

Chapter 9



9 Extending the Point Distribution Model

9.1 Introduction

Thus far, statistical models of deformation have been considered where the vector \mathbf{x} consists of related features such as the co-ordinates of a connected contour, the vertices of a surface or the grey level intensity of each pixel of an image. The principle relies upon the variation of elements with regard to others and attempts to generalise the **relative** movement of the constituent components. It therefore holds that similar statistical linkage of features could be achieved even if they lie within different co-ordinate frames and represent quite different elements providing that there is still some linear relationship between the various elements. This chapter will discuss the use of this technique to link together related information from differing sources. Section 9.2 will discuss combining shape information with abstract parameters and using this to infer unseen information from examples of shape. Section 9.3 will present the application of this technique to inferring the shape and position of a human body from an image sequence. Finally conclusions will be drawn.

9.2 Combining Features Statistically

9.2.1 A Linear PDM with an Abstract Parameter

The linear 3D PDM of an eigenGlass, as constructed in chapter 9, provides an ideal example to demonstrate the hypothesis that related information can be combined into a PDM. It has already been shown that this PDM is essentially the same as the 2D contour of the glass due to the rotational symmetry of the object. Thus, the two dimensional vector that describes the glass profile can be used as a training vector for PCA and the final reconstructed model swept around the central axis to attain the full 3D model. This training vector \mathbf{x} describes the shape of the glass for each example in the training set. However, additional parameters can be concatenated to the vector for each example in the hope that some mapping which links the shape with other features can be achieved. For each training example an abstract parameter MF was estimated. The parameter corresponds to the masculinity or femininity of a specific training example. This provides a rather subjective scale but provides an illustrative demonstration that a link between shape and aesthetic appearance can be achieved. Figure 9.2.1 shows each training example with the corresponding MF parameter estimated, $0 < MF < 1$, where 0 corresponds to feminine and 1 to masculine.

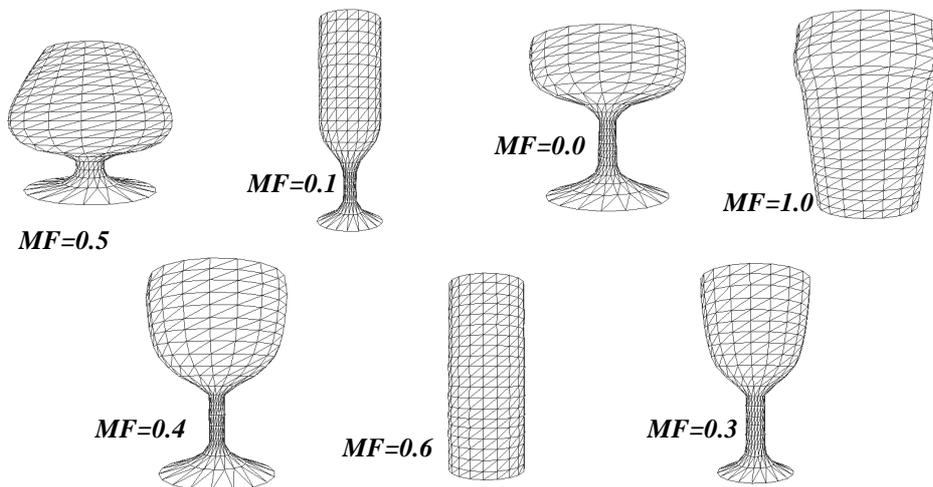


Figure 9.2.1 - MF Parameter for eigenGlass Training Set

For each n -dimensional training vector \mathbf{x} a new training vector is constructed by concatenating the MF parameter to the existing vector producing an $n+1$ dimensional vector $\mathbf{x}' = (\mathbf{x}, MF) = (x_1, y_1, x_2, y_2, \dots, MF)$.

After PCA has been performed on the training set the resulting PDM can be used to reconstruct new drinking vessels of various shapes along with a corresponding MF value. Figure 9.2.2 shows the primary mode of variation of the eigenGlass model from the mean shape along with the corresponding MF value.

