DEFORMABLE STRUCTURAL MODELS

Steven Bergner[*]

GrUVi Lab
Simon Fraser University

Stephan Al-Zubi, Klaus Tönnies[†]

Department of Simulation and Graphics
Otto-von-Guericke University Magdeburg, Germany

ABSTRACT

A hierarchical framework for the recognition of complex deformable shapes is developed. In extension to traditional approaches an additional layer of control is introduced to guide the local search for subshapes. This is realized by incorporating knowledge about their spatial relationships. A new technique of expectation maps is applied to allow simultaneous shape searches to influence each other. Furthermore, these maps are used to assess spatial coherence among shapes. Thus, the occurrence of well matched shapes at some places in the image may suggest searches for related shapes at other positions. An application to classify species in ant image databases shows promising initial results.

1. INTRODUCTION

The automated recognition of complex shapes is of great interest for numerous applications. An area where the analysis of complex deformable shapes plays a fundamental role is the discipline of systematics. Different species are classified and named according to their morphological similarities. As more species are discovered and catalogues are growing, more sophisticated tools are needed to assist the researchers. The new structural deformable model presented here is applied to a set of images of ant databases as they can be publicly accessed over the Internet[1]. The images are taken in colour from standardized perspectives using macro-photography. Thus, although the animals are three-dimensional, it is possible to solve the problem of classification applying a 2D shape search.

A common problem of deformable shape search approaches is their dependency on initialisation. A local search converges within a certain radius. Hamarneh et al. [1] have discussed an additional *brain* layer to add 'self awareness' to their deformable organism. This layer employs a number of techniques for artificial intelligence (AI) to guide the shape search out of mislead approaches. Our approach is inspired by this idea of introducing further layers of control. A complex shape is split into multiple sub-shapes. Thus, it is possible to separately analyse the *typical deformation* behaviour of the sub-shapes and their *spatial relationships*. This *structural knowledge* can now be used to guide the sub-shape searches and to relate them in a way so they can assist each other.

2. RELATED WORK

Techniques for shape representation can be distinguished into *dynamic* and *statistical* models. The first is employing a dynamic, mostly physically based, system to model deformation. A popular method are active contours or so called *snakes* introduced by Terzopoulos et al. in 1988 [2]. A contour is formed to minimise an energy potential made of internal constraints of rigidity and stiffness and external influences of image features and landmarks. The characteristics of a shape may be represented using finite element models (FEM) as discussed by Pentland and Sclaroff [3]. They make use of free vibration modes of a given deformable template to obtain a more significant (low-frequency) representation. A general advantage of dynamic models is an intuitive deformation behaviour. It allows for instant use without any training. Nevertheless, a *correct* treatment of deformation relies on an appropriate adjustment of the physical parameters, which is still an open problem.

The most influential representative among the class of *statistical models* is the *active shape model* (ASM) developed by Cootes et al. [4]. The set of training data consists of landmarked example images. It is analysed for the statistical variation modes of the point distribution yielding a point distribution model (PDM). It compactly represents typical deformations and inherently provides a measure for shape comparison. Active appearance models (AAM) are extending the method to incorporate texture variation into the model [5]. Statistical models are a good way of analysing of what is *typical* about a shape. A drawback of such models is reliance on training data.

Structural models combine basic shape representations on a higher level. The FORMS framework by Zhu and

[*] sbergner@cs.sfu.ca

[†] {stephan,klaus}@isg.cs.uni-magdeburg.de

[1] Two sources used here are from the Museum of Comparative Zoology at Harvard University, to be found at http://mcz-28168.oeb.harvard.edu/mcztypedb.htm, and AntWeb by the Californian Academy of Sciences located at www.antweb.org

Yuille [6] is arranging deformed worms and circles to form the contour of an object. The characteristics of the shapes are described, but their spatial relationships are restricted to connection points along the medial axis. A drawback of their method is that the shape matching is only working on readily extracted silhouettes. Also, the shapes of the parts are not influencing each other. A recent development by Al-Zubi et al. [7] is addressing this problem of structural covariation. Their approach of *active shape structural models* (ASSM) combines the advantages of structural and statistical shape representations. A single shape is statistically evaluated in terms of other shapes it relates to in the structure. This *shape context* is used to narrow statistical variation of shape constituents based on the role they play in the composition. The incorporation of structural knowledge into statistical evaluation allows to extract more descriptive local characteristics of single shapes. Thus, the model is capable of comparing and characterising shapes based on structural and shape variability, which makes it a very promising approach for many domains. The approach presented in this paper is different in terms of the underlying shape model. We are using a dynamic model to increase the instant usability of the method without extensive training. A fusion of the statistical and the dynamic approach to a combined structural model remains subject to future research.

