

MIRACL at LifeCLEF 2014: Multi-organ observation for Plant Identification

Hanen Karamti, Sana Fakhfakh, Mohamed Tmar, Walid Mahdi, and
Faiez Gargouri

MIRACL Laboratory, City ons Sfax, University of Sfax, B.P.3023 Sfax TUNISIA
karamti.hanen@gmail.com,
sanafakhfakh@yahoo.fr,
mohamed.tmar@isimsf.rnu.tn,
walid.mahdi@isimsf.rnu.tn,
faiez.gargouri@fsegs.rnu.tn

Abstract. ImageCLEF 2014 has a challenge based on analysis for identifying plants. This article describes our first participation to the multi-image plant observation queries task of PlantCLEF 2014. The task will be evaluated as a plant species retrieval task based on multi-image plant observations queries. The goal is to retrieve the correct plant species among the top results of a ranked list of species returned by the evaluated system.

In this paper, we present two method. Our first method is purely visual and entirely automatic, using only the image information. One should mention that the total time spent with preparing this submission was only about three week. The results were accordingly fairly poor. The challenge of our second method is to identify plant species based on combination of textual and structural context of image. Indeed, we have used the meta-data in our system for exploring the image characteristics. Our approach is based on a modern technique for exploitation of structure of XML document. Also, the results were accordingly fairly poor.

Although our results are not quite promising as compared to other participant groups, they can still guide our work in this field for some conclusions reached.

Keywords: plant observation, feature extraction, ImageClef, XML, image retrieval.

1 Introduction

The plant identification has become an important and challenging research area since it is estimated that approximately one half of world plant species is still not cataloged. Among such unidentified species one may find, for instance, the healing of a disease or a plant that can cooperate in the equilibrium of the ecosystem around it.

Despite the importance of studies related to the description and categorization of plants, this task remains still a difficult for a botanist which is limited to a specific amount of information about the vegetal. Furthermore, among the information which may be collected, the most relevant for the botanist analysis are flowers and fruits. However, it turns out that in most cases these elements are observed only in specific periods of the year. This is a complicated issue given that the observation may not be possible when these characteristics are noticeable.

A solution for this impasse is the identification of plant by observing its different organs, such as it has been demonstrated in [1]. Indeed, botanists usually observe simultaneously several organs like the leaves and the fruits or the flowers in order to disambiguate species which could be confused if only one organ were observed. Moreover, if only one organ is observed, such as the bark of a deciduous plant during winter where nothing else is observable, then the observation of this organ with several photos related to different point of views could be more informative than only one point of view.

Thus, image analysis based on computational tools is a worthwhile approach in order to help the botanist or even provide by itself a reliable outcome for the classification task. In this context, ImageCLEF the Combined Lab Evaluation Forum (CLEF) hosts an annual competition on identification plant species.

In this year, ImageCLEF organizes a new challenge dedicated to botanical data (called PlantCLEF 2014), the species identification task won't be image-centered but observation-centered. The aim of the task will be to produce a list of relevant species for each observation of a plant of the test dataset, i.e. one or a set of several pictures related to a same event: one same person photographing several detailed views on various organs the same day with the same device with the same lightening conditions observing one same plant. The main novelties compared to the last years are the following: An explicit multi-image query scenario, User ratings on image quality, a new type of view called '*Branch*' additionally to the 6 others views (*Scan*, photos of *Flower*, *Fruit*, *Stem*, *Leaf* and *Entire views*), and basically more species (the number of species will be this year about 500, which is an important step towards covering the entire flora of a given region).

In this context, the following paper describes our (team MIRACL) approach for the LifeCLEF, task on multi-image plant observation queries. Our focus for this endeavor was on three main points: (i) image retrieval based on content that requires a thorough knowledge of low-level features of image. (ii) Image retrieval based on context based on information extracted in proximity of image for example: title or caption of image, textual description, etc. (iii) the combination of textual and visual features.

The following Section 2 describes the materials and methods used. In Section 3, we describe the results of our experiments. Finally, some conclusions reached are presented in Section 4.