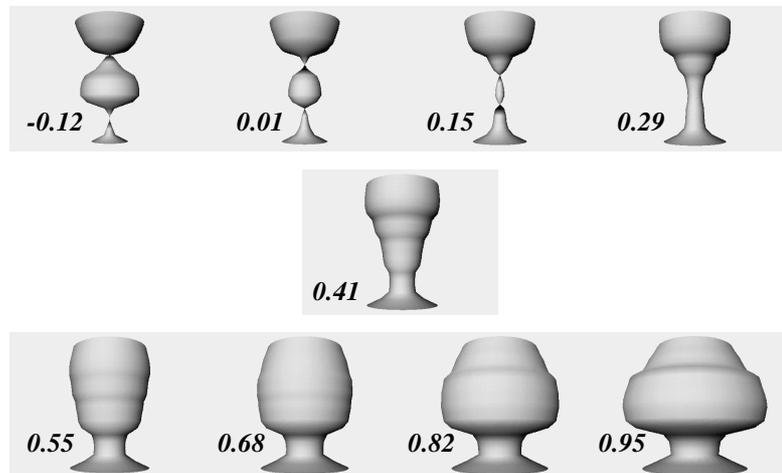


Figure 9.2.2 - Primary mode of variation of Augmented eigenGlass PDM

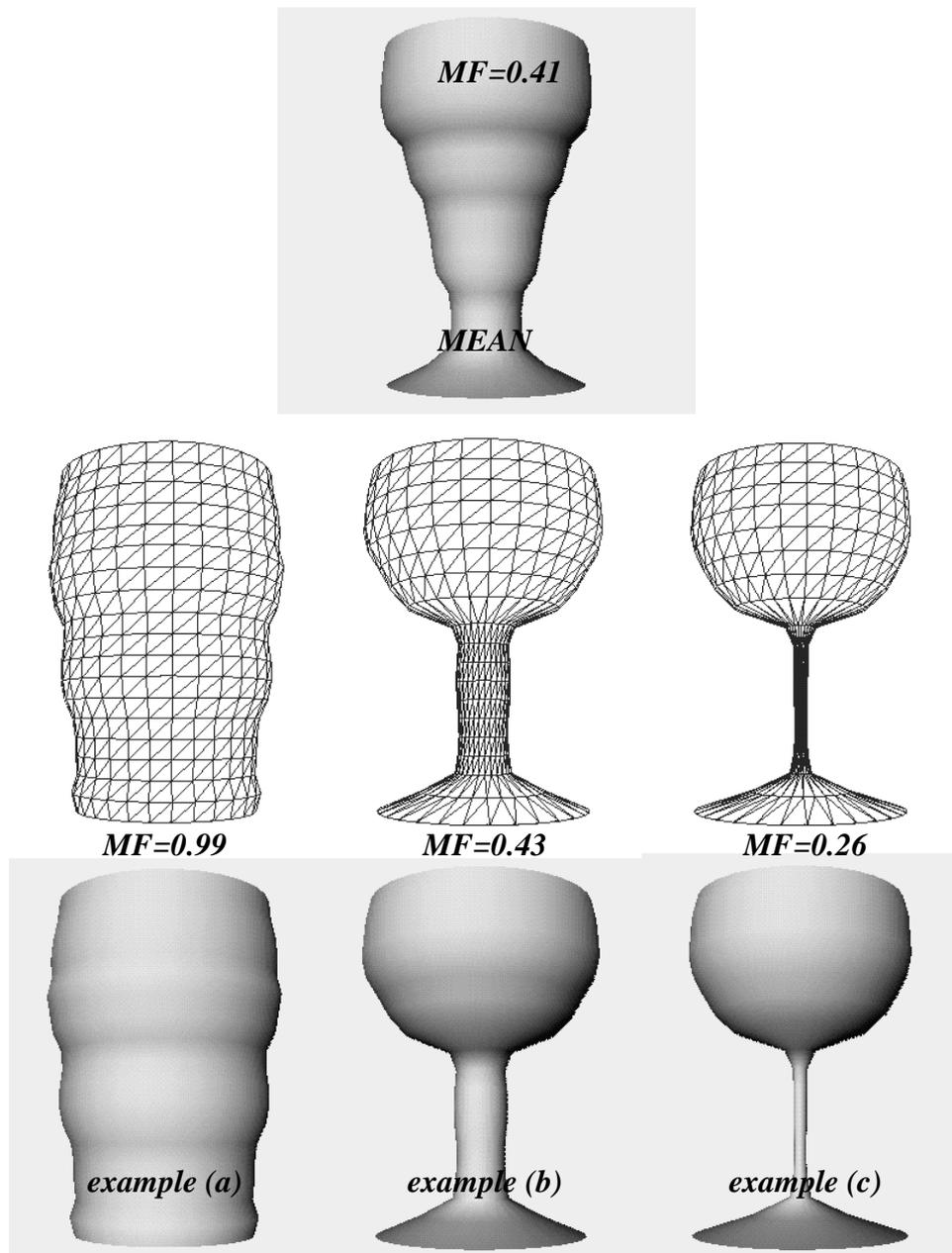


Figure 9.2.3 - Reconstructed glasses and *MF* value from Augmented eigenGlass PDM

Figure 9.2.3 shows the results of reconstructing various glass types from the eigenGlass model along with the corresponding *MF* value. This is achieved by manipulating the weighting parameters of the model. As the overall shape changes, so the additional *MF* parameter changes accordingly. It can be seen that the pint glass produces a high *MF* value which corresponds to the training set. Similarly the wine glass example c) produces a low *MF* value, demonstrating that the PDM has successfully achieved some mapping between the elements. As

with all PDMs, the ability of the final model to reproduce examples from the training set is augmented by the ability to generalise the shape information and produce *unseen* shapes, not present within the training set. When this is done an *MF* value is also produced and by observing this parameter it is possible to draw some conclusion about what the PDM has encoded.

The model demonstrates a high correlation between the size (volume) of the glass and the *MF* value. This is to be expected, as the high *MF* examples were the larger types of glass. However, example (b) shows the results of attempting to make a 'more' masculine wine glass and results in a thicker stem. So it could also be concluded that the more delicate the stem of a glass the more feminine its appearance. This would seem a fair assumption given that in the training examples the two extremities of *MF* were a pint Beer glass and a Champagne glass where the major difference between the examples was the stem thickness (see Figure 9.2.1).

This is an extremely subjective example but demonstrates how additional information can be incorporated into a PDM.

9.2.2 Scaling Issues and Eigen Entropy

One of the important issues when elements are to be combined for statistical analysis is that of scaling. If an element contains too much variation across the training set (due to the incorrect scaling of that component) then that element will bias the PCA and dominate the principal modes of variation. In some cases this may be desirable, e.g. when it is intended that the primary mode correlates directly to the variation of a specific feature. However, more often, this is an undesirable effect.

The premise of the PDM is that the largest variation of the training set should be represented within the eigenvector corresponding to the largest eigenvalue. By artificially biasing the PCA with an incorrect scaling the information content of the PDM is destroyed.

If the eigenGlass model is considered, the construction of the training vector should contain a scaling parameter α where

$$\text{Equation 9.2-1} \quad \mathbf{x}' = (\mathbf{x}, \alpha MF) = (x_1, y_1, x_2, y_2, \dots, \alpha MF).$$

When the training set is assembled this additional parameter can be scaled appropriately to ensure incorrect dominance does not occur. However, it is generally not apparent what this scaling value α should be for any particular example.

Sumpter, Boyle and Tillett [Sumpter 97] proposed a method for estimating the scaling of parameters by calculating the eigen entropy (E) of the normalised eigen vectors (p), and estimating the value α which maximises this entropy $E(\alpha)$,

$$\text{Equation 9.2-2} \quad E = E(\alpha) = -\sum_{i=1}^{n+1} p_i \log_2(p_i), \text{ where}$$

$$\text{Equation 9.2-3} \quad p_i = \frac{\lambda_i}{\sum_{i=1}^{n+1} \lambda_i}, \quad E(\alpha) \leq \log_2(n+1),$$

$$\text{and } E(\alpha) \rightarrow 0 \text{ as } \alpha \rightarrow \infty$$

Figure 9.2.4 shows the results of performing this procedure upon the eigen glass example. From this graph it can be seen that the optimum eigen entropy is achieved with a scaling of around $\alpha = 137$.

