3D Human Body Modeling Using Range Data

Koichiro Yamauchi[1,2] Bir Bhanu[1] Hideo Saito[2]
[1] *Center for Research in Intelligent Systems University of California, Riverside, CA 92521, USA*
bhanu@cris.ucr.edu
[2] *Keio University, Yokohama, 223–8522, JAPAN*
{*yamauchi, saito*}*@hvrl.ics.keio.ac.jp*

Abstract

For the 3D modeling of walking humans the determination of body pose and extraction of body parts, from the sensed 3D range data, are challenging image processing problems. Real body data may have holes because of self-occlusions and grazing angle views. Most of the existing modeling methods rely on direct fitting a 3D model into the data without considering the fact that the parts in an image are indeed the human body parts. In this paper, we present a method for 3D human body modeling using range data that attempts to overcome these problems. In our approach the entire human body is first decomposed into major body parts by a parts-based image segmentation method, and then a kinematics model is fitted to the segmented body parts in an optimized manner. The fitted model is adjusted by the iterative closest point (ICP) algorithm to resolve the gaps in the body data. Experimental results and comparisons demonstrate the effectiveness of our approach.

1. Introduction

Precise estimation of 3D body parts of articulated human body and its pose from range data captured by either a multi-camera video system or a projector-camera range system is a challenging problem. Kinematic model of an articulated structure has been used for modeling of human body extensively in security and healthcare applications.

Kehl and Gool [1] presented a method for markerless tracking of full body pose from five camera views. The system performs volumetric reconstruction using edges and color information. The articulated model built from super ellipsoids is matched against the image edges and color to overcome ambiguous situations such as touching limbs or strong occlusion. Caillette *et al.* [2] presented a full body tracker based on Monte-Carlo Bayesian framework. The volumetric reconstruction method follows shape-from-silhouette paradigm from four cameras. The appearance model represented by Gaussian blobs is fitted onto voxels using the K-means algorithm. The voxels are assigned to the nearest blob using Mahalanobis distances between blobs and voxels using both color and position. Horaud *et al.* [3] introduced a new metric to register a model surface to body data. The model surface of ellipsoids are used to represent body parts, and the body data recovered by six cameras include point and normal vectors. The metric is defined by the Euclidean distance from the ellipsoid-point to the data-point under the constraint that the ellipsoid-normal and the data-normal are parallel.

All the above methods are based on minimizing the distance between the data and the model. The *first* difficulty with these methods is that a body model is directly fitted to the entire body data. We advocate that segmentation of human body provides a priori knowledge that can improve model fitting [4]. It is preferable to fit a model for a body part to the data from the corresponding body part. The *second* difficulty is to deal with the effect of holes caused by self-occlusions and grazing angle views. A common approach to overcome the holes is to fill them with a surface patch [5]. However, this produces undesirable bridges between the two legs and in the areas near and under the arms.

In this paper we present a new 3D human body modeling technique that uses 3D range data and addresses the above problems. The entire human body is decomposed into six regions by a parts-based segmentation approach that exploits a *priori* information and avoids the effect of holes. The kinematic model is fitted into the segmented body parts, and then the fitted model is refined by the iterative closest point (ICP) algorithm. Our approach provides a high degree of accuracy compared to the conventional direct fitting approach.

The contributions of this paper are: (a) Parts-based

1051-4651/10 $26.00 © 2010 IEEE
DOI 10.1109/ICPR.2010.849

segmentation that incorporates a priori knowledge and overcomes the problems caused by the holes. (b) Unlike direct model fitting a two-step linear and non-linear registration approach for fitting models to 3D data body parts. (c) Modeling results are shown using 3D real range image data acquired from a range camera projection system.

2. Technical Approach

Fig. 1 shows the block diagram of the proposed approach. We first measure 3D human body data by a range-based projector-camera system with human in the standing or walking postures. Following these data collection we separate the human body data into six regions, and then 3D human body model is fitted to the segmented body parts in a top-down hierarchy from head to legs. The body model is refined by the ICP algorithm during the optimization process.

