

A Generic System For Classifying Variable Objects Using Flexible Template Matching

A.Lanitis, C.J.Taylor and T.F.Cootes

Dpt. of Medical Biophysics, University of Manchester,
email:lan@wiau.mb.man.ac.uk

Abstract

A technique for classifying variable objects using flexible template models is described. A model representing each class of object is generated from training examples and during recognition each model is fitted to the input image. The object represented by the model that maximises a fit measure is recognised as the input. The method has been tested on plant seeds, handprinted characters and human faces; quantitative results are presented.

1 Introduction

The work described in this paper is concerned with classifying variable objects, based on their shape. The main problem encountered when dealing with variable objects is that shapes representing the same class of objects are not identical. We have addressed this problem by using Point Distribution Models [4] to model the ways in which object shapes vary. During training, examples of shapes from each class of object are collected so that the allowable shape deformations for each object are learnt. A flexible model representing each class is generated; these models can be deformed to fit to new examples of the objects. During recognition all models are fitted to the input image and the object represented by the model which maximises a fit measure is recognised as the input. A block diagram of the method is shown in figure 1.

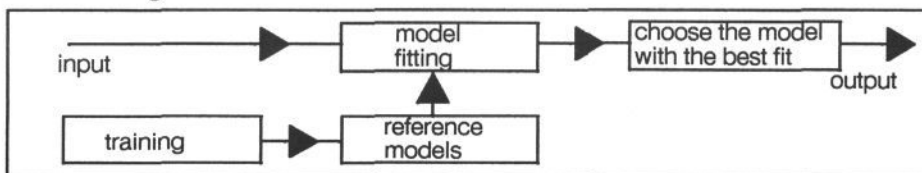


Fig 1 A general diagram of the system

2 Background

The work described in this paper is a continuation of the work reported by Rushall [11]. He described the use of Chord Length Distribution (CLD) models [1] for classifying plant seeds. CLD models representing several classes of seed were generated; during recognition Genetic Algorithms were employed to fit the models to seed outlines. The probability that a particular model could have generated a given seed outline was calculated and the class of seed represented by the model that maximised the probability was recognised as the input.

A similar method has been developed by Hinton et al [9,12], for recognising handwritten digits. In their original work they had an 8 point model representing each digit. During fitting the points moved from their mean position

in an attempt to fit to the digit presented and a deformation energy for the model was calculated. The model with the least deformation energy was recognised as the input. However, because of the lack of specificity of the models used, they could not rely just on the model fitting to recognise digits correctly. In their latest work they combine a neural network approach with the model matching. The neural network classifies using information about how well the model fits the input image and the value of the deformation energy.

3 Point Distribution Models

Point Distribution Models (PDM's) [4] are flexible models developed in an attempt to model shape variations of variable objects. So far they have been used for automatic image interpretation. Applications include locating both industrial [4,5] and biological [6] objects within images. However, this is the first time that PDM's have been used for classification. PDM's are generated from a set of training shapes belonging to the class of objects we wish to model. On each training shape a number of landmark points are placed at some key positions. For example figures 2 and 3 show a possible training set and the location of the control points used to build a model representing the number '7'. Based on the co-ordinates of the control points, the mean

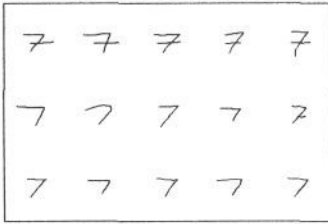


Fig 2 Example of a training set for the number 7

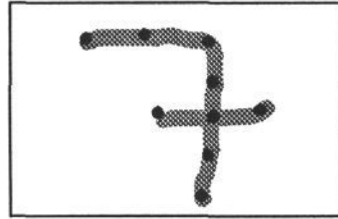


Fig 3 Location of the landmark points for the number 7

shape(M) and the deviation of each point from the mean position are recorded. A Principal Component Analysis of the covariance matrix of the deviations is performed in which the dominant eigenvectors(K) and eigenvalues are calculated. Often points move in highly correlated ways so most of the variability is captured by the first few eigenvectors. Any shape(X_i) within the training set can be approximated by

$$X_i = M + Kb \quad (1)$$

By varying the weights of the most important eigenvectors (the shape parameters, b) instances of the object are generated; if the weights are kept within some reasonable limits (about 3 standard deviations from the mean), all shapes produced are plausible examples of the class. These models are *specific* to a certain class of objects and they are unlikely to deform to approximate the shape of an object that does not belong to that class.

