Towards an Online Image-Based Tree Taxonomy

paul.de.zeeuw,elena.ranguelova,eric.pauwels@cwi.nl

Abstract. This paper reports on a first implementation of a webservice that supports image-based queries within the domain of tree taxonomy. As such, it serves as an example relevant to many other possible applications within the field of biodiversity and photo-identification. Without any human intervention matching results are produced through a chain of computer vision and image processing techniques, including segmentation and automatic shape matching. A selection of shape features is described and the architecture of the webservice is explained. Classification techniques are presented and preliminary results shown with respect to the success rate. Necessary future enhancements are discussed. Benefits are highlighted that could result from redesigning image-based expert systems as web services, open to the public at large.

1 Introduction

The pervasiveness of broadband Internet and mobile telephony has created an unprecedented connectivity between people and computational devices such as computers, mobile phones, digital camera's and GPS units. As a consequence, a growing number of initiatives is harnessing this infrastructure to set up new communities and exploit novel opportunities for large-scale interaction and participation. As Internet access thresholds continue to fall, the public at large is slowly being transformed from passive content consumers into active and avid content producers. Indeed, the likes of Wikipedia, Flickr, and YouTube have demonstrated beyond any doubt the viability of *"crowd sourcing"* development projects (a.k.a *Peer Productions*, e.g. Yahoo Answers) in which a comprehensive, high quality product emerges as the result of modest contributions from literally thousands or even millions of participants. This goes to show that there is a tremendous amount of talent and resources "out there" of people who both have the means and the aspiration to contribute to online communities that have captured their interest.

In this paper we report on our ongoing efforts to set up a (to the best of our knowledge, first!) *web-based tree taxonomy searchable by image query.* More specifically, we have created a webservice[1] that aspires to assist users in identifying trees by uploading a photograph of one of their leaves. This photograph is then processed by the web server and matched against a database of exemplars

---

[1] http://biogrid.project.cwi.nl/projects/leaves_v2/

of known species. As a result, a web page is created showing the, say, ten most similar exemplars along with species information, inviting the user to make the final choice and provide feedback. If the user considers the determination to be successful, the image is retained as a new exemplar for future queries. If the user deems the identification to be unsuccessful, the image is forwarded to an expert-biologist for a second opinion.

We envisage that in the near future it will become possible to point the camera in your mobile phone at a plant or tree, take a snapshot of one of the leaves and send it as an MMS message to a designated phone number, such as `1-800-whichtree`, say. Half a minute later you receive an sms serving up both the Latin and common name for the tree, as well as the link to the Wikipedia page where more information can be found. Moreover, it is our explicit intention to open this webservice up to the public at large so that *in addition to querying* everyone can *contribute*o to the exemplar database by uploading information and images of tree species.

1.1 Previous and Related Work

Although there are a number of online tree taxonomies available, the proposed web service is, in our opinion, innovative to the extent that it supports image-based queries and therefore contributes to the small but growing collection of applications that try to extend Internet search beyond the classic keyword paradigm. The best-known web-applications that support similar input modalities are focusing primarily on face recognition, such as FaceIt, myheritage.com or Riya.

It is clearly apparent that, given the pervasiveness of digital cameras, biologists are waking up to the possibilities of computer-assisted photo-identification, and a number of stand-alone systems are under development (cf. [1,3,6,7]). However, to the best of our knowledge the proposed website is the first to offer a vision-based taxonomy system as a web service.

