View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by CWI's Institutional Repository

Engineering Applications of Artificial Intelligence 22 (2009) 26–31

Contents lists available at ScienceDirect

Engineering Applications of Artificial Intelligence

journal homepage: www.elsevier.com/locate/engappai

Computer-assisted tree taxonomy by automated image recognition

Eric J. Pauwels [a,*], Paul M. de Zeeuw [a], Elena B. Ranguelova [b]

[a] CWI, Kruislaan 413, 1098 SJ Amsterdam, The Netherlands
[b] TNO, Brassersplein 2, 2612 CT Delft, The Netherlands

ARTICLE INFO

Article history:
Received 24 April 2008
Accepted 26 April 2008
Available online 18 July 2008

PACS:
87.57.N
87.57.nm
87.63.lj

Keywords:
Image retrieval
Photo-identification
Shape features
Moment invariants
Shape matching
Biodiversity

ABSTRACT

We present an algorithm that performs image-based queries within the domain of tree taxonomy. As such, it serves as an example relevant to many other potential applications within the field of biodiversity and photo-identification. Unsupervised matching results are produced through a chain of computer vision and image processing techniques, including segmentation and automatic shape matching. The matching itself is based on a nearest neighbours search in an appropriate feature space. Finally, we briefly report on our efforts to set up a webservice to allow the general public to perform such queries online.

© 2008 Elsevier Ltd. All rights reserved.

1. Introduction

Digital cameras and connectivity have become commonplace. As a result, 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 in which a comprehensive, high quality product emerges as the result of modest contributions from literally thousands or even millions of participants. This observation provided the starting point for this paper: we intend to explore whether it is possible to set up a webservice that allows interested lay people to upload an image of a tree leaf and get instant information on its genus and species. The underlying architecture is relatively straightforward and essentially based on case-based reasoning (CBR) (Perner, 2002, 2005): when a query image is uploaded the database is searched for similar images for which the genus- and species-information is available. A shortlist of the most similar are returned for final inspection by the submitter. At this point the user has the option to pick the most similar and criticize the outcome (critique mode). The latter would make the CBR cycle complete, provided the information is used in reasoning and learning strategies for improvement of the retrieval.

However, it is still a challenge to make image retrieval such a reliable method that it can be presented to the public at large. In this paper we summarize our efforts to create a sufficiently expressive feature set that captures most of the salient visual characteristics used to classify leaf shapes. We will also briefly touch upon the webservice that aims to open up this facility to the public.

1.1. Previous and related work

Several biological multimedia databases are already online. The vast majority of them allow for search by metadata, e.g. see Hepp and Gurk (2004) which provides excellent pictures and metadata on leaves. Also a number of stand-alone systems are under development (cf. Hillman et al., 2003; Mizroch et al., 1990; Ranguelova and Pauwels, 2005; Van Tienhoven et al., 2007) relying on computer-assisted photo-identification. However, the combination of an online biological multimedia database with search facilities through computer-assisted photo-identification is rare. The only important example known to us is a whale shark photo-identification library by Norman and Holmberg (2005). The search facility uses photographs of the skin patterning behind the gills of each shark (and scars) to distinguish between individual

* Corresponding author.
E-mail addresses: eric.pauwels@cwi.nl (E.J. Pauwels), paul.de.zeeuw@cwi.nl (P.M. de Zeeuw), elena.ranguelova@tno.nl (E.B. Ranguelova).

0952-1976/$ - see front matter © 2008 Elsevier Ltd. All rights reserved.
doi:10.1016/j.engappai.2008.04.017

animals. Such areas are marked by unique spot patterns. An algorithm originally used to map stars, photographed by the Hubble telescope, has been adapted to identify the unique patterns.

A modest example with respect to tree taxonomy has been reported in de Zeeuw et al. (2007). To the best of our knowledge the webservice as proposed in this paper is the first one in his kind to offer online computer-assisted photo-identification in tree taxonomy.