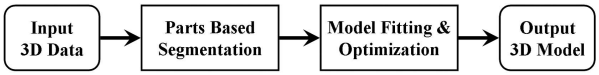


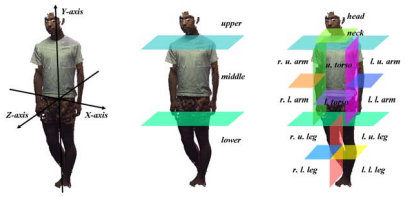
Figure 1. Block diagram of our approach.

2.1. Segmentation of Range Data into Body Parts

The human body data are segmented into six regions: head/ neck, torso, right arm, left arm, right leg, and left leg. It can be written as $\mathbf{x}_0 = \{\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4, \mathbf{x}_5, \mathbf{x}_6\}$. The subscript, reg, indicates the region number. Fig. 2 shows body axes, three segments, and six major regions which include a total of twelve body parts. Here, *r.* and *l.* indicate right and left, *u.* and *l.* indicate upper and lower (e.g. *r. l. arm* is right lower arm). In the following we present a fully automatic parts-based segmentation method.

Body Axes: The principal component analysis is applied to determine the coronal axis (X-axis), vertical axis (Y-axis), sagittal axis (Z-axis), and centroid (O) of a human body in the world coordinate system (O-X-Y-Z). Here, the 3D point, $\mathbf{x}$ on the surface of a human body, is denoted by $[x, y, z]$, the 3D point, $\mathbf{y}$, in the center of the cross-section is denoted by $[x', y', z']$, and the 3D point, $\mathbf{z}$, on the contour of the cross-section is denoted by $[x'', y'', z'']$.

Decomposition of Human Body: We divide body data into three segments: *upper*, *middle*, and *lower* using the cross-sectional areas along Y-axis. The upper segment in height larger than m includes head and heck, the middle segment in height between m and n includes torso and arms, and the lower segment in height smaller than n includes legs. The cross-sectional area in height of the centroid, g, is denoted by s_g and the cross-sectional areas in height m, n are denoted by s_m, s_n are given by $s_m = \xi_1 s_g$, $s_n = \xi_2 s_g$. Here, ξ_1, ξ_2 are height parameters. They are estimated by searching for the similar values in a vertical direction. Therefore, the body data is separated into the three segments.



(a) (b) (c)

Figure 2. Segmentation of human body. (a) Body axes. (b) Three segments/levels. (c) Twelve body parts.

Head and Neck: The upper segment includes only head and neck, $\mathbf{x}_1$, in our database. If a subject gets her/his hands higher than the middle segment, the upper segment includes 3D points of arms. In such cases, they are removed by specifying the appropriate range. The discriminant function which separates head and neck is given by

$$p(\mathbf{x}_1|\mathbf{x} \in upper\ segment(\mathbf{x})) = \text{sgn}\{(1+\zeta_1)\frac{\|\mathbf{z}-\mathbf{y}\|}{\|\mathbf{x}-\mathbf{y}\|} - 1\} - 1 \quad (1)$$

where ζ_1 is a height parameter. If the function provides zero value, $\mathbf{x}$ is assigned to $\mathbf{x}_1$.

Torso and Arms: The middle segment includes torso, $\mathbf{x}_2$, right arm, $\mathbf{x}_3$, and left arm, $\mathbf{x}_4$. Modeling methods which restrict a subject's pose with the arms held obliquely upward or downward often fail when applied to the subject whose arms are close to torso. Therefore, we divide the middle segment by using the cross-sections. The discriminant function which separates torso and two arms is given by

$$p(\mathbf{x}_2, \mathbf{x}_3, \mathbf{x}_4|\mathbf{x} \in middle\ segment(\mathbf{x})) = (x - x')\Big\{\text{sgn}\{(1+\zeta_2)\frac{\|\mathbf{z}-\mathbf{y}\|}{\|\mathbf{x}-\mathbf{y}\|} - 1\} - 1\Big\} \quad (2)$$

where ζ_2 is a height parameter. The output zero, positive, and negative value assign $\mathbf{x}$ to $\mathbf{x}_2$, $\mathbf{x}_3$, $\mathbf{x}_4$ respectively.

