Acquisition of Articulated Human Body Models Using Multiple Cameras

Aravind Sundaresan and Rama Chellappa*

Center for Automation Research, Dept. of Electrical and Computer Engineering
University of Maryland, College Park, MD 20742-3275, USA
{aravinds, rama}@cfar.umd.edu

Abstract. Motion capture is an important application in different areas such as biomechanics, computer animation, and human-computer interaction. Current motion capture methods typically use human body models in order to guide pose estimation and tracking. We model the human body as a set of tapered super-quadrics connected in an articulated structure and propose an algorithm to automatically estimate the parameters of the model using video sequences obtained from multiple calibrated cameras. Our method is based on the fact that the human body is constructed of several articulated chains that can be visualised as essentially 1-D segments embedded in 3-D space and connected at specific joint locations. The proposed method first computes a voxel representation from the images and maps the voxels to a high dimensional space in order to extract the 1-D structure. A bottom-up approach is then suggested in order to build a parametric (spline-based) representation of a general articulated body in the high dimensional space followed by a top-down probabilistic approach that registers the segments to the known human body model. We then present an algorithm to estimate the parameters of our model using the segmented and registered voxels.

1 Introduction

The task of motion capture can be divided into a number of systematically distinct stages: initialisation, pose estimation and tracking. There exist a number of algorithms to estimate the pose using images captured from a single or multiple cameras [1]. Some of the problems encountered, especially in the monocular case, are the segmentation of the image into different, possibly self-occluding body parts and the complex articulated structure of the human body which results in wide range of body part configurations or poses. It is, therefore, often necessary to use a human body model to deal with the large number of body segments and to guide the tracking and pose estimation processes especially in bio-mechanical and clinical motion capture applications.

Krahnstoever and Sharma [2] address the issue of acquiring structure, shape and appearance of articulated models directly from monocular video using a

* This research was funded in part by NSF ITR 0325715.

F.J. Perales and R.B. Fisher (Eds.): AMDO 2006, LNCS 4069, pp. 78–89, 2006.

single camera and hence has has limited scope for complete human body model estimation. Mikic et al. [3] propose a model acquisition algorithm using voxels that starts with a simple body part localisation procedure based on template fitting and growing, and uses prior knowledge of average body part shapes and dimensions. Kakadiaris and Metaxas [4] present a Human Body Part Identification Strategy (HBPIS) that recovers all the body parts of a moving human based on the spatio-temporal analysis of its deforming silhouette using input from three mutually orthogonal views. The subject, however is required to follow a specified protocol of movements. Anguelov et al. [5] describe an algorithm that automatically decomposes simple objects into approximately rigid parts and obtains the underlying articulated structure given a set of meshes describing the objects in different poses. Cheung et al. [6] also describes a model acquisition algorithm where the kinematics is estimated using correspondence. Chu et al. [7] describe a method for estimating pose using isomaps [8] to transform the voxel body to its pose-invariant intrinsic space representation and obtain a skeleton representation.

We model the human body as comprising of several rigid body segments that are connected to each other at specific joints forming 1-D kinematic chains originating from the trunk as described in Section 2. These chains can be visualised as 1-D curves embedded in 3-D space. We exploit the 1-D nature of the chains and transform the voxel coordinates to a domain where we are able to extract the 1-D structure. We are thus able to register each voxel to its position along the chain for a set of frames that capture the subject in different poses. The model estimation algorithm involves locating the joint locations and estimating the shape parameters of the different body segments as well as the implicit estimation of the pose. We first estimate the joint locations and limb lengths from the skeletons and then compute the super-quadric parameters of the body segments from the voxels using the segmentation and registration results. While human dimensional variability is enormous across different demographics and sexes, it is not arbitrary. We can, therefore, use our prior knowledge of the approximate ratios between the stature and different long bones, as well as the joint location in our model acquisition algorithm. We describe the model estimation algorithm in Section 3, and the experiments in Section 4. Our algorithm is different from that of Chu et al. [7], in that we use Laplacian eigenmaps [9] in order to simultaneously segment and extract the one-dimensional structure of the human body. Belkin and Niyogi [9] describe the construction of a representation for data lying in a low dimensional manifold embedded in a high dimensional space. We obtain much better segmentation and explicitly compute the position of each voxel along the articulated chain that it belongs to. This step enables us to acquire the shape and joint model. Some other techniques for dimensionality reduction and reducing shape to pose invariant structure can be found in Elad and Kimmel [10], manifold charting [11] and Locally Linear Embedding [12]. However, we choose Laplacian eigenmaps as they best serves the purpose of extracting the 1-D nature of the curves. There is also a similarity to skeletal representation algorithms [13] that we expound on in Section 4.