2. Features

2.1. Introduction

In image retrieval *features* are used as a way of translating particular image characteristics into (vectors of) numerical quantities as the latter are more amenable to mathematical analysis and classification. From the literature on botanic taxonomy we learn that experts will use essentially three visual characteristics to classify leaves: their overall *shape*, the contour details of the leaf *margins*, and the overall structure of their vein pattern (the so-called *venation*). However, in this paper we will restrict ourselves to shape characteristics as they are most accessible to lay persons. Notice that this also means that we are not using other visual characteristics such as colouration or texture as they are too unreliable.

Since the input to our algorithm are digital images of leaves on a relatively uniform background, we have chosen to start by (automatically) segmenting the leaf from its background, thus creating a binary image (called *mask*), that accurately delineates the outline of the leaf under scrutiny (for some examples, see Fig. 3). The segmentation is based on a watershed algorithm and makes use of the assumption that the leaf itself is positioned roughly in the centre of the image. It first uses an automatically generated threshold to obtain a coarse segmentation which is then fed into a watershed algorithm for fine-tuning. For more details on the segmentation we refer to de Zeeuw et al. (2007, Section 3.1).

In the next two sections we discuss in some detail the features that we draw upon for shape matching. Section 2.2 expounds a number of features that have a clear geometrical interpretation and have been introduced to capture specific visual characteristics. In Section 2.3 we extend the feature set to shape invariants based on moments as they have proven their usefulness in other applications. However, their geometric interpretation (in terms of salient visual characteristics) is less clear and—given this paper's space constraints—we will not attempt to elaborate on this issue.

The size of database currently at our disposal contains 209 images, representing 22 different genera.

2.2. Shape features

In this section we introduce and describe the shape features we have employed. We start from the observation that, at least to a non-expert, a leaf's two most important shape characteristics are its elongation (long or slender versus roundish) and whether or not it is lobed or indented. To capture these visual characteristics we have implemented a number of relatively straightforward features (commonly encountered in the literature) that quantify these aspects. In addition, we will introduce features based on Hu's moment invariants (Hu, 1962). For ease of reference, we will denote the region defined by the binary mask as R and its boundary contour as ∂R. The diameter D of the region R is the maximal distance between two of its points:

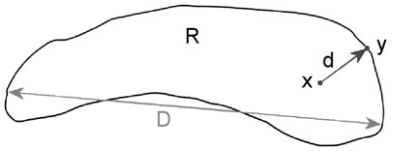

Fig. 1. Definition of the maximal exit distance (denoted by d) for the region enclosed by a contour of diameter D.

$D = \max\{d(p,q) \mid p,q \in R\}$ (see Fig. 1). We start by introducing some features that measure the slenderness of leaves.

Eccentricity: This is the eccentricity of the ellipse with identical second moment as the binary leaf image. Its value ranges from 0 to 1.

Aspect ratio: To compute the aspect ratio we first find two points p and q on the boundary ∂R that realize the diameter D, i.e. $d(p,q) = D$. Next we find the boundary point that realizes the maximal distance to the line segment $\overline{pq}$ and denote this by $D^{\perp}$. The aspect ratio is then defined to be the ratio of these two diameters:

$$\alpha = \frac{D^{\perp}}{D}.$$

Clearly, α-values close to zero are indicative of a very elongated structure.

Elongation: From each point inside the region R we measure the minimal (exit) distance $d_{\min}$ to the boundary contour ∂R and we denote by d_{me} its maximal value over the region. More precisely (see Fig. 1)

$$d_{me} = \max_{x \in R} d(x, \partial R).$$

Elongation ℓ is then defined as

$$\ell = 1 - \frac{2d_{me}}{D},$$

where D is the diameter of the region. Notice that ℓ varies between 0 and 1 where the lower limit is reached for a circular region. Notice also that the fraction $2d_{me}/D$ is the ratio of the diameter of the *largest inscribed* circle to the *smallest circumscribed* circle. In this sense, elongation "extends" the concept of eccentricity from ellipses to more general closed curves.

To quantify the amount of indentation we again introduce a number of related concepts.