As mentioned above, a problem of deformable shape searches is their tendency to get distracted by local minima. For that it is possible to investigate more elaborate *search strategies*. Hill et al. [8] have successfully employed genetic algorithms (GA) to match a deformable model to medical data. Their results show convergence independent of the random initialisation. A recent approach by Felzenszwalb [9] employs *dynamic programming* to efficiently perform a full search obtaining a globally optimal solution for a given energy function. Although, these methods yield very promising results, they are restricted to simple shape descriptions.

3. STRUCTURAL DEFORMABLE MODELS

Our approach is designed to deal with complex arrangements of several shapes that are combined through higher order structural relationships. Dynamic models can be made even more powerful by introducing structural knowledge about spatial relationships to other shapes into the control layer. This further reduces the search space. At the same time it requires more guidance so structure and deformation correspond to each other. The architecture of our system for detecting ant shapes, as shown in Fig. 1, is chosen to make use of well established solutions as much as possible. The data is processed by a set of feature extractors or so-called *sensors*. They are attached to nodes of an FEM shape model that is performing a (local) shape matching. The two topmost layers of the scheme are the new concepts that will be discussed below. The *finder* control the evolution of shapes in order to escape local minima. Structural knowledge is applied to connect finders. This allows them to communicate and adjust each others search focus. Combinations of subshapes form more complex objects. These are evaluated by accumulating quality measures for the subshapes and structural relations between them. The output of the system are several structural interpretations and their ratings.

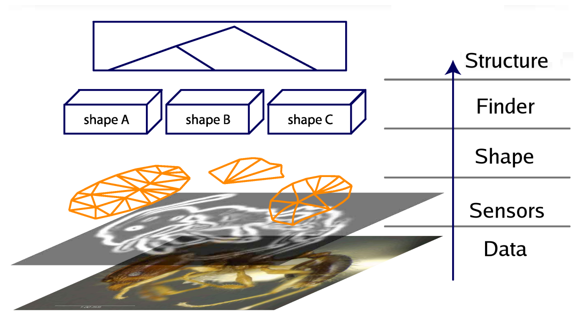

Fig. 1. The framework is formed as a hierarchy of different layers of processing.

The system is operating on 2D colour images. The *feature extraction* is set up as a number of cascaded processing steps. Basic methods that are applied here are statistical colour classification, edge extraction and curvature detection. The operations may be computed on different levels in Gaussian scale space. Coarser scales of resolution are used to distribute features over a larger radius and to thus support convergence of the shape.

The *shape matching* is performed using FEM shape descriptions. The nodes of this model are mass points that have a certain feature sensor attached. Nodes at the contour are being attracted by edges, inner nodes are attracted towards larger distribution of ant skin colour in the image. The gradient of the feature intensity at the node position is posing a force at the node. The shape of the model is maintained by the edges between the nodes that are acting as springs. Once the model is placed inside the image internal shape forces and external image forces begin to act on the model. The whole system comes to rest as the forces balance out (reach equilibrium). The physical parameters involved are set to values that yield reasonable default behaviour of the shape. Details of parameter selection can be found in [10]. We note that in general the dynamic model satisfactorily finds a shape in the image if it is placed sufficiently close to the target structure. Thus, shape matching is not guaranteed to succeed from any arbitrary initialisation.

3.1. Stochastic search using expectation maps

Using a dynamic model has the advantage that it is operational without any further training. The drawback is its strong dependency on the initialisation, which requires an additional guidance mechanism. Thus, a global search is performed for possible initialisations on a coarser scale. This is done in the four-dimensional space of (c_x, c_y, s, ϕ), where c_x and c_y are the coordinates of the centroid $\mathbf{c}$ of a shape, s is an indication of the size of the shape, and ϕ is the angle of its orientation. A representation of a shape in this space is called its *property vector*.