2 Human Body Model

The human body model that we use is illustrated in Fig. 1 (a) with the different body segments as well as joints labelled. Each of these body segments has a coordinate frame attached to itself. The body segment can be described by an arbitrary shape in terms of the coordinates of this frame, and in our case is modelled using a tapered super-quadric. We choose tapered super-quadrics for their simplicity and versatility [14]. Some of the other shape models used to model human body segments are cylinders, CAD models and ellipsoids [3]. The tapered super-quadric (Fig. 10a) is described in equation (1), and is characterised by five scalar parameters x_0, y_0, z_0, d, and s. If sliced in a plane parallel to the xy plane, the cross section is an ellipse with parameters αx_0 and αy_0, where α is a scalar (Fig. 1 (c)). The length of the segment is z_0 as shown in Fig. 1 (d). The scale parameter, s, denotes the amount of taper, and the "power" parameter, d, denotes the curvature of the radial profile, $r(z)\sqrt{x_0 y_0}$, along the z-axis. For e.g., $d = 2, s = 0$, is an ellipsoid, $d = \infty, s = 0$ is a right-elliptical cylinder and $d = \infty, s = -1$ is a right-elliptical cone.

$$\left(\frac{x}{x_0}\right)^2 + \left(\frac{y}{y_0}\right)^2 = \left(1 + s\frac{z}{z_0}\right)\left(1 - \left(1 - 2\frac{z}{z_0}\right)^d\right) = r^2(z) \text{ for } 0 \leq z \leq z_0 \quad (1)$$

A joint between two body segments is described as a vector in the coordinate frame of the parent body segment connecting the origin of the parent segment coordinate frame to the origin of the child segment. The pose of the child segment is described in terms of the rotational parameters between the child coordinate frame and the parent coordinate frame. The pose of the model, $\boldsymbol{\varphi}$, is a vector of the position and orientation of of the base-body (6 degrees of freedom) and the joint angles of the various articulated body segments (3 degrees of freedom for each joint). We observe that the joint locations cannot be easily obtained, even manually, from a single pose.

3 Model Acquisition Algorithm

We begin with grey-scaled images captured from multiple cameras. Simple background subtraction is performed on the images to obtain binary silhouettes (Fig. 2). We perform space carving using the binary silhouettes from the cameras and the calibration data to obtain a voxel representation where each voxel block is of size 30mm$\times$ 30mm$\times$ 30mm, which we find to be an acceptable compromise between complexity and accuracy. In the first part of the algorithm, we segment the voxels and obtain a parametric representation for the different articulated chains as well as register the chain to the body model. We then compute a skeletal representation of the subject for a set of key frames where the registration is successful. In the second part of our model acquisition algorithm, we estimate a simple stick model for the subject and progressively improve the model to finally obtain the parameters of the complete super-quadric-based model.