2 Material and Methods

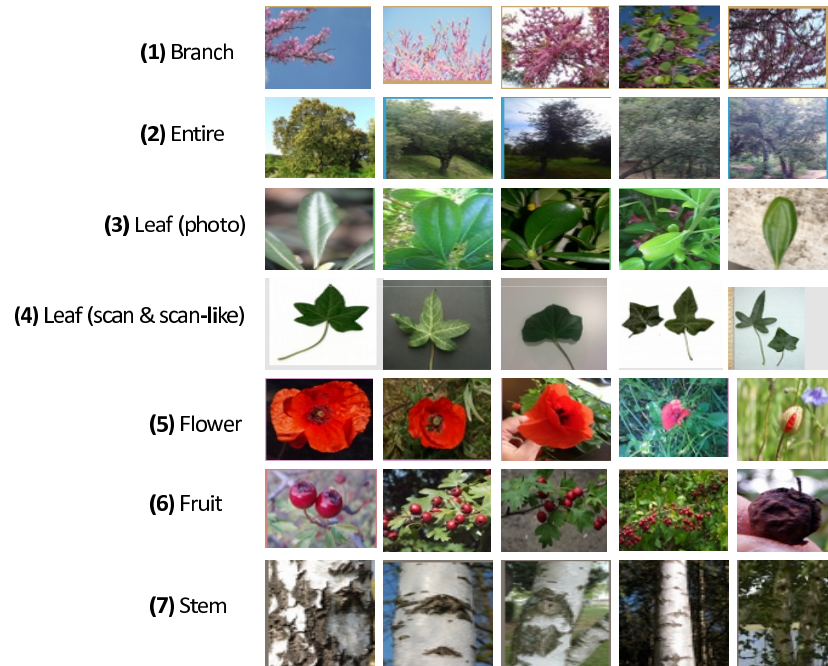


Fig. 1. Sample images for the database of PlantCLEF 2014

The task will be based on Pl@ntView dataset which focuses on 500 herb, tree and fern species centered on France (some plants observations are from neighboring countries). This database (shows figure 1) is maintained by the French project *PlantNet* (INRIA, CIRAD, Telabotanica) and the French CNRS program *MASTODONS*.

It contains more than 60000 pictures belonging each to one of the 7 types of view reported into the meta-data, in a xml file (one per image) with explicit tags [8]. For (1), (2), (3), (5), (6) and (7), the images are taken directly from the trees. For (4), the images are oriented vertically along the main natural axis and with the petiole visible collected using flatbed scanners. These images haven't an uniform background.

Each image has an xml file associated that contain the following meta-data:

- **ObservationId**: the plant observation ID from which several pictures can be associated
- **FileName**

- **MediaId**
- **View Content:** Branch, Entire, Flower, Fruit, Leaf, LeafScan, Stem
- **ClassId:** the class number ID that must be used as ground-truth. It is a numerical taxonomical number used by Tela Botanica
- **Species:** the species names (containing 3 parts: the Genus name, the Species name, the author(s) who discovered or revised the name of the species)
- **Genus:** the name of the Genus, one level above the Species in the taxonomical hierarchy used by Tela Botanica
- **Family:** the name of the Family, two levels above the Species in the taxonomical hierarchy used by Tela Botanica
- **Date:** (if available) the date when the plant was observed,
- **Vote:** the (round up) average of the user ratings on image quality
- **Locality:** (if available) locality name, most of the time a town
- **Latitude & Longitude:** (if available) the GPS coordinates of the observation in the EXIF metadata, or, if no GPS information were found in the EXIF, the GPS coordinates of the locality where the plant was observed (only for the towns of metropolitan France)
- **Author:** name of the author of the picture, And if the image was included in previous plant task:
- **Year:** ImageCLEF2011, ImageCLEF2012, ImageCLEF2013, PlantCLEF2014 when the image was integrated in the benchmark
- **IndividualPlantId2013:** the plant observation ID used last year during the ImageCLEF2013 plant task,
- *ImageID2013:* the image id.jpg used in 2013.

