of retrieved species rather than a single brute determination. Each test image was attributed with a score between 1 and 0 equals to the inverse of the rank of the correct species. 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:

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

$U$ : 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

$s_{u,p,n}$ : score between 1 and 0 equals to the inverse of the rank of the correct species 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*, *Photogragh*), 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.

## 4   Participants and techniques

A total of 11 groups submitted 30 runs, which confirms the successful participation rate of the previous year task on a new topic. Participants are mainly academics, specialized in computer vision and multimedia information retrieval, coming from all around the world: Australia (1), Brazil (1), China (1), France (3), Germany (2), India (1), Italy (1) and Turkey (1). We list below the participants and give a brief overview of the techniques they used to run the plant identification task. We remind here that ImageCLEF benchmark is a system-oriented evaluation and not a formal evaluation of the underlying methods. Readers interested by the scientific and technical details of any of these methods should refer to the ImageCLEF 2012 working note of each participant.

**ArTe-Lab (1 run)** [14] In their preliminary evaluations, these participants experiment several features and classification methods. They were guided by the wish to design a mobile application and thus aimed to have a good compromise between accuracy and computing time and cost. They retained the Pyramid of Histograms of Orientation Gradients (PHOG) and a variant of the HAAR descriptor evaluated as the most stable and robust features in their preliminary tests. They chose a multi-class probability estimation based on Support Vector Machine (SVM) classifiers with a one-vs-one, or pairwise comparisons, strategy. They choose to train a single SVM model for all 3 categories of images, without distinguish *Scan*, *Scan-like* and *Photograph*.

**Brainsignals (1 run)** [16] The aim of these participants was to focus less on possible computer-vision related techniques (especially without costly image preprocessing and feature extraction), and the much as possible to deal with machine learning techniques. First, they binarized the images using the Otsu algorithm, in order to extract two features: global shape features (from lateral projections), and local features (histograms of content types of small rectangular patches of bits, with and without sub-sampling). A Random Forest Classifier[8] was then used with 100 trees trained with several cross-validation steps on the

whole training data without distinguish *Scan*, *Scan-like* and *Photograph* categories.

**BTU DBIS (4 runs)** [**10**] This team used a rather non-typical approach not relying on classification techniques. Logical combinations of low-level features are expressed in a query language, the Commuting Quantum Query Language[21], and used to assess a document's similarity to a species. The principle is to combine similarity predicates as found in information retrieval and relational predicates common in databases. They tested 4 distinct methods without distinguishing the 3 categories *Scan*, *Scan-like* and *Photograph*. Their preliminary evaluations suggested to use MPEG7 Color Structure Descriptor eventually combined with GPS information. The first two runs explored two Query By Example based approaches, with or without GPS information. As a third approach, they combined the second approach with a Top-k Clustering (or "K-methoid") techniques which exploits relationships within the top-k results. The last run used a pure clustering image approach with GPS information, including the whole train and test datasets.

**IFSC USP (3 runs)** [**11**] Mainly focused on automatic shape boundary analysis, these participants adapted their approach according the image content and obtained the best results for the *Photograph* with a semi-automatic approach. Full-automatic approaches were used for *Scan* and *Scan-like*, but also on *Photograph* for one run. They started by an automatic leaf contour extraction with Otsu's method on *Scan* and *Scan-like*, while they used a k-means clustering algorithm in the RGB colorspace for *Photograph*. The semi-automatic approach on *Photograph* needs that a user marks leaf and background on regions automatically detected by a Mean Shift algorithm. Then, a merging process labelled gradually all content of the picture with these 2 parts. Contour were described with complex network[5], volumetric fractal dimension [**?**] and geometric parameters. Gabor filters, and local binary patterns were used too. They included for each run the GPS information. Classification was performed by using Linear Discriminant Analysis. The first run used all samples from the 3 images categories for the training stage, which paid off comparing to the second run where only *Scan* and *Scan-like* were used. Only Gabor and GPS features were used for the automatic run on *Photograph* because contour extraction were sufficiently reliable.

**Inria IMEDIA (3 runs)** [**6**] These participants used distinct approaches. For *Scan* and *Scan-like*, they applied a late fusion of an approach based on shape boundary features and a large-scale matching approach. The second run used an advanced shape context descriptor combining boundary shape information and local features within a matching approach. These two runs used a top-K decision rule as classifier, while the third used a multi-class SVM technique on contour based descriptors with a one-vs-one schema and a linear kernel. For *Photograph*, local features around constrained Harris points in order to reduce the impact

of the background, but automatic segmentation with a rejection criterion was attempted in order to extract shape features when possible. Two first runs used matching and top-K classifier. Last run used a method were Harris points were associated to bounding boxes detected as images' interesting zones; each box proposed a classification based on a multi-class SVM and a voting scheme unifying their responses. All preliminary evaluations used cross validation without splitting images from a same individual plants.

**LIRIS ReVes (3 runs) [12]** The method used by these participants is clearly oriented mobile application and is designed to cope with the challenges of complex natural images and to enable a didactic interaction with the user. The system first performs a two-step segmentation of the image, fully automatic on plain background images, and only guided by a rough coloring on photographs. It relies on the evaluation of a leaf model representing the global morphology of the leaf of interest, which is then used to guide an active contour towards the actual leaf margin. High-level geometric descriptors, inspired by the criteria used by botanists, are extracted on the obtained contour, along with more generic shape features, making a semantic interpretation possible. All these descriptors are then combined in a Random Forest classification algorithm[8].