A finder is applying one specific shape model to the image. Several instances of the model are simultaneously moving on the image. As time progresses bad shapes are removed and new ones are spawned according to a certain probability distribution. To evaluate what is a fitness of an instance we compute a *quality of fit* function (QOF). It consists of two components, the data fit and the shape fit. The data fit is evaluated as average sensor response at each node. The shape fit is computed as a ratio of the sum of all edge lengths over the sum of all spring rest lengths. This provides a size independent measure of deformation. The resulting quality value is mapped to the interval $[0, 1]$ to allow for better comparison among different shapes.

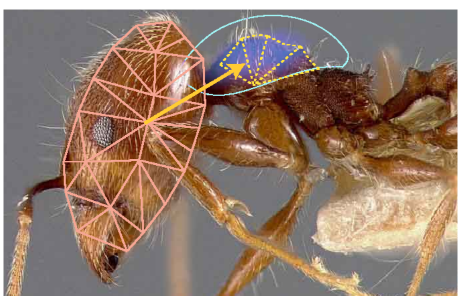

Fig. 2. Expectation maps are passed to the finder to indicate areas of higher anticipation. The image is showing *Pheidole carribaea sloanei*.

To generate property vectors for new shape searches we use an *expectation map* that is passed to the finder. An example of such a map is given in Fig. 2. Based on the position of the head, a search area for the back can be estimated. The dotted line shows the resulting match for the back. Expectation maps are also used to *rate spatial relationship* between shapes as used below. This makes them a convenient means to exchanges information among different shape finders. As long as nothing is known about the quality function a uniform distribution is used to generate new model instances. As experience increases this information may be refined by narrowing the probability distribution. This is comparable to the technique of importance sampling.

The input to a finder is a prototype shape model and an expectation map. As the search proceeds a *list of winners* can be obtained. Winners are selected based on their QOF. Selecting too many candidates is not problematic as long as the correct shape match is contained in the set. The existing method may be replaced by other methods such as ASM if a QOF can be computed for this representation.

3.2. Representing structural knowledge

Using the expectation maps it is possible to incorporate structural knowledge into the search that spatially relates the shapes to each other. The structural representation that we have applied to our test datasets is depicted in Fig. 3. It consists of the shape constituents, the spatial relationships (indicated by the arrows) and interpretation paths (shown as dotted lines). For each of the shapes a separate finder is employed. The spatial relationships are obtained statistically and are refined as more training data is confirmed by the user.

As the search process evolves, the lists of winners from the finders are used to generate expectation maps for adjacent shape finders. This way they are influencing each other to focus on certain ranges in search space. This yields an autonomous cooperative search for higher order structures. As the finders are running simultaneously the overall structural match improves over time. A possible combination of shape constituents is rated using a weighted additive combination of ratings: $q_p = \sum_i \omega_i q_i + \sum_i \theta_i r(e_{i0}, e_{i1})$, where q_i is the QOF for a shape, $r(e_{i0}, e_{i1})$ is the rating for the spatial relationship between two shapes. This is indicated by the start and end indices e_{i0} and e_{i1} of an edge e_i. Finally, the weights ω_i and θ_i are putting different emphasis on shapes and connections. For each of the possible interpretations an optimal combination of shape candidates is obtained with respect to the above rating function. This is done by performing an exhaustive search that is accelerated using dynamic programming.

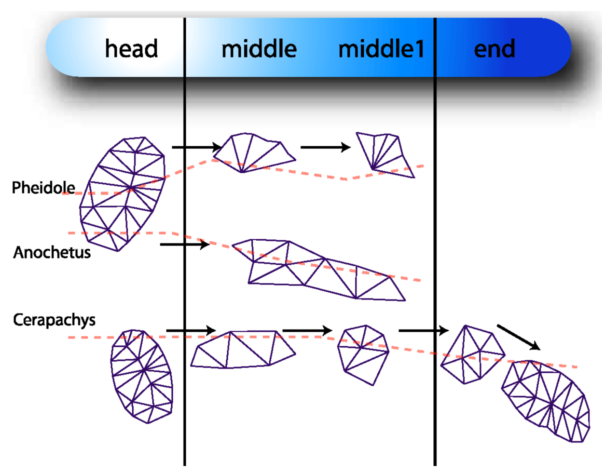

Fig. 3. **Ant graph** showing all important parts of the structural representation.

