EVALUATING STATISTICAL SHAPE MODELS FOR AUTOMATIC LANDMARK GENERATION ON A CLASS OF HUMAN HANDS

A. N. Angelopoulou [a, *], A. Psarrou [a]

[a] Dept. of Artificial Intelligence & Interactive Multimedia, University of Westminster, Watford Road, Harrow, HA1 3TP – {agelopa, psarroa}@wmin.ac.uk

KEY WORDS: Statistics**,** Automation, Classification, Recognition, Image, Segmentation

ABSTRACT:

In this article, we present an evaluation of the application of statistical shape models for automatic landmark generation from a training set of deformable shapes and in particular, from a class of human hands models. The human hand is a dynamic object with considerable changes over time and variations in pose. A human being can easily recognize a hand despite its variations (e.g. skin tone, accessories, etc.) and put it in the context of an entire person. It is a visual task that human beings can do effortlessly, but in computer vision, this task is a complicated one. While a number of different techniques have been proposed, ranging from simple edge-detection algorithms to neural networks and statistical approaches, the development of a robust hand extraction algorithm is still a difficult task in computer vision. Human hand extraction is the first step in hand recognition systems, with the purpose of localizing and extracting the hand region from a complex and unprepared environment. This paper presents work in progress toward the segmentation and automatic identification of a set of landmark points. The landmarks are used to train statistical shape models known as Point Distribution Models (PDMs). Our goal is to enable automatic landmark identification using a context free approach of human hands' grey-scale still images held in a database. Our method is a combination of previously applied methods in shape recognition. In this paper we describe the motivation of our work, the results of our method applied on still images of examples of human hands and the extension of the method for building Active Appearance Model (AAM) using automatically extracted data for the recognition of deformable models in augmented reality systems.

1. INTRODUCTION

The aim of our research is to extract, using automatic methods, *landmark points* which are used to train a statistical flexible template known as Point Distribution Model (PDM) introduced by Cootes *et. al*. (1995), for the statistical analysis of 2D models from a set of deformable shapes. These models with the true location of the underlying shape can be used to built an Active Appearance Model (AAM), which in future work will be used for the tracking of 3D human hands in an augmented reality system.

A landmark point is a point of correspondence on each object of the class; it identifies a salient feature such as high curvature and is present on any object of the class. Dryden and Mardia (1998), discriminate landmarks as anatomical, mathematical and pseudo-landmarks. In our method we use mathematical landmarks, points located on an object based on high curvature or extreme points. These landmarks can be generated manually or automatically by applying different constrains. The manually correspondence is both laborious and subjective. While it gives good results in 2D images it would be impossible in the labelling of 3D images. On the other hand, the automatic correspondence can be more reliable, less time consuming, objective and can be applied to 3D images, but it works with constrains proposed by a number of authors (Hicks, 2002; Hill, 2000). This paper compares and addresses the problems of the manual and the automatic correspondence, reviews existing approaches and describes a simple and efficient method for automatic landmarking. In section 2, we review past work to automate the model building process. Section 3 outlines the method used to automate the construction of statistical shape models from a training set of human hands. Section 4 presents experimental results of applying the method to the training set. Section 5 concludes our method and suggests further extensions.

2. BACKGROUND

The motivation of our work is the automatic identification of landmark points from a training set of human hands. Baumberg and Hogg (1993) describe a system, which generates flexible shapes models from walking pedestrians using automatic landmark extraction. Landmarks are generated by computing the principal axis of the boundary, identifying a reference pixel of the boundary where the axis crosses the boundary, and by generating a number of equally spaced points along the boundary. While the process is satisfactory the parameterisation of the process is arbitrary and is described only for 2D shapes. Hicks and Bayer (2002) describe a system that automatically extracts landmark features from biological specimens, and is used to build an Active Shape Model (ASM) of the variations in the shape of the species. Their approach is based on identifying shape features such as regions of high curvature that can be used to establish point correspondences with boundary length interpolation between these points. While this method works well for diatom species where the heights and the relative position of the contour curvature local maxima and minima changes a little, it is unlikely that it will be generally successful for shapes such as hands where there are a lot of variations in the shape. Hill and Cris (2000) present an auto-landmarking framework, which employs a binary tree of corresponded pairs of shapes to generate landmarks automatically on each of a set

[*] A. N. Angelopoulou is with the Research Laboratory in Computer Vision, University of Westminster, London, UK.