Solidity: For a given region R we determine its convex hull $H(R)$ and compare their areas (see Fig. 2 left top panel):

$$S = \frac{\text{area}(R)}{\text{area}(H(R))}.$$

Stochastic convexity: Recall that a region R is called convex when for every two points $p,q \in R$ the line segment connecting these two points (denoted by $\overline{pq}$) lies entirely in R. We extend this notion by giving it a stochastic twist: we define the *stochastic convexity* to be equal to the *probability* that the line-segment $\overline{pq}$ spanned by two randomly chosen points $p,q \in R$ will be contained in R. Although this parameter might be difficult to compute analytically, it is quite straightforward to get an accurate estimate based on sampling. Notice how this feature is subtly different from the solidity defined above. Both are a function of the difference $H(R)\backslash R$ between the region and its convex hull, but for the stochastic convexity the spatial orientation of this difference is also important as it will impact on the likelihood of intersection with $\overline{pq}$.

Isoperimetric factor: If the closed contour ∂R of length $L(\partial R)$ encloses a region R of area $A(R)$, the isoperimetric factor is defined

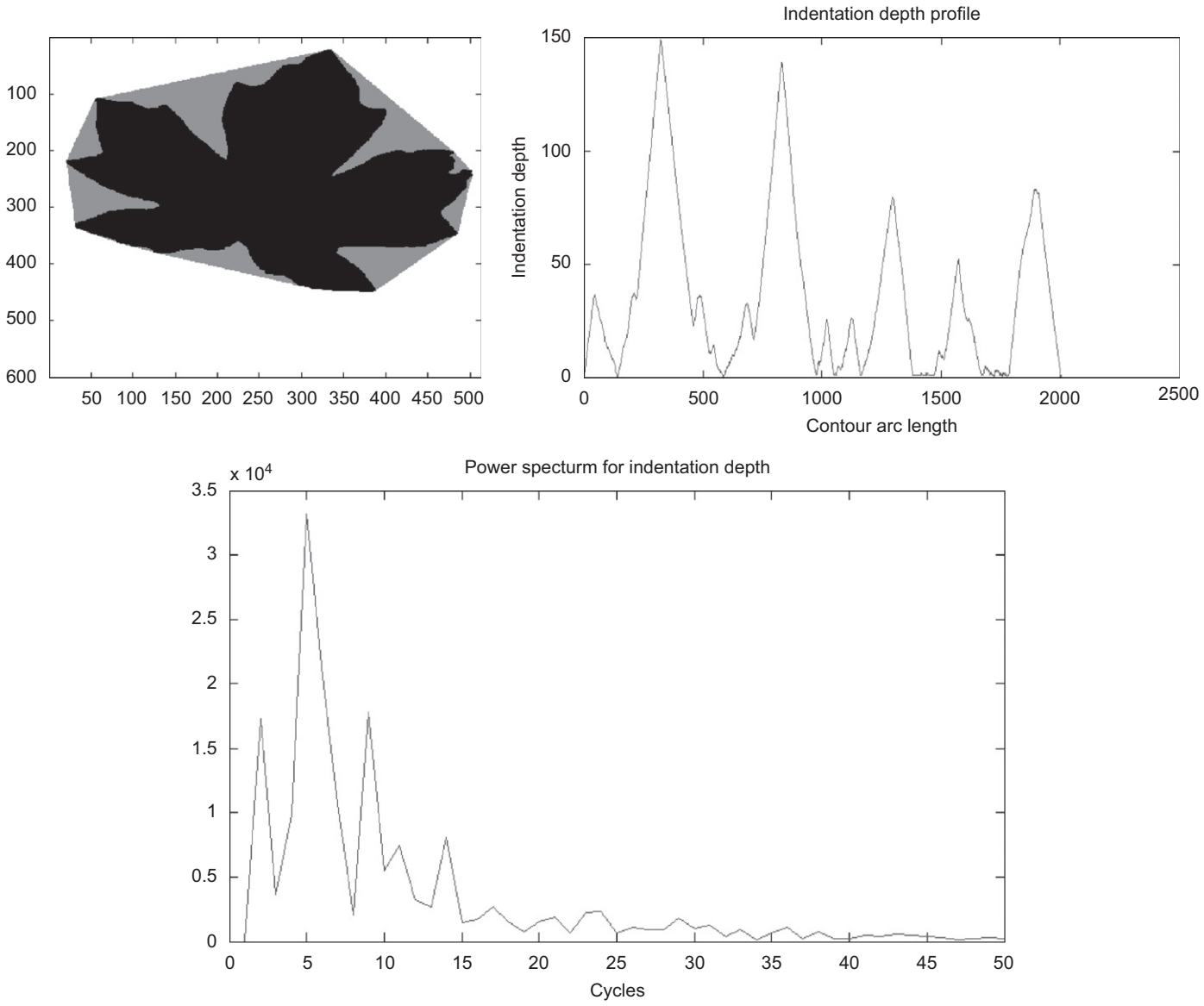

Fig. 2. Top left. Binary image of a leaf (black) with convex hull (grey). Top right. Indentation depth: The distance from the leaf contour to the convex hull as a function contour arc-length, i.e. the distance covered along the leaf's perimeter. Bottom: Power spectrum of the indentation depth function. The peak at frequency 5 is quite dominant reflecting the fivefold symmetry in the leaf.

as (a.k.a. circularity cf. Jähne, 2005)

$$\frac{4\pi A(R)}{L(\partial R)^2}.$$