2 Architecture of the Webservice

Broadly speaking the webservice is designed as a 2-tier system: The front-end allows the user to upload query images which are then shipped to the back-end server for processing and matching. The results are included in a webpage which is then transferred to the front-end for display and feedback. The system therefore comprises the following main components:

– **Database of exemplars on back-end.** Predictably, one starts by creating a database of *exemplars* that encapsulate the domain knowledge — for the taxonomy application we will often refer to them as *exemplars*. This database contains, for each of the target tree species, information such as the common name (e.g. *white oak*), the genus and species as specified in the Linnaeus binomial nomenclature (e.g. *Quercus alba*), a link to Wikipedia (if available), as well as one or more relevant photographs. Associated with
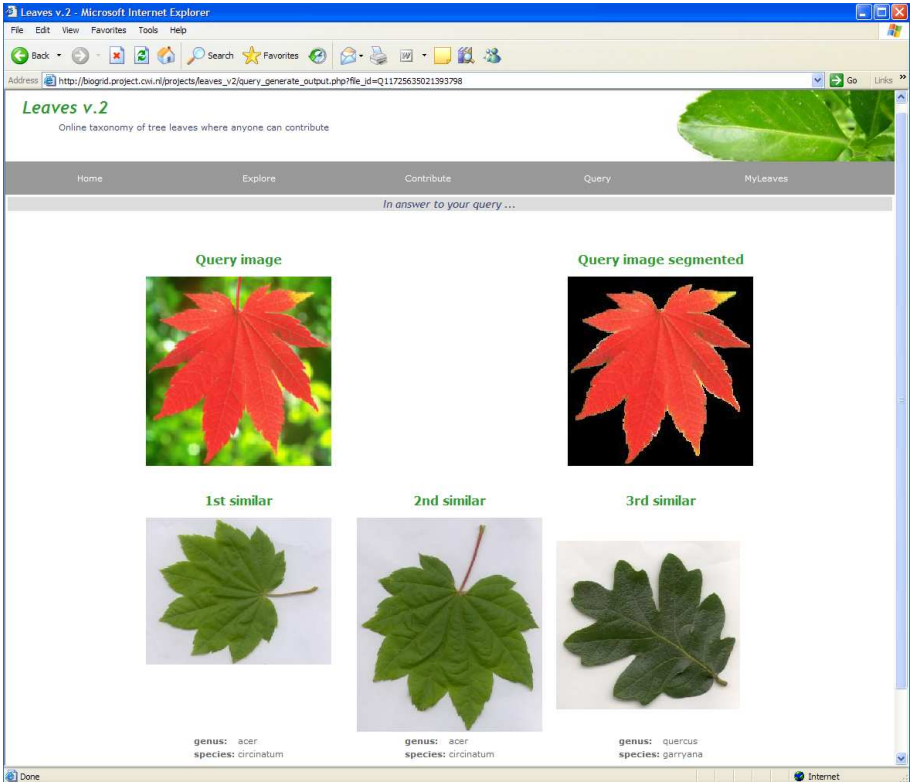
Fig. 1. Webpage generated by the taxonomy webservice in response to a submitted query image (top left). The result of the automatic segmentation is shown top right. The most similar images in the cases-database are displayed on the second row, together with relevant metadata such as genus and species. The displayed shortlist of most similar leaves offers the user the possibility to pick – as a final selection – the leaf that best matches his query image.

    each photograph is a set of automatically computed numerical features that characterize the shape of the corresponding leaf (for more details on these features, see section 3).

– **Front-end.** The front-end is a straightforward webpage that allows the user to upload an image of the query leaf. To improve performance we request users to adhere to certain standards (e.g. the leaf should be photographed against high contrast background) which greatly increase the reliability of the automatic image segmentation.

– **Processing on back-end.** Uploading a query image triggers a sequence of algorithms that *(i)* segment the leaf from the background and extract the result as a binary mask, and *(ii)* computes ten numerical shape features (for more details, see Section 3). The results are then checked against the pre-computed exemplar features and the most similar ones are shortlisted.

– **Feedback .** Once the similarity-based shortlist is available, a response web-
page is compiled displaying images of the n best matches (where n typically
ranges between 3 and 10). Each image has a caption detailing the genus,
species and common name of the corresponding tree. Showing a ranked short-
list of images (see Fig. 1) allows the user to perform a final visual check and
discard any obvious mismatches.