**LSIS DYNI (3 runs) [19]** These participant used a modern computer vision framework involving feature extraction coupled with Spatial Pyramidal Matching for local analysis and large-scale supervised classification based on linear SVM with the 1-vs-all multi-class strategy. For all submitted runs, they do not consider the three image categories. They notably obtained good results on *Scan* and *Scan-like* without any segmentation and/or specific pre-processing, and they obtained the 3 best results for *Photograph* among all full-automatic approaches with a significant difference. The differences between the runs concerns the choice of features and their combination (in late fusion schema). They used a Multiscale and Color extension of the Local Phase Quantization[17], dense Multiscale (Color) Improved Local Binary Patterns with Sparse Coding of patches[24], and dense SIFT feature with Sparse Coding too. More they combined the descriptors, more they obtained better performances.

**Sabanci-Okan (2 runs) [25]** Mainly focused by automatic shape boundary analysis, these participants adapted their approach according the image content and obtained the best results for *Scan* and *Scan-like*. They used a balanced combination of 2 kinds of classifications, a Local Features Matcher and SVM classifiers. Only *Scan* and *Scan-like* images were used as a training dataset. They used two distinct segmentation methods: Otsu's method for *Scan*, *Scan-like*, and two methods based on watershed transform for *Photograph* (full or semi-automatic). The Local Feature Matcher exploited local features (shape context, HOG, local curvature) from salient points on the boundary of the leaf. The SVM classifiers used texture and optionally shape features if test image were from the category *Scan*, or *Scan-like* or interactively segmented *Photograph*. For automat-

ically segmented *Photograph* test images, the leaf boundary extractions were not sufficiently reliable for using shape information, and thus they used texture features for re-training the classifiers. These participants are among the rare ones to mention that they used cross validation without splitting images from a same individual plants in order to avoid over fitting problems.

**The Who IITK (4 runs)** [4] One originality is that these participants redefined 3 categories taking into account morphological properties of leaves: *Simple leaf in Scan or Scan-like*, *Compound leaf in Scan or Scan-like* and *Photograph*. For the "Simple" category, they developed an automatic segmentation method based on Otsu working on two steps in order to remove shadows, noisy background and the *petiole* of the leaf since it can affect the shape. For the "Compound" category, they used a combination of polynomial curve fitting and minimum-bounding-ellipse scores in order to extract the contour of the "most" representative *leaflet*. For *Photograph*, they developed a GrabCut [20] based interactive segmentation technique. Shapes were then described with complex network [5] and geometric parameters, and extracted specialized features on margin and venation. In order to fully automatise the process, they designed a *Simple/Compound* leaf classifier for test images. Note that on some "complex" photos where contour was too difficult to extract, like *bipinnate* compound leaves, they chose to use SURF with a bag-of-word technique. Random Forest Classification [8] for each category was finally used for the species prediction. The run giving the best result used in addition the Smote algorithm [13] designed to deal with unbalanced classification problems.

**ZhaoHFUT (3 runs)** [26] These participants investigated a linear Spatial Pyramidal Matching using Sparse coding (ScSPM) since this kind of approach achieve very good performance in object recognition. Indeed, SPM extends popular bag-of-features approach by taking account the spatial order of local appearances. They used the same structure in [24] and they tested thought the 3 runs the spatial encoding of dense SIFT, dense flip-SIFT descriptors (a SIFT extension considering leaf symmetry), and a late fusion of the two approaches. The sparse encoded features are pooled among several scales as spatial pyramid histogram representations, and these last representations are used as input of SVM classifiers tuned by cross-validations. The species predictions are formed by normalizing the degrees of attribution in different SVMs.

Table 2 attempts to summarize the methods used at different stages (feature, classification, subset selection,...) in order to highlight the main choices of participants. This table should be used in next section on result analysis, in order to see if there is some common techniques which tend to lead to good performances.

| Team | Segmentation | Features | Classification | Training subset(s) | S/C | IP |
|---|---|---|---|---|---|---|
| ArTe-Lab | × | PHOG HAAR-like | SVMs, one-vs-one | All | × | × |
| Brain-signals | Otsu-like | Lateral proj. shape Histo. of local binary content | Random Forests | All | × | × |
| BTU DBIS | × | MPEG7 CSD (+GPS) | Query By Example | All | × | × |
| IFSC USP | AllScans: Otsu Photo: Auto. k-Means Photo: Interactive MeanShift | Numerous shape and texture, GPS | Linear Discriminant Analysis | All AllScans | × | × |
| Inria IMEDIA | AllScans: Otsu Photo: Otsu+abort criterion | Harrislike+constraints+SURF +Local shape and texture Shape context, DFH shape | Large scale matching +knn classifier SVM one-vs-one | All AllScans | × | ✓ |
| LIRIS ReVes | Guided active contour | Geometric model parameters, Numerous shape and texture, Morphological features | Random Forests | Scan Scan-like Photo. | ✓ | × |
| LSIS DYNI | × | (Multiscale local (color) texture, SIFT) + Sparse coding Spatial Pyramidal Matching | Linear SVM 1-vs-all multi-class strategy | All | × | × |
| Sabanci Okan | AllScans: Otsu Photo: Semi-auto watershed | Numerous shape and texture, Morphological features | SVM classifiers | AllScans | × | ✓ |
| TheWho IITK | AllScans: Otsu Photo: interactive GrabCut | Numerous shape and texture, Morphological features, SURF | Random Forest (+Smote) | AS-Simple, AS-Compound, Photo. | ✓ | × |
| Zhao-HFUT | × | (flip-)SIFT + Sparse coding Spatial Pyramidal Matching | Linear SVM 1-vs-all multi-class strategy | All | × | × |

**Table 2.** Approaches used by participants. In training subset column *All* means that the 3 categories where not distinguished, while *AllScans* means *Scan* and *Scan-like*. Column S/C indicates if participants treated distinctively Simple and Compound. Column IP indicates if participants avoid to split images from a same Individual Plant during evaluation on training dataset.