of example shapes. In order to solve the pairwise correspondence problem they use a polygon-based correspondence algorithm which locates a pair of matching sparse polygonal approximations, one for each of boundaries, by minimising a cost function using a greedy algorithm. While the algorithm works well with different classes of objects, it assumes that the objects are represented by closed boundaries. Furthermore, the algorithm was not tested on objects with multiple closed boundaries, e.g., faces. Recently Davies *et. al*. (2002) have described a method for automatically building statistical shape models by using the Minimum Description Length (MDL) principle. The MDL is obtained from information theoretic considerations and the model order is defined as the model that minimises the description length, e.g. the model that encodes the vector observations in the most efficient way (Walter, 2002). In their method each shape is mapped onto a corresponding sphere where a given number of landmarks is first selected. The positions of the landmarks are then altered by parameterisation functions before selecting the parameterisations that build the best model. The best model is defined as the one, which minimises the description length of the training set, and its quality is regulated by an objective function that evaluates the quality of the PDM. This is a very promising method for measuring the model quality of a statistical shape model and results show better PDMs via manual landmarking. However, due to the very large number of function evaluations this optimisation method is computationally expensive.

3. OUTLINE OF THE METHOD

In our system we used 20 close related grey images of human hands recorded using a digital camera at a resolution 1024x768. Four people contributed with five images each of their right hand. In our experiments three main stages have been conducted:

1. Image segmentation or outline extraction using thresholding and the Canny Edge Detector to obtain the foreground (shape of the hand).
2. Freeman chain code 8-connectivity boundary descriptor to obtain automatically the coordinate of the boundary pixels and the direction of the boundary. Minimum Perimeter Polygon (MPP) is used to identify curvature descriptions.
3. Point Distribution Model (PDM) to describe the hand shapes and their variations based on the position of the automatic landmark points.

3.1 Extracting the hand shape

We are interested in extracting the outline of the hand shape. The method to separate the shapes from the background was to select a threshold

$$T=T[x,y,f(x,y)] \tag{1}$$

where f(x,y) is the grey level of point (x,y). The threshold image g(x,y) is defined as:

$$g(x,y)=\begin{Bmatrix} 1\,if\,[f(x,y)>T] \\ 0\,if\,[f(x,y)\le T] \end{Bmatrix} \tag{2}$$

where 1 and 0 corresponds to the distinction between the background and the foreground. For edge detection we used the Canny Edge detector which find edges by looking for local maxima of the gradient of f(x,y).

The gradient $g(x,y)=[G_X^2+G_Y^2]^{1/2}$ and edge direction $a(x,y)=\tan^{-1}(Gy/Gx)$ are computed at each point. A vector T=[T1 T2] containing two threshold is used to find the strong and the weak edge pixel and an edge linking is performed by including the weak pixels to the strong. By using those two detectors we managed to extract the shape of the hand regardless of the grey level difference between the foreground and the background.

3.2 Obtaining the Boundary coordinates of the shapes

For contour-based shape representation and description we chose an 8-connectivity derivative Freeman chain code, which is based on the fact that an arbitrary curve is represented by a sequence of small unit length vectors and a predefined set of possible directions. During the encoding successive contour points are adjacent to each other. The chain code was used as a numbered sequence that represents relative directions of boundary elements measured in a counter-clockwise 45° direction changes. The representation is based on 8-connectivity and the direction of each component is coded by the numbering scheme seen in Figure 1.

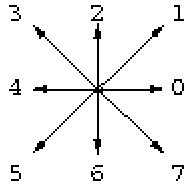

Figure 1. Chain code in 8-connectivity.

The chain code was used to derive the boundary length of the hand shapes and their direction. The entire vertical and horizontal steps have unit length while the length of the diagonal steps is $\sqrt{2}$. We calculated the contour length of the chain code as the number of vertical and horizontal components plus $\sqrt{2}$ times the number of diagonal components. The diameter of the shape boundary *b* is defined as:

$$D(b)=\max_{i,j}[D(p_i,p_j)] \tag{3}$$

where D is the *Euclidean distance* measure between p_i and p_j and is defines as:

$$D(p_i p_j)=\sqrt{(x_i-y_i)^2+(x_j-y_j)^2} \tag{4}$$

where (x_i, y_i, x_j, y_j) coordinates of the p_i and p_j points. One constrain of the chain codes is that in order to work properly the boundary should be a closed boundary, and a starting point should be defined. In order to choose an initial starting point on the closed boundary, which will have the associated parameter value, u=1 we searched where the vertical and the horizontal principal axes, the axes that pass though the centroid of the boundary points, intersect. The horizontal extension of the intersection meets the thumb, which was selected as the starting point of the chain code. It is assumed that this point is fixed for all shapes. This is reasonable since we are dealing with shapes where no occlusion occurs. With Freeman chain code we managed to found changes in code direction which indicates a corner in the boundary. By analysing direction changes following a clockwise direction through the boundary we determine and mark the convex and concave vertices based on Sklansky's approach (1972). In our method a vertex of a polygon is defined to be a convex if its angle is in the range $0^o \prec \theta \prec 180^o$; otherwise the vertex is a concave. We obtain the polygon by connecting all the convex points and delete all points that are outside the polygon. Doing this we obtain a uniform boundary with curvature defined where a change of the slope occurs and the control points approximately uniformly spaced along the curvatures.

3.3 Point Distribution Model (PDM)

The point distribution model (PDM) is a recent development in shape description and its roots are based on the development of active contour models from Kass, Witkin and Terzopoulos (1987b). It is most useful in describing features that have well understood shape and are non-rigid. This method was first develop by Cootes *et. al*. (1992) and has been applied to numerous examples including electrical resistors, faces and bones within the hand.

3.3.1 Labelling the training set. Previously we described how we could automate the labelling of the training set. Let us denote the number of shapes from the training set {S} and *n* the landmark coordinate points for each of the {S} shapes. The vector describing the *n* landmark points of the {Si} shape in the training set would be:

$$\mathbf{x}_i = [x_{i_0}, x_{i_1}, x_{i_2}, ...x_{i_{n-1}}, y_{i_0}, y_{i_1}, y_{i_2}, ...y_{i_{n-1}}]^T \qquad (5)$$

3.3.2 Aligning the training set. Before any alignment takes place, we assume that each shape has been normalised such that the centre-of-gravity is at the origin. Figure 2 shows unaligned and un-normalised shapes. We achieved the required alignment (scaling, rotating and translation) by using Generalised Procrustes Analysis. Details of the method are given by Cootes *et. al*. (1995). The result of this is shown with the Procrustes mean shape superimposed in Figure 3.

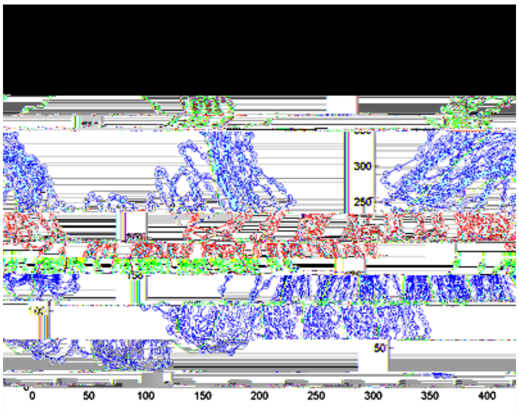

Figure 2. Unaligned and un-normalised shapes.

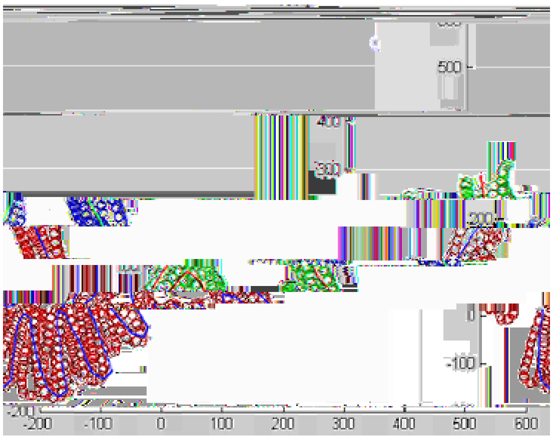

Figure 3. Aligned shapes with the centre-of-gravity removed and the mean shape superimposed.

3.3.3 Capturing the statistics. The outcome of the alignment is *M* aligned shapes $x_1, x_2, ...x_M$, and we proceed to determine the mean shape $\mu_{\mathbf{x}}$. Each shape is given by *n* coordinate pairs, $\mathbf{x}_i = [x_{i_0}, x_{i_1}, x_{i_2}, ...x_{i_{n-1}}, y_{i_0}, y_{i_1}, y_{i_2}, ...y_{i_{n-1}}]^T$ so, the mean shape is given by:

$$\mu_{\mathbf{x_i}} = \frac{1}{M}\sum_{i=1}^{M} \mathbf{x_i} \qquad (6)$$

The modes of variation of the training set can be found by applying the principal component analysis (PCA). For PCA to work properly we subtract the mean from each of the data dimensions. The mean subtracted is the average across each dimension. This produces a data set which mean is zero. The deviation from the mean is given by:

$$d\mathbf{x}_i = \mathbf{x}_i - \mu_{\mathbf{x}_i} \qquad (7)$$

The covariance matrix for the *2n* x *2n* landmark points is given as:

$$C_{\mathbf{x}} = \frac{1}{M}\sum_{i=1}^{M} d\mathbf{x}_i d\mathbf{x}_i^T \qquad (8)$$