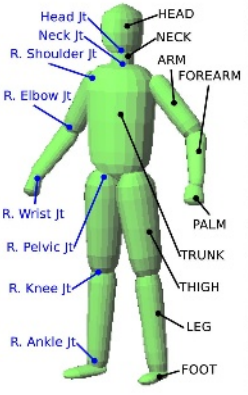


Fig. 1. Model

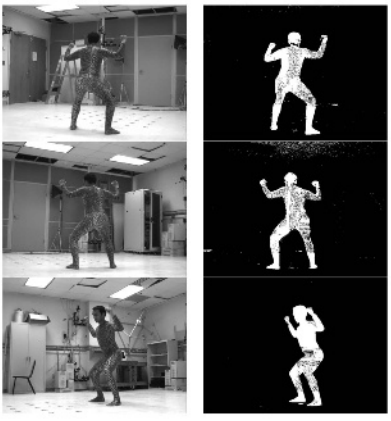


Fig. 2. Images and Silhouettes

Fig. 3. Voxels

3.1 Segmenting and Registering the Articulated Chains

Our key observation is that the human body can be visualised as consisting of 1-D continuous articulated chains embedded in 3-D space. We observe this in Fig. 3, in which we can identify the five articulated chains: the head and four limbs, attached to the trunk, the sixth segment. Our objective is to extract the 1-D structure and the position of each voxel along the chain using a parametric form, and thus segment the different articulated chains that have these 1-D structure. The articulated structure of these chains, however, make it difficult to segment them in normal 3-D space. We use Laplacian Eigenmaps to extract the structure of the underlying 1-D curve. Our objective is *not* to preserve geodesic distances between points [7] or reduce the dimensionality of the data [9], but to extract the one-dimensional manifold structure.

It is known that the Laplacian Eigenmap preserves local information optimally in a certain sense as described in Belkin and Niyogi [9]. Given a data set of k nodes (voxel coordinates), we construct a weighted graph $G = (V, E)$, with edges connecting two nodes if they are neighbours. We consider the problem of mapping the weighted graph to a $k \times m$ matrix $Y = [\boldsymbol{y}^{(1)}, \ldots, \boldsymbol{y}^{(m)}] = [\boldsymbol{y}_1, \ldots, \boldsymbol{y}_k]^\mathsf{T}$, where the i^{th} row, $\boldsymbol{y}_i^\mathsf{T}$, provides the embedding for the i^{th} node. A reasonable criterion for choosing a "good" map is to minimise, under appropriate conditions, the objective function given by $\sum_{i,j} \|\boldsymbol{y}_i - \boldsymbol{y}_j\|^2 W_{ij}$ (which imposes a penalty if vertices connected by an edge are not close to each other) subject to $Y^\mathsf{T} Y = I$ (which removes an arbitrary scaling factor). Standard methods show that the solution is provided by the matrix of eigenvectors corresponding to the k lowest non-zero eigenvalues of the generalised eigenvalue problem $L\boldsymbol{y} = \lambda \boldsymbol{y}$, where $W_i = \sum_j W_{ij} = \sum_j W_{ji}$, and $D = \text{diag}(W_1, W_2, \ldots, W_k)$, and $L = D - W$.

The concept of neighbours is natural because the voxels are positioned in a uniform spatial grid and constructing the adjacency graph is intuitive. We place an edge between two voxels if they are neighbours connected by a face, an edge

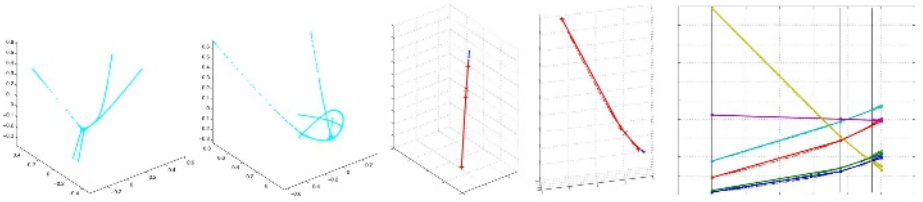
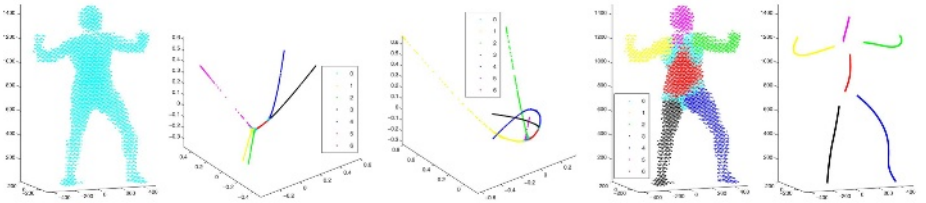


Fig. 4. Extracting the 1-D curves in Eigenspace

or a corner. We thus obtain a sparse graph of size $k \times k$. We only consider voxels that belong to the biggest connected component in the graph. We compute $d = 6$ eigenvectors corresponding to the d smallest non-zero eigenvalues and thus embed the graph in a 6-D Euclidean space. We choose $d = 6$, because we wish to segment six different articulated chains. The graph embedding in the 6-D space is illustrated in Fig. 3.1(a-b). A close examination of the plot reveals six one-dimensional curves that we expect to correspond to the six articulated chains described earlier. We observe that the "bending" effect of articulation has been removed, as we would expect, because joint angles do not in general affect the computation of the adjacency matrix.