3   Features for Shape Matching to Exemplars

3.1   Image Segmentation: Segregating Foreground from Background

Prior to computing the features detailed below, both exemplar and query im-
ages are first segmented into actual leaf (foreground) and a background. In what
follows, we will use the term *mask* to refer to the resulting binary image that
specifies the foreground pixels. Note that we can think of leaves as flat objects
(something everyone who assembled a herbarium book can relate to) which sim-
plifies the analysis of the image considerably as we can restrict our attention to
measures for 2D shapes.

Automatic leaf segmentation proceeds through a number of steps. First, the
colour image is converted to gray scale in several different manners (using the
RGB and the HSV values). Then, each gray-level representation is segmented
using a gray-level histogram to which a mixture of Gaussian density is fitted.
The local minimum of the density is then used as a data-driven threshold for
segmentation. In this manner, several initial segmentations are obtained. For
each binary segmentation the number of 1 - connected components is computed
and the best initial segmentaiton is chosen as the one with minimum number.

The next step of the algorithm uses the initial segmentation to guide a *water-
shed transformation* on the best gray-level representation of the original image.
The watershed transformation is a powerful and well-established mathematical
morphology tool for image segmentation which has been used in many appli-
cations [5]. Any grey-level image can be considered as a topographical surface.
Flooding this surface from its minimum while preventing the merging of water
coming from difference sources, will result in a partitioning of the image into
catchment basins associated with each minimum. The boundaries between the
catchment basins are the watershed lines. If we apply this transformation to the
gradient of an image, we should obtain catchment basins corresponding to ho-
mogeneous grey-level regions. It is well-known however, that the transform tends
to produce an over-segmentation due to the local variations in the gradient. A
marker-controlled transformation is a solution to this problem: The gradient im-
age is modified via morphological reconstruction [5] in order to keep only the
most significant gradient edges in the areas of interest between the markers. The
biggest connected component from the chosen initial segmentation is used as
the *foreground marker* and the image boundaries as the *background marker*. As
a result of this step one gets the leaf boundaries.

The last step is the stem detection and removal. The stem is considered as a
significant deviation from the main leaf shape. All such deviations are detected

using the *top-hat transform* [5] of the binary segmentation. The detected deviations are the "teeth" of the leaf margin and the stem. The stem is singled out by imposing the additional restriction for large eccentricity and major axis. After the stem removal, the remainder is used as the binary leaf shape mask.

3.2   Shape Matching

Shape – even planar shape – is notoriously tricky to characterize accurately in a short sequence of numerical features. In order to cope with the considerable variations encountered in the dataset, we have implemented a number of features, each tailored to capture specific shape aspects. The idea is that combining them will produce a more discerning similarity measure. Below we briefly discuss the selection of features that are currently being used. It is likely that this set will be expanded in future versions of the search engine. All of them are computed on the binary image (a.k.a. mask) that results from the segmentation. This means that all internal structure, such as colour, texture and (most importantly) vein-structure has been discarded. This is an obvious weakness in the current system that we intend to remedy in a subsequent version. We also assume that the stem has been pruned so that only the intrinsic leaf shape remains. For ease of future reference we denote by L the resulting 2-dimensional shape, and by ∂L its contour.

Solidity (Sol). To measure the extent to which a leaf is lobed, we compute its *solidity* which is defined by comparing the area of the leaf to the area of its convex hull ($CH(L)$) to obtain a number between 0 and 1:

$$\text{Sol} = \frac{\text{area}(L)}{\text{area}(CH(L))}.$$

Isoperimetric factor (IF). This is another measure that roughly captures how winding (oscillatory) the contour is. If the perimeter is defined as $\ell = \text{length}(\partial L)$ and $A = \text{area}(L)$ then IF is defined as

$$IF = \frac{4\pi A}{\ell^2} \leq 1.$$

Equality prevails if and only if the contour is a circle.

Eccentricity (X). The third straightforward measure we employ is the eccentricity of ellipse with identical second moment as the leaf shape L.