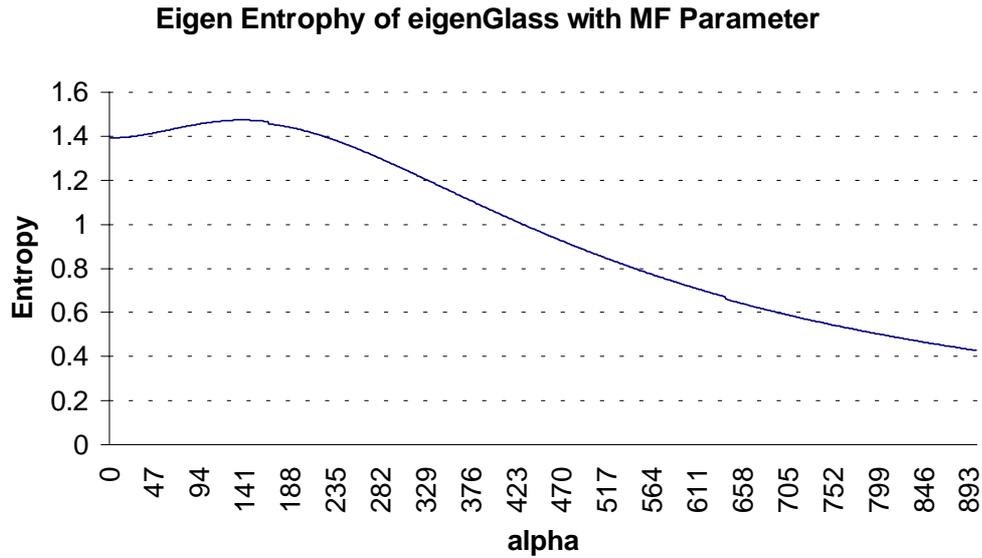


Figure 9.2.4 - Graph of eigen entropy for varying parameter scaling

If PCA is now repeated upon the eigenGlass example with $\alpha = 137$, a new PDM is constructed. It is hoped that the *MF* parameter presents increased significance within the primary modes of variation. This hypothesis can be confirmed by examining the eigenvalues and the variance of *MF* for models constructed with $\alpha = 1$ and 137.

Figure 9.2.5 shows the histogram of normalised eigenvalues in percentile form (see chapter 3.2) for the eigenGlass example with the two aforementioned *MF* scalings. As would be expected the addition of this parameter and its increased significance within the primary modes (for $\alpha = 137$) has removed some of the information content from the primary modes of variation, with a small increase in the significance of the latter modes. However, the resulting model still retains 99% of the variance within the first four modes so the information content is preserved, unlike $\alpha \rightarrow \infty$ which results in only a single mode of variation (due to the dominance of *MF* over the PCA), destroying the information content of the model.

Graph Showing the Contribution of eigen Vectors to the Total Deformation

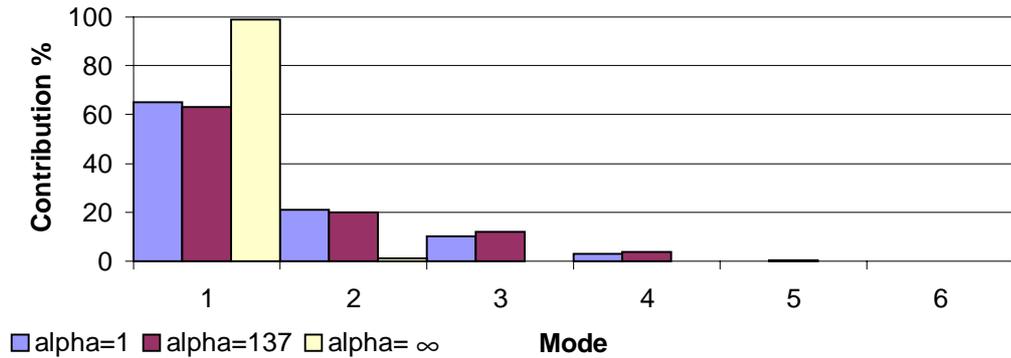


Figure 9.2.5 - Graph demonstrating the normalised eigen values for the eigenGlass example with different parameter scaling

Graph Showing the Variance of the MF Parameter for PDMs with Different Alpha Scalings

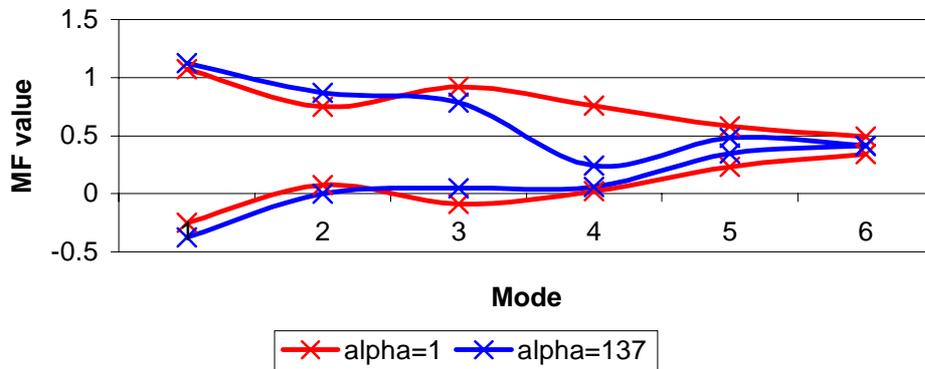


Figure 9.2.6 - Graph demonstrating the increased variance in eigenGlass example for correct parameter scaling

Figure 9.2.6 demonstrates this increase in the variance of the MF parameter by plotting the bounds of the variance for each of the primary modes of the two PDMs. Both variances are based around the mean MF and regress to this mean as the contribution of a mode diminishes. It can be seen for $\alpha = 137$ that the variance of MF is increased within the first two modes, with a significant reduction of this variance in the latter modes. This demonstrates that the

increased scaling has forced the statistical correlation into the primary three modes while more evenly distributing the overall variance of the model.