# 5 Results

## 5.1 Global analysis

Figures 14, 15 and 16 present the normalized classification scores of the 30 submitted runs for each of the three image types. Note that two colors are used in the graphs to distinguish fully automatic methods from humanly assisted methods. Table 3 finally presents the same results but with detailed numerical values (in that case a run is labelled as automatic only if all image types were processed fully automatically).
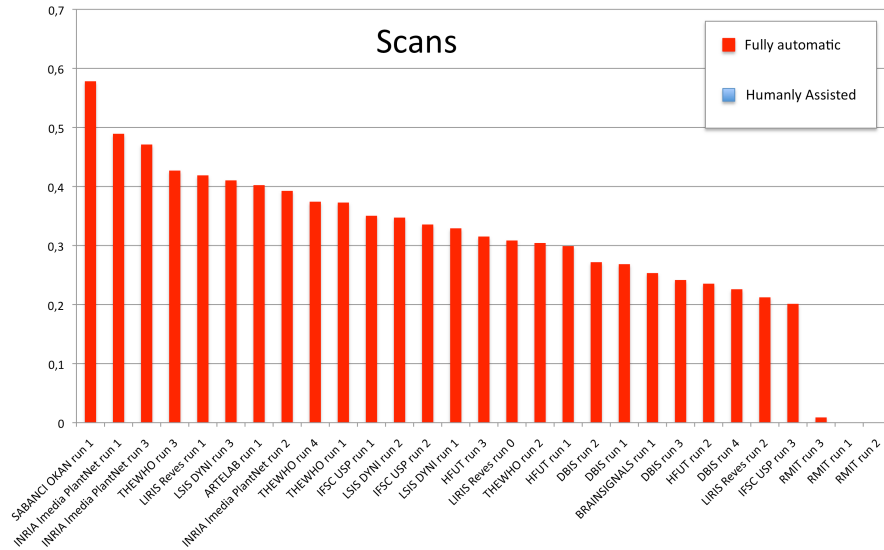


**Fig. 14.** Scores for *Scan* images

**Best runs:** Although scores are more uniform across the three image types, there is still no run this year achieving the best performances on the three types. There is even no run in the top-3 of each image type. Sabanci-Okan runs achieved very good results on both *Scan* (1st) and *Scan-like* (2nd). TheWho runs achieve good results across the 3 categories but do not obtain the best performances on any of them. IFSC USP run1 obtained the best performance on *Photograph* category using semi-supervised segmentation but gets modest performances on *Scan*. Similarly, INRIA run 2 obtained the best performance on *Scan-like* but gets mitigated performances on *Scan* and *Photograph*. Finally, the 3rd run of LSIS DYNI is interesting even if it is positioned only at the 9th rank when looking at the average score across all categories: it actually obtained the best fully automatic performances on *Photograph* category which is actually the most challenging task. A normalized recognition rate of 32% on that

| Run name | retrieval type | run-type | Scans | Scan-like | Photos | Avg |
|---|---|---|---|---|---|---|
| SABANCI OKAN run 2 | Visual | HA | **0,58** | 0,55 | 0,22 | **0,45** |
| THEWHO run 3 | Visual | HA | 0,43 | 0,4 | 0,49 | 0,44 |
| SABANCI OKAN run 1 | Visual | Auto | **0,58** | 0,55 | 0,16 | 0,43 |
| IFSC USP run 1 | Mixed | HA | 0,35 | 0,41 | **0,51** | 0,42 |
| LIRIS Reves run 1 | Visual | HA | 0,42 | 0,51 | 0,33 | 0,42 |
| INRIA Imedia PlantNet run 1 | Visual | Auto | 0,49 | 0,54 | 0,22 | 0,42 |
| INRIA Imedia PlantNet run 2 | Visual | Auto | 0,39 | **0,59** | 0,21 | 0,40 |
| THEWHO run 4 | Mixed | HA | 0,37 | 0,35 | 0,43 | 0,38 |
| LSIS DYNI run 3 | Visual | Auto | 0,41 | 0,42 | 0,32 | 0,38 |
| THEWHO run 1 | Mixed | HA | 0,37 | 0,34 | 0,43 | 0,38 |
| INRIA Imedia PlantNet run 3 | Visual | Auto | 0,47 | 0,46 | 0,15 | 0,36 |
| IFSC USP run 2 | Mixed | HA | 0,34 | 0,43 | 0,3 | 0,36 |
| LSIS DYNI run 2 | Visual | Auto | 0,35 | 0,35 | 0,29 | 0,33 |
| ARTELAB run 1 | Visual | Auto | 0,4 | 0,37 | 0,14 | 0,30 |
| LSIS DYNI run 1 | Visual | Auto | 0,33 | 0,32 | 0,24 | 0,30 |
| THEWHO run 2 | Mixed | HA | 0,3 | 0,25 | 0,24 | 0,27 |
| LIRIS Reves run 0 | Visual | HA | 0,31 | 0,26 | 0,19 | 0,25 |
| HFUT run 3 | Visual | Auto | 0,32 | 0,26 | 0,11 | 0,23 |
| HFUT run 1 | Visual | Auto | 0,3 | 0,24 | 0,09 | 0,21 |
| DBIS run 2 | Visual | Auto | 0,27 | 0,17 | 0,17 | 0,21 |
| DBIS run 1 | Mixed | Auto | 0,27 | 0,18 | 0,17 | 0,20 |
| DBIS run 3 | Mixed | Auto | 0,24 | 0,16 | 0,16 | 0,19 |
| BRAINSIGNALS run 1 | Visual | Auto | 0,25 | 0,22 | 0,05 | 0,17 |
| HFUT run 2 | Visual | Auto | 0,24 | 0,17 | 0,09 | 0,17 |
| DBIS run 4 | Mixed | Auto | 0,23 | 0,14 | 0,13 | 0,17 |
| IFSC USP run 3 | Mixed | Auto | 0,2 | 0,14 | 0,12 | 0,16 |
| LIRIS Reves run 2 | Visual | Auto | 0,21 | 0,15 | 0,05 | 0,14 |
| RMIT run 3 | Visual | Auto | 0,01 | 0 | 0 | 0,00 |
| RMIT run 1 | Visual | Auto | 0 | 0 | 0,01 | 0,00 |
| RMIT run 2 | Visual | Auto | 0 | 0 | 0 | 0,00 |
| Mean | | | 0,32 | 0,31 | 0,21 | 0,28 |
| Median | | | 0,32 | 0,29 | 0,18 | 0,26 |
| **Max** | | | **0,58** | **0,59** | **0,51** | **0,56** |