Moment Invariants for Shape Characterization. The next set of measures are less straightforward. Hu's invariants [2], based on centralized moments, serve as a classic tool for recognizing geometrical shapes. An image is regarded as a density distribution function f. A central moment $\mu_{pq}(f)$ of f is given by

$$\mu_{pq}(f) = \iint_{\mathbb{R}^2} (x - x_c)^p (y - y_c)^q f(x, y) \, dx \, dy, \tag{1}$$

where p and q are non-negative integers and (x_c, y_c) is the center of mass. Such a central moment is said to be of $(p+q)$th order. In our case a binary foreground mask plays the role of the image of f which equals 1 inside the leaf region and 0 outside of it. By their definition it is immediate that central moments are translation invariant. Hu constructed polynomials with variables μ_{pq} in such a way that the outcome is invariant under rotations and reflections (the latter apart from sign). Two polynomials are built with second-order moments, four polynomials with third-order moments and one combines second-order and third-order moments.

$$I_1 = \mu_{20} + \mu_{02}, \tag{2}$$

$$I_2 = (\mu_{20} - \mu_{02})^2 + 4\mu_{11}^2, \tag{3}$$

$$I_3 = (\mu_{30} - 3\mu_{12})^2 + (3\mu_{21} - \mu_{03})^2, \tag{4}$$

$$I_4 = (\mu_{30} + \mu_{12})^2 + (\mu_{21} + \mu_{03})^2, \tag{5}$$

$$I_5 = (\mu_{30} - 3\mu_{12})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) + (3\mu_{21} - \mu_{03})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2), \tag{6}$$

$$I_6 = (\mu_{20} - \mu_{02})((\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2) + 4\mu_{11}(\mu_{30} + \mu_{12})(\mu_{21} + \mu_{03}), \tag{7}$$

$$I_7 = (3\mu_{21} - \mu_{03})(\mu_{30} + \mu_{12})((\mu_{30} + \mu_{12})^2 - 3(\mu_{21} + \mu_{03})^2) - (\mu_{30} - 3\mu_{12})(\mu_{21} + \mu_{03})(3(\mu_{30} + \mu_{12})^2 - (\mu_{21} + \mu_{03})^2). \tag{8}$$

We elaborate briefly on the numerical computation of the moments. Using the values of the image pixels we construct an interpolating function based on piecewise constant approximation. The piecewise constant basisfunctions have their support on squares centering around the pixels. Furthermore, the rectangular domain of an image is scaled in the sense that the size of the short side is equal to 1. As the supports of the basisfunctions are square, the size of the longer side of the domain follows at once. Hereby we can now perform the integration in (1) numerically.

So far, the expressions defined by (2)–(8) using (1) are invariant under translation, rotation and reflection (provided we ignore the sign of I_7). For shape invariance we still need to enforce similitude invariance, that is, after a mere change in dimensions of an object (leaf) it is identified as the same. Such invariance can be obtained by normalizing the moments μ_{pq}. Dilations (changes in size) by a scalar $\alpha > 0$ of the whole image or of objects in an image against a neutral background result in new central moments given by [2]

$$\mu'_{pq} = \alpha^{p+q+2}\mu_{pq}. \tag{9}$$

It follows in particular that $\mu'_{00} = \alpha^2\mu_{00}$, and also $\mu'_{20} + \mu'_{02} = \alpha^4(\mu_{20} + \mu_{02})$. Combining this result with Eq. (9) yields

$$\frac{\mu'_{pq}}{(\mu'_{00})^{(p+q+2)/2}} \frac{\mu_{pq}}{\mu_{00}^{(p+q+2)/2}}, \quad \frac{\mu'_{pq}}{(\mu'_{20} + \mu'_{02})^{(p+q+2)/4}} \frac{\mu_{pq}}{(\mu_{20} + \mu_{02})^{(p+q+2)/4}}$$

respectively. As we recall that both μ_{00} and $\mu_{20} + \mu_{02}$ are invariants with respect to rotation and reflection this demonstrates how to normalize the moments to achieve invariance under dilation. The first choice leads to the following new set of invariant generators [4]

$$I'_1 = I_1/\mu_{00}^2, \ I'_2 = I_2/\mu_{00}^4, \ I'_3 = I_3/\mu_{00}^5, \ I'_4 = I_4/\mu_{00}^5,$$
$$I'_5 = I_5/\mu_{00}^{10}, \ I'_6 = I_6/\mu_{00}^{10}, \ I'_7 = I_7/\mu_{00}^7. \tag{10}$$

The second choice leads to a different but similar result. It may be more suitable (as a starting point) in case the density distribution corresponds to wavelet detail coefficients, see [4].