From Figure 9.2.5 and Figure 9.2.6 it can be concluded that the parameter scaling increases the correlation between shape and parameter without destroying the information content of the resulting PDM. However, another important consideration is 'how has this affected the primary modes of shape deformation?'. This can be answered by comparing the deformation of the original eigenGlass PDM to this new weighted model.

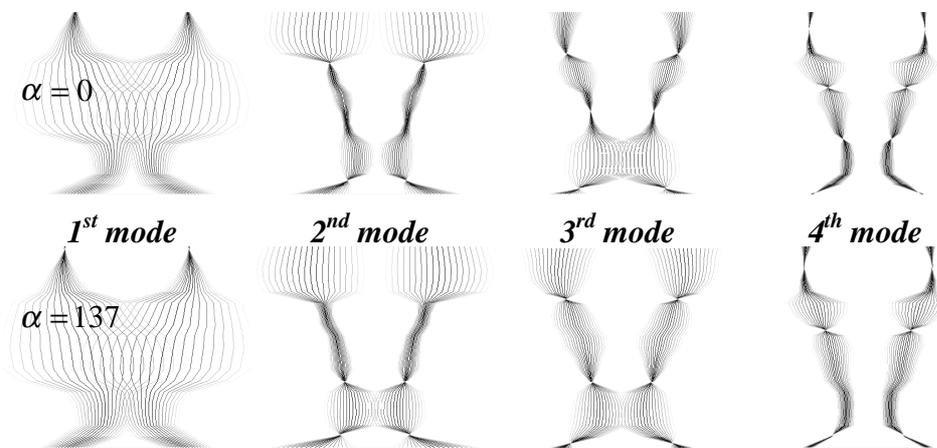


Figure 9.2.7 - Primary modes of eigenGlass PDM with different alpha scalings

It can be seen from Figure 9.2.7 that the increased significance of the *MF* parameter has done little to effect the overall deformation of the eigenGlass shape. It has increased the shape deformation to accommodate the *MF* parameter which shows that, although a correlation is being achieved, it is not a simple linear correlation. This could be addressed by using a non-linear model as previously developed, this will be discussed in more detail in Section 9.3.

9.2.3 Statistical Inference

It has been shown how additional information can be incorporated into a PDM which does not necessarily have to lie within the same co-ordinate frame as the shape deformation. It has also been shown how this information can be statistically linked to the other features of the model. When a shape is

reconstructed, so the additional parameters of the model are estimated due to the statistical linkage that occurs between the elements during PCA. However, what is desirable is to be able to use this model to estimate the parameters for unseen objects or even predict shapes that correspond to specific parameter values.

Using the matrix form of a linear PDM the shape \mathbf{x} of a model is equal to the mean shape plus the weighted sum of the eigenvectors

$$\text{Equation 9.2-4} \quad \mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$$

where \mathbf{x} is the shape vector, $\bar{\mathbf{x}}$ is the mean shape, $\mathbf{P} = (\mathbf{v}_1, \mathbf{v}_2, \dots, \mathbf{v}_t)$ is a matrix of the first t eigenvectors and $\mathbf{b} = (b_1, b_2, \dots, b_t)^T$ is a vector of weights.

Given a new shape \mathbf{x}' , the closest allowable shape from the model is constructed by finding \mathbf{b} such that

$$\text{Equation 9.2-5} \quad \mathbf{b} = \mathbf{P}^{-1}(\mathbf{x}' - \bar{\mathbf{x}}) \text{ and } -3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}$$

The closest allowable shape can then be reconstructed as

$$\text{Equation 9.2-6} \quad \mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}$$

If the eigenGlass example is now considered, it is feasible that given a new 'unseen' glass example (\mathbf{x}') the PDM could be used to estimate a value for MF . As the PDM has encoded a statistical link between the shape and parameter this model can be used to predict this estimate. However, the two elements have different dimensionality. The unseen example has dimensionality of $2n$, where the PDM has a dimensionality of $2n+1$. The new example \mathbf{x}' could be converted to a $2n+1$ vector by the addition of a zero, and the vector then reconstructed using the procedure above. However, in finding the closest allowable shape from the PDM, weighting parameters would be extracted that best fit the shape and provide an MF of zero. For non-linear mappings where the correlation between these elements is complex and the linear formulation of the PDM is over

generalising, this mapping will lead to unreliable results, i.e. **the zero parameter will bias the reconstruction**. As the number of unknown parameters increases this zero bias will begin to dominate the reconstruction and the resulting reconstructed vector will begin to degrade. Instead the model must be reduced to the dimensionality of the vector.

This is achieved by taking the matrix \mathbf{P} which is a $2n+1 \times j$ matrix of eigen vectors and extracting a smaller matrix \mathbf{P}' which is a $2n \times j$ matrix.

$$\mathbf{P} = \begin{pmatrix} v_{0,0} & v_{0,1} & v_{0,2} & \cdots & v_{0,j} \\ v_{1,0} & v_{1,1} & v_{1,2} & \cdots & v_{1,j} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ v_{2n,0} & v_{2n,1} & v_{2n,2} & \cdots & v_{2n,j} \\ v_{2n+1,0} & v_{2n+1,1} & v_{2n+1,2} & \cdots & v_{2n+1,j} \end{pmatrix}$$

\mathbf{P}'

MF element of each eigen vector
is discarded to construct \mathbf{P}'

This is done by discarding the elements of each eigen vector which correspond to the unknown elements of the model (in this case the *MF* parameter). A similar procedure must be performed on the mean shape $\bar{\mathbf{x}} \in \mathfrak{R}^{2n+1}$ by discarding the unknown parameter to obtain $\bar{\mathbf{x}}' \in \mathfrak{R}^{2n}$. The weightings which produce the shape can then be calculated in a similar manner with the reduced dimensional model, where

