AUTOMATIC LANDMARKING FOR BUILDING BIOLOGICAL SHAPE MODELS

Y. Hicks, D. Marshall, R.R. Martin, P.L. Rosin

Department of Computer Science
Cardiff University
Wales, UK
email: *y.a.hicks, dave,*
ralph, paul.rosin @cs.cf.ac.uk

M.M. Bayer, D.G. Mann

Royal Botanic Garden
Edinburgh
Scotland, UK
email: *M.Bayer,*
D.Mann@rbge.org.uk

ABSTRACT

In this article we present a new method for automatic landmark extraction from the contours of biological specimens. Our ultimate goal is to enable automatic identification of biological specimens in photographs and drawings held in a database. We propose to use Active Appearance Models for visual indexing of both photographs and drawings. Automatic landmark extraction will assist us in building the models. Below we describe the results of using our method on drawings and photographs of examples of diatoms, and present an Active Shape Model built using automatically extracted data.

1. INTRODUCTION

The ultimate aim of our research is to develop methods for automatic identification of biological specimens from photographs and drawings held in a database. Automatic identification of objects described in visual form and held in a database is also referred to as visual indexing. Biological specimens are frequently described in visual form for taxonomic and other purposes. Vast catalogues of specimen material have been accumulated over many years in the form of microscope slides, drawings and photographs. Recently, efforts have been made to digitize such data for electronic storage, transmission and processing.

There has been some recent work in indexing between images in biological databases [1, 9]. However, there is a clear need to extend such indexing capabilities; the inclusion of biological drawings is a natural extension. The aims of research are novel in that they seek to incorporate taxonomic drawings as a prime source of taxonomic data and to develop methods to enable indexing between digital photographic images and drawings stored in a biological database (Figure 1).

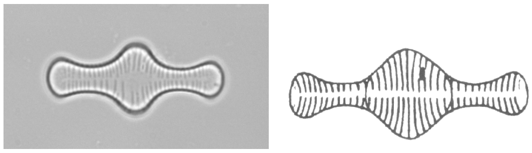

Fig. 1. A photograph and a drawing of *Tabellaria flocculosa* diatom specimens

We propose to transform the high-dimensional image spaces of both photographs and drawings into lower-dimensional spaces where only relevant features are represented, and use these spaces for visual indexing. Active Shape Models (ASMs) and Active Appearance Models (AAMs) [2] have become very popular in the last ten years for representing shapes and appearance inside the shapes in a compact way, significantly reducing the dimensionality of the original image space. These models can also be used for resynthesising examples of valid shapes and identifying instances of models in images. They therefore appear to be a logical choice for our task.

We propose to build Active Appearance Models of images and drawings of the biological specimens and find the mapping between the models

0-7803-7622-6/02/$17.00 ©2002 IEEE

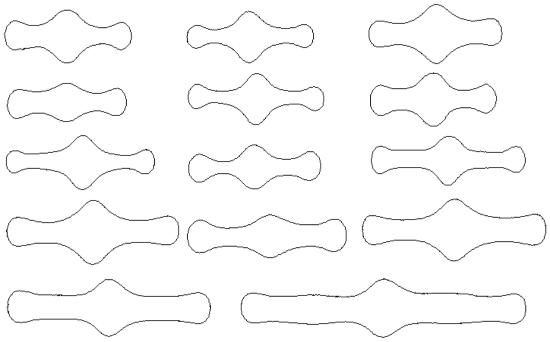

Fig. 2. Variation of the shape of *Tabellaria flocculosa* over its life cycle.

of photographs and the models of drawings. We aim to produce a generic system that can operate across a wide range of biological species. We have selected three taxonomic groups as test cases for our system: diatoms, desmids, and water mites (Acari); so far we have been working with the diatom taxonomic group.

In this paper we present a novel method for automatic extraction of landmark features on the contours of two-dimensional biological samples. Landmarking is a necessary step for building ASMs of the samples. We then present the results of application of the method to photographs and drawings of a diatom species, and build an ASM for the species describing the variation in the shape of the species over its life cycle (Figure 2).