The maximal value of 1 is attained when R is a circular region, while a long curvy contour that snakes around a narrow area will yield a small value.

Maximal indentation depth: For each point on the leaf countour we determine the distance to the convex hull (see Fig. 2, left). By expressing this distance as a function of the curve length along the contour we arrive at the indentation function depicted in the right panel of Fig. 2. We then define the maximal indentation depth $D_{\max}$ as the maximum of this function normalized by the contour length.

Indentation spectrum: As a next step we compute the Fourier spectrum of the indentation function depicted in the top right panel in Fig. 2 and the result (up to frequency 50) is shown in the bottom panel. Clearly, the peak at frequency 5 (cycles) corresponds to the fivefold symmetry of the leaf. To quantify the spread of the spectrum we compute F_{p80} which is the smallest frequency at which the cumulated energy exceeds 80% of the total energy in the spectrum (of the first 50 frequencies).

Lobedness: Finally, we define a "lobedness" feature L as

$$L \equiv D_{\max}^2 F_{p80}. \tag{1}$$

This feature was handpicked to show a clear correlation with the degree to which a leaf is lobed, as is borne out by the examples in Fig. 3 that visualize the intuitive geometrical interpretation of the "lobedness" feature.

2.3. Moment invariants

We present a short recap of seven moment invariants designed by Hu (1962) and how they can be improved upon to serve as elements of a seven-dimensional feature vector (further details can be found in Oonincx and de Zeeuw, 2003, Section 5, de Zeeuw,

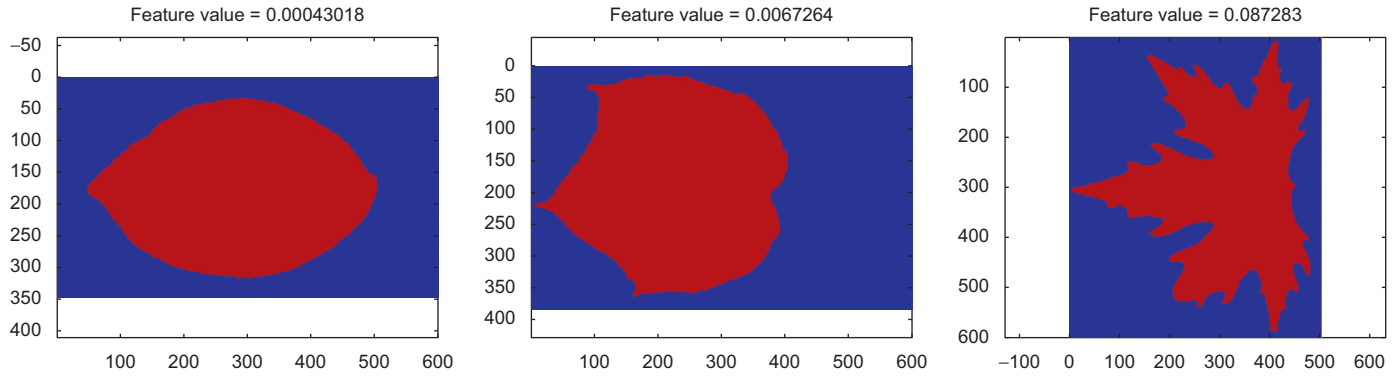

Fig. 3. From left to right: leaves of low, medium and high lobedness.

