ELSEVIER

Image and Vision Computing 18 (2000) 729–737

image AND vision COMPUTING

www.elsevier.com/locate/imavis

Non-linear statistical models for the 3D reconstruction of human pose and motion from monocular image sequences

R. Bowden*, T.A. Mitchell, M. Sarhadi

Brunel University, Uxbridge, Middlesex UB8 3PH, UK

Received 9 December 1998; received in revised form 23 August 1999; accepted 25 October 1999

Abstract

This paper presents a model based approach to human body tracking in which the 2D silhouette of a moving human and the corresponding 3D skeletal structure are encapsulated within a non-linear point distribution model. This statistical model allows a direct mapping to be achieved between the external boundary of a human and the anatomical position. It is shown how this information, along with the position of landmark features such as the hands and head can be used to reconstruct information about the pose and structure of the human body from a monocular view of a scene. © 2000 Elsevier Science B.V. All rights reserved.

Keywords: Human body tracking; Non-linear point distribution model; Statistical model; Pose reconstruction

1. Introduction

The human vision system is adept at recognising the position and pose of an object, even when presented with a monocular 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 information, encapsulated within a point distribution model (PDM) [5] can be used to locate and track a body. 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 work investigates the feasibility of such an approach to the reconstruction of 3D structure from a single camera view. 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 within the image plane.

2. Point distribution models

PDMs have proven themselves an invaluable tool in image processing. The *classic formulation* combines local edge feature detection and a model-based approach to provide a fast, simple method of representing an object and how its structure can deform. For a 2D contour, each pose of the object is described by a vector $\mathbf{x}_i \in \mathbb{R}^{2n} = (x_1, y_1, \ldots, x_n, y_n)^T$, representing a set of points specifying the object shape. A training **E** with a set of N vectors is then assembled for a particular model class. The training set is aligned (using translation, rotation and scaling) and the mean shape calculated. To represent the deviation within the shape of the training set, principal component analysis (PCA) is performed on the deviation of the example vectors from the mean, using eigenvector decomposition on the covariance matrix **S** of **E** [5] where

$$\mathbf{S} = \frac{1}{N}\sum_{i=1}^{N}(\mathbf{x}_i - \bar{\mathbf{x}})(\mathbf{x}_i - \bar{\mathbf{x}})^{\mathrm{T}}$$

and

$$\bar{\mathbf{x}} = \frac{1}{N}\sum_{i=1}^{N}\mathbf{x}_i$$

PCA projects the data into a linear subspace with a minimum loss of information by multiplying the data by the eigenvectors of the covariance matrix (**S**). By analysing the magnitude of the corresponding eigenvalues, the

* Corresponding author.
E-mail address: richard.bowden@brunel.ac.uk (R. Bowden)

PII: S0262-8856(99)00076-1

minimum dimensionality of the space on which the data lies can be calculated and the information loss estimated [2].

The t unit eigenvectors of $\mathbf{S}$ corresponding to the t largest eigenvalues supply the variation modes; t will generally be much smaller than N, thus giving a very compact model and it is this dimensional reduction that will facilitate non-linear analysis. A deformed shape $\mathbf{x}$ is generated by adding weighted combinations of v_j to the mean shape

$$\mathbf{x} = \bar{\mathbf{x}} + \sum_{j=1}^{t} b_j \mathbf{v}_j$$

where b_j is the weighting for the jth variation vector. Suitable limits for b_j are $\pm 3\sqrt{\lambda_j}$, whereλ_j is the jth largest eigenvalue of $\mathbf{S}$ [13]. This provides a compact mathematical model on how the shape deforms.

The formulation of the PDM can also be expressed in matrix form [5]:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb}$$

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

This mathematical model is used to constrain the shape of the PDM when applied to an image. To locate an object, a contour is placed near to the desired feature in the image plane. The fitting process is an iterative one, where by the contour makes small steps within the image to find a natural resting place. The model uses suggested movements from control points (using edge detection or grey level matching). Movement of the model is then allowed through the relocation of the model within the image plane using translation, rotation, and scaling. Deformation of the model is also permitted by finding the closest allowable shape as determined by the bounds of the mathematical model of deformation. Given a new shape $\mathbf{x}'$, the closest allowable shape from the model is constructed by finding $\mathbf{b}$ such that

$$\mathbf{b} = \mathbf{P}^{\mathrm{T}}(\mathbf{x}' - \bar{\mathbf{x}}) \text{ since } \mathbf{P}^{\mathrm{T}} = \mathbf{P}^{-1}, \text{ and } -3\sqrt{\lambda_i} \le b_i \le 3\sqrt{\lambda_i}.$$

The closest allowable shape can then be reconstructed as

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb}.$$

2.1. Linear principal component analysis

Typically PDMs are generated by hand to ensure the correct identification of landmark points around a contour. Automated techniques are common place, but often require user intervention in the identification of landmarks to reduce the non-linearity introduced by non-optimum landmark point assignment. Further, non-linearities are often present within a model where parts of the contour display rotational characteristics in the image plane around some pivotal point. Non-linear models when represented by the linear mathematics of PCA manifest themselves as unrobust models, which allow deformation not presented within the original dataset [2,9,10]. These problems become more acute when the move to 3D is considered.

It has been proposed by Kotcheff and Taylor [12] that non-linearity introduced during assembly of a training set could be eliminated by automatically assigning landmark points in order to minimise the non-linearity of the corresponding training cluster. This can be estimated by analysing the size of the linear PDM that represents the training set. The more non-linear a proposed formulation of training set, the larger the PDM needed to encompass the deformation. This was demonstrated using a small test shape and scoring a particular assignment of landmark points according to the size of the training set (gained from analysis of the principal modes and the extent to which the model deforms along these modes, i.e. the eigenvalues of the covariance matrix [12]). This was formulated as a minimisation problem, using a genetic algorithm. The approach performed well, however at a heavy computation cost.

As the move to larger, more complex or 3D models, is considered, where dimensionality of the training set is high, this approach becomes unfeasible. A more generic solution is to use accurate non-linear representations and recent years has seen a wealth of publications on this subject (discussed in the next section).

2.2. Non-linear principal component analysis

Sozou et al. [9] first proposed using polynomial regression to fit high order polynomials to the non-linear axis of the training set. Although this compensates for some of the curvature represented within the training set, it does not adequately compensate for high order non-linearity which manifests itself in the smaller modes of variation as high frequency oscillations of shape or non-linearity which produces discontinues shape spaces. In addition, the order of the polynomial to be used must be selected and the fitting process is time consuming.

Sozou et al. [10] further proposed modelling the non-linearity of the training set using a backpropagation neural network to perform non-linear principal component analysis. This performs well, however, the architecture of the network is application specific, and also the training times and the optimisation of network structure are time consuming.

Heap and Hogg [6] suggested using a log polar mapping to remove non-linearity from the training set. This allows a non-linear training set to be projected into a linear space, where PCA can be used to represent deformation. The model is then projected back into the original space. Although a useful suggestion for applications where the only non-linearity is pivotal and represented in the paths of the landmark points in the image plane, it does not provide a solution for the heavier non-linearity generated by other sources. What is required is a means of modelling the non-linearity accurately, but with the simplicity and speed of the linear model.

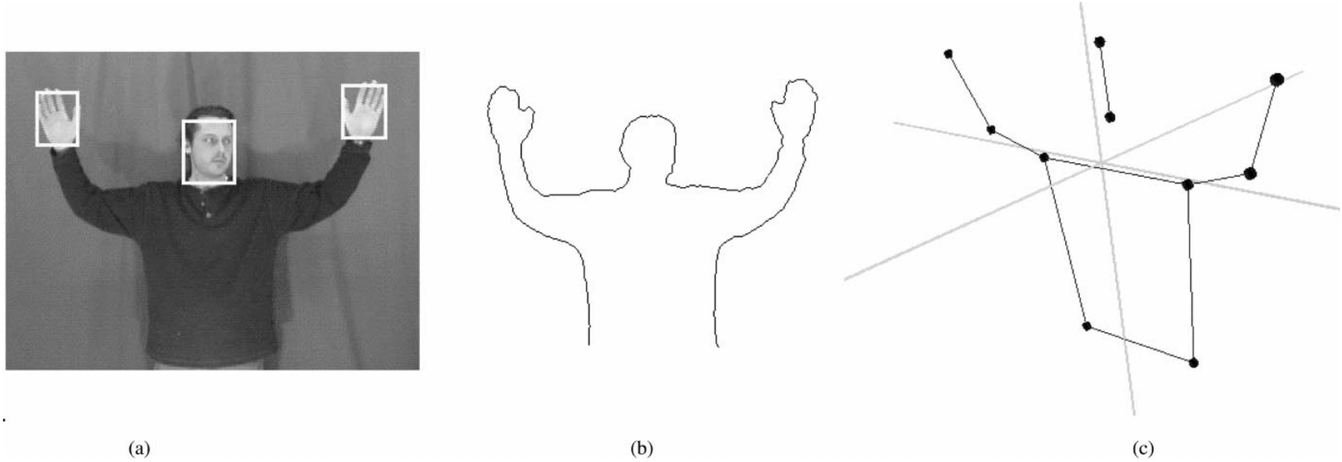

Fig. 1. (a) Position of head and hands V_H. (b) Body contour V_C. (c) Corresponding 3D model V_S.

Several researchers have proposed alternatives, which utilise non-linear approximations, estimating non-linearity through the combination of multiple smaller linear models [2–4,7]. These approaches have been shown to be powerful at modelling complex non-linearity in extremely high dimensional feature spaces [2]. It has also been proposed that analysis of the path taken by the training set through the non-linear feature space can be used to provide a probabilistic model of how the model varies through shape space [7].

3. Building 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 Fig. 1). Each of these is 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 $V_H = (x_1, y_1, \ldots x_3, y_3)$. 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 $V_C = (x_1, y_1, \ldots x_{400}, y_{400})$. This is far larger than necessary, but the dimensionality will be reduced prior to non-linear analysis. 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 from occlusion and cluttered backgrounds. It can also be used to estimate an initial location and shape for the body contour.

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 axis of S. This can be done by minimising the eigen entropy as proposed by Sumpter et al. [11]. However, as all three components exist within the same co-ordinate frame and are directly linked, such scaling is unnecessary.

3.1. Hand and head position estimation

Colour information is an important attribute within an image, which is often discarded. McKenna, Gong and Raja have demonstrated that in a Hue–Saturation space, human skin occupies a relatively small cluster and can be used to segment a human head from a complex noisy scene [8]. Using a gaussian mixture model to represent this colour space they have shown how multiple models for individuals can be used to probabilistically label an image and determine the most likely person present. Azarbayejani and Pentland [1] have used similar methods to automatically segment both the hands and head from stereo image pairs, and using this, calculate their position and trajectories in 3D space.

By taking example images of human skin and plotting the constituent points in hue saturation space, a single gaussian cluster is sufficient to provide a reliable probabilistic measure for the location of skin clusters in the image frame. By performing PCA on this colour cluster approximate bounds are calculated. If a sample pixel from a new image is within the Hue-Saturation bounds of the gaussian cluster then that pixel is marked as a probable location, producing a binary image. By performing erosion then dilation, noisy points are removed, and clusters of marked skin points consolidated to blobs. A simple blobbing algorithm is then used to calculate approximate locations of skin artefacts within the image.

Fig. 2 shows a sample image frame after processing. The results from the blobbing algorithm are used to calculate the centre and approximate size of the skin artefacts. This is used to place a cross over the segmented features for demonstration purposes. This procedure is repeated for each image in the training sequence to extract the trajectories of the head and hand as the human moves.

3.2. Shape extraction

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 '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 simple 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° of a vertical line drawn from the centroid of the shape. Two further points are positioned at the horizontal extremities of the contour. These landmarks are then used to resample the contour to 400 points. This simple landmark point identification results in non-linearity within the training set. The problems associated with this are discussed in the section on non-linear estimation.

3.3. Introducing 3D information

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.

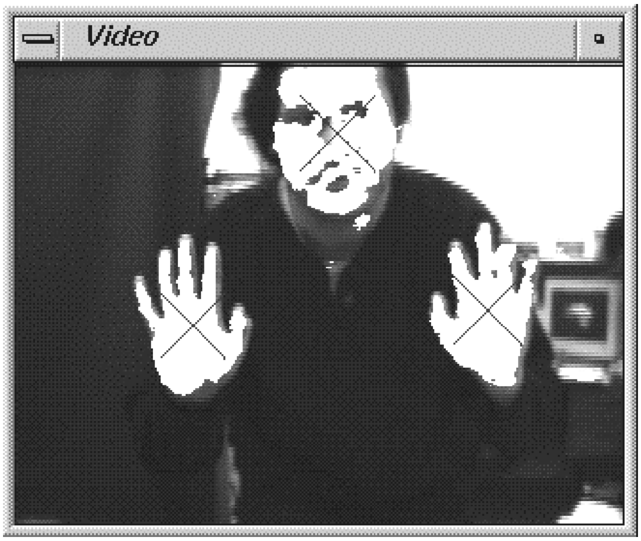

Fig. 2. Blobs of skin artefacts.

3.4. The linear PDM

Once these separate feature vectors are extracted, they are concatenated to form an 836-dimensional vector that represents the total pose of the model. A training set of these vectors is assembled which represents the likely movement of the model. Fig. 3 shows a sample of training images along with the corresponding contour and skeletal models in two dimensions.

A linear PDM is now constructed [5] from the training set and its primary modes of variation are shown in Fig. 4.

Fig. 4 demonstrates the deformation of the composite PDM. The crosses are the locations of the hands and head. It can be seen that although the movements of the three elements are closely related, the model does not accurately represent the natural deformation of the body. As the model approaches the outer bounds of the modes the linearity of the variation causes the separation of the constituent elements. The possible shapes generated by the combination of 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.

3.5. Non-linear estimation

Any analysis performed upon this training set is time consuming due to its high dimensionality and size. This dimensionality is, therefore reduced by projecting the dataset down into a lower dimensional space while preserving the important information, its shape. The linear PDM model, which is unsuitable for modelling the deformation, is invaluable in this sense. Although the Linear PDM contains deformation uncharacteristic to a human, it is capable of representing the original deformation contained in the training set in a more compact form [2].

After PCA, it is calculated that the first 84 eigenvectors that corresponding to the 84 largest eigenvalues, encompass 99.99% of the deformation contained in the training set. By projecting each of the sample shapes down onto these vectors, and recording their distance from the mean, a new vector $\mathbf{r}_i \in \mathbb{R}^{84} = (e_1, \ldots, e_{84})$, $i = 1, \ldots, N$ is constructed, where

$$e_j = \mathbf{v}_j \cdot (\mathbf{x} - \bar{\mathbf{x}})$$

or alternatively using the first 84 eigenvectors from **P** where

$$\mathbf{r}_i = \mathbf{P}^{\mathrm{T}}(\mathbf{x}_i - \bar{\mathbf{x}}), \quad i = 1 \ldots N$$

From this vector it is a simple back projection to reconstruct the model. Upon visual observation of the original vector and the reconstructed vector it can be seen that only the first 40 eigenvectors are necessary to represent the shape accurately. These 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 from 836 to 40, which significantly reduces

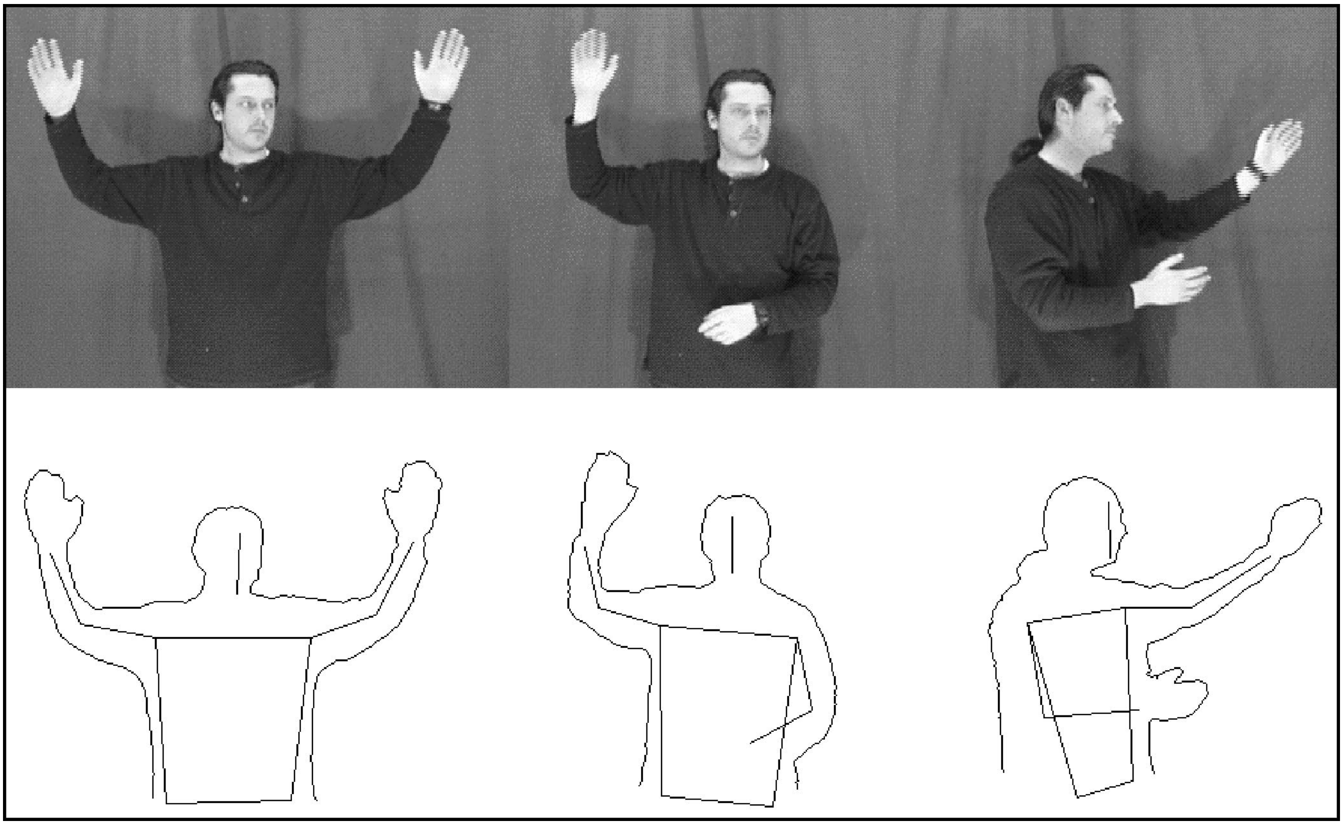

Fig. 3. Sample training images, corresponding contour and skeletal models.

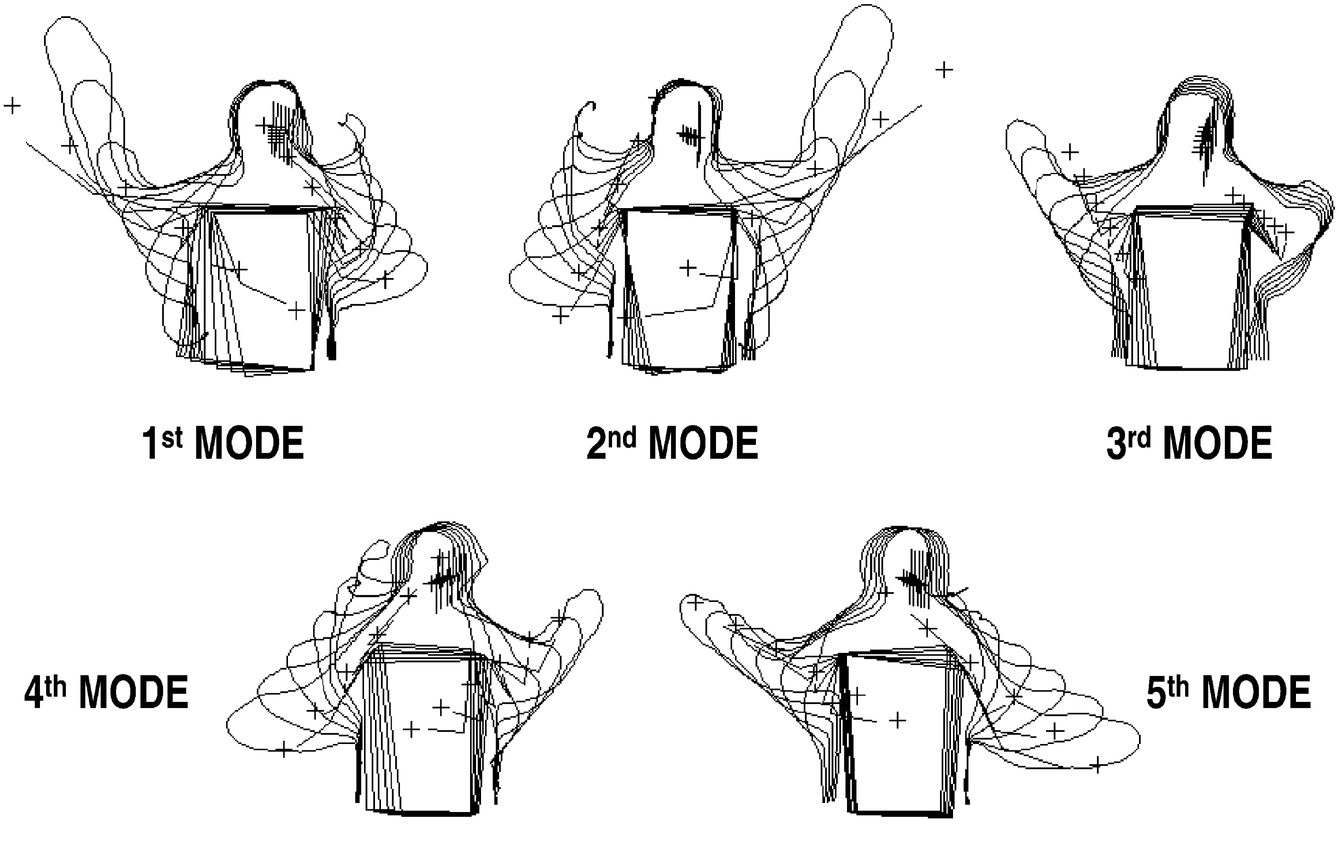

Fig. 4. Primary modes of variation on the linear PDM.

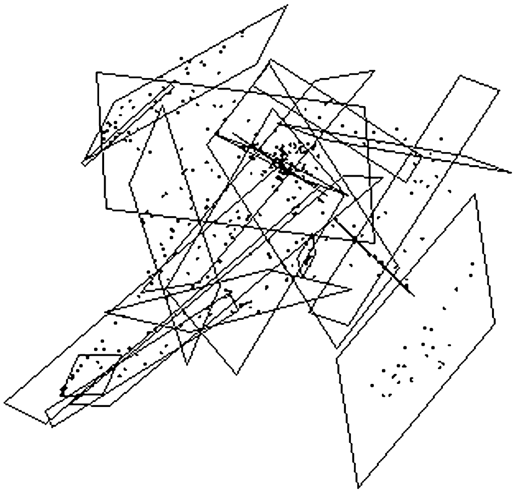

Fig. 5. Clusters in reduced shape space.

the computation time required for further analysis. This vast reduction in dimensionality is a direct consequence of the large number of samples from the body contour, sampling the contour at fewer points would produce equally effective results. However, even with a lower initial sampling the training set will still contain redundant dimensionality that can be discarded thorough this projection (as this is the premise of the linear PDM [5]).

In this lower dimensional space the information about the shape of the training set and how the model moves throughout it is preserved allowing further statistical analysis. Using a *k*-means clustering algorithm, the space can be segregated into sub areas, which estimate the non-linearity [2]. Using standard cluster analysis the natural number of clusters can be 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 hyperdimensional volume.

Fig. 5 shows the training set after dimensional reduction using the linear PDM, projected into two 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 of the sub-PCA models. The skew appearance of these bounds is due to the visual projection of the 40-dimensional space down to just two dimensions. 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 modelled to encompass 99.9% of the deformation within that cluster.

4. Applying the PDM to an image

4.1. Initialising the PDM

Upon initialisation the first step is to locate the position of the head and hands. This can be done via the procedure described earlier, which although is computationally expensive, need not be repeated during every iteration. Once this is 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, nine possible contours can be produced. If the highest blob is assumed to be the head then this is simplified to two possible contours. The contour that iterates to the best solution provides the initial state from which tracking proceeds.

4.2. Tracking with the PDM

Once initialised the components must be fitted to the image separately. The contour is attracted to high intensity

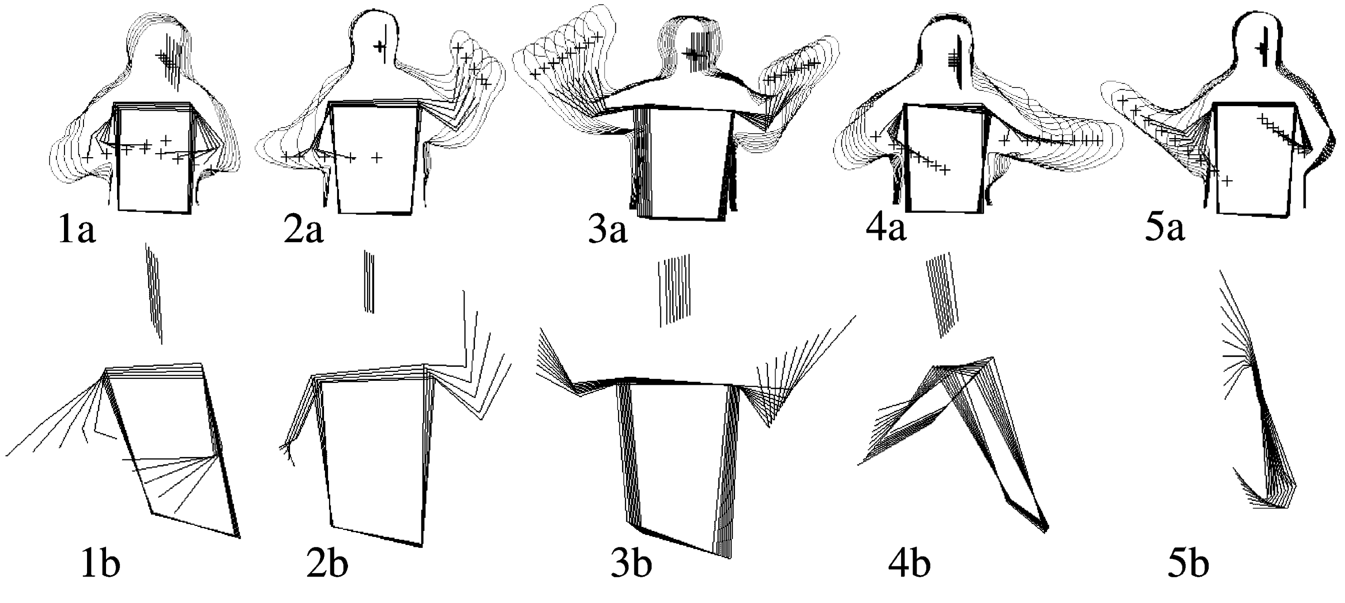

Fig. 6. How the model deforms.

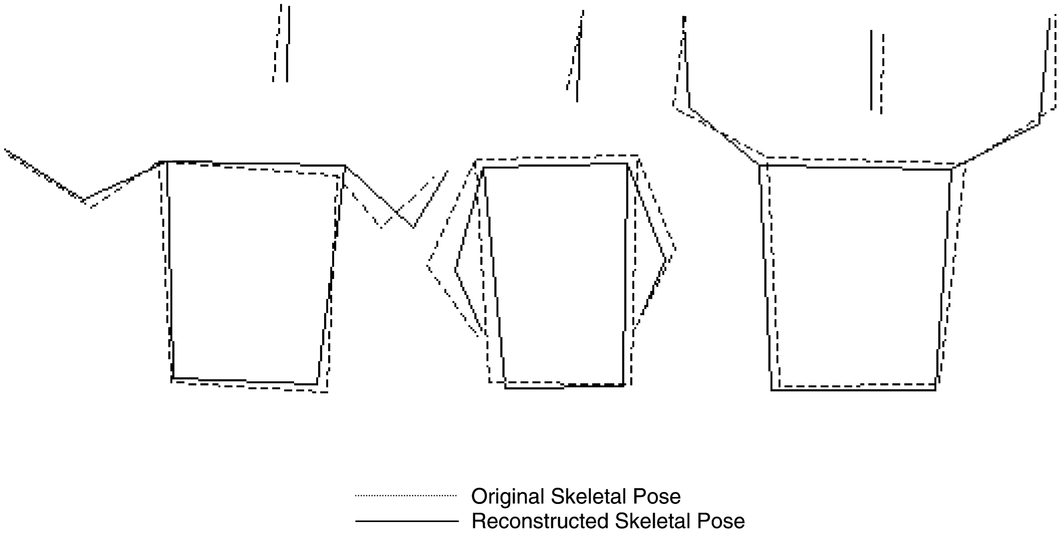

Fig. 7. Reconstructed poses from the model.

gradients within the image using local edge detection. A local search is made along the normal to the contour at each key point. The contour is then moved to the highest intensity edge located along that normal.

The hand and head positions are used as centres in a single iteration of a *k*-means-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. This nearest neighbour approach to moving the location of the head and hands provides robustness to noise generated by the colour thresholding process.

Once both the contour and the head and hands have been moved to the their next position in the image frame, the closest point within the PDM is located. This is done, by projecting the model down into the reduced space, and finding the closest cluster to that point. Once done, the closest allowable shape is located within the bounds of that cluster (as modelled with PCA). The resulting closest shape is then projected back up into the original dimensionality and used as the start for the next iteration of the procedure.

4.3. Reconstruction of 3D shape and pose

As the shape deforms to fit with the image, the third element of the model, the skeleton, also deforms. As the non-linear constraints are applied to the model (by finding the nearest allowable shape), the skeletal element of the model is 'pulled or dragged' through the PDM space towards its new position. By plotting this 3D skeleton, its movements mimic the motion of the human in the image frame.

Fig. 6 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

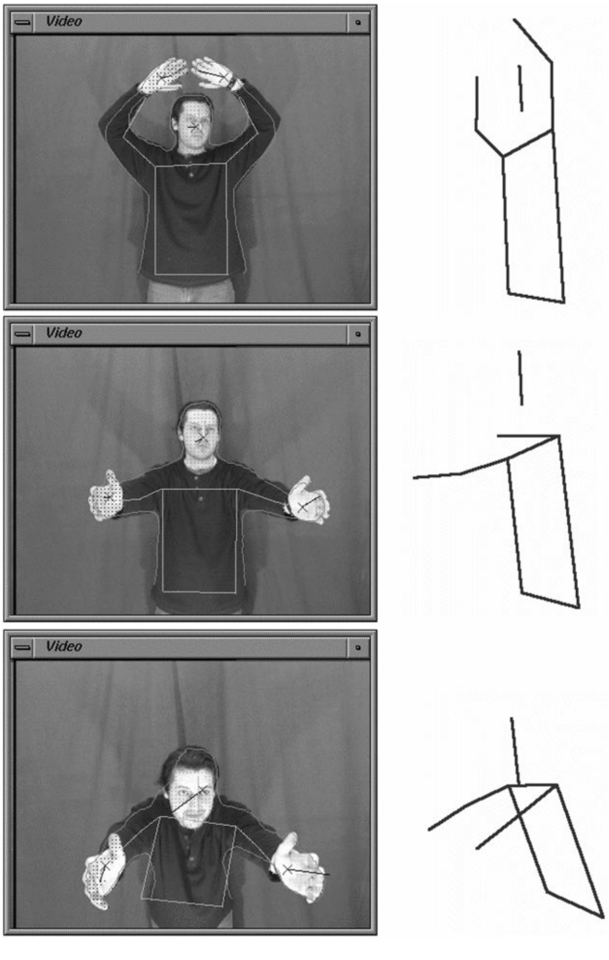

Fig. 8. Tracking the human body and reconstructing pose.

for that cluster. The 3D skeletal diagram (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 visualise better 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. No constraints are placed upon the movement of the skeletal model to ensure joints conform to physical constraints (i.e. elbow joints bending the correct way) as this is unnecessary. As the training set consisted of only valid deformation for the body, the PDM has generalised this information and cannot produce invalid movement of the skeleton (this is only true for the constrained non-linear model).

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 cluster based non-linear principal component analysis ability at modelling the non-linearity.

Fig. 7 shows the original model pose from the training set in grey (dashed) 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. The PDM generalises these inconsistencies and smooth errors introduced during training. This factor leads to the final model producing mean skeletal shapes which have been 'learnt' from the original training set and hence produces the inconsistencies observed in Fig. 7.

Fig. 8 shows the non-linear PDM tracking the human body along with the corresponding reconstructed skeletal structure. The first frame shows the human raising his hands above his head and this is mimicked by the reconstructed 3D stick model. As the hands are brought down and forward so the 3D model reflects this. Finally, as the human leans forward, the corresponding movement of the reconstructed model can be seen. Although this does not provide evidence of the accuracy of the technique, it clearly demonstrates that the model is capable of inferring 3D information from a monocular image for complex articulated objects.

5. Conclusions

This paper has demonstrated how the 3D structure of an object can be reconstructed from a single view of its outline, using an internal model of shape and movement. The approach 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. The model appears to reconstruct 3D pose accurately. However, due to the acquisition of skeletal position, no ground truth information is available to test its accuracy. The next stage of this work is to assess this accuracy. Being able to reconstruct 3D pose 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.

6. Future work

A full body model needs to be constructed for the generic tracking and pose estimation of humans. During its construction accurate skeletal information must be acquired to ensure usability of the resulting model. Key point trajectories will be acquired using electromagnetic sensors to provide training information during image acquisition for contour extraction. This will also provide the necessary ground truth data required, to assess, the accuracy of the final model. The incorporation of temporal constraints on the model will also increase reliability and robustness. It is also possible to implement a stereoscopic or quad camera system to increase accuracy and further disambiguate the contour fitting process. Inclusion of information about the orientation of the skeleton added during training would allow the estimation of contour shape in other image frames from a multi-camera system. Alternatively multiple 2D projections of the contour in different image plans could be concatenated with a single 3D model to provide a direct mapping between all the elements.

Acknowledgements

This work was funded by an EPSRC case studentship in conjunction with Foster Findlay Associates Ltd.

References

[1] A. Azarbayejani, A. Penland, Real-time self calibrating stereo person tracking using 3D shape estimation from blob features, ICPR'96, Vienna, Austria, 1996.

[2] R. Bowden, T.A. Mitchell, M. Sahardi, Cluster based non-linear principal component analysis, IEE Electronics Letters 33 (22) (1997) 1858–1859.

[3] C. Bregler, S. Omohundro, Surface learning with applications to lip reading, Advances in Neural Information Processing Systems, 6, 1994.

[4] T.F. Cootes, C.J. Taylor, A mixture model for representing shape variation, in: A.F. Clark (Ed.), British Machine Vision Conference

1997, British Machine Vision Association, Essex, UK, 1997, pp. 110–119.

[5] T.F. Cootes, C.J. Taylor, D.H. Cooper, J. Graham, Active shape models—their training and application, Computer Vision and Image Understanding 61 (1) (1995) 38–59.

[6] T. Heap, D.C. Hogg, Automated pivot location for the cartesian–polar hybrid point distribution model, in: D. Pycock (Ed.), British Machine Vision Conference 1995, British Machine Vision Association, Birmingham, UK, 1995, pp. 97–106.

[7] T. Heap, D.C. Hogg, Improving specificity in PDMS using a hierarchical approach, in: A.F. Clark (Ed.), British Machine Vision Conference 1997, British Machine Vision Association, Essex, UK, 1997, pp. 80–89.

[8] S. McKenna, G. Gong, Y. Raja, Face recognition in dynamic scenes, in: A.F. Clark (Ed.), British Machine Vision Conference 1997, British Machine Vision Association, Essex, UK, 1997, pp. 140–151.

[9] P.D. Sozou, T.F. Cootes, C.J. Taylor, E.C. Di-Mauro, A non-linear generalisation of pdms using polynomial regression, in: E. Hancock (Ed.), British Machine Vision Conference 1994, British Machine Vision Association, York, 1994, pp. 397–406.

[10] P.D. Sozou, T.F. Cootes, C.J. Taylor, E.C. Di Mauro, Non-linear point distribution modelling using a multi-layer perceptron, in: D. Pycock (Ed.), British Machine Vision Conference 1995, British Machine Vision Association, Birmingham, UK, 1995, pp. 107–116.

[11] N. Sumpter, R.D. Boyle, R.D. Tillett, Modelling collective animal behaviour using extended point distribution models, in: A.F. Clark (Ed.), British Machine Vision Conference 1997, British Machine Vision Association, Essex, UK, 1997, pp. 242–251.

[12] A.C.W. Kotcheff, C.J. Taylor, Automatic Construction of Eigenshape Models by Gentic Algorithms, Lecture Notes in Computer Science, 1230 1997 pp. 1–14.

[13] R. Bowden, A.J. Heap, D.C. Hogg, Real time hand tracking and gesture recognition as a 3D input device for graphical applications, in: P.A. Harling, A.D.N. Edwards (Eds.), Progress in Gestural Interaction, Springer, London, 1997, pp. 117–129.