Finally, it is clear that the *shape* of the foreground in the binary image f should be invariant under a change in luminosity; in mathematical parlance: $f \longmapsto \lambda f$ where $\lambda > 0$. As a scalar multiplication of the distribution function f does not affect the center of mass, it follows from (1) that

$$\mu_{pq}(\lambda f) = \lambda \mu_{pq}(f), \text{ for all } \lambda \neq 0. \tag{11}$$

One observes that the feature vector I' defined element by element through (10)

$$(I'_1, \ I'_2, \ I'_3, \ I'_4, \ I'_5, \ I'_6, \ I'_7)$$

then changes into

$$(\lambda^{-1}I'_1, \ \lambda^{-2}I'_2, \ \lambda^{-3}I'_3, \ \lambda^{-3}I'_4, \ \lambda^{-6}I'_5, \ \lambda^{-4}I'_6, \ \lambda^{-6}I'_7)$$

which is an undesirable result. (The result when moments would not be normalized would be equally undesirable.) To overcome this inhomogeneous change in the feature vector we use the following operator

$$R_p(u) = \text{sign}(u)|u|^{1/p}, \text{ with } p \in \mathbb{N} \text{ and } u \in \mathbb{R}. \tag{12}$$

When applied to an invariant I_k it produces again an invariant. It is a "legal" operation that invariants can be subjected to, i.e., neither their invariance properties nor their discriminative power are lost. We define the homogenized feature vector as

$$\tilde{I}' = (I'_1, \ R_2(I'_2), \ R_3(I'_3), \ R_3(I'_4), \ R_6(I'_5), \ R_4(I'_6), \ R_6(I'_7)). \tag{13}$$

This feature vector $\tilde{I}'$ now satisfies the *homogeneity condition* [4,8] in that a rescaling of the luminosity now affects all components in a homogeneous fashion:

$$f \longmapsto \lambda f \quad \Longrightarrow \quad \tilde{I}' \longmapsto \lambda^{-1}\tilde{I}'. \tag{14}$$

In addition, it turns out that hereby all elements operate in the same order of magnitude and that Mahalanobis's method is superfluous.

4   Results and Discussion

4.1   Populating the Exemplar Database Through Webharvesting

For the above outlined system to be successful, it needs to have access to a sufficiently large and comprehensive database of exemplars (cases). In order to create this with minimal effort we have taken recourse to webharvesting. More precisely, using Wikipedia we have first compiled a long list of the (Latin) Linnaeus classification (i.e. *genus* and *species*) of the tree species we are interested in. This list is fed into a programme that submits each combination into Google's *Image Search* and collects all the images that are returned. Relatively straightforward image processing software then winnows down this collection by rejecting all pictures that lack a convincing oval-shaped foreground. In most cases this prunes the collection down to a few percent of the original "harvest". The final selection is done by a human supervisor who reject everything except images where the well-defined foreground corresponds to a leaf. These images then go in the exemplar database and the Linnaeus classification (i.e. the *genus* and *species* that were used as search terms) are inserted as metadata. Once these images have been added to the exemplar database, the residing feature agents jump into action and compute the necessary features so that these new exemplars can be compared to any incoming image queries.

As mentioned earlier we have realised a first implementation of the above outlined webservice. To date we have compiled a small database of exemplars which comprises 23 unique genus-species combinations harvested from the web. For each of these genus-species combinations we have on average 5 to 10 exemplar images, adding up to 146 images in total. All the images have been segmented and the above-mentioned 10 shape parameters have been computed (i.e. solidity, isoperimetric factor, eccentricity and 7 moment invariants). We have then tested the two classification tools which we describe next.