2002 and also de Zeeuw et al., 2007, Section 3.2). An image is regarded as a density distribution function f. In our case of a binary segmentation mask f takes on a particularly simple form:

$$f(x) = \begin{cases} 1 & \text{if } x \in R, \\ 0 & \text{otherwise.} \end{cases}$$

The (p,q) central moment $\mu_{pq}(f)$ of f is given by

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

where p and q are non-negative integers and (x_c, y_c) is the centre of mass. By their very definition central moments are translation invariant. However, since shape characteristics should enjoy additional invariances (viz. rotations, reflections and dilations) Hu (1962) constructed seven polynomials (the so-called *moment invariants*) in the variables μ_{pq} that are also invariant under rotations and reflections (the latter up to a sign). Two polynomials are built with second-order moments, four polynomials with third-order moments and one combines second- and third-order moments. The seven polynomials constitute a feature vector $I \in \mathbb{R}^7$. In Oonincx and de Zeeuw (2003, Section 5) and de Zeeuw (2002) it is demonstrated how to normalize these moments to achieve invariance under dilation too. There also the *homogeneity condition* is introduced which requires that a rescaling of the luminosity f affects the elements in a homogeneous fashion (and therefore can be compensated for)

$$f \longmapsto \lambda f \Longrightarrow \tilde{I} \longmapsto \lambda^{-1}\tilde{I}, \tag{3}$$

where $\tilde{I}$ denotes the normalized version of I. The invariance requested above is a natural requirement as the shape of a region in a binary image remains unaffected by the constant used to differentiate the region's interior from its exterior. It turns out that hereby all elements operate in the same order of magnitude and that using $\tilde{I}$ as a feature vector makes sense.

2.4. Feature robustness and internal consistency

The list of features enumerated in Section 2 clearly harbours a lot of overlap. This is borne out by the correlation matrix (computed on a database of 209 images, comprising 22 different genera) displayed in Fig. 4 where the features have been ordered to highlight the natural clustering due to similarity. More precisely, we ordered the features as follows: (1) *aspect ratio*, (2) *eccentricity*, (3–9) *seven moment invariants*, (10) *elongation*, (11) *maximal indentation depth*, (12) *solidity*, (13) *stochastic convexity*, (14) *lobedness*, (15) *isoperimetric factor and finally* (16) F_{p80}. A cursory inspection of the correlation matrix shows that the features naturally cluster in a number of blocks. Features 1–4

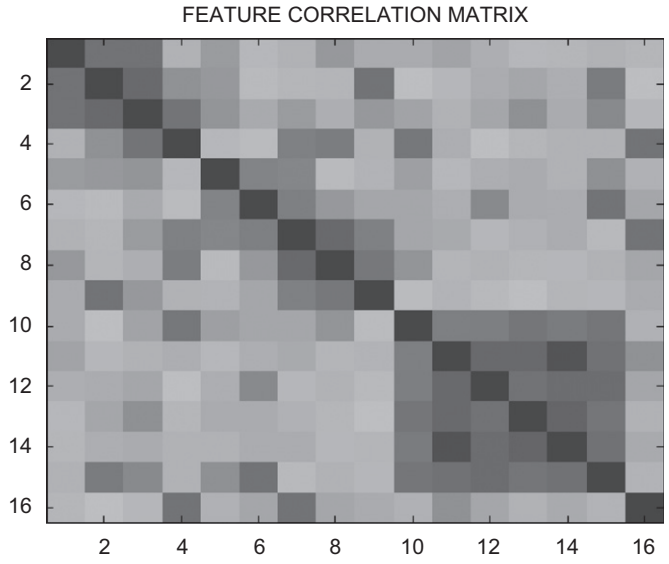

Fig. 4. Matrix depicting the correlation between the features defined in Section 2. Reddish signifies a high and bluish a low correlation. Columns from left to right and rows from top to bottom are ordered as follows: (1) aspect ratio, (2) eccentricity, (3–9) seven moment invariants, (10) elongation, (11) maximal indentation depth, (12) solidity, (13) stochastic convexity, (14) lobedness, (15) isoperimetric factor and (16) F_{p80}.