4 Fitting The Models

Point Distribution Models can be used as Active Shape Models [2] to fit to new examples in images in an iterative local optimization scheme. The

model is placed on the image and is allowed to interact dynamically until it fits to the object in the image. At each model point a profile perpendicular to the boundary is extracted and a new preferred position for that point is estimated along the profile. Various ways of defining the new preferred position for each point have been developed [2,3,4,5]; the most common method is to choose the strongest edge along the profile since in many cases control points are located on boundaries. The key to the method is that model points *do not* move *individually* to the suggested new positions. The scale, translation, rotation and shape parameters are adjusted so that the model moves and deforms in an attempt to minimise the overall distance between all the current model points and the new set of preferred positions for each point, keeping the shape parameters within some limits so that the model always represents a legal example of the class.

5 Classification

Classification is based on the goodness of fit for each model. A Point Distribution Model representing each class of object to be classified is generated; during classification all models are fitted to the input image. The quality of fit of each model is assessed using a fit measure which shows how well the model example corresponds to the object. This can be defined according to the particular application, but in general it must have two terms;

- A *support measure* which is an indication of how well the model is supported by image data.
- An *explained evidence* measure which indicates how well the model explains the image data.

Only the correct model is likely to be able to maximise both measures.

6 Applications

The system has been tested in classifying plant seeds, handwritten characters and human faces. All three applications are difficult, the last two being classical problems in computer vision. The aim of our experiments is to prove that it is feasible to approach these problems and other similar ones with our technique.

6.1 Classifying plant seeds

Classifying plant seeds is a suitable application for the system described since plant seeds belonging to the same class are not identical. For this experiment we used only the shape of the boundary to classify seeds. No other information (i.e size, colour or depth) was utilised. Seeds belonging to 6 different classes were used; pumpkin, popcorn, sunflower, barley, lentil and rice. During testing we presented the system with seed outlines, belonging to the 6 classes mentioned, in an attempt to classify them. As shown in figure 4 boundaries of some classes are very similar so that even for humans is very difficult to recognise the seeds given only their shapes.

6.1.1 Training. Point Distribution Models representing the boundary of each class of seeds were generated. Sixteen training shapes from each class

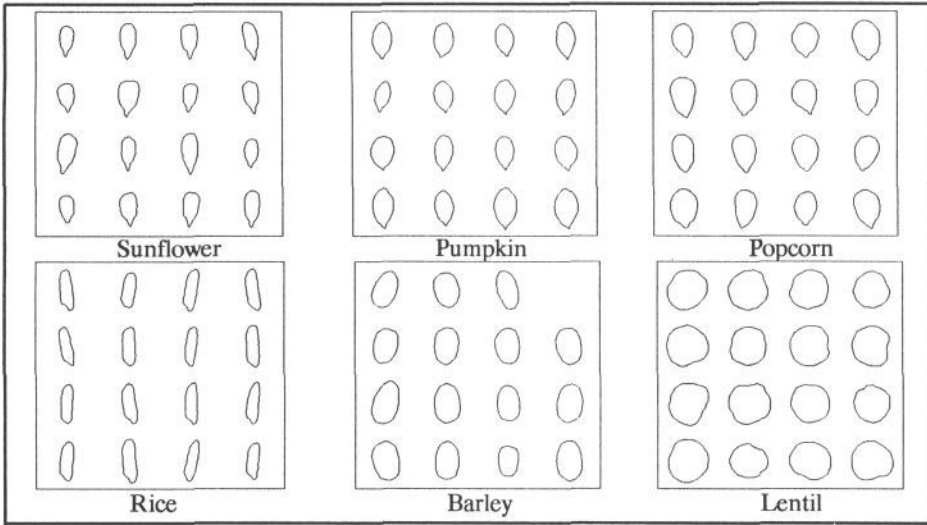


Fig 4 Example of boundaries representing each class of six seed classes

were used; each one was represented by 40 points. For all the models most of the variability is explained with the first 3 eigenvectors showing that there are only 3 important modes of variation.

6.1.2 Fitting The Models. For this experiment we assumed that the orientation of the plant seeds was known so that we could apply the local optimization technique directly. The new suggested position for each model point was the strongest nearby edge, since we wanted to locate the boundary. Models fitted to the boundaries of seeds are shown in figure 5.