**Table 3.** Normalized classification scores for each run and each image type. HA = humanly assisted, Auto=Full automatic. Top results per image type are highlighted in bold

category makes an automatic real-world application more realistic (whereas last year's automatic runs on photographs did not exceed 20% and with less species).

**Impact of the number of species:** A first general remark is that the scores are globally lower than the one obtained during the 2011 campaign whereas the new evaluation metric of the 2012 campaign is more advantageous for participants (only the top-1 specie contributed to the score within the old metric whereas the
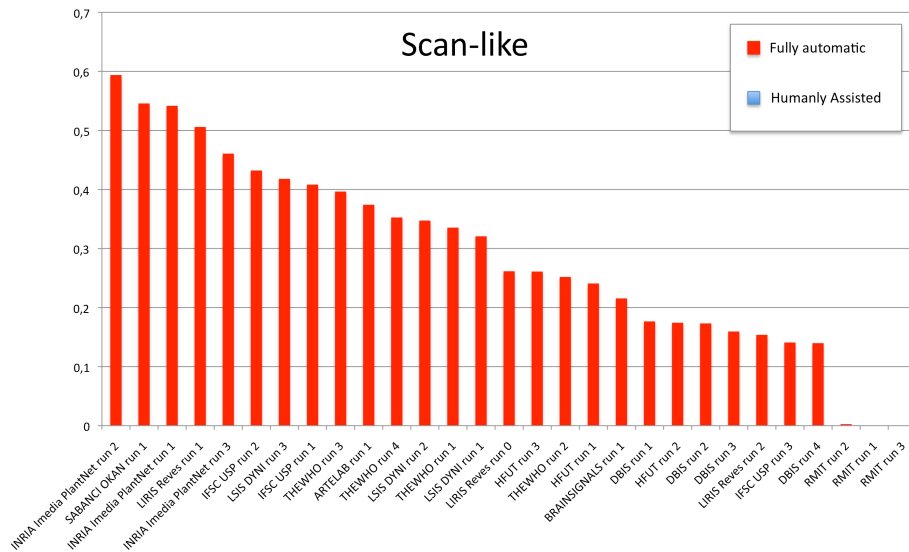
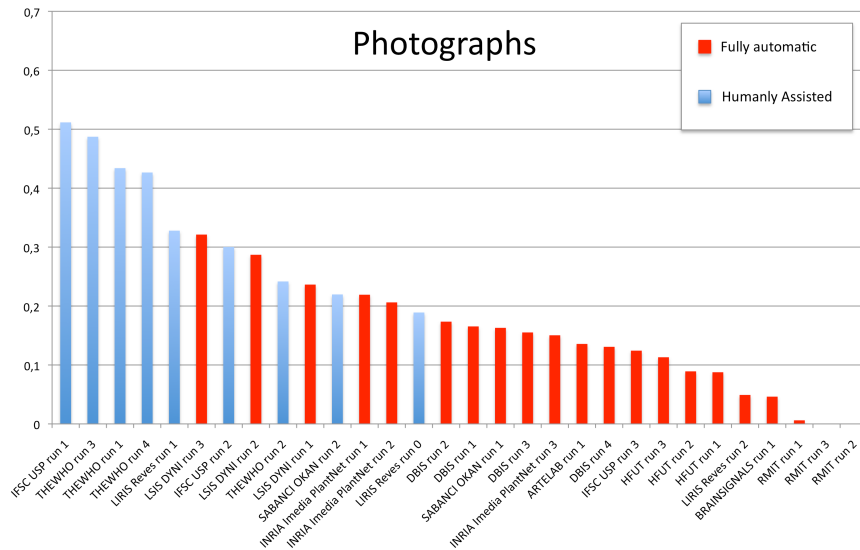**Fig. 15.** Scores for *Scan-like* images



**Fig. 16.** Scores for *Photograph* images

new one allows retrieving the right species at larger ranks). That means that the difficulty of the task increased, and that the technological progress achieved by the participants did not compensate the increased difficulty. The main change between the two campaigns being the number of species covered by the task (from 70 to 126 species), it is very likely that this growth explains the overall lower performances. More classes actually involve more confusion. From a

botanical point of view, the 2012 dataset actually contains more pairs of morphologically similar species. The species *Betula pendula Roth* was for instance quite well recognized in many runs during the 2011 campaign but is this year often confused with other species, probably *Betula pubescens Ehrh.* or *Betula utilis var. jacquemontii* sharing several morphological attributes.