The full database is split into training and testing dataset. The train dataset has 47815 images (1987 of '*Branch*', 6356 photographs of '*Entire*', 13164 of '*Flower*', 3753 '*Fruit*', 7754 of '*Leaf*', 3466 '*Stem*' and 11335 '*scans*' and '*scan-like*' pictures of '*leaf*').

The test dataset results in 8163 plant-observation-queries. These queries are based on 13146 images (731 of '*Branch*', 2983 photographs of '*Entire*', 4559 of '*Flower*', 1184 '*Fruit*', 2058 of '*Leaf*', 935 '*Stem*' and 696 '*scans*' and '*scan-like*' pictures of '*leaf*').

2.2 Methods

In this section, we will present our methods.

Feature Extraction Two main functionalities are supported of feature extraction (*run 1*): data extraction processing and query processing. The data extraction processing is responsible for extracting appropriate features from images and storing them into their feature vectors. This process is usually performed offline. A feature vector V_I of an image I can be thought of as a list of low-levels features (C_1, C_2, \dots, C_m) , where m is the number of features.

We have used three descriptor to features extraction:

- A color layout descriptor (CLD) [3] is designed to capture the spatial distribution of color in an image. The feature extraction process consists of two parts: grid based representative color selection and discrete cosine transform with quantization. The image is divided into an 8×8 grid. An 8×8 pixel image is created, with each pixel given the representative color from the corresponding grid area in the image. The 8×8 matrix is transformed with the discrete cosine transform. Finally, a zigzag scan is performed on the matrix. The resulting matrices (one for each color component) make up the descriptor.
- A Edge Histogram Descriptor (EHD) [2] is a texture descriptor proposed for MPEG-7 expresses only the local edge distribution in the image
- A Scalable Color Descriptor (SCD) [4], the histogram is generated by color quantizing the image into 256 bins in the HSV color space, with 16 bins for hue, and 4 bins each for saturation and value. The image on the right is an example of color images with corresponding histograms. The left image and center image are more similar to each other based on the color histograms, compared to the image on the right. This descriptor is useful when searching for similarities between images.

The C_i represents the combination of C_{CLD} , C_{SCD} and C_{EHD} of feature i . The query processing, in turn, extracts a feature vector from a query and applies a metric (Euclidean distance equation 1) to evaluate the similarity between the query image and the database images [7].

$$S_{visuel_I} = dist_{Euclidean}(V_I, V_{I_i}) = \sqrt{\sum_{i=1}^m (C_I - C_{I_i})^2} \quad (1)$$

The similarites scores of the queries results builds a score vector. A score vector S_{visuel_I} of an image I can be thought of as a set of scores ($S_{visuel_1}, S_{visuel_2}, \dots, S_{visuel_n}$), where n is the dimension of database images.

System using Structure of XML document In this section, we are interested in Context-based Image retrieval techniques, and more precisely in Image retrieval based on textual and structural context in XML documents. Image context is composed all textual information surrounding the image. For retrieve image presentated in Figure 2, we can use text surrounding image such as document title, image name, image caption, etc

There are other sources of evidence that were used as visual descriptors, information from link around the image, structure of XML document. Indeed, We focus on XML documents don't have a homogeneous structure. What makes the structure as new source of evidence. The textual context remains insufficient in most of time. In this context, [5] say: "Ignore the document structure is to ignore its semantics". The idea is to calculate the relevancy score of image element based on information from the textual and structural context to answer a specific

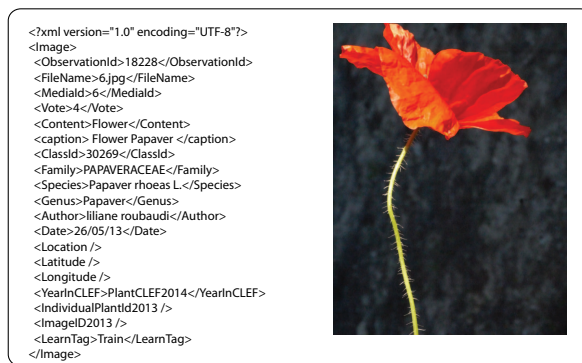


Fig. 2. Example of a image element context.

information needs of user, expressed as query composed of set of keywords. And seeking the most appropriate manner to combine two sources of evidence: text and structure. Our main inspiration is to use the structure to involve each textual information depending on its position in XML document, that is textual information that gives the best possible description of image element.