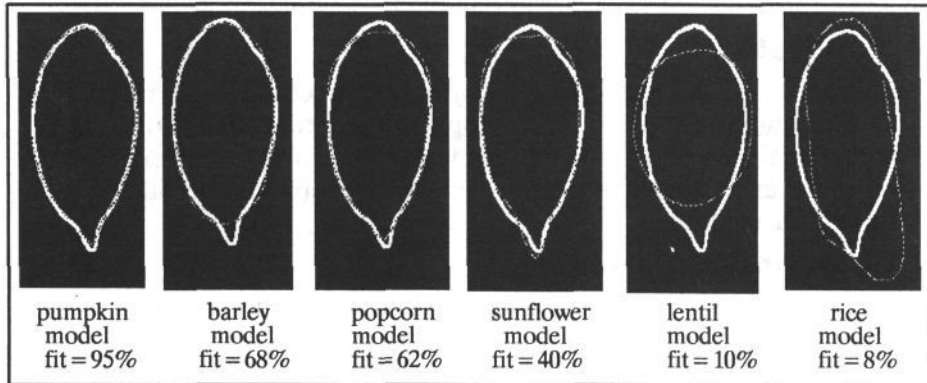


Fig 5 Example of different seed models fitted to a pumpkin outline

6.1.3 Classification For this application a support measure alone is adequate for assessing the fit since models cannot maximise the support measure on the closed seed boundaries without maximising the explained evidence at the same time. The support measure(X_s) was calculated using

$$X_s = \frac{(\text{model points}) \cap (\text{dilated boundary image points})}{(\text{model points})} \quad (2)$$

Where *model points* are all the points produced by interpolating between the model control points and *boundary image points* are all the points in the boundary of the binarised seed.

6.1.4 Results. The system was tested on 396 seeds (66 from each class); 96% of the examples were recognised correctly. In the test set we included the seeds used for training (96 seeds); the rest were new examples. Table 1 is a confusion matrix showing how the examples were classified. Most of the mis-classifications involve confusion between pumpkin, sunflower and popcorn seeds which look very similar.

input	recognised as					
	pumpkin	popcorn	sunflower	barley	lentil	rice
pumpkin	94%	6%				
popcorn	3%	94%	2%		1%	
sunflower	4%	6%	90%			
barley				100%		
lentil					100%	
rice						100%

Table 1 The confusion matrix for the plant seed classification experiment

6.2 Optical Character Recognition of Handwritten Postcodes

A more difficult problem is that of recognising handwritten postcode characters. In this application the shape variability is enormous and the number of classes is quite large (36 if we try to recognise letters and numerals). Additional complications arise because many characters are similar to each other (i.e O and D, B and 8, Z and 2).

We have trained and tested our method using real postcode characters taken from envelopes (see acknowledgements). Software to binarise and isolate the postcode characters was already available so our work focused on the recognition algorithm.

6.2.1 Training. Point Distribution Models representing all upper case letters and numerals were generated (35 all together as the model for 0 and O were the same). The training examples were thinned so that the models generated represented the skeletons of the characters. Twenty-five examples were used to train each model; on each training shape 20 control points were located. In all cases the first 4 eigenvectors of the covariance matrix represented more than 90% of the total variability suggesting that most of the characters have only 4 main modes of variation. The effects of varying the first 4 shape parameters for model '7' are shown in fig 6. By changing all the shape parameters simultaneously a wide range of possible shapes representing the number 7 can be produced. This model should be able to deform and fit to any shape representing character 7 that can be found as a postcode character.

6.2.2 Fitting The Models. Each model was used as an active shape model in this case. The input image was thinned and model points were attracted towards the bright spots of the image i.e towards the points on the skeleton.

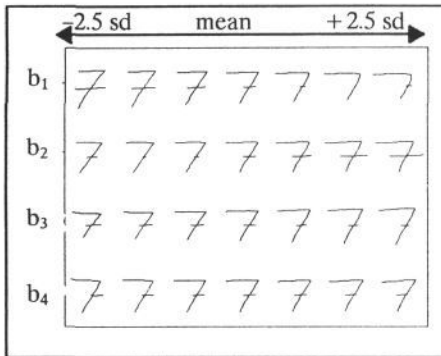


Fig 6 The effects of the first 4 shape parameters for the model '7'