The modes of variation can be derived by applying an eigen-decomposition of the Covariance matrix $C_{\mathbf{x}}$ such that:

$$C_{\mathbf{x}}\mathbf{p}_i = \lambda_i \mathbf{p}_i \qquad (9)$$

where λ_i ($\lambda_i \geq \lambda_{i+1}$) is the i^{th} eigenvalue of $C_{\mathbf{x}}$ and $\mathbf{p}_i$ is the associated i^{th} eigenvector. It can be shown that the eigenvectors of the covariance matrix corresponding to the largest eigenvalues describe the most significant modes of variation. Most of the variation can be described by a small number of modes, let us say *t*. One method for calculating *t* is to calculate the sum of the λ_i and choose *t* such that:

$$a(\lambda_T = \sum_{i=1}^{2n} \lambda_i) \le \sum_{i=1}^{t} \lambda_i \qquad (10)$$

where $0 \le a \le 1$ will govern how much of the variation seen in the training set can be represented by a small number of t modes. Any shape in the training set can be approximated using the mean shape and a weighted sum of the principal components from the t modes.

$$\mathbf{x} = \mu_{\mathbf{x}} + \mathbf{P}_t \mathbf{b}_t \qquad (11)$$

where

$$\mathbf{P}_t = (\mathbf{p}_1 \mathbf{p}_2 ... \mathbf{p}_t) \qquad (12)$$

is the matrix of the first t eigenvectors and

$$\mathbf{b}_t = (\mathbf{b}_1 \mathbf{b}_2 ... \mathbf{b}_t)^T \qquad (13)$$

is a vector of weights for each eigenvector. The eigenvectors are orthogonal so equation (11) can be written as:

$$\mathbf{b}_t = \mathbf{P}^T (\mathbf{x} - \mu_{\mathbf{x}}) \qquad (14)$$

Since the variance of $\mathbf{b}_i$ over the training set will be the associated eigenvalue λ_i we might expect that the limits should be in the order of:

$$-3\sqrt{\lambda_i} \le b_i \le 3\sqrt{\lambda_i} \qquad (15)$$

where we see that most of the population is in the order of $3\boldsymbol{\sigma}$ from the mean. This allows us to generate plausible shapes that are not part of the training set. Summarising, the PCA analysis has given us the original shapes in terms of their differences and similarities. In other words it has identified the statistical patterns in the data. Since the variations can be performed with the most significant eigenvectors we can reduce the dimensionality of the data and describe the variations with fewer variables (Hamarneh, 1998).

3.3.4 Back-projection. In order to get the original data back we need to add the mean of the original data. So, the new generated back-projected data will be given by:

$$\mathbf{b}' = \mu_{\mathbf{x}} + (\mathbf{P}^T \mathbf{P}(\mathbf{x} - \mu_{\mathbf{x}})) \qquad (16)$$

The result of this process is shown in Figure 4.

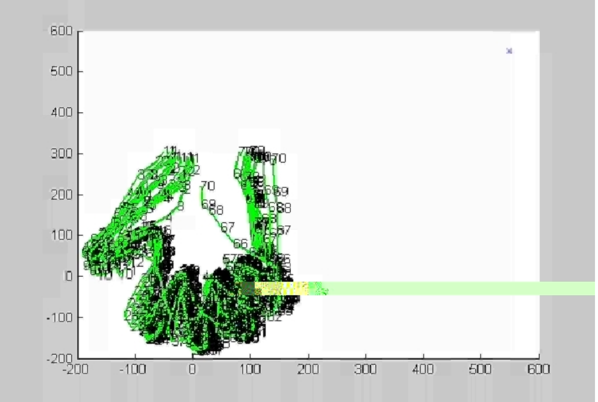

Figure 4. Back-projection of the original data annotated with the landmark points

4. RESULTS

In our experiments we used 20 training shapes with 70 landmark points (hence 140 parameters) and found that the first four modes accounted for 92% of the variance of the training data. The variance of each mode is as follows: mode 1 (m=1) with variance (v=1.34), mode 2 (m=2) with variance (v=1.03), mode 3 (m=3) with variance (v=0.6), and mode 4 with variance (v=0.43). Some of the significant modes of variation together with how the training data are arranged in the PCA space are shown in Figure 5.

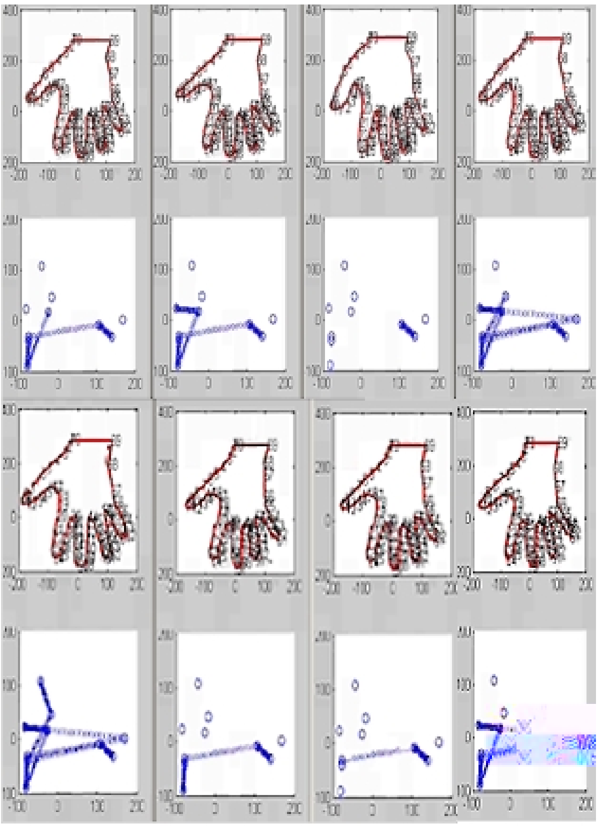

Figure 5. The first four modes of variation of the automatically generated model of the hand outlines.

5. CONCLUSIONS

In this article we have presented a new method for automatic landmark detection on the contour of hand shapes. The method is based on Freeman chain code for changes in code direction, which indicates a corner in the boundary and on Minimum Perimeter Polygon approximation for defining curvatures where a change of the slope occurs with the control points approximately uniformly spaced along the curvatures. Success of the whole procedure is suggested by distinct modes that generate eligible shape variations. However, the above method considered only hands represented by closed boundaries and non-occluding boundaries. These problems are part of our current research together with extending the method to high-level 3D hand shape variations.

References

Baumberg A., Hogg D., 1994. *Learning Flexible Models from Image Sequences*. Proc. Third European Conf. Computer Vision, pp. 299-308.

Cootes T., Taylor C., Cooper D., and Graham J, 1995. *Active shape models - their training and application*. Computer Vision and Image Understanding, 61(1): pp. 38-59.

Cootes T., *et. al*, 1992. *Active Shape Models –'smart snakes'*. In D C Hogg and R D Boyle, editors, Proceedings of the British Machine Vision Conference, Leeds, UK, pp. 266-275.

Davies R., *et. al.,* 2002. *A Minimum Description Length Approach to Statistical Shape Modeling.* IEEE Trans Med. Imaging, vol. 21, pp. 525-537.

Dryden L., and Mardia K., 1998. *Statistical Shape Analysis.* John Wiley & Sons.

Hamrneh G., *et. al*., 1998. A*ctive Shape Models: Modelling Shape and Gray Level Variation.* Swedish Symposium on Image Analysis, SSAB 1998, pp. 125-128.

Hicks Y. *et. al*., 2002. *Automatic Landmarking for Building Biological Shape Models.* Int. Conf. Image Processing, Rochester, USA, vol. 2, pp. 801-804.

Hill A. *et. al*., 2000. *A Framework for Automatic Landmark Identification Using a New Method of Nonrigid Correspondence.* IEEE Trans. Pattern Anal. Machine Intell., vol. 22, pp. 241-251.

Kass *et. al,* 1987b. *Snakes: Active Contours Models.* In 1st International Conference on Computer Vision, London, England, pp. 259-268, IEEE, Piscataway, NJ.

Sklansky J., et. al., 1972. *"Minimum Perimeter Polygons of Digitized Silhouettes"*. IEEE Trans. Comput., vol. C-21, no. 3, pp. 260-268.

Walter M., 2002. *Automatic Model Acquisition and Recognition of Human Gestures.* PhD Thesis, University of Westminster, London.