2. PREVIOUS WORK

ASMs are built from a set of labelled examples. Each example is labelled with the same number of points, usually positioned in places of visual significance, called landmarks. There is a requirement for precise correspondence between positions of landmarks in different examples. Detection of landmarks is a non-trivial problem and is frequently performed by hand. Several methods for automatic landmark detection have been introduced recently. Hill *et al.* [5] proposed an algorithm where they iteratively place the pairs of samples in correspondence and replace each pair by their average until there is only one sample left. Then landmarks detected on the average sample are used to generate the landmarks on all samples. However, this method is not suitable for our purpose because not all landmarks are present in all samples. Some of them are very subtle and could disappear when averaging while we would like to retain them in the model (Figure 2). Davies *et al.* [3] presented an optimisation approach based on information theory to extract an optimal set of landmarks for a set of examples. The drawback of this approach is its complexity and computational intensity.

We now present a method for automatic landmark selection on the contour of diatoms for use in ASMs, build an ASM of a diatom species and show that the method is applicable for use in both photographs and drawings.

3. MODELLING BIOLOGICAL SPECIMEN CONTOURS

The shape of a specimen is one of the most important features used in classification of biological samples. In this project we consider taxonomic groups in which one standard specimen view is used for classification purposes. In this case the contour of the specimen often contains most of the information about its shape. Previous research showed that automatic classification of diatom specimens based on their shape produces good results [4].

There are different ways to describe a closed contour. In the past, methods based on Fourier descriptors and morphological curvature scale spaces as well as other methods have been used to describe biological specimens [7, 8]. We are interested in a method that would allow us to reconstruct the contour from its description. Previous research [8] has shown that the contour curvature changes little for a diatom species during the life cycle and that this can be used for distinguishing between different species (Figure 3). Fourier descriptors, on the other hand, change significantly as the shape of diatoms changes. ASMs allow reconstruction of the samples, so we decided to build an ASM of the contours and automatically choose the landmarks for our ASM at the positions of curvature extrema on the contour of biological specimens.

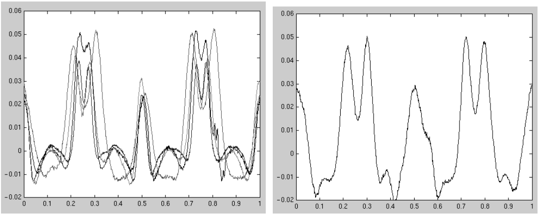

Fig. 3. **left:** Scaled contour curvatures of several *Tabellaria flocculosa* specimens from photographs; **right:** contour curvature extracted from a drawing of a *Tabellaria flocculosa* specimen.

4. AUTOMATIC LANDMARK DETECTION

The diatom contour curvature profile does not change much for the same species over its life cycle, but there is some variation. The heights and relative positions of the major extrema only change a little; some less significant extrema may appear, merge or disappear (Figure 3). We wish to model both the features appearing only in some of the samples and the features appearing in all of the samples, so we build a template T of typical curvature extrema for a species and include in it both types of features. The template contains descriptions for each of the curvature extrema that can occur on the contour. It includes for each extremum its average (over the samples) position along the contour ($\bar{l}_i$), average height ($\bar{h}_i$) and average contour tangent direction ($\bar{a}_i$) at extreme's position. The template assists us in matching the extrema of curvature found in different samples, which is necessary for landmarking.

We start processing the digital photographs of diatoms by extracting the contour of the specimen, which is a difficult problem in its own right due to the diffraction effect around the diatom specimens (Figure 1). We deal with it by subsequently applying automatic thresholding, area closing and area filling operations. After that we extract all the contour chains from the image and assume that the chain with the largest enclosed area is the contour of the diatom. Extracting the contour from a drawing of a diatom is relatively straightforward. After this we smooth the extracted contours, which are represented by their chain-codes, with a Gaussian, and differentiate them, which gives us the contour curvatures as shown in Figure 3.

In the next step we build the template as mentioned above. First of all we find all significant curvature extrema in all available samples from the chosen species. We do this by building a morphological scale space using the method of Leymarie and Levine [6]. This method has been previously successfully applied to describing diatom contours in [8]. Then we match the curvature extrema from different samples on the basis of their height, relative positions to each other and the corresponding contour tangent directions. We count the number of occurrences of each extremum through all the samples and disregard those with low count. The extrema appearing in many samples correspond to the consistent features in the curvature plot. We choose the number of such landmarks (N) as the number of points we want to extract from each diatom contour. Then we store the average values describing the curvature extrema for a species in the template $T = \{\bar{l}_i, \bar{h}_i, \bar{a}_i\}_{i=1}^{N}$.

The next step is to put into correspondence contour coordinates and tangent directions at the places of selected landmarks in all diatom samples. We do so by matching extrema of the sample curvature to the extrema in the template in the same way as above. If an extremum from the template is not matched to any of the extrema on the sample then we use its average relative contour position from the template to extract the corresponding contour point coordinates and tangent.