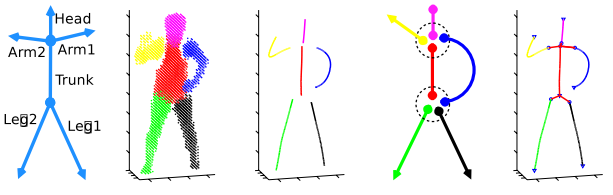
We observe that each articulated chain is an 1-D curve in 6-D space irrespective of the thickness of the body segment in normal 3-D space. This is a result of using the Laplacian eigenmap transformation and we observe that the 1-D nature of the curve is preserved even in higher dimensions (Fig. 3.1 (c)). This is an advantage over geodesic distance preserving algorithms [7] as we can easily fit 1-D splines to the data.

We describe a completely unsupervised algorithm to segment the voxels in eigenspace into 1-D curves. We represent the voxels in terms of an 1-D parameter in this eigenspace by fitting a cubic smoothing spline function to the data according to the following algorithm. All the computations are in the 6-D eigenspace. We begin each spline with a "pivot" node that is farthest from all existing spline segments. There are two kinds of curves, those that are connected at one end and those that are connected at both. The "pivot" node is at the free end or the middle in the first and second cases respectively. We create a cluster by adding nodes that are closest to the "pivot" node. We compute the principal axis for the cluster and the projection of each node on the principal axis (site value t). Thus, for each node, $\boldsymbol{y}_i$, in the cluster we obtain its site value t_i. We can compute a smoothing spline $\boldsymbol{f}(.)$ to minimise $\sum_i e_i$, where $e_i = \|\boldsymbol{f}(t_i) - \boldsymbol{y}_i\|^2$. We grow the curve by adding nodes that are close to each end. The principal axis used to compute the site value is recomputed locally using nodes at the growing end. The growth is terminated when the error of new nodes exceeds a fixed threshold, $CL\sqrt{d}$, where $C = 0.005$, L is the length of the average spline in eigenspace (set to 1 as we have normalised the eigenspace such that $\boldsymbol{y}_i \in [0, 1]^6$) and d is the dimension of the space. We now have six spline segments as shown in Fig. 5(b-c). Fig. 5 (d) presents the segmentation results in the normal 3-D



(a) Normal (b) Eigenspace (1-3) (c) Eigenspace (4-6) (d) Segmented (e) Skeleton

Fig. 5. The splines are colour coded according to their index. (d) and (e) denote the voxels colour-coded according to the index of the spline segments they belong to.



(a) Model (b) Seg. (c) Skeleton (d) Graph (e) Conn.

Fig. 6. Matching computed graph with body model

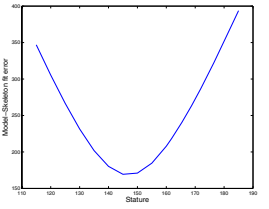


Fig. 7. Model fit error

space and we note that the segmentation is fairly accurate. Unclassified voxels are labelled 0. As noted earlier each node ($\boldsymbol{x}_i$ in normal 3-D space and $\boldsymbol{y}_i$ in eigenspace) has a site value t_i. This value denotes the position of the node along the 1-D curve and can be used to compute the skeleton in Fig. 5 (e) using a 3-D smoothing spline with the set of pairs $(t_i, \boldsymbol{x}_i)$.