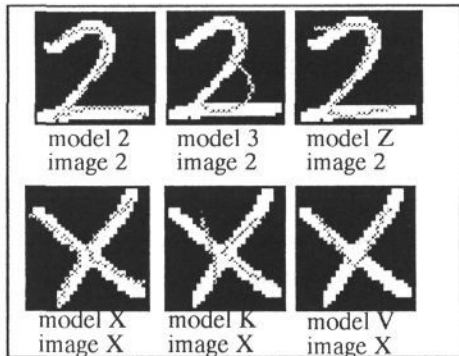


Fig 7 Examples of character models fitted to images

However, because of the complexity of the shapes it is important that the initial placement of the model is as close as possible to the image; if this is not the case model points may get attracted by the wrong skeleton points resulting in a poor fit. To overcome this problem the scale and centre of gravity of the input character was estimated and a coarse exhaustive search around these values performed. For each possible starting pose the local optimization technique was applied until the fit measure was optimised. Example of models fitted to characters are shown in figure 7. It is important to note that all examples represent a plausible instance of the modelled character and models *do not* deform to fit adequately to a character outside their class.

6.2.3 Classification During classification all models are fitted to the input character. For this application it is essential to use an explained evidence measure (X_e) as well as a support measure (X_s) since there are many cases in which wrong models maximise only one of the two measures. For example in figure 7 model V can have a very high support measure when fitted to character X but it explains only half of the image. The two measures are defined as follows:

$$X_s = \frac{(\text{model points}) \cap (\text{dilated skeletonised image points})}{(\text{model points})} \quad (3)$$

$$X_e = \frac{(\text{dilated model points}) \cap (\text{skeletonised image points})}{(\text{skeletonised image points})} \quad (4)$$

Where *model points* are all the points produced by interpolating between the model control points and *skeletonised image points* are all the points in the skeleton of the binarised character.

6.2.4 Results The system was tested on the test set of the data set (about 5900 characters). The results obtained (see table 2) are very promising and demonstrate that the classification method described can be used successfully for recognising handwritten characters. Other published attempts to classify characters from the same database using a single algorithm give correct classifications between 65%–75% [7]. Using parallel multi-layer clas-

sifier architectures[8] results up to 98% have been reported. Experiments to define the performance of humans in recognising postcode characters from the same database have also been performed[10]. For the full alphanumeric set the rate of correct classifications was about 90% suggesting that the data base used for our experiments contains characters which are very difficult to recognise.

	performance of humans	correct classifications	correct model within best 3	correct model not within best 3
alphanumeric	90%	86%	97%	3%
letters	not quoted	90%	97%	3%
numerals	95%	93%	98%	2%

Table 2 Results for the character recognition

6.3 Recognising Human Faces

The experiments described so far involve binary images. In the final experiment we dealt with grey-level images; we chose to use face images since automatic face recognition is regarded as a challenging problem in computer vision. For this initial experiment we used face images taken from 10 individuals to train the system. We then presented the system with new images of each subject and attempted to classify them. Typical face images used in the experiment are shown in figure 11. The only information used to classify face images was the quality of fit of each model. In order to assess the fit a grey level modelling technique was employed. The models represented only the inner characteristics of faces and no information about the hairline was utilised. The face models used consisted of several sub-parts, each part representing a distinct feature like the outline, mouth, nose, eyes, eyebrows and ears.

6.3.1 Training. During training Point Distribution Models representing each individual's face were generated. Ten training images were used to train each model; on each training image 144 points were located. Each training set contained images of a person with different expressions. The models generated could be deformed to approximate an individual's face with any expression. A typical set of training shapes and the main modes of variation are shown in figures 8 and 9 respectively.

6.3.2 Fitting The Models. During fitting, models were used as active shape models. In this case the calculation of the new suggested position for each model point was more complicated since model points do not always lie on strong edges. To overcome this problem a grey level modelling technique [3,4,5] was employed. During training each training shape was overlaid on the corresponding training image and profiles perpendicular to the boundaries were extracted. The mean (M) and the covariance matrix (S) of the normalised derivatives of the profiles were calculated for each model point so that a grey level model of the expected profile at each model point was generated. During fitting profiles perpendicular to the boundary at each model point were extracted. The profile model was run along the extracted profile