measure eccentricity or slenderness of the region (it can be shown mathematically that the first two moment invariants are in fact simple functions of the eccentricity). Blocks 3–9 represents the moment invariants which clearly add independent information although there is some internal redundancy. Features 10–15 are all involved in quantifying the amount of indentation. Notice that on our database the elongation feature (10) which was designed to quantify eccentricity seems to function primarily as an indentation measure (although there is also as strong correlation with the eccentricity block).

The importance of the above observations is twofold. Firstly, we intentionally introduced redundancy in the feature set to improve robustness. The idea is that it is often prohibitively costly to check whether features accurately capture the visual characteristics they have been designed for (i.e whether they have successfully bridged the so-called *numerical gap*). We therefore use the redundancy in the feature set to flush out problem cases. Indeed, if two features—designed to quantify the same visual characteristic albeit using complementary methods—yield divergent results, then that image is earmarked for visual inspection by

a human expert. This way we can quickly search the database for outliers that might be indicative of aberrant shapes or (perhaps even more importantly) flaws in the assumptions subsumed by the features and leading to the acknowledgement of unforeseen but valid shapes. It is a critique mode by expert users, see also Perner (2005).

The elucidation of the dependencies among the features provides the means to reduce the complexity of the classifier. This is tackled in the next section.

3. Recognition based on classification

The scenario envisaged in the proposed application calls for a query image to be uploaded to the system, whereupon it is segmented and processed by the feature extractors. The task of the classifier is to find the k most similar images in the (flat) database of exemplars and to present them to the query submitter for inspection and a final decision. In mathematical parlance, this amounts to searching the feature space for the k-nearest neighbours (NN) of the query. For the similarity measure we use the Euclidean distance. We construct a k-NN classifier as follows:

(1) We proclaim the classification of query leaf to be successful if at least one of the k images (where k is typically 10) that are returned by the k-NN search has the same *genus* as the query. Notice that this deviates from the more common criterion that uses majority voting; however, given that the final decision is taken by a human after inspecting the returned results, we believe the proposed measure to be more appropriate.
(2) We first rank the individual features based on their performance on the above criterion for $k = 1$ (i.e. nearest neighbour search).
(3) Next, we increase the list of features involved in the classification by incrementally adding new features based on their ranking in the above test, thereby taking care to skip features that are strongly correlated with the ones already included.

The above is in line with the approach by Perner (2002) who uses redundant features and selects the most distinguishing ones through decision tree learning among the whole set of features. If the constructed classifier cannot come up with good accuracy or if the classifier cannot distinguish between two classes then there is evidence that there must be constructed a new feature.

By systematically searching the subsets suggested by the above strategy it turned out that the best performing nearest neighbour classifier uses only three features (incidently, one from each main block): (2) *eccentricity*, (5) a *moment invariant* and (11) *maximal indentation depth*. Fig. 5 shows the resulting performance: the horizontal axis lists the number (k) of nearest neighbours that have been retrieved while the vertical axis shows the success rate, i.e. the fraction of such neighbour groups that contained at least one leaf of the same genus. In about 70% of the cases, the actual nearest neighbour is of the same genus, and if we search for the same genus among the first three neighbours, then we get a success rate of about 80%, which rises further to almost 90% when we allow the user to inspect the 10 nearest neighbours. We should also point out that the success rate is significantly lower if we use all features in the k-NN search, which clearly indicates that the redundancy in the feature set—although useful for flushing out outliers and non-conforming shapes—hampers the actual classification.