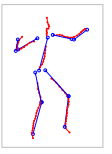
Once we have the skeleton segments as in Fig. 5 (e), we would like to register them to the different segments presented in Fig. 6 (a). Each spline segment consists of a curve connecting two nodes. We can estimate the probability of a connection between nodes of different segments based on the distance between the nodes in eigenspace. We can also estimate the probabilities of a spline segment being an arm or a leg, for example, by examination of the properties of the spline in normal space such as its length and thickness. We choose that permutation of body segments that has the highest probability. In most cases, the registration is straightforward, but however, in poses like in Fig. 6 (b), there are ambiguities. We resolve these ambiguities by selecting that set of connections between segments that has the highest probability (Fig. 6 (e)). We can also identify cases where the number of segments is less than six due to segmentation errors.

3.2 Estimating Human Body Model

The human body model parameters cannot be reliably estimated from a single pose. We, therefore, select a set of $N(= 20)$ key frames from the sequence that




Fig. 8. Fit of initial skeleton model



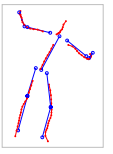
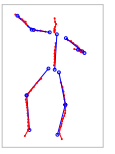
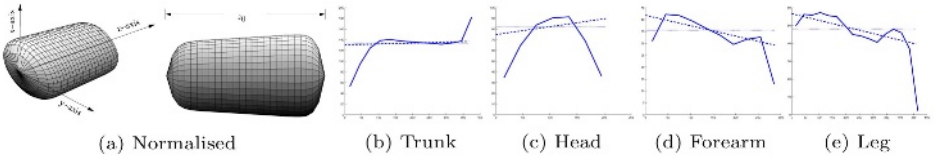



Fig. 9. Optimised model

have been registered successfully. We estimate an initial skeleton-based model from the set of frames and progressively refine model parameters and increase model complexity. For e.g., we begin with a skeleton model and progress to super-quadric model. We use techniques that leverage our knowledge of the human body structure and use a top down approach. The stature (or height) of the subject is a key parameter that is strongly related to a number of human body model parameters, such as the lengths of long bones in the body [15]. Anthropometric studies have been performed on certain demographic groups to study the relationship between stature and the long bones in the body. These studies indicate that we can estimate the lengths of the large body segments for an "average" human subject (model skeleton) from the stature. For our initial model we construct the model skeleton (including joint locations and limb lengths) for an "average" human subject for a range of stature values. We fit line segments, corresponding to the trunk, neck, head, forearms, arms, thighs and leg segments to the voxel-skeleton in Fig. 6 (c), using the known lengths of the body segments. We also obtain an approximate estimate of the pose in the process. We identify the limbs on the left and the right by examination of their positions with respect to the trunk and also by examination of the joint angles between the limbs of the legs. We then compute the distance between the points on the voxel-skeleton and the line segments of the model skeleton obtained from the stature. The skeleton model fit error corresponding to the stature is computed and summed across key frames for each stature value in the range to determine the stature parameter that best fits the voxel skeletons. We note that there is a clear minimum in the error versus stature plot in Fig. 7 and we select the model skeleton corresponding to the minimum error stature value as our initial estimate. The computed skeleton of a few key frames with the model super-imposed on them are presented in Fig. 8. The two sets of parameters we are interested in estimating are the pose parameters (joint angles) and the body structure (joint locations). We can express the fit error as a function of the joint locations ($\boldsymbol{X}$) and the joint angles ($\boldsymbol{\varphi}$). We minimise the fit error by varying $\boldsymbol{X}$ while keeping $\boldsymbol{\varphi}$ fixed, and vice versa (varying $\boldsymbol{\varphi}$ while keeping $\boldsymbol{X}$ fixed), using optimisation techniques. $\boldsymbol{X}$ and $\boldsymbol{\varphi}$ are allowed to vary within a small region around $\boldsymbol{X}_0$ and $\boldsymbol{\varphi}_0$ respectively. The skeletons of a few key frames with the optimised model and pose super-imposed on them are presented in Fig. 9.