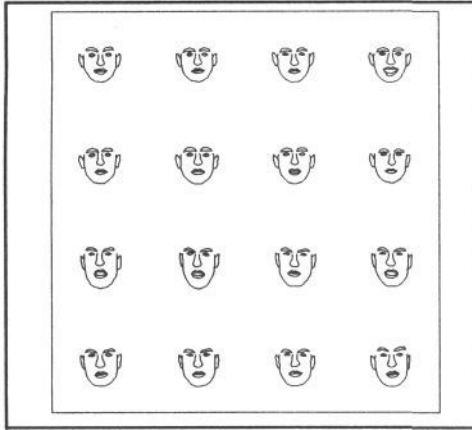


Fig 8 Training shapes for two individual face models. (top two and bottom two rows)

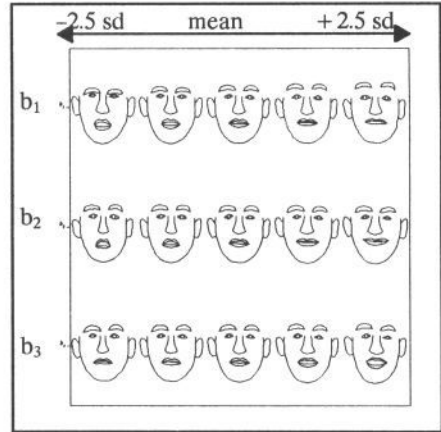


Fig 9 The effects of the first 3 shape parameters in a face model

and the best correspondence was found by minimising the Mahalanobis distance(d), where g is a subsection of the extracted profile and

$$d = (g - M)^T S^{-1} (g - M) \quad (5)$$

In order to avoid the inversion of the matrix in this equation an approximation to this expression was used (see [3]). Examples of models fitted to face images are shown in figure 10.

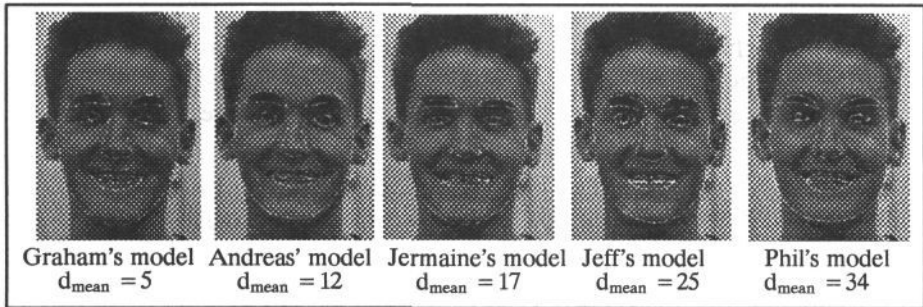


Fig 10 Face models fitted to Graham's face image

6.3.1 Classification The quality of fit of each model was assessed using the same method as that which defines the new suggested position for model points. At each instance of the model the Mahalanobis distance(d_i) between the profile at each point(i) and the model profile for that point is calculated. The best fit is given by the model that minimises

$$d_{mean} = \frac{1}{N} \sum_{i=0}^{i=N-1} d_i \quad (6)$$

where N is the number of points in the model.

In order to cope with 'disguises' (e.g glasses, changed hairstyle) the mean distance for each part of the model is calculated individually and the mean and standard deviation of all the parts is calculated. If the distance for a part is not within 2 standard deviations of the mean then the part is not used for

the calculation of the total distance. For example when Phil's test image (see figure 11) was presented to the system the fit measure for the eyes and the ears was not taken into account.

6.3.1 Results. The system was tested on 120 test images (12 from each individual). For 85% of the cases the individual was recognised correctly. In the test set distractions and changes in the lighting conditions were introduced in order to test the robustness of the system. For example people wore spectacles or sun-glasses, were unshaven or had different hairstyles to those they had for the training images. It should be noted that in the training images nobody was wearing spectacles and all were clean shaven. Examples of images from the training and test sets are shown in figure 11.