Equation 9.2-7 $\mathbf{b}' = \mathbf{P}'^{-1}(\mathbf{x}' - \bar{\mathbf{x}}')$ and $-3\sqrt{\lambda_i} \leq b_i \leq 3\sqrt{\lambda_i}$

However, as only the dimensionality of the eigen vectors was changed and not the number of eigen vectors, \mathbf{b}' has the same dimensionality as $\mathbf{b} \in \mathfrak{R}^j$. The weighting vector \mathbf{b}' can therefore be placed directly into Equation 9.2-4 to reproduce the shape \mathbf{x} , by

Equation 9.2-8 $\mathbf{x} = \bar{\mathbf{x}} + \mathbf{P}\mathbf{b}'$

The closest allowable shape vector $\mathbf{x} \in \mathfrak{R}^{2n+1}$ to $\mathbf{x}' \in \mathfrak{R}^{2n}$ has now been reconstructed. However, the additional information in \mathbf{x}' , contains the missing *MF* information which has been estimated from the available shape information and the *a priori* information contained within the model of about shape and how this relates to the *MF* parameter.

9.3 Extending the Model to Inferring Human Motion

9.3.1 Introduction

The human vision system is adept at recognising the position and pose of an object, even when presented with a monoscopic view. In situations with low lighting conditions in which only a silhouette is visible, it is still possible for a human to deduce the pose of an object. This is through structural knowledge of the human body and its articulation.

A similar internal model can be constructed mathematically which represents a human body and the possible ways in which it can deform. This is the premise of model based vision, and as has been previously shown, this deformation can be learnt using a Point Distribution Model. By introducing additional information to the PDM that relates to the anatomical structure of the body, a direct mapping between skeletal structure and projected shape can be achieved.

This section uses the previously presented techniques to statistically combine the 2D silhouette of a human body projected onto the image frame with the 3D pose of the body. To further aid the tracking and reconstruction process, additional information about the location of both the head and hands is combined into the model. This helps disambiguate the model and provides useful information for both its initialisation and tracking.

9.3.2 Constructing a Combined Non-linear Point Distribution Model for a Human

The point distribution model is constructed from three components: the position of the head and hands within the image frame; the 2D contour which represents the shape of the body silhouette; and the 3D structure of the body (see Figure 9.3.1). Each of these components are generated separately from the training image sequence and then concatenated to provide a training vector representing all these attributes.

The relative position of the head and hands is represented as the location of these features in the image frame. When concatenated, this generates a six dimensional feature vector $V_H=(x_1,y_1,...x_3,y_3)$. The body contour, once extracted from the image, is resampled to a list of 400 connected points. These are concatenated into an 800 dimensional feature vector $V_C=(x_1,y_1,...x_{400},y_{400})$. Lastly the skeletal structure of the 3D model is represented by 10 3D points which produce a 30 dimensional feature vector V_S . The relative location of the hands and head helps to disambiguate the contour during tracking. It can also be used to estimate an initial location and shape for the body contour.

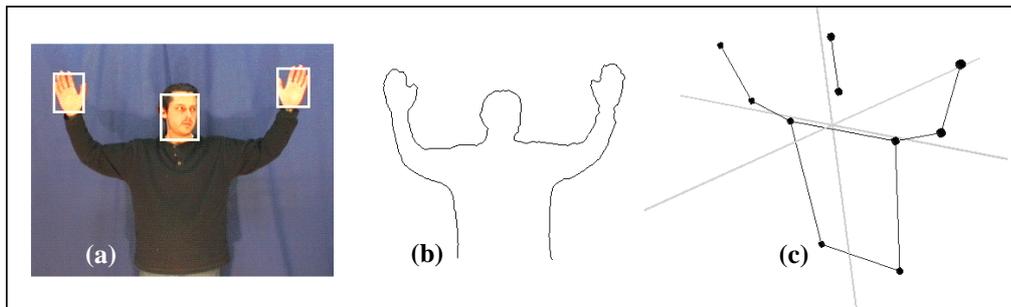


Figure 9.3.1 Composite elements of human body PDM

**(a) Position of head and hands V_H (b) Body Contour V_C
(c) Corresponding 3D model V_S**

The position of the head and hands is extracted from the training image sequences using the Hue-Saturation colour thresholding technique described in Chapter 4.

For the purpose of simple contour extraction from the training set, shape extraction is facilitated through the use of a blue screen and chroma keying. This allows the background to be simply keyed out to produce a binary image of the body silhouette. As the figure always intersects the base of the image at the torso, an initial contour point is easily located. Once found, this is used as the starting point for a contour tracing algorithm which follows the external boundary of the silhouette and stores this contour as a list of connected points. In order to perform any statistical analysis on the contour, it must first be resampled to a fixed length. To ensure some consistency throughout the training set, landmark points are set at the beginning and end of the contour. A further landmark point is allocated at the highest point along the contour within 10 degrees of a vertical line drawn from the centroid of the shape. Two further points are positioned at the leftmost and rightmost points of the contour. This simple landmark point identification results in non-linearity within the model. The problems associated with this are discussed in Section 9.3.5.

The 3D skeletal structure of the human is generated manually. Co-ordinates in the xy (image) plane are derived directly from the image sequence by hand labelling. The position in the third dimension is then estimated for each key frame.

9.3.3 Scaling the Model

When combining information for statistical analysis via PCA it is important that constituent features ($V_H V_C V_S$) are scaled to ensure that any particular feature does not dominate the principal axes. This can be done by calculating the eigen entropy as discussed earlier (section 9.2.2). However, as all three components exist within the same co-ordinate frame and are directly linked, such a scaling should be unnecessary.

This assumption can easily be tested by formulating the vector \mathbf{x} as the weighted combination of the components where $\mathbf{x} = (V_C, \alpha V_H, \beta V_S)$. Using the same procedure as described earlier, the eigen entropy is calculated for

$0 < \alpha, \beta < \infty$ and suitable scaling values determined by maximising the entropy of the resulting PDM.