4.2   Classification Trees

Classification trees seemed a first obvious choice for the classification of leaves (no pun intended!). The full 10-dimensional feature vector of section (3.2) was used to predict class-membership (running from 1 through 23 as there are 23 unique genus-species combinations). However, when we tested performance using cross-validation, the prediction success turned out to be disappointingly low 42%. For that reason we switched to a nearest neighbour classifier described in the next section.

4.3   Nearest Neighbour Classification

Since the 10-dimensional feature-vector is an amalgamation of qualitatively different characteristics (dimension-wise $10 = 1+1+1+7$), we decided that it was best to first compute distances in each space separately, and then produce a resulting distance by computing an (empirically optimized) linear combination

of these partial results. We settled for the straightforward (1-dimensional) Euclidean distance for the solidity, eccentricity and isoperimetric features (denoted by d_{Sol}, d_X and d_{IF} respectively). Further, because of (14) we opted for the (normalized) cosine distance in the 7-dimensional space of homogenized moment invariants:

$$d_{HM}(\overrightarrow{x}, \overrightarrow{y}) = \frac{2}{\pi} \arccos \left(\frac{|< \overrightarrow{x}, \overrightarrow{y} >|}{\|\overrightarrow{x}\|.\|\overrightarrow{y}\|} \right).$$

All distances are within the $0-1$ range, simplifying comparison and combination.

We proceed by making the assumption that the comprehensive distance is a straightforward linear combination of the above:

$$d_{LC} = d_{HM} + \alpha d_{Sol} + \beta d_X + \gamma d_{IF}.$$

The values for the weight parameters are determined by systematically searching for the combination that produces the best results on the exemplar database, i.e. each exemplar is used as a test-image and assigned to the same class as its nearest neighbour (in the d_{LC}-metric). The discriminative power of d_{HM} turns out to be predominant but even though $\alpha, \beta, \gamma \ll 1$ the other distances cannot be dispensed with.

We have estimated the classification accuracy by simulating the results that would be displayed on the webpage. More precisely, for each exemplar we have computed the 10 nearest d_{LC}-neighbours as these would be displayed on the webpage if the selected exemplar was submitted as a query. The results are shown in Fig. 2. If we insist that classification is only successful if the most
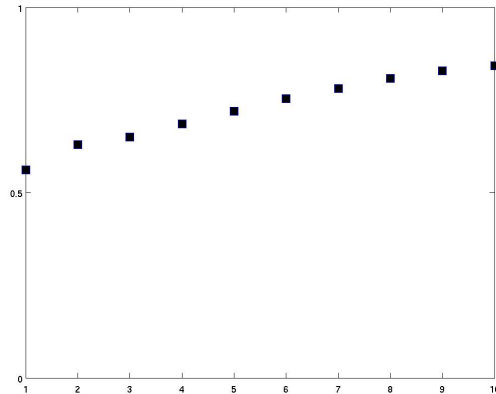
Fig. 2. Query success rate in terms of the number (k) of retrieved nearest neighbours. If we allow the user to inspect the ten most similar images, then the success rate is slightly higher than 85%. See main text for more details.

similar has the correct genus and species, then the success rate is about 53%. However, this is unduly pessimistic as not one but ten nearest neighbours are