In *run 2*, we propose a automatic method in the field of image retrieval that takes into account the structure as a source of evidence and its impact on search performance. We present a new source of evidence dedicated to image retrieval based on the intuition that each textual node contains information that describes semantically a image element. And the participation of each text node in the score of a image element varies with its position in there XML document. To compute the geometric distance, we initially place the nodes of each XML document in an Euclidean space to calculate the coordinates of each node. Then, we compute the score of a image element depending on the distance between each textual node [6]. Figure 3 present our indexing system based on textual and structural context for image.

The result will be represented by a vector S_{score_I} . This vector of an image I can be thought of as a set of scores $(S_{score_1}, S_{score_2}, \dots, S_{score_n})$, where n is the dimension of database images.

Combination of Textual and Visual features From our two vectors S_{score} and S_{visuel} , we calculate a score that is a linear combination of scores for each modality:

$$S_{score_i} = \begin{pmatrix} S_{score_1} \\ S_{score_2} \\ \vdots \\ S_{score_n} \end{pmatrix} \quad S_{visuel_i} = \begin{pmatrix} S_{visuel_1} \\ S_{visuel_2} \\ \vdots \\ S_{visuel_n} \end{pmatrix} \quad S_i = \begin{pmatrix} S_1 \\ S_2 \\ \vdots \\ S_n \end{pmatrix}$$

$$S_i = score(S_{visuel_i}, S_{score_i}) = \alpha S_{visuel_i} + (1 - \alpha) S_{score_i} \quad (2)$$

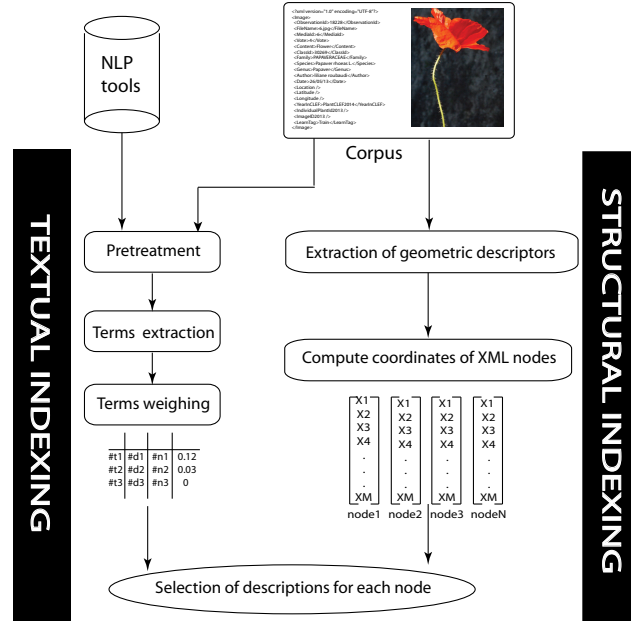


Fig. 3. Architecture of our indexing model (textual and structural context).

This score (equation 2) is a dot product between two vectors representing the visual features and the textual features. The parameter α is used to weight the amount of information conveyed by each modality.

3 Experiments and Results

3.1 Evaluation Metric

The metric evaluation S was related to the rank of the correct species in the list of retrieved species as follows [8]:

$$S = \frac{1}{U} \sum_{u=1}^U \frac{1}{P_u} \sum_{p=1}^{P_u} \frac{1}{N_{u,p}} \sum_{n=1}^{N_{u,p}} S_{u,p,n} \quad (3)$$

Where U is the number of users.