Legs: The lower segment includes right leg, $\mathbf{x}_5$, and left leg, $\mathbf{x}_6$. Both legs are distributed separately when they

are projected to the X-Z plane. So, we can separate them by using the 2D coordinates $[x, z]$ in the K-means clustering algorithm. First, two initial cluster centers are computed from x minimum and maximum values. It proceeds to classify all the 3D points in this segment according to the nearest center, and then recomputing the two centers. The algorithm iterates this procedure until convergence. The discriminant function which separates the two legs is given by

$$p(\mathbf{x}_5, \mathbf{x}_6 | \mathbf{x} \in lower\ segment(\mathbf{x})) = \text{sgn}\Big\{\frac{\|\mathbf{x} - \boldsymbol{\mu}_2\|}{\|\mathbf{x} - \boldsymbol{\mu}_1\|} - 1\Big\} \qquad with \quad y = 0 \tag{3}$$

where $\boldsymbol{\mu}_1$, $\boldsymbol{\mu}_2$ are the arithmetic means of 3D points corresponding to the right leg and the left leg. They are computed by the 2D coordinates in right leg and left leg cluster, respectively. If the function provides a positive value, $\mathbf{x}$ is assigned to $\mathbf{x}_5$, and if the function provides a negative value, $\mathbf{x}$ is assigned to $\mathbf{x}_6$.

2.2. Model Fitting and Optimization

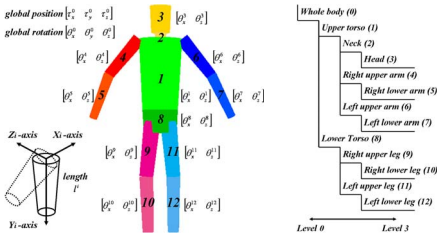
Kinematic Model: We use a kinematic model which represents the articulated structure of the human body as shown in Fig. 3. Our simplified 3D skeleton model consists of 12 body parts and 11 joints with a total of 40 degrees of freedoms, where a 6D vector, **p**, represents the global position and rotation, a 22D vector, **q**, represents the joint angles, and a 12D vector, **r**, represents the lengths of the body parts.

The body part, i, shares a parent node based on a kinematic tree. It is transformed by two angles, $[\theta_x^i, \theta_z^i]$, and one length, l^i, in the local coordinate system (O_i-X_i-Y_i-Z_i). Here, we define the neck such that it does not have the rotational joint; it is a fixed body part. The kinematic model is approximated by a set of tapered cylinders to express human shape variations among the subjects.

Model Fitting: The segmentation is useful for coarse registration, because it is unreasonable to fit a kinematic model to articulated objects without any previous knowledge. The previous knowledge allows an automatic model fitting and a reduction in the computational cost. Therefore, we will fit the model to body data by using the segmented regions.

The distance between 3D data of a segmented region, $\mathbf{x}_{reg,j}$, and 3D model of the tapered cylinder, $\mathbf{y}_{i,j}$, is linearly minimized as

$$d_c = \frac{1}{M}\sum_{j=1}^{M} \| \mathbf{x}_{reg,j} - \mathbf{R}\mathbf{y}_{i,j} \| \tag{4}$$



(a) (b)

Figure 3. Kinematic model. (a) Body model approximated by 12 tapered cylinders. (b) Hierarchical structure.

where **R** is 3×3 rotation matrix including two angles θ_x^i, θ_z^i and M is the number of points of the segmented region. The tapered cylinders can be fitted by determining two angles and one length in the order of levels 1, 2, 3 of the hierarchical structure. With regard to the head and neck, the parameters are estimated from the distribution of 3D points in the X-Y plane and Y-Z plane, respectively because the data of head hair and lower head region cannot be captured.

Optimization: The Interactive Closest Point (ICP) algorithm provides fine registration by minimizing the distance between the body data and kinematic model [6]. The key steps in the algorithm are (a) Uniform sampling of points on both shapes. (b) Matching each selected point to the closest sample in the other shape. (c) Uniform weighting of point pairs. (d) Rejecting pairs with distances lager than some multiple of the standard deviation. (e) Point-to-point error metric. (f) Standard select-match-minimizes iteration.

The distance between 3D data of the entire body, $\mathbf{x}_{0,j}$, and 3D model of the tapered cylinder, $\mathbf{z}_{i,j}$, to be iteratively minimized is

$$d_f = \frac{1}{N}\sum_{j=1}^{N} \| \mathbf{x}_{0,j} - \mathbf{R}\mathbf{z}_{i,j} \| \tag{5}$$

where **R** is 3×3 rotation matrix including two angles θ_x^i, θ_z^i and N is the number of corresponding points. The tapered cylinders are adjusted by changing the two angles in the order of the levels 1, 2, 3 of the hierarchical structure.