**Impact of the image type:** Another global remark is that the scores of the 2012 campaigns are more uniform across the different image types than the year before. So that it is not true anymore that the performances are degrading with the complexity of the acquisition image type. This might be due to several reasons: (i) many species in the training set are relatively to other types more populated with *Scan-like* images than last year, (ii) several participants did use semi-supervised approaches to extract leaf contours in the *Photograph* category, (iii) participants may have work more on improving their methods for *Scan-like* and *Photograph* (since scans were identified more easily in the 2011 campaign).

**About full automatic approaches and using semi-supervised segmentation:** Every runs were performed with full automatic methods on *Scan* and *Scan-like*, with or without automatic image segmentation, revelling that all teams conceive that the problem is solvable in a long term. But with a max of 0.58, and average and a median around $0,32$ for *Scan*, there is still an important room of improvement especially when all tree species from a given flora (like around the 500 to 600 tree species in France) will be considered.

Concerning *Photograph*, results demonstrated that a quite accurate leaf contour extraction in all training dataset performed with semi-supervised segmentation approaches enable to tend towards equivalent performances for *Scan* and *Scan-like* like for the IFSC-USP or TheWho teams. However it is not the case for Sabanci-Okan semi-automatic run, maybe because they performed semi-supervised segmentation on test-images only while learning models on *Scan* and *Scan-like*. In the other side, IFSC-USP team obtained lower results on *Scan* and *Scan-like* than on *Photograph* which maybe highlights that accurate leaf contour is required. Moreover, it is difficult to conceive an approach needing tens or even hundreds thousands of interactive leaf segmentations of a dataset on a entire flora.

**About matching and shape boundary approaches:** The most frequently used class of methods is *shape boundary analysis* (16 runs among 30 used), exclusively or not) which is not surprising since state-of-the-art methods addressing leaf-based identification in the literature are mostly based on leaf segmentation and shape boundary features [7, 18, 9, 23, 5]. Last year, one important remark was that more generic image retrieval approaches like matching could provide better results especially on *Scan*. But, rather than opposing these two approaches, several participants combined them with success. This suggests that modelling leaves as part-based rigid and textured objects might be complement to shape boundary approaches that do not characterize well margin details, ribs or limb

texture.

**Impact of the training subsets and cross-validations:** Considering test images of *Scan* and *Scan-like* categories, we can say it was preferable to exclude *Photograph* images from training dataset and to not distinguish *Scan* and *Scan-like* during preliminary evaluations. But for *Photograph* test images, this rule seems not to be confirmed if we look the best performances obtained automatically by LSIS DYNI who used all images in the training dataset. Moreover, one can note that Sabanci-Okan and Inria teams seem to have take benefit from the respect of individual-plant during their cross-validation. Indeed, they mentionned that they do not split images from one same individual plant, in order to avoid overfitting problems, because images from a one same plants (and thus one same event) are very similar. This approach could be one key of the success of these runs independently to the relevance of the methods.

**About using metadata:** Using geo-tags to help the identification has been tried again this year by some participants (IFSC USP, THEWHO and DBIS). Unfortunately, this year again, the results show that adding GPS information in the identification process is likely to degrade the performances. Best runs of TheWho and DBIS are actually the ones that did not use any additional metadata complementary to their baseline visual-based technique. The main reason is probably that the geographic spread of the data is limited (French Mediterranean area). So that most species of the dataset might be identically and uniformly distributed in the covered area. Geo-tags would be for sure more useful at a global scale (continent, countries). But at a local scale, the geographical distribution of plants is much more complex. It usually depends on localized environmental factors such as sun exposition or water proximity that would require much more data to be modelled.

We have to note that some potentially relevant information were not or very few explored. The content tags (*Leaf*, *Picked leaf*, *Leafage*) for instance were only used by Inria Team. Date could have been used in order to identify evergreen plants and improve performances on these species. None of teams explored either the hierarchical taxonomy structure which could be a source of improvement. But some participants like The Who and ReVes LIRIS teams, created their own metadata about the leaf organisation (*Simple* or *Compound*) in order to treat differently these two categories of leaves, which seems to have paid off for improving performances.

## 5.2   Performances per species

This section aims to analyse, considering the results of the various approaches, how and which part of the task were difficult. We analyse which species, images and kind of leaves according to morphological attributes, were easy or very hard to identify.

We attempt here to evaluate which species are more difficult to identify than

others. We focus only on *Scan* in order to limit view conditions bias. We computed some basic statistics (mean, median, max) on the score by species over the runs of all participants. Figure 17 shows these statistics by species with a decreasing order of the max score values. We can note that few species enable to obtain good performances over all approaches. For around one third there is a least one approach which performed perfectly on all images of theirs associated species. Most of them were species yet contained in the previous year dataset, and thus most of the time with more individual-plants and images. Figure 18 shows in addition how are related max scores to the number of individual plant by species and it indicates clearly how more a species has some individual plants (as existing species previous year) more the identification of the species is reachable by most of the methods. For species with a small number of individual plants it is more difficult to understand precisely the score variations. They can be due to morphological variations but also to different view conditions (we will see it next subsections).

Coming back to figure 17, for the other two-thirds of species, none of the tested methods perform perfectly. However, by comparing max and mean scores, we can say that there is each time on method performing distinctly better than the other. It is encourageing for all teams, beacuse it means there is for all these species a room of improvement.