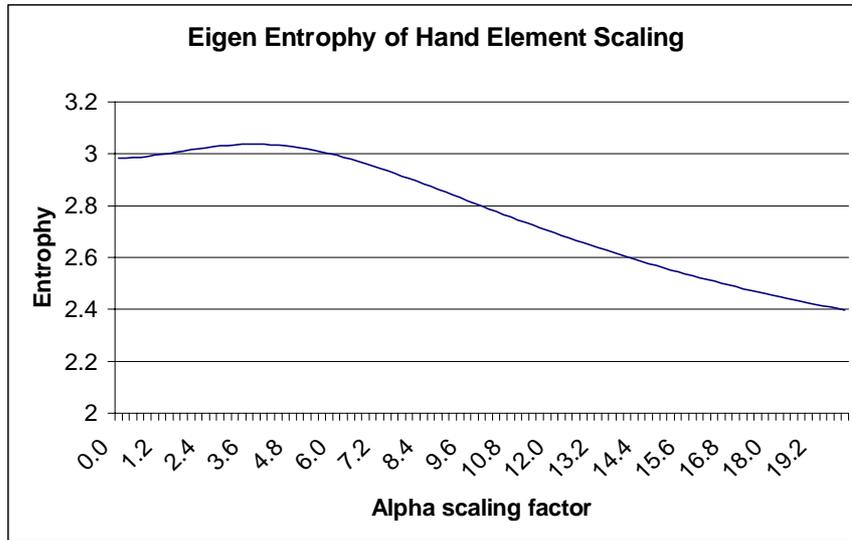


Figure 9.3.2 - Graph showing eigen entropy of hand element in composite body PDM

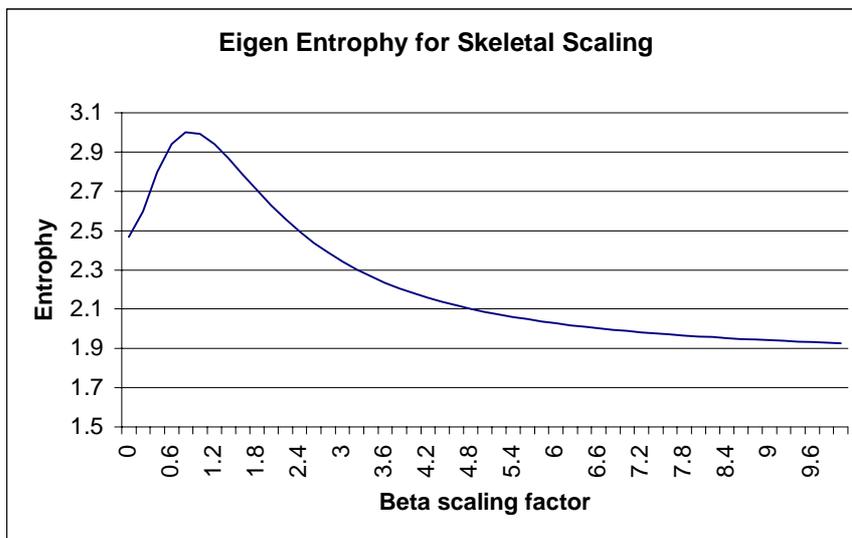


Figure 9.3.3 - Graph showing eigen entropy of skeletal element in composite body PDM

From Figure 9.3.2 it can be seen that the optimum scaling for V_H is around 4. Figure 9.3.3 shows that the skeletal element does not need scaling as the greatest

entropy is achieved when $\beta = 1$. This confirms the assumption that scaling is unnecessary as all the elements lie within the same (image) co-ordinate frame.

9.3.4 The Linear PDM

Once these separate feature vectors are assembled, they are concatenated to form an 836 dimensional vector which represents the total pose of the model. A training set of these vectors is assembled which represents the likely movement of the model. Figure 9.3.4 shows a sample of training images along with the corresponding contour and skeletal models in 2D.

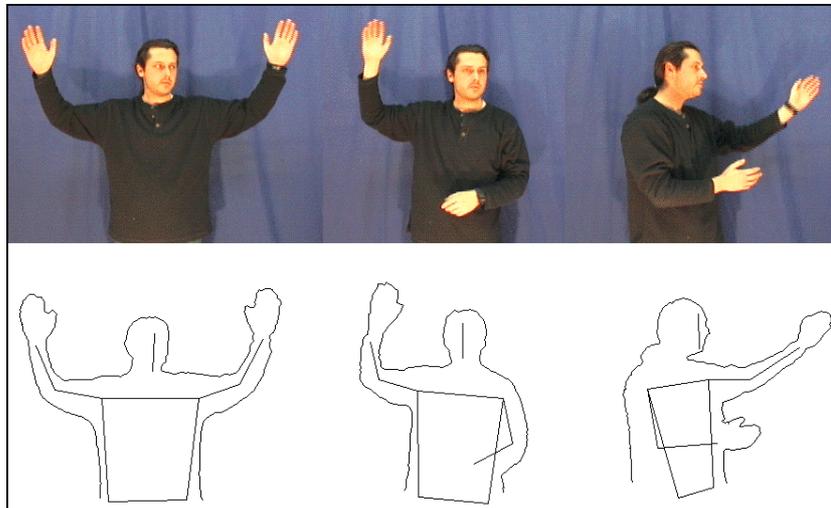


Figure 9.3.4 - Sample training images and corresponding contour and skeletal models

A linear PDM is now constructed from the training set and its primary modes of variation are shown in Figure 9.3.5.

After PCA is performed, it is calculated that the first 84 eigenvectors, which correspond to the 84 largest eigenvalues, encompass 99.99% of the deformation contained in the training set.

Figure 9.3.5 demonstrates the deformation of the composite PDM. The crosses are the locations of the hands and head. It can be seen that although the movement of the three elements are closely related, the model does not

accurately represent the natural deformation of the body. The shapes generated by the primary modes of variation are not indicative of the training set due to its inherent non-linearity. In order to produce a model that is accurate/robust enough for practical applications, a more constrained representation is required.