3. Experimental Results

Data and Parameters: The body data of three subjects are captured by the range projector-camera system [7]. The system, consisting of nine projector-camera pairs, captures range data of the entire body in ∼2 seconds with 640 × 480 pixels images, 3 mm depth resolution,

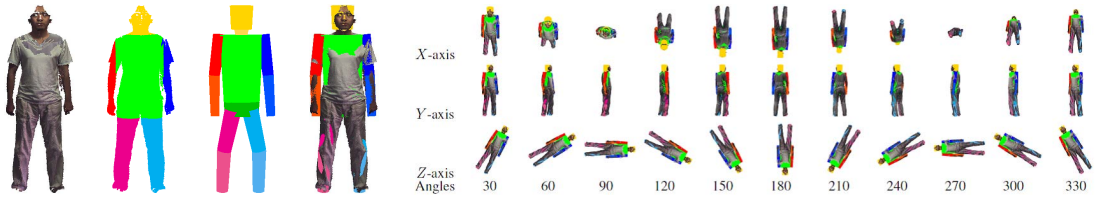


Figure 4. Fitted kinematic models by proposed approach.

Table 1. Performance comparison of two fitting approaches with/without segmentation in parts of the body.

Parts#	1	3	4	5	6	7	8	9	10	11	12
Direct fitting approach [mm]	25.9	47.3	30.4	32.1	17.1	17.1	20.3	14.9	25.4	20.0	**32.1**
Proposed approach [mm]	**24.8**	46.8	**28.9**	**23.6**	16.5	17.6	20.2	14.9	**19.7**	**15.1**	37.3

and the measurement accuracy within 2 mm. The numbers of measurement points are approximately 1/2 to one million. For all of experiments in this paper we used $\xi_1 = 0.25$, $\xi_2 = 0.5$, and $\zeta_1 = 1.0$. The value of ζ_2 is from 0.13 to 0.14 depending on the subject and the pose.

Results: Fig. 4 shows the results of human body modeling by the proposed two-step fitting using parts-based segmentation. From left to right, the first figure shows the body data in the standing posture. The numbers of measurement points is approximately one million. The second figure shows the segmented body parts. The colored regions correspond to torso, head/neck, right arm, left arm, right leg, and left leg. The segmentation facilitates model fitting and reduce the computational cost. The third figure shows the kinematic model and the fourth figure shows the kinematic model superimposed on the body data. The fitted model provides joint angles and lengths of body parts. The fifth figure shows the fitted model rotating about X-axis, Y-axis, and Z-axis, at every 30 degrees.

Comparisons: Table 1 is the results of comparison of the two approaches. The fitting error is defined as the Euclidian distance between the data and the model. We observe that our approach achieves better performance than the direct fitting approach, especially for the five body parts, i.e. upper torso (#1), right upper arm (#4), right lower arm (#5), right lower leg (#10), left upper leg (#11) the results are significantly improved. The average error for direct fitting and the proposed approach for all the body parts are 25.7mm and 24.1mm, respectively. The experiments clearly demonstrate the effectiveness of the proposed technique and the value of image segmentation.

4. Conclusions

We proposed a method for 3D human body modeling using parts-based segmentation. The segmentation exploits the previous knowledge and avoids the effect of the holes in the data. In our approach, the body data is decomposed into six regions, and then the kinematic model is fitted to the segmented body parts. The fitted model is refined by the ICP algorithm outwardly from the center of a human body. Through rigorous experiments we verified that our approach is better than the direct fitting approach.

References

[1] R. Kehl and L. V. Gool, "Markerless tracking of complex human motions from multiple views," *Computer Vision and Image Understanding*, vol. 104, no. 2, pp. 190–209, 2006.

[2] F. Caillette, A. Galata, and T. Howard, "Real-time 3-D human body tracking using learnt models of behavior," *Computer Vision and Image Understanding*, vol. 109, no. 2, pp. 112–125, 2008.

[3] R. Horaud, M. Niskanen, G. Dewaele, and E. Boyer, "Human motion tracking by registering an articulated surface to 3D points and normals," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 1, pp. 158–164, 2009.

[4