shown on the webpage, and the user is invited to manually pick the best match. This means that the query will be successful if the correct genus and species are found among the 10 nearest neighbours (i.e. $k = 10$ in Fig. 2). It turns out that for this less stringent success-criterion, the success rate is approximately 85%. The complete distance matrix is shown in Fig. 3.
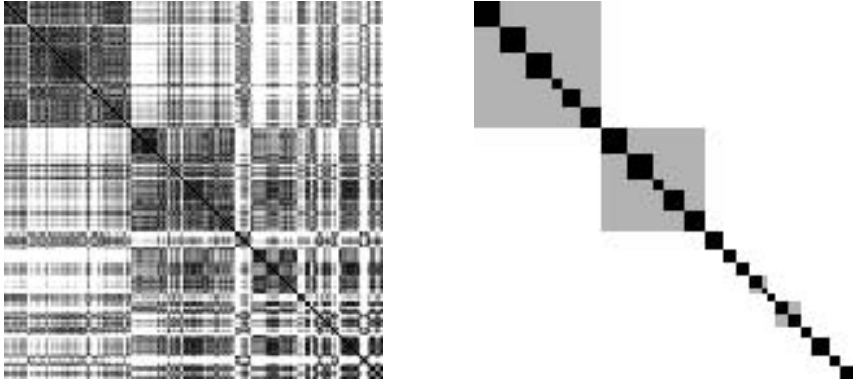
Fig. 3. *Left:* Schematic representation of the (146×146) d_{LC}-distance matrix for the database of exemplars. Bright points represent large distances while dark shades indicate similarity. *Right:* The corresponding class delineation: black blocks corresponds to leaves that have identical genus and species, while gray shades indicate identical genus but different species within that genus. Ideally, the d_{LC} matrix on the left should look very similar to the classification groundtruth depicted on the right.

5   Conclusions and Future Directions

In this paper we have reported on a first implementation of a webservice that supports image-based queries within the domain of tree taxonomy. We have argued that thanks to computer vision and image processing it is now feasible to gain good segmentation and recognition results with little or no input from a human supervisor. This opens the door to efficient searches through large databases of photographic material and therefore allows us to tackle classification problems for which the domain knowledge is primarily encoded in visual form. These developments are effectively extending the scope of image-based search task where traditionally, most efforts have been focused on face recognition. Clearly, photo-identification of plant and animal species (individual animals even) opens up a vast and exciting new application domain, the relevance of which is beyond dispute given the current concerns about the conservation and biodiversity.

The preliminary results that are obtained seem acceptable but clearly leave considerable scope for improvement. Apart from employing a much larger dataset, we see at least two directions that remain to be explored. First, we could add more sophisticated shapes measure to fine-tune the global distance d_{LC}. It is obvious from Fig. 3 that the current version of that distance is far from optimal. Secondly,

and more importantly, classification of leaves also depends on their vein-structure, something which we completely neglected. These issues will be addressed in a forthcoming follow-up paper.

Acknowledgments

References

1. Hillman, G., et al.: Computer-assisted photo-identification of flukes using blotch and scar patters. In: Proceedings of 15th Biennial Conference on the Biology of Marine Mammals (December 2003)
2. Hu, M.K.: Visual pattern recognition by moment invariants. IRE Transactions on Information Theory IT-8, 179–187 (1962)
3. Mizroch, S., Beard, J., Lynde, M.: Computer Assisted Photo- Identification of Humpback Whales. In: Hammond, P., Mizroch, S., Donovan, G. (eds.) Individual Recognition of Cetaceans, pp. 63–70. International Whaling Commission, Cambridge (1990)
4. Oonincx, P.J., de Zeeuw, P.M.: Adaptive lifting for shape-based image retrieval. Pattern Recognition 36, 2663–2672 (2003)
5. Soille, P.: Morphological Image Analysis. Springer, Heidelberg (2003)
6. Ranguelova, E., Pauwels, E.J.: Saliency Detection and Matching Strategy for Photo-Identification of Humpback Whales. In: GVIP05. International Conference on Graphics, Vision and Image Processing, Cairo, Egypt, pp. 81–88 (December 2005)
7. Van Tienhoven, A., den Hartog, J., Reijns, R., Peddemors, V.: A computer-aided program for pattern-matching of natural marks on the spotted raggedtooth shark carcharias taurus. Journal of Applied Ecology 44(2), 273–280 (2007)
8. de Zeeuw, P.M.: A toolbox for the lifting scheme on quincunx grids (lisq). CWI Report PNA-R0224, Centrum voor Wiskunde en Informatica, Amsterdam (2002)