DAEDALUS at ImageCLEF 2011 Plant Identification Task: Using SIFT Keypoints for Object Detection

Julio Villena-Román[1,3], Sara Lana-Serrano[2,3], José Carlos González-Cristóbal[2,3]

[1] Universidad Carlos III de Madrid
[2] Universidad Politécnica de Madrid
[3] DAEDALUS - Data, Decisions and Language, S.A.
jvillena@it.uc3m.es, slana@diatel.upm.es,
josecarlos.gonzalez@upm.es

Abstract. This paper describes the participation of DAEDALUS at ImageCLEF 2011 Plant Identification task. The task is evaluated as a supervised classification problem over 71 tree species from the French Mediterranean area used as class labels, based on visual content from scan, scan-like and natural photo images. Our approach to this task is to build a classifier based on the detection of keypoints from the images extracted using Lowe's Scale Invariant Feature Transform (SIFT) algorithm. Although our overall classification score is very low as compared to other participant groups, the main conclusion that can be drawn is that SIFT keypoints seem to work significantly better for photos than for the other image types, so our approach may be a feasible strategy for the classification of this kind of visual content.

Keywords: Plant identification task, image retrieval, Scale-Invariant Feature Transform, SIFT, keypoints, classifier, training, test, Pl@ntLeaves.

1 Introduction

This paper describes the participation of DAEDALUS research team at the Plant Identification task [1], a new pilot task within ImageCLEF 2011 whose objective is to research on the application of image retrieval technologies for identifying plant species. Specifically, this first year the focus is on tree species identification based on leaf images. Leaves are easily observable and the most studied organ in the computer vision community, although they are known to not be the only discriminant key between tree species.

The task is evaluated as a supervised classification problem over 71 tree species from the French Mediterranean area used as class labels, based on visual content from Pl@ntLeaves dataset, published under a creative commons license within the Pl@ntNet project [2], containing 3070 leaf scans, 897 leaf pictures with a white uniform background (referred as scan-like pictures) and 2469 leaf pictures in natural conditions (taken on the tree) provided by Telabotanica [3], a French social network of amateur and expert botanists.

In addition to the image file itself, the dataset contains a series of meta-data attributes apart from the full taxon name (species, genus, family…) and French or English vernacular names (common names), including the acquisition type (scan, pseudoscan or photograph), content type (single leaf, single dead leaf or several leaves on tree visible in the picture), date, locality and GPS coordinates, and information about the author, all encoded in XML files. An example is shown in Figure 1.

Taxon: Kingdom: Plantae, Sub-kingdom: Tracheobionta, Division: Magnoliophyta, Class: Magnoliopsida, Sub-class: Rosidae, Order: Fagales, Family: Fagaceae, Genus: Castanea, Species: sativa Mill.

French vernacular name: Châtaignier
English vernacular name: Sweet Chestnut

Author: Henry Brisse

Date: 08/13/02

Locality: Gard Bréau

Figure 1. Example of one picture in the dataset.

A part of Pl@ntLeaves dataset is provided as training data whereas the remaining part is used later as test data. The training data finally results in 4004 images and the test data results in 1432 images. The goal of the task is to associate the correct tree species to each test image. Each participant was allowed to submit up to 3 runs built from different methods. As many species as possible could be associated to each test image, sorted by decreasing confidence score.

In the following sections we will describe our approach, the experiments that we submitted, the results that we achieved on this task, and some preliminary conclusions.

2 Our Approach

We approach this task with the construction of a classifier based on keypoints that represent objects within the images, extracted using Lowe's Scale-Invariant Feature Transform (SIFT) algorithm [4] [5].

The fundamentals of SIFT algorithm are to extract interesting points for a given training image that model the objects depicted in it, so that those objects can be identified in a given test image containing many other objects. To perform reliable recognition, those features extracted from the training image must be detectable under changes in image scale, noise and illumination. In addition, the relative positions between these features in the original scene should not change from one image to

another. Such interesting points usually lie on high-contrast regions of the image, such as object edges.

Our classifier is trained by first extracting SIFT keypoints from all images in the training set. Each set of keypoints is stored in a database, associated to the tree species that corresponds to such training image.

The number of extracted keypoints can be controlled by scaling the image resolution. Image resolution must not be very high as it is the larger scale keypoints that are most reliable and this is also much more efficient than processing large images. According to Lowe, an image of size 500 pixels square will typically give over 1000 keypoints depending on image content, which is plenty for most applications. For this purpose, each training image is rescaled to a width of 200 pixels. Moreover, as required by the Lowe's implementation that is used to obtain the SIFT keypoints [6], images are converted to greyscale PGM format prior to the extraction.

Once all the training set is processed, an object is recognized in a test image by individually comparing each feature from the test image to this database and finding candidate matching features based on Euclidean distance of their feature vectors. Test images are also downscaled, in this case to a width of 400 pixels to be able to find more keypoints, and then also converted to greyscale PGM format.

From the full set of matches, subsets of keypoints that agree on the object and its location, scale and orientation in the new image are identified to filter out good matches. The same criteria as proposed by Lowe is used [6], in which matches are identified by finding the 2 nearest neighbours of each keypoint from the training image among those in the test image, and only accepting a match if the distance to the closest neighbour is less than 0.6 of that to the second closest neighbour. This threshold can be adjusted up to select more matches or down to select only the most reliable.

Then the probability that a particular set of features indicates the presence of an object is computed, given the accuracy of fit and number of probable false matches. Object matches that pass all these tests are supposed to be identified as correct with high confidence.

The output of the SIFT classifier provides a list of training images sorted by relevance. To get the matching among training images and classification labels, the relevance of the top-ranked training image for each classification label is selected as the relevance for such label.

3 Experiments and Results

Although we initially planned different experiments changing the image downscaling and the object acceptance thresholds, we finally submitted just one run to be evaluated due to lack of time when carrying out the experiments.

For the same reason, we had to discard our initial idea to build three different specific classifiers based on acquisition type.

Apart from the image itself and the taxon name in the training set, no use of any other metadata information was made.

The primary metric used by the organizers to evaluate the submitted runs is a classification rate on the 1st species returned for each test image. Each test image is attributed with a score of 1 if the 1st returned species is correct and 0 if it is wrong. An average score is then computed on all test images. As a simple mean will introduce some bias due to the different number of images of the same individual plant and the number of pictures provided by each contributor to the Pl@ntLeaves dataset, the final metric is defined as an average classification score S:

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

where **U** is the number of users (who have at least one image in the test data), **Pu** is the number of individual plants observed by the u-th user, **Nu,p** is the number of pictures taken from the p-th plant observed by the u-th user and **Su,p,n** is classification score (1 or 0) for the n-th picture taken from the p-th plant observed by the u-th user. An average classification score S is computed separately for each type (scan, scan-like or photo) to isolate and evaluate its impact.

The results achieved in our experiment are shown in Table 1.

Table 1. Results (by classification score).

Run	Scans	Scan-like	Photos	Mean
daedalus_run1	0.043	0.025	0.055	0.041

In general, those figures are very low and results are a bit disappointing. However, an interesting point shown in the table is that the top values are achieved for natural photos. As a preliminary interpretation, we think that this may be because of the fact that SIFT keypoints strongly rely on contrast changes in images (such as colour gradients or edges), and natural pictures represent more realistic conditions.

Furthermore, another possible explanation may be the fact that the training and test dataset are not evenly balanced among the three acquisition types and not even between them, as shown in Table 2. Our conclusion is that we should have built three different classifiers, one for each type of image.

Table 2. Distribution of image types in training and test datasets

Type	Training	Test	Difference
Scans	58.1%	51.7%	-6.4%
Scan-like	17.1%	14.7%	-2.4%
Photos	24.8%	33.6%	+8.8%

A detailed analysis considering more than the 1st result is presented in Table 3. This table shows, for each classification label (tree species), the number of test images

where the label was returned (independently of its position in the result list) and the average position of that label in the result list.

Table 3. Detailed analysis by classification label.

Tree species	Average Position	Identified Images
Acer campestre	5.86	28
Acer monspessulanum	2.59	27
Acer negundo	9.26	19
Acer platanoides	12.70	10
Aesculus hippocastanum	17.00	1
Arbutus unedo	12.88	16
Betula pendula	13.33	3
Broussonetia papyrifera	8.80	45
Castanea sativa	5.00	1
Celtis australis	17.14	7
Cercis siliquastrum	8.95	38
Corylus avellana	20.00	7
Cotinus coggygria	4.29	28
Crataegus monogyna	12.46	41
Diospyros kaki	12.00	2
Eriobotrya japonica	6.80	10
Ficus carica	14.71	7
Fraxinus angustifolia	19.00	1
Ginkgo biloba	7.24	50
Ilex aquifolium	16.15	13
Juglans nigra	16.00	1
Juglans regia	2.00	1
Laurus nobilis	7.63	19
Nerium oleander	2.40	10
Olea europaea	2.59	32
Paliurus spina-christi	6.86	7
Pistacia lentiscus	13.40	15
Pistacia terebinthus	22.00	1
Pittosporum tobira	3.20	25
Platanus x	5.40	5
Punica granatum	6.75	4
Quercus coccifera	4.00	1
Quercus ilex	12.18	45
Quercus pubescens	18.20	5
Rhamnus alaternus	7.24	50
Robinia pseudoacacia	10.50	2
Syringa vulgaris	2.85	20
Viburnum tinus	16.23	47
Vitex agnus-castus	9.60	5

Our classifier was able to find the valid label for 649 test images (45.1% of the training set), in the 8.9th position on average. No test image was identified for the following tree species: *Alnus glutinosa*, *Fagus sylvatica*, *Fraxinus ornus* and *Magnolia grandiflora*.

Finally, Figure 2 shows the comparison of all 21 runs submitted by all 8 groups.

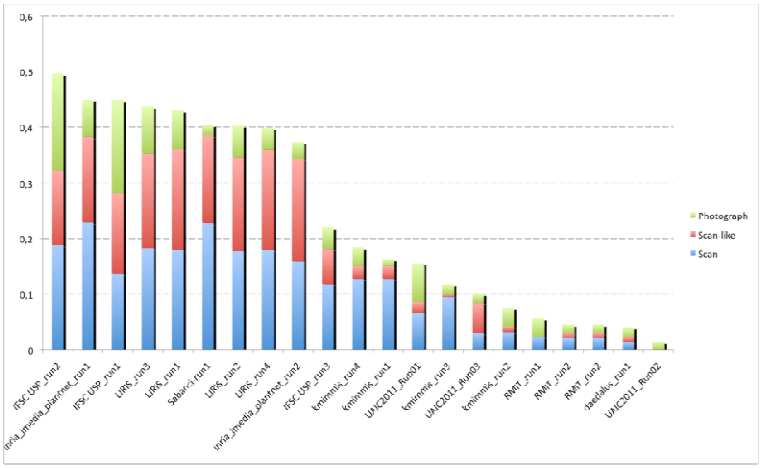

Figure 2. Overall results (by classification score).

Our group is the last one in the overall ranking because of the low performance for scans and especially for scan-like images. However our results for natural photos outperform the best ranked experiment from two other groups, as shown in Figure 3. This reinforces the idea that SIFT keypoints may be a valuable strategy for natural photos.

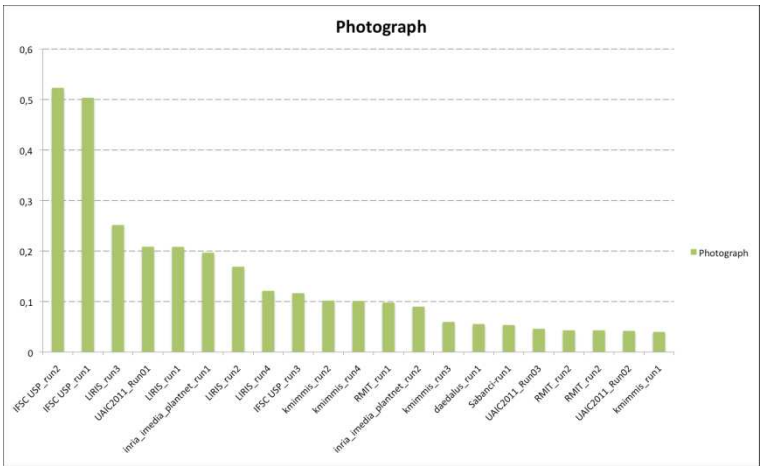

Figure 3. Results for photographs (by classification score).

4 Conclusions and Future Work

Despite the poor overall classification score, the main preliminary conclusion that can be drawn is that SIFT keypoints seem to work better for natural photos rather than scan and scan-like images, and our experiment has been able to outperform the best experiment by other groups in this type.

For future participations, we will definitely build specific classifiers for each image type. Moreover, we will try other alternatives to SIFT that are less demanding to compute and may handle colour images, such as SURF keypoints.

Acknowledgements

This work has been partially supported by several Spanish research projects: MA2VICMR: Improving the access, analysis and visibility of the multilingual and multimedia information in web for the Region of Madrid (S2009/TIC-1542), MULTIMEDICA: Multilingual Information Extraction in Health domain and application to scientific and informative documents (TIN2010-20644-C03-01) and BUSCAMEDIA: Towards a semantic adaptation of multi-network-multiterminal digital media (CEN-20091026). Authors would like to thank all partners for their knowledge and support.

References

1. Goëau, Hervé; Bonnet, Pierre; Joly, Alexis; Boujemaa, Nozha; Barthelemy, Daniel; Molino, Jean-François; Birnbaum, Philippe; Mouysset, Elise; Picard, Marie. The CLEF 2011 plant image classification task. *CLEF 2011 working notes*, Amsterdam, The Netherlands, 2011.
2. Pl@ntNet Project. http://www.plantnet-project.org/ [online August 2011].
3. Tela Botanica, The French Botany Network. http://www.tela-botanica.org/ [online August 2011].
4. Lowe, David G. Object recognition from local scale-invariant features. *Proceedings of the International Conference on Computer Vision*, vol 2. pp. 1150-1157, 1999.
5. Lowe, David G. Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60, 2, pp. 91-110, 2004.
6. Demo Software: SIFT Keypoint Detector. http://www.cs.ubc.ca/~lowe/keypoints/ [online August 2011].