4. APPLICATION TO ANT DATABASES

Ant databases are a good example to test the structural model. It would be desirable to build models for single species. Unfortunately, since most databases have only one or very few examples for each species, it was necessary to generalise among distinct genera. We restricted ourselves to three different genera of ants *Pheidole*, *Anochetus*, and *Cerapachys*. The goal is to identify a genus by looking for significant morphological properties. Fig. 4 shows the output of the algorithm for different images. A test run on 75 images of class *Pheidole* has been performed and resulted in 84% correctly classified images. This result has to be interpreted with care. The operation has not been performed on the full database. The species used as input were chosen for clear appearance of the depicted species. So the detection ratio applies to images were all significant features are present. Misclassification is due to a bad response to the crucial feature detectors. Furthermore, similarities in the shape and structure of the ants may increase their probability to become misclassified. In that case more distinct characteristics, such as the antennae of an ant, may be added to increase significance of the structural description.

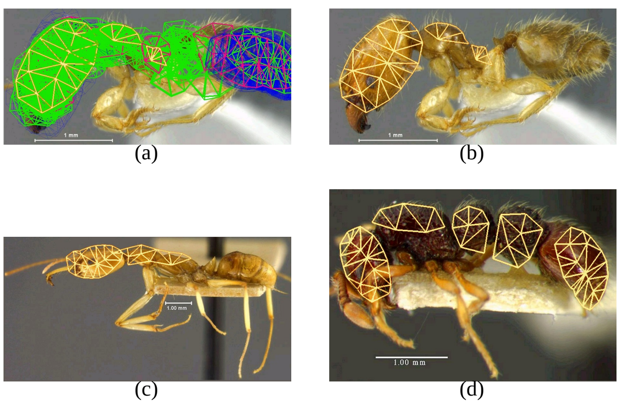

Fig. 4. Multi-structure search showing the best rated interpretation. a,b) *Pheidole subarmata*, c) *Anochetus haytianus* d) *Cerapachys vitiensis*

5. SUMMARY AND OUTLOOK

Our approach is to split up the complex shape into an arrangement of subshapes and spatial relationships. Using expectation maps concurrent shape searches can influence each other forming an autonomous structural search. The algorithm returns multiple interpretations of the image along with a confidence measure that can be used for subsequent classification.

In future, we will explore the potential of using the control element of the model for learning model parameters from examples as well as from user input. The model will be used to experiment with high level semantic information in a Content Based Image Retrieval system.

6. REFERENCES

[1] G. Hamarneh, T. McInerney, and D. Terzopoulos, “Deformable organisms for automatic medical image analysis,” in *Proc. of MICCAI’01*, Utrecht, The Netherlands, October 2001, pp. 66–75.

[2] M. Kass, A. Witkin, and D. Terzopoulos, “Snakes: Active contour models,” *International Journal of Computer Vision*, vol. 1, no. 4, pp. 321–331, 1988.

[3] A.P. Pentland and S. Sclaroff, “Closed-form solutions for physically-based modeling and reconstruction,” *IEEE Trans. Patt. Anal. Mach. Intell.*, vol. 13, no. 7, pp. 715–729, July 1991.

[4] T. F. Cootes, C. J. Taylor, D. H. Cooper, and J. Graham, “Active shape models - Their training and application,” *Computer Vision and Image Understanding*, vol. 61, no. 1, pp. 38–59, January 1995.

[5] T. F. Cootes, G. J. Edwards, and C. J. Taylor, “Active appearance models,” *Lecture Notes in Computer Science*, vol. 1407, pp. 484–499, 1998.

[6] S. Zhu and A. Yuille, “FORMS: A flexible object recognition and modeling system,” *Int. J. Comp. Vision*, vol. 20, no. 3, pp. 187–212, 1996.

[7] S. Al-Zubi and K. D. Tönnies, “Extending active shape models to incorporate a-priori knowledge about structural variability,” in *Proceedings of 24th DAGM Symposium*, 2002, vol. 2449, pp. 338–344.

[8] A. Hill and C. J. Taylor, “Model-based image interpretation using genetic algorithms,” *Image and Vision Computing*, vol. 10, no. 5, pp. 295–300, 1992.

[9] P. F. Felzenszwalb, “Representation and detection of deformable shapes,” in *Proc. of IEEE Conf. on Comp. Vision and Pattern Recognition*. June 2003, pp. 102–108, IEEE.

[10] S. Bergner, “Structural deformable models for robust object recognition,” M.S. thesis, Dept. of Simulation and Graphics, Otto-von-Guericke University, December 2003, http://wwwisg.cs.uni-magdeburg.de/bv/theses/thesis_bergner.pdf.