### 5.3 Performances per image

To qualitatively assess which kind of images causes good results and which one makes all methods failed, we sorted all test pictures by the rate of succesfull runs in which they were correctly identified (where the correct species was retrieved at first rank). Figure 19 attempts to represent that for each image category and with a decreasing order. First we can note that that *Scan* and *Scan-like* have more or less the same distribution relatively to their number of images, confirming that these two categories have a same level of difficulty in spite of the great variety of the methods experimented through the run. Moreover, the figure confirms that *Photograph* is definitely the most difficult category: a handful of photographs has a rate of successful runs over 0.5 while at the same time around one third of *Scan* and *Scan-like* images have it. 260 images were not correctly identified by any run with a majority of unconstrained photos (63 *Scan*, 27 *Scan-like*, 168 *Photograph*.

Figure20 displays 4 very well identified images with at least a rate of 0.8 successful run). They all are very standard leaf images similar to the one found in books illustrations.

Concerning *Photograph*, the three images with the higher rate of successful runs are showed in first line in figure21. One explication of these relative good rates can be that the associated species contains after 3 years numerous images and individual plants now, but also that these pictures were well framed, with a close-up on the whole leaf. In comparison photos from the same species in second line of figure21 have a very low, even 0, rate of successful runs: we can note that
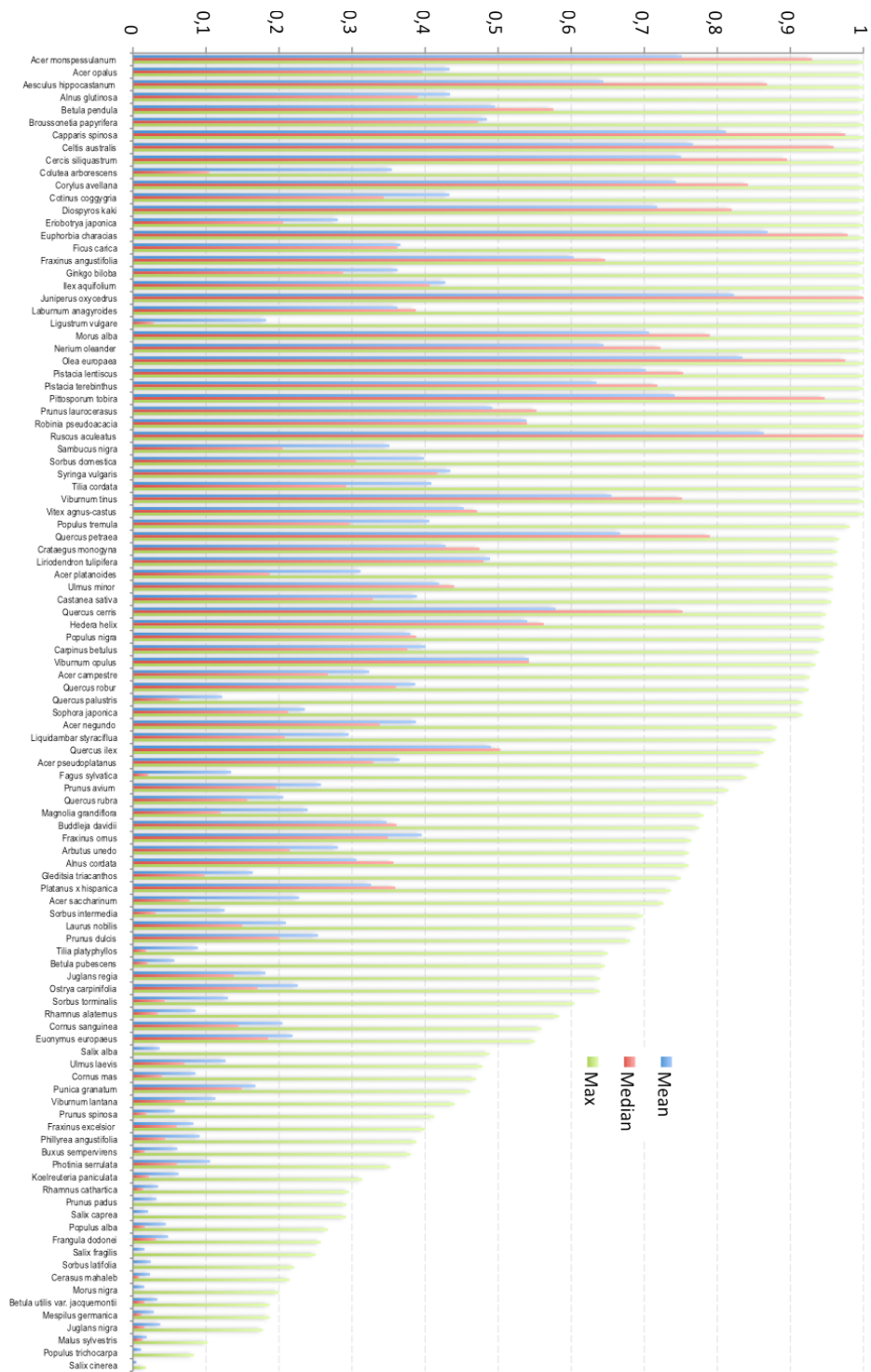
**Fig. 17.** Mean, median and max over the runs of all participants for each specie - scan test images only
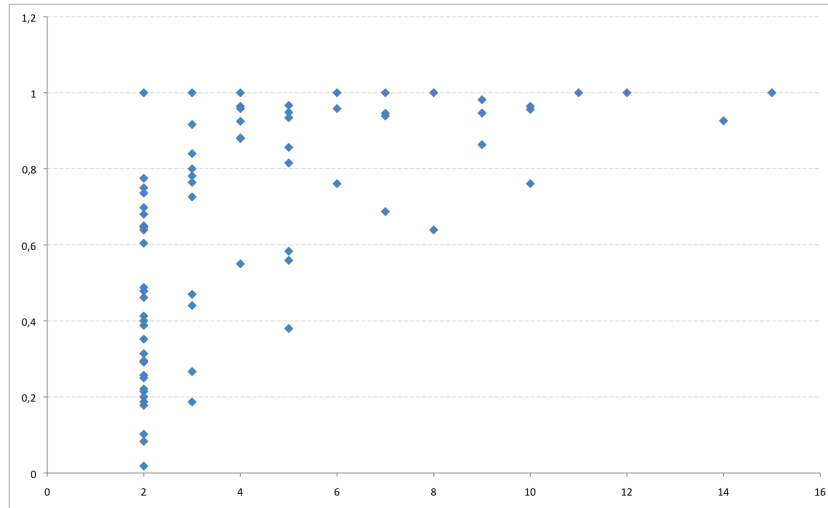
**Fig. 18.** Max score over the runs by number of individual-plants for each species (scan test images only).
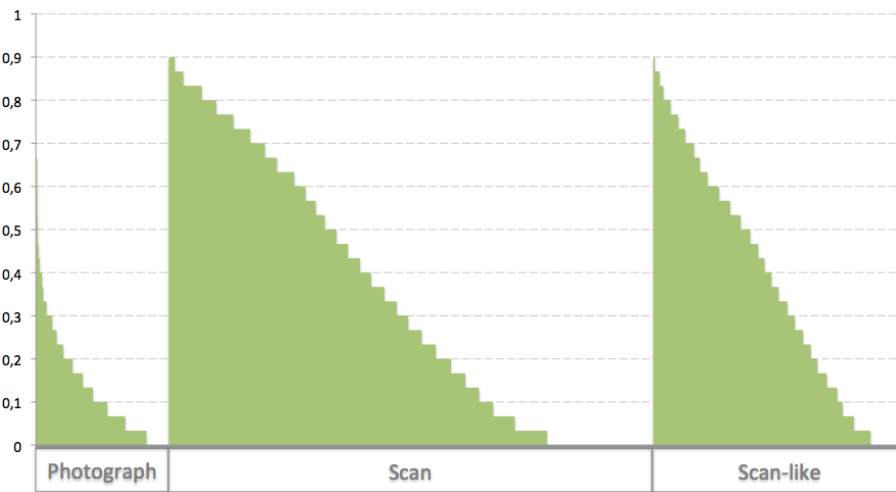


**Fig. 19.** Distribution of the rates of successful runs for each test-image, re-arranged by the three image categories and with a decreasing order.

these pictures are not well framed, too small, too dark, or with a part of the leaf out of the frame...

However, we have to be careful with the "easy" or "difficult" images ccording to the methods used in the runs. We notice that numerous of the "easiest"

**Fig. 20.** 4 well identified images (more than 24/30 successful runs).

| | | | |
|---|---|---|---|
| **Easy** | 0.66 | 0.43 | 0.43 |
| **Difficult** | 0.03 | 0.03 | 0.0 |
| Sp. | *Cercis siliquastrum L.* | *Corylus avellana L.* | *Pittosporum tobira (Thunb.) W.T. Aiton* |

**Fig. 21.** The 3 photos with the highest rate of successful runs, and 3 photos wih very low ones from the same species.

images can be considered "too much easy" because on some species we can have images from a same "event" in the training and the test dataset. Indeed, we remind that we split the Pl@ntLeaves II dataset with the respect of individual plants, i.e. 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. However, some individual plants from a same species are closely related by one same "event" because they

were observed the same day, by the same contributor, and images were collected with the same device, with similar pose, framing and conditions. Moreover, the individual plants from a same "event" are observed at the same place and can be parent tree and/or present the same morphological attributes in function of the environment. In other words, these individual plants can be considered as one same individual-plant, and Since we not split the dataset with the respect of the notion of "event", it still have some bias between test and training datasets, as as it can be observed in figure 22.

test image 10921
indiv. 209
12/10/11

train image 5483
indiv. 42
12/10/11

train image 6358
indiv. 103
12/10/11

**Fig. 22.** One test image identified by a high rate of 0.9 successful runs, but, which can be considered "too easy" because the only individual plants of the associated species *Rhus coriaria L.* in the training dataset were observed by a same contributor (DB), the same day (12/10/11), and probably at the same place in Monferrier-sur-Lez.

This notion of "event", instead of "individual-plant", should be considered in a next plant identification task.

## 6  Conclusions

This paper presented the overview and the results of ImageCLEF 2012 plant identification testbed following the pilot one in 2011. The number of participants increased from 8 to 11 groups showing an increasing interest in applying multimedia search technologies to environmental challenges (in 8 distinct countries). The main change between the two campaigns was the number of species covered by the task which was increased from 70 to 126 species. The technological progress achieved by the participants did unfortunately not compensate this increased difficulty since the overall performances degrade. This shows that scaling state-of-the-art plant recognition technologies to a real-world application with thousands of species might still be a difficult task. On the other side, a positive point was that the performances obtained on unconstrained pictures of leaves with non-uniform background were better than last year. The use of semi-supervised segmentation techniques has notably been a key element towards boosting performances showing that interactivity might play an important role

towards successful applications, typically on mobile phones. But, the best automatic method also provided very honourable results showing that the door is not closed for a fully automatic identification. Interestingly this method was a generic visual classification technique without any specificity related to plants. Regarding the other image types, the best method on scans used a combination of many leaf boundary shape and texture features trained with a SVM classifier. The best run on pseudo-scans (photographs with uniform backgrounds) was achieved by another technology purely based on shape-context features that are already known to provide good performances for leaf recognition in the literature. Finally, using GPS data in the recognition process was still not beneficial probably because the geographic spread of the data is limited. A possible evolution for a new plant identification task in 2013 is to extend the task to multiple organs including flowers, trunks and fruits.

## Acknowledgements

## References

1. The icl plant leaf image dataset, http://www.intelengine.cn/English/dataset
2. Tropicos (August 2012), http://www.tropicos.org
3. Agarwal, G., Belhumeur, P., Feiner, S., Jacobs, D., Kress, J.W., R. Ramamoorthi, N.B., Dixit, N., Ling, H., Mahajan, D., Russell, R., Shirdhonkar, S., Sunkavalli, K., White, S.: First steps toward an electronic field guide for plants. Taxon 55, 597–610 (2006)
4. Arora, A., Gupta, A., Bagmar, N., Mishra, S., Bhattacharya, A.: A Plant Identification System using Shape and Morphological Features on Segmented Leaflets. In: Working notes of CLEF 2012 conference (2012)
5. Backes, A.R., Casanova, D., Bruno, O.M.: A complex network-based approach for boundary shape analysis. Pattern Recognition 42(1), 54 – 67 (2009)
6. 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: Working notes of CLEF 2012 conference (2012)
7. Belhumeur, P., Chen, D., Feiner, S., Jacobs, D., Kress, W., Ling, H., Lopez, I., Ramamoorthi, R., Sheorey, S., White, S., Zhang, L.: Searching the world4s herbaria: A system for visual identification of plant species. In: ECCV, pp. 116–129 (2008)
8. Breiman, L.: Random forests. Mach. Learn. 45(1), 5–32 (Oct 2001), http://dx.doi.org/10.1023/A:1010933404324
9. 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)

10. Böttcher, T., Schmidt, C., Zellhöfer, D., Schmitt, I.: BTU DBIS' Plant Identification Runs at ImageCLEF 2012. In: Working notes of CLEF 2012 conference (2012)
11. Casanova, D., ao Batista Florindo, J., Gonçalves, W.N., Bruno, O.M.: IFSC/USP at ImageCLEF 2012: Plant identification task. In: Working notes of CLEF 2012 conference (2012)
12. Cerutti, G., Antoine, V., Tougne, L., Mille, J., Valet, L., Coquin, D., Vacavant, A.: ReVeS Participation - Tree Species Classification using Random Forests and Botanical Features. In: Working notes of CLEF 2012 conference (2012)
13. Chawla, N.V., Bowyer, K.W., Hall, L.O., Kegelmeyer, W.P.: Smote: Synthetic minority over-sampling technique. Journal of Artificial Intelligence Research 16, 321–357 (2002)
14. Gallo, I., Zamberletti, A., Albertini, S.: Fast Tree Leaf Image Retrieval using a Probabilistic Multi-class Support Vector. In: Working notes of CLEF 2012 conference (2012)
15. Goeau, H., Joly, A., Selmi, S., Bonnet, P., Mouysset, E., Joyeux, L.: Visual-based plant species identification from crowdsourced data. In: Proceedings of ACM Multimedia 2011 (2011)
16. Grozea, C.: Brainsignals Submission to Plant Identification Task at ImageCLEF 2012. In: Working notes of CLEF 2012 conference (2012)
17. Heikkila, J., Ojansivu, V., Rahtu, E.: Improved blur insensitivity for decorrelated local phase quantization. In: Proceedings of the 2010 20th International Conference on Pattern Recognition. pp. 818–821. ICPR '10, IEEE Computer Society, Washington, DC, USA (2010), http://dx.doi.org/10.1109/ICPR.2010.206
18. 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)
19. Paris, S., Halkias, X., Glotin, H.: Participation of LSIS/DYNI to ImageCLEF 2012 plant images classification task. In: Working notes of CLEF 2012 conference (2012)
20. Rother, C., Kolmogorov, V., Blake, A.: grabcut": interactive foreground extraction using iterated graph cuts. ACM Trans. Graph pp. 309–314 (2004)
21. Schmitt, I.: Qql: A db&ir query language. The VLDB Journal 17(1), 39–56 (Jan 2008), http://dx.doi.org/10.1007/s00778-007-0070-1
22. Söderkvist, O.J.O.: Computer Vision Classification of Leaves from Swedish Trees. Master's thesis, Linköping University, SE-581 83 Linköping, Sweden (September 2001), liTH-ISY-EX-3132
23. Yahiaoui, I., Hervé, N., Boujemaa, N.: Shape-based image retrieval in botanical collections. In: Advances in Multimedia Information Processing - PCM 2006, vol. 4261, pp. 357–364 (2006)
24. Yang, J., Yu, K., Gong, Y., Huang, T.: Linear spatial pyramid matching using sparse coding for image classification. In: in IEEE Conference on Computer Vision and Pattern Recognition(CVPR (2009)
25. Yanikoglu, B., Aptoula, E., Tirkaz, C.: Sabanci-Okan at ImageCLEF 2012 Plant Identification Competition: Combining Features and Classifiers (2012)
26. Zheng, P., Zhao, Z.Q., Glotin, H.: ZhaoHFUT at ImageCLEF 2012 Plant Identification Task. In: Working notes of CLEF 2012 conference (2012)