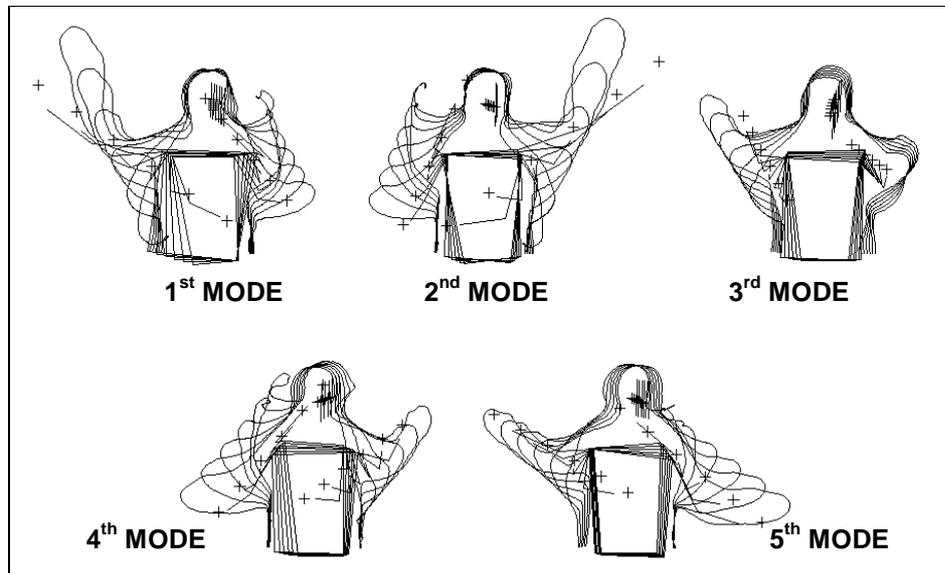


Figure 9.3.5 - Primary modes of variation on the linear PDM

9.3.5 Non-Linear Estimation

As described in chapter 6, to perform non-linear estimation upon the dataset the linear model is first used to reduce the dimensionality. 99.99% of the deformation is contained within the first 84 eigenvectors. However, the primary 40 modes of deformation encompass 99.8% of the deformation. Projecting the entire training set down into this lower dimensional space achieves a dimensional reduction of 836 to 40, which significantly reduces the computation time required for further analysis.

Performing cluster analysis upon the dimensionally-reduced dataset, the natural number of clusters is estimated to be 25. By performing further PCA on each of the 25 clusters, the shape of the model can be constrained by restricting the shape vector to remain within this volume. These constraints upon *shape space* are applied in the same manner as described in earlier chapters.

Figure 9.3.6 shows the training set after dimensional reduction gained from the initial linear PDM, projected into 2 dimensions. The bounding boxes represent the 25 clusters that best estimate the curvature. These bounding boxes are the bounds of the first and second modes of deformation for each linear patch (cluster). The number of modes for each cluster varies according to the complexity of the training set at that point within the space. All clusters are constructed to encompass 99.9% of the deformation within that cluster.

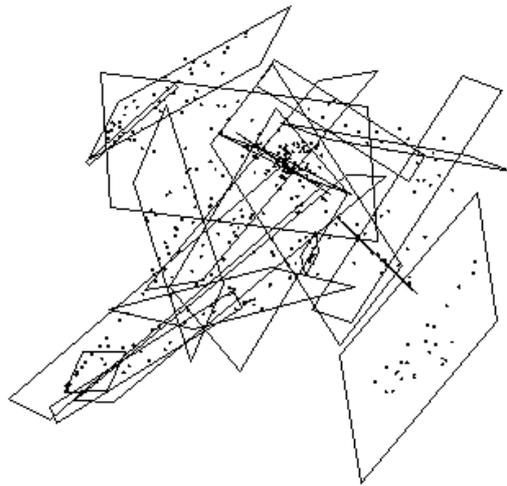


Figure 9.3.6 - Clusters in reduced shape space

9.3.6 Initialising the PDM

Upon initialisation the first step is to locate the position of the head and hands. This can be done by colour thresholding the entire image which, although computationally expensive, does not need to be repeated on every iteration. Once done these positions can be used to initialise the PDM and give an initial guess as to the shape of the contour to be found. As is it not clear which blobs correspond to which features, three possible contours are produced. The contour that iterates to the best solution provides the final state from which tracking proceeds.

9.3.7 Tracking with the PDM

Once initialised the two components must be fitted to the image separately. The contour is attracted to high intensity gradients within the image using local edge detection (chapter 3). The hand and head positions are used as centres in a single iteration of a kmeans-clustering algorithm on the segmented binary skin image. This is possible due to the assumption that the model will not change significantly from the last image frame.

9.3.8 Reconstruction of 3D Shape and Pose

As the shape deforms to fit with the image so the third element of the model, the skeleton, also deforms. By plotting this 3D skeleton, its movements mimic the motion of the human in the image frame.

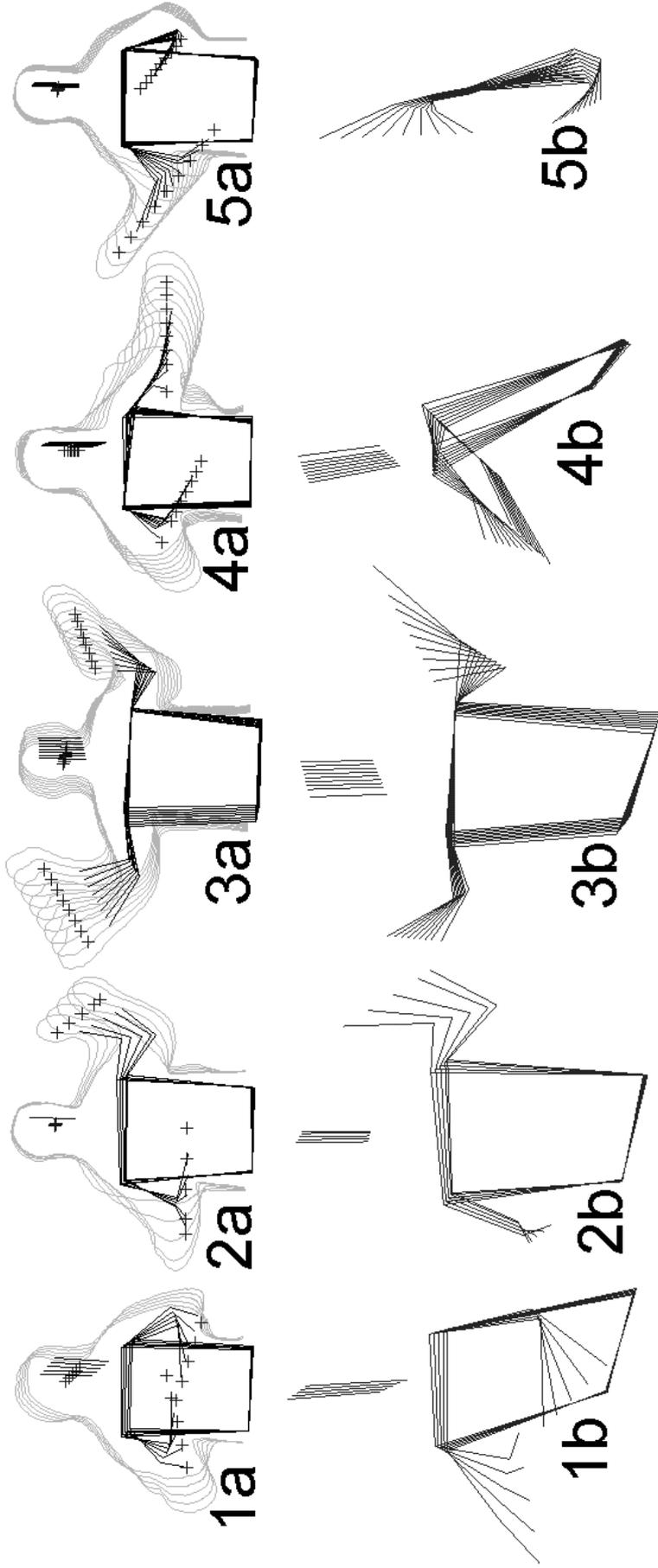


Figure 9.3.7 - How the Model Deforms

Figure 10.3.7 demonstrates the correspondence between the body contour and skeletal structure. Each contour image (a) is generated from a different sub cluster of shape space. The deformation corresponds to the largest mode of deformation for that cluster. The 3D skeletal diagrams (b) correspond to the relevant contour (a), and demonstrate the movement of the skeleton. The orientation of these skeletal models has been changed in order to better visualise the movement in 3D. Skeleton (1b) demonstrates the arms moving in the z direction corresponding to the change in contour (1a) around the elbow region. Contour (4a) represents a body leant toward the camera with moving arms. Skeleton 4b shows the corresponding change in the skeleton with the shoulders twisting as the arms move. The Skeleton 5b is a plan view showing the movement of the hands.

All model points move along straight lines due to the linear clusters used to approximate the non-linear shape space. However, all poses of the models are lifelike human silhouettes, demonstrating the CSSPDM's ability at modelling the non-linearity.

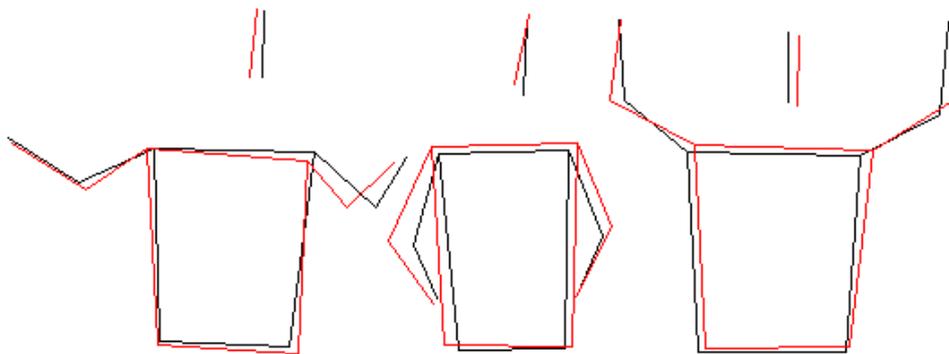


Figure 9.3.8– Reconstructed poses from the model

Figure 9.3.8 shows the original model pose from the training set in red with the reconstructed skeletal model in black. It can be seen that the original and reconstructed models are similar in pose and position with the length of limbs preserved, further demonstrating the absence of non-linear effects. However, as the constraints on shape space are increased, so the performance degrades. Inconsistencies in the original and reconstructed models and the deterioration

under heavy constraints can be attributed to the hand labelling of the training set. During hand labelling it is impossible to provide consistent models of the skeletal structure throughout the training set. This factor leads to the final model producing mean skeletal ‘smoothed’ shapes which have been ‘learnt’ from the original training set and hence produces the inconsistencies observed in figure 1.3.8.

9.4 Conclusion

This section has shown how information can be statistically linked through PCA to produce point distribution models which contain multiple perspectives of data. These perspectives do not have to lie in the same co-ordinate frame and may be related but abstract in nature. By concatenating features, ensuring that incorrect biases do not occur, models can be constructed which not only learn about shape and deformation and how this relates to other aspects of an object, but also to predict these aspects or other missing information from that which is available.

It has been shown how these techniques for statistical inference can be applied to the extraction of 3D structure of an object, given only a monoscopic view of its outline. The technique uses computationally inexpensive techniques for real time tracking and reconstruction of objects. It has also been shown how two sources of information can be combined to provide a direct mapping between them. Being able to reconstruct 3D pose of a human from a simple contour has applications in surveillance, virtual reality and smart room technology and could possibly provide an inexpensive solution to more complex motion capture modalities such as electromagnetic sensors and marker based vision systems.