The next step is to obtain the super-quadric parameters given the joint locations and angles. We estimate the super-quadric parameters for the trunk, head, arm, forearm, thigh, and leg, as these body segments are large enough to be estimated using the resolution and quality of the voxels that we possess. On any



(a) Normalised (b) Trunk (c) Head (d) Forearm (e) Leg

Fig. 10. Radial profiles of different body segments: The solid line is the median radial profile. The dotted line is the super-quadric radius with scale parameter set to zero. The dashed line is the super-quadric radius with estimated scale parameter. The x-axis of the plots is the distance in mm along the z-axis of the body segment coordinate system. The y-axis of the plots is the radius value also in mm.

articulated chain, we know the position of each voxel along the chain. Using this knowledge and the estimated joint locations, we can segment each articulated chain into the different body segments that make up the chain. Using the estimated joint angles, we can also compute the position of the coordinate frame attached to the body. For a given body segment, we can thus normalise the pose using the body coordinate frame, so that the body segment is positioned at the origin and aligned with the z-axis as in Fig. 10 (a). We compute the area of the cross-section of the voxels, A_z, (plane parallel to the xy-plane) at different points along the z-axis. We assume that the cross-section is elliptical and find the parameters (x_0, y_0) (in (1)), from the area using the relation $A = \pi x_0 y_0$. A circle of equal area would have radius $r = \sqrt{x_0 y_0}$. We compute the radius of the equivalent circle, at different points along the z-axis, as $r_z = \sqrt{A_z/\pi}$, which we refer to as the radial profile (Fig. 10 (a)). We compute the radial profile in all the key frames for each body segment and use the median radial profile. The median radial profiles for some of the body segments are presented in Fig. 10 (b-e). We can compute the length, radius and the scale parameter of the body segment from the radial profile. If we wish to determine the parameters x_0 and y_0 of the super-quadric, we obtain the xy-histogram, $I(x, y)$, a function whose value at (x_i, y_i) is given by the number of voxels that have x and y coordinates given by x_i and y_i respectively. We find the values of x_0 and y_0 that maximise the function,

$$\sum_{(x,y)\in S_{x_0,y_0}} I(x,y), \text{ where } S_{x_0,y_0} = \left\{(x,y) : (\frac{x}{x_0})^2 + (\frac{y}{y_0})^2 < 1\right\},$$

and satisfy the constraint, $x_0 y_0 = r^2$. The model composed of super-quadric segments computed above is presented in Fig. 11(b).

We refine the pose using the super-quadric body segments and the voxels directly instead of the voxel-skeleton. The objective is to obtain the pose that maximises the overlap between the super-quadric model and the voxels. The pose is refined by bounded optimisation of the pose parameter to minimise the "distance" between the voxels and the super-quadric model. This "distance" measures the distance of each voxel from the centre of the body segment closest to it. The distance vector, $\boldsymbol{e} = [e_1, e_2, \cdots, e_N]^\mathsf{T}$, where $e_i = \min\left(e_i^{(1)}, e_i^{(2)}, \cdots, e_i^{(J)}\right)$

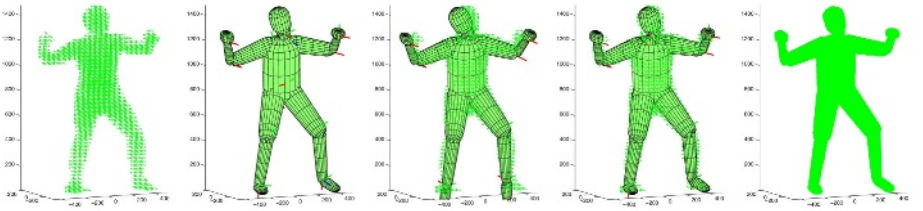


Fig. 11. The model (b) constructed from initial estimate of the quadratic parameters compared with the voxels (a), and super-imposed with voxels before pose refinement (c) and after (d). (e) is the model in voxel representation.