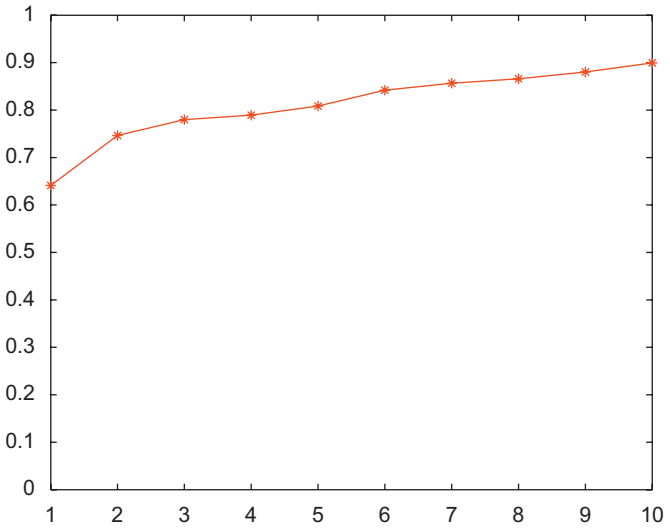

Fig. 5. Query success rate. The horizontal axis lists the number (k) of nearest neighbours that are retrieved, while the vertical axis shows the success rate, i.e. the fraction of such neighbour groups that contained at least one leaf of the same genus.

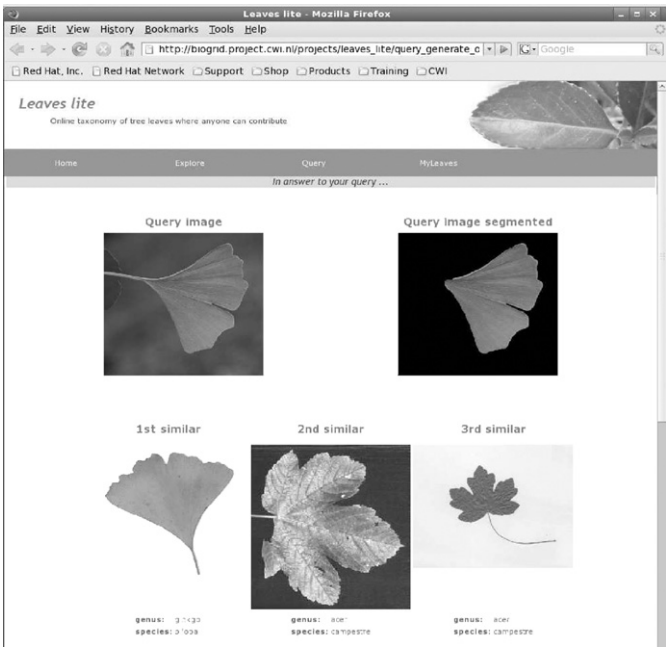

Fig. 6. Part of a webpage generated by the taxonomy webservice in response to a submitted query image (top left). The result of the automatic segmentation is shown top right. The most similar images in the cases-database are displayed on the second row, together with relevant metadata such as genus and 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.

4. Webservice

We have implemented a tentative webservice[1] that aspires to assist users in identifying trees by uploading a photograph of a leaf. The service is designed as a two-tier system. The front-end allows the user to upload query images and the back-end server performs the matching. Hereafter, a web page is created showing the 10 most similar exemplars along with the names of their

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

genus and species, see Fig. 6. The system comprises the following main components.

- *Database at back-end*: Obviously we need a database of leaf images (*cases* or *exemplars* in CBR terminology) that is supposed to encapsulate the domain knowledge. The database contains for each leaf information regarding the genus and species specified by the Linnaeus binomial nomenclature (e.g. *Quercus alba*), the common name (e.g. *white oak*) as well as one or more relevant photographs. Associated with each photograph is a set of automatically computed (numerical) shape features (see Section 2.2).
- *Query* (*image*) *upload facility at front-end*: The front-end provides a webpage that allows the user to upload an image of his leaf of interest. To improve performance it is advised either to photoscan or to photograph the leaf against a relatively homogeneous, contrasting background.
- *Matching at back-end*: Uploading a query image triggers a sequence of algorithms. Firstly, the leaf needs to be segmented from the background. Secondly, we compute a range of numerical shape features (see Section 2.2) on the (binary) segmented image. Thirdly, the query features are matched against the features from leaves in the database (see Section 3).
- *User feedback facility*: The 10 most similar leaves are shortlisted on a webpage with additional information like name of genus, species, etc. allowing the submitter to conduct a final visual inspection. It is a critique mode by non-expert users, see also Perner (2005). It helps to point out deficiencies in the automated recognition.

5. Conclusions

In this paper we have outlined the architecture and algorithms underlying a webservice that allows a user to upload a leaf image for image-based identification, see Fig. 6. The submitted image is automatically segmented to segregate the leaf from the background, whereupon the leaf shape is characterized using a number of numerical features. Classification is based on nearest neighbours matching against a database of exemplars to create a shortlist which is presented to the submitter for double checking. The current exemplars database is still incomplete and patchy but we plan on expanding the database of (visual) domain knowledge by allowing the public at large to contribute additional exemplars (i.e. image and appropriate metadata). Other future work involves CBR enhancements for image related systems, like learning new features, similarity learning, case refinement and case generalization.

Our web-based image recognition provides an example, albeit specialized, of how to find and browse images using their content rather than just by metadata. As such and within the context of the semantic web, it might contribute to search-engines of the future.

Acknowledgements

This work was partially supported by project NWO 613.002.056 *Computer-assisted identification of cetaceans* and by FP6 Network of Excellence MUSCLE.

References

Hepp, C., Gurk, C., 2004. Baumkunde, Einheimische und exotische Baumarten und Sträucher in einer Gehölze-Datenbank. ⟨http://www.baumkunde.de⟩.

Hillman, G., et al., 2003. Computer-assisted photo-identification of flukes using blotch and scar patterns. In: Proceedings of 15th Biennial Conference on the Biology of Marine Mammals, December 2003.

Hu, M.K., 1962. Visual pattern recognition by moment invariants. IRE Transactions on Information Theory IT-8, 179–187.

Jähne, B., 2005. Shape presentation and analysis. In: Digital Image Processing, Part IV, ISBN 978-3-540-24035-8 (Print) 978-3-540-27563-3 (Online) pp. 515–532.

Mizroch, S., Beard, J., Lynde, M., 1990. Computer assisted photo- identification of humpback whales. In: Hammond, P., Mizroch, S., Donovan, G. (Eds.), Individual Recognition of Cetaceans. International Whaling Commission, Cambridge, pp. 63–70.

Norman, B., Holmberg, J., 2005. The ECOCEAN whale shark photo-identification library. ⟨http://www.whaleshark.org⟩.

Oonincx, P.J., de Zeeuw, P.M., 2003. Adaptive lifting for shape-based image retrieval. Pattern Recognition 36, 2663–2672.

Perner, P., 2002. Image mining: issues, framework, a generic tool and its application to medical-image diagnosis. Journal of Engineering Applications of Artificial Intelligence 15 (2), 193–203.

Perner, P., 2005. Case-based reasoning for image analysis and interpretation. In: Chen, C., Wang, P.S.P. (Eds.), Handbook on Pattern Recognition and Computer Vision, third ed. World Scientific Publisher, Singapore, pp. 95–114.

Ranguelova, E., Pauwels, E.J., 2005. Saliency detection and matching strategy for photo-identification of humpback whales. In: International Conference on Graphics, Vision and Image Processing—GVIP05. Cairo, Egypt, December 2005, pp. 81–88.

Van Tienhoven, A., den Hartog, J., Reijns, R., Peddemors, V., 2007. 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.

de Zeeuw, P.M., 2002. A toolbox for the lifting scheme on quincunx grids (lisq). CWI Report PNA-R0224, Centrum voor Wiskunde en Informatica, Amsterdam.

de Zeeuw, P.M., Ranguelova, E., Pauwels, E.J., 2007. Towards an online image-based tree taxonomy. In: Perner, P. (Ed.), Advances in Data Mining—Theoretical Aspects and Applications, Lecture Notes in Artificial Intelligence, vol. 4597. Springer, Heidelberg, pp. 296–306.