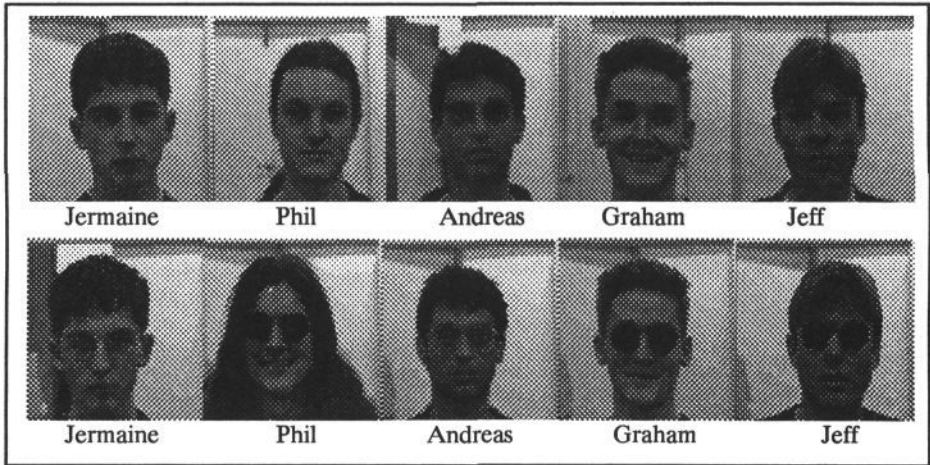


Fig 11 Face images used in the classification experiment. Top row are examples from the training set, bottom row are from the test set.

7 Discussion

A system for classifying variable objects has been presented. We have tested the system on three difficult problems and the results obtained are very promising. An important point is that although it is a generic method the results compare well with results obtained when using techniques dedicated to solve a particular problem.

The main disadvantage of the system is that when the number of classes is large it becomes impractical to train the models. In that case a single model representing all classes may be employed. For example in the face recognition problem a generic model may be built so that instead of modelling just the shape variations due to changes in expressions we also model variations between different individuals. This approach is under development.

8 Acknowledgements

Andreas Lanitis is funded by a University of Manchester research studentship and an ORS award. We would like to thank Dr A.C.Downton, Department of Electronic Systems Engineering, University of Essex and The Post Office Research Centre for providing the database containing handprinted

postcodes. Also members of the Department of Medical Biophysics, especially Dr A.Hill and Mr G.Charters for their help and advice. Finally all those who kindly volunteered to provide face images for our experiments.

9 References

- [1] T.F.Cootes, D.H.Cooper, C.J.Taylor and Graham J. "A Trainable Method of Parametric Shape Description" in Proceedings of the British Machine Vision Conference, 1991.
- [2] T.F.Cootes and C.J.Taylor, "Active Shape Models - "Smart Snakes" ", in Proceedings of the British Machine Vision Conference, 1992.
- [3] T.F.Cootes and C.J.Taylor, "Active Shape Model Search Using Local Grey-Level Models: A Quantitative Evaluation ", in Proceedings of the British Machine Vision Conference, 1993.
- [4] T.F.Cootes, C.J.Taylor, D.H.Cooper and J.Graham, "Training Models of Shape From Sets of Examples", in Proceedings of the British Machine Vision Conference, 1992.
- [5] T.F.Cootes, C.J.Taylor, A.Lanitis, D.H.Cooper and J.Graham, "Building and Using Flexible Models Incorporating Grey Level Information", in Proceedings of the Fourth International Conference on Computer Vision, 1993.
- [6] T.F.Cootes, A.Hill, C.J.Taylor and J.Haslam, "The Use of Active Shape Models For Locating Structures in Medical Images", in Proceedings of the 13th International Conference on Information Processing in Medical Imaging, 1993.
- [7] A.C.Downton, R.W.S.Tregidgo and E.Kabir, "Recognition and Verification of Handwritten and Handprinted British Postal Addresses", International Journal of Pattern Recognition and Artificial Intelligence vol 5, pp 266-291, 1991.
- [8] M.C. Fairhurst and K.D. Cowley, "Parallel Multi-layer Architectures of Increasing Hierarchical Order" Pattern Recognition Letters, vol 14, pp 141-145, 1993.
- [9] G.E.Hinton, C.K.I.Williams and M.D.Revow, "Adaptive Elastic Models for Handprinted Character Recognition", Advances in Neural Information Processing Systems 4, Morgan Kauffman, San Mateo CA, 1992.
- [10] J.Nellis, "The Application of Artificial Neural Networks and Associated Technologies to the Task of Hand-Printed Character Recognition", PhD Thesis, Dpt. of Electrical Engineering, Brunel University, 1993.
- [11] D.Rushall, "Shape Classification Using CLD", Msc Thesis submitted to the Department of Computer Science, University of Manchester, 1991.
- [12] C.K.I.Williams, M.D.Revow and G.E.Hinton, "Hand-printed Digit Recognition Using Deformable Templates", in Spatial Vision in Humans and Robots, eds. L.Harris and M.Jenkin, Cambridge University Press, 1992.