and $e_i^{(j)}$ is the distance of the i^{th} voxel with respect to the j^{th} body segment and given as

$$e_i^{(j)} = \begin{cases} \exp\left(r_i^j - q_i^j\right), & \text{if } 0 \le z_i^j \le z_0^j \\ \exp\left(r_i^j + p_i^j\right), & \text{otherwise} \end{cases}, \text{ where } p_i^j = \min\left(\left|z_0^j - z_i^j\right|, \left|z_i^j\right|\right),$$

$$r_i^j = \sqrt{\left(\frac{x_i^j}{x_0^j}\right)^2 + \left(\frac{y_i^j}{y_0^j}\right)^2} \text{ and } q_i^j = \sqrt{\left(1 + s^j \frac{z_i^j}{z_0^j}\right)\left(1 - \left(1 - 2\frac{z_i^j}{z_0^j}\right)^d\right)}.$$

(x_i^j, y_i^j, z_i^j) are the voxel coordinates in the coordinate system of the j^{th} body segment and $(x_0^j, y_0^j, z_0^j, s^j, d^j)$ are the super-quadric parameters of the j^{th} body segment. Although the distance function appears complicated it is just a measure of how close the voxel is to the central axis of the super-quadric. The refined pose is the pose that minimises $\|\boldsymbol{e}\|$. The pose of the subject before and after optimisation is presented in Fig. 11 (c) and (d) respectively.

4 Experiments and Conclusion

We use 15 calibrated cameras in our experiments positioned around the subject and pointing towards the centre of the capture volume. The images are 484×648 grey-level with 8-bit depth. The frequency of the capture is 3 frames per second. The units in the experiments are millimetres. The background subtraction algorithm does not perform very well on grey-scale images and as a result the voxel reconstruction is not of good quality at times. The algorithm is fairly robust to such errors and rejects frames where registration fails due to missing body segments or when the pose is not suitable.

We conducted experiments on four male subjects with different body mass, stature and BMI (body mass index). The same algorithm parameters were used in all the cases. Twenty key frames (where registration was successful) were used to estimate the model parameters as well as the pose at each time instant.

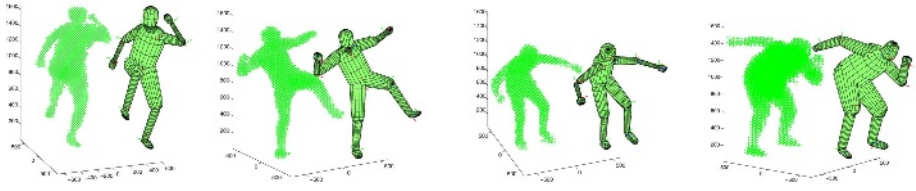


Fig. 12. Estimated models and corresponding voxels for different subjects

The results are illustrated in Fig. 12. We constructed a synthetic voxel image for each of the key frames using the estimated model and pose. We use the synthetic voxels (illustrated in Fig. 11 (e)) in order to evaluate the algorithm with respect to the data voxels (Fig. 11 (a)). We also estimate the model parameters using the synthetic voxels as *input*, so that we can compare the original pose and estimated pose. The pose errors were computed at 24 major joint locations as the absolute difference between the original and estimated joint angle values for all the key frames used in the model estimation algorithm. The results are tabulated in Table 1.

Table 1. Model and pose estimation error for real and synthetic data: Let N_D and N_M be the number of data voxels and voxels from the estimated models, and N_I is the number of intersecting voxels. Model Fill Ratio (FR) is N_I/N_M and Data FR is N_I/N_D. Note that each voxel is 1 cm^3. Volume is in m^3. Pose error is in degrees.

Experiment	Subject	Data		Model		Data/Model	Pose Error	
		Vol.	FR	Vol.	FR	Vol. Ratio	Mean	Median
Synthetic	A	0.083	0.909	0.081	0.935	1.030	5.7	2.2
	B	0.065	0.923	0.065	0.934	1.012	8.6	2.0
	C	0.057	0.858	0.054	0.902	1.052	7.0	2.2
	D	0.127	0.871	0.117	0.947	1.088	8.4	4.0
Real	A	0.088	0.766	0.083	0.812	1.059	–	–
	B	0.073	0.773	0.065	0.865	1.119	–	–
	C	0.063	0.690	0.057	0.765	1.111	–	–
	D	0.146	0.748	0.127	0.856	1.145	–	–

We have addressed the problem of model acquisition in great detail and provided the results of experiments conducted on different subjects. No prior measurements of the subjects were used. The only prior data used was a simple graph-based model of an "average" human subject and an approximate relation between the stature of an average human subject and the length of the long bones, as well as approximate locations of the shoulder, neck and pelvic joints with respect to the trunk. We have provided a systematic algorithm that aims to build a human body model in intuitive stages. We first perform segmentation and registration of the articulated chains that the human body is composed of. We have introduced a method to extract the different articulated chains that are part of the human body and also parameterise each voxel on the chain according

to its distance from the joint. The latter is a key contribution and an important step in accurately estimating the skeleton and the body model parameters. In the next step, a skeleton is computed from the voxels and the parameters of the human body model are estimated and refined by computing the fit with the skeleton obtained from the voxels. We then compute the super-quadric parameters for each body segment and refine the pose and body model parameters using the super-quadric parameters and the voxels directly. We use distance measures between the model skeleton and the skeleton computed from the voxels to optimise the pose and joint locations and presented a method to obtain an initial estimate of the parameters of super-quadric segments.

Our method has advantages over other algorithms [7,13,3] in that we explicitly extract the 1-D nature of the structure using the Laplacian eigenmap transformation in high dimensional space. We also explicitly model the continuous 1-D structure using d-dimensional splines of a single parameter. We are thus able to segment the limbs at the joints using the spline-fit error as a natural indicator of when to stop growing the spline. We then use a probabilistic registration method in order to handle complex poses where the limbs may touch other body parts as in Fig. 6 (b-e) which are not considered in other methods such as [7]. The explicit modelling of the body segments as splines in the eigenspace domain also helps in creating the skeleton in the normal 3-D space. We then show that our segmentation and registration algorithm can be exploited to estimate the human body model parameters. We estimate the probability that the segmented body parts match our model, so that we can discard frames that have missing limbs due to possible errors in the voxel reconstruction.

References

1. Moeslund, T., Granum, E.: A survey of computer vision-based human motion capture. CVIU (2001) 231–268
2. Krahnstoever, N., Sharma, R.: Articulated models from video. In: Computer Vision and Pattern Recognition. (2004) 894–901
3. Mikic, I., Trivedi, M., Hunter, E., Cosman, P.: Human body model acquisition and tracking using voxel data. International Journal of Computer Vision **53** (2003)
4. Kakadiaris, I.A., Metaxas, D.: 3D human body model acquisition from multiple views. In: Fifth International Conference on Computer Vision. (1995) 618
5. Anguelov, D., Koller, D., Pang, H., Srinivasan, P., Thrun., S.: Recovering articulated object models from 3-D range data. In: Uncertainty in Artificial Intelligence Conference. (2004)
6. Cheung, K., Baker, S., Kanade, T.: Shape-from-silhouette of articulated objects and its use for human body kinematics estimation and motion capture. In: IEEE CVPR. (2003) 77–84
7. Chu, C.W., Jenkins, O.C., Mataric, M.J.: Markerless kinematic model and motion capture from volume sequences. In: CVPR (2). (2003) 475–482
8. Tenenbaum, J.B., de Silva, V., Langford, J.C.: A global geometric framework for nonlinear dimensionality reduction. Science **290** (2000) 2319–2323
9. Belkin, M., Niyogi, P.: Laplacian eigenmaps for dimensionality reduction and data representation. Neural Comput. **15** (2003) 1373–1396

10. Elad, A., Kimmel, R.: On bending invariant signatures for surfaces. IEEE Transactions on Pattern Analysis and Machine Intelligence **25** (2003) 1285–1295
11. Brand, M.: Charting a manifold. In: Neural Information Processing Systems. (2002)
12. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290** (2000) 2323–2326
13. Brostow, G., Essa, I., Steedly, D., Kwatra, V.: Novel skeletal representation for articulated creatures. In: European Conference on Computer Vision. (2004)
14. Badler, N.I., Phillips, C.B., Webber, B.L.: Simulating Humans. Oxford University Press, Oxford, UK (1993)
15. Ozaslan, A., M. Yasar Iscan, Inci Oxaslan, H.T., Koc, S.: Estimation of stature from body parts. Forensic Science International **132** (2003) 40–45