After extracting the landmarks from the contours we have a vector of size $3N$ for each diatom: $2N$ elements are the x and y landmark coordinates and N are the corresponding tangent directions. We align the landmark vectors using translation and rotation and normalise the vector elements through all the samples by subtracting their mean and dividing by their standard deviation.

Finally we perform principal component analysis (PCA) on the collected data. Previous research [7] has showed that the distribution of diatom shape data has a strong nonlinear functional relationship. Therefore we adopt a nonlinear PCA approach where we represent our data as a mixture of Gaussians. For illustrative purposes in the example in this paper we have chosen to use two clusters and find the main modes of variation in these two clusters (Figure 4).

We modelled the shape variation of two

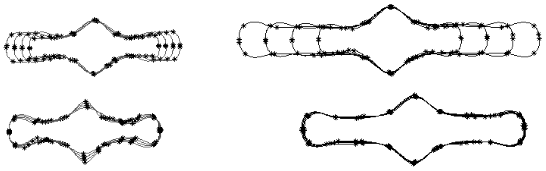

Fig. 4. **left column:** main modes of variation in the first cluster of the contour ASM; **right column:** main modes of variation in the second cluster.

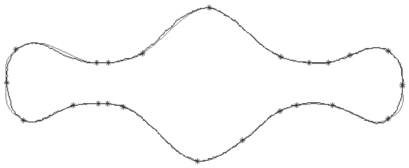

Fig. 5. Superimposed original and reconstructed contours of a drawing of a *Tabellaria flocculosa* specimen together with automatically extracted landmarks

species. The data for the first species, *Tabellaria flocculosa*, consisted of 16 digital photographs. The data for the second species, *Sellaphora pupula*, consisted of 200 photographs.

We found the curvature of the diatom contour in a single drawing as well. It was very similar to those extracted from the photographs and contained the same distribution of curvature extrema (Figure 3). Thus we were able to match the landmarks in the drawing to the extrema in our specimens, and extract the corresponding contour point coordinates and tangent directions.

We are able to reconstruct very closely the original contours of the drawn and photographed diatoms from the landmark coordinates and corresponding tangent directions at the landmarks using hermite cubic splines (Figure 5). We intend to use this idea in future as part of a process to automate the production of drawings.

5. CONCLUSIONS

In this article we have presented a new method for automatic landmark detection on the contours of biological specimens with application to building ASMs of diatom contours. We have showed both photographic and drawn specimen data map well into ASMs. We can resynthesise the shape of the specimens well from relatively few points automatically chosen at the places of extremal curvature. We are also able to model the shape variation over the life cycle of a diatom species.

6. REFERENCES

[1] H. du Buf *et al.* Diatom identification: a double challenge called ADIAC. In Proc. 10th Int. Conf. on Image Analysis and Processing, Venice, pp. 734-739, Sep. 27-29, 1999.

[2] T.F.Cootes *et al.* Active shape models—their training and application. Computer Vision and Image Understanding, pp.38-59, Vol.61, No. 1, January, 1995.

[3] R.H.Davies *et al.* An information theoretic approach to statistical shape modelling. In Proc. BMVC'2001, Manchester, pp. 3-12, Sep.10-13, 2001.

[4] S.Fischer *et al.* Feature based retrieval of diatoms in an image database. TR IAM-00-001, Institute of Computer Science and Applied Mathematics, University of Bern, 2000.

[5] A.Hill *et al.* A framework for automatic landmark identification using a new method of nonrigid correspondence. IEEE PAMI, pp. 241-251, Vol.22, No.3, March 2000.

[6] F.Leymarie and M.D.Levine. Curvature Morphology. TR-CIM-88-26, Computer Vision and Robotics Laboratory, McGill University, Montreal, Quebec, Canada, 1988.

[7] D.Mou and E.Stoermer. Separating Tabellaria (Bacillariophyceae) shape groups: a large sample approach based on Fourier descriptor analysis. Journal of Phycology 28, pp. 386-395, 1992.

[8] M.H.F.Wilkinson *et al.* Diatom contour analysis using morphological curvature scale spaces. In Proc. ICPR'2000, Barcelona, Sep. 3-7, pp. 656-659, 2000.

[9] P.J.D.Weeks *et al.* Species identification of wasps using principal component associative memories. Image and Vision Computing, 17, pp.861-866, 1999.