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).

3.2 The Results

Miracl team has submitted three automatic runs to the multi-image plant observation queries. And the details are as follows [9]. In the first run, we extracted visual descriptors from the images, used texture, color and edge model for generating the feature vectors. In the second run, we changed the evidence source of image in hope of taking general context of image into account the textual and structural information in XML document.

In the third run, also the last run, we fused the classifiers of the two runs above to form a new kind of degree of attribution.

Table 1. Results for all runs in plant collection

Run name	Retrieval type	Score
Miracl Run 1	Visual	0,063
Miracl Run 2	Context(Textual and Structural)	0,051
Miracl Run 3	Visual + Context	0,047

The results are showed in Table 1. From it we can find that the first run achieves the best results in all kinds of pictures. That's to say, the visual scheme usually can reach a better result for combination of different visual descriptors advantages.

The run for the contextual descriptors is in the middle place of the three runs and the one for flip combination descriptors gets the worst results which is in contradiction with our expectation. So we can speculate that the reasons for this phenomenon may have a relation to the number of training images in certain species which have very few images and the meta data of image are similar and relatively short which does not identify a species.

4 Conclusion

This work is our first try in the plant identification task. Although the results are not quite promising, we still reach some conclusions and gain some experience.

(1) At first, as there are some species which have only a few images, the bias for these species may be emphasized and combing some retrieval methods is a good choice.

(2) Many speacies have the same observation and the same *id_class*. In fact, we have a problem to product the run. Each run contains between 10 to 200 species for each queries.

While our solution was not performing up to the best ones in this competition (the idea score is 0.47), it is fast, accurate enough for non-critical applications and has potential for improvement. We intend to proceed developing this solution further and plan a mobile phone implementation of it.

References

1. Goëau, Hervé and Bonnet, Pierre and Barbe, Julien and Bakic, Vera and Joly, Alexis and Molino, Jean-François and Barthélemy, Daniel and Boujemaa, Nozha: Multi-organ Plant Identification. In :Proceedings of the 1st ACM International Workshop on Multimedia Analysis for Ecological Data (MAED '12). pp.41–44 (2012)
2. Park, Dong Kwon and Jeon, Yoon Seok and Won, Chee Sun: Efficient Use of Local Edge Histogram Descriptor. In :Proceedings of the 2000 ACM Workshops on Multimedia. pp.51–54 (2000)
3. Kasutani, E. and Yamada, A.: The MPEG-7 color layout descriptor: a compact image feature description for high-speed image/video segment retrieval. In :Proceedings. 2001 International Conference on Image Processing. pp.674 - 677, vol.1 (2001)
4. Cieplinski, Leszek: MPEG-7 Color Descriptors and Their Applications. In :Proceedings of the 9th International Conference on Computer Analysis of Images and Patterns (CAIP '01). pp.11–20 (2001)
5. Schlieder, T., Holger, M.: Querying and Ranking XML Documents. In :Journal of the American Society for Information Science and Technolog. pp.489–503, vol.53 (2002)
6. FAKHFAKH, S., Tmar, M., MAHDI, W.: A New Metric for Multimedia Retrieval in Structured Documents. In:15th International Conference on Enterprise Information Systems, ICEIS 2013, pp.240–247 (2013)
7. Hanen Karamti: Vectorisation du modèle d'appariement pour la recherche d'images par le contenu. In: CORIA 2013, pp.335-340 (2013)
8. Joly, Alexis and Müller, Henning and Goëau, Hervé and Glotin, Hervé and Spampinato, Concetto and Rauber, Andreas and Bonnet, Pierre and Vellinga, Willem-Pier and Fisher, Bob. LifeCLEF 2014: multimedia life species identification challenges. In: Proceedings of CLEF 2014 (2014)
9. Goëau, Hervé and Joly, Alexis and Bonnet, Pierre and Molino, Jean-François and Barthélemy, Daniel and Boujemaa, Nozha. LifeCLEF Plant Identification Task 2014. In: CLEF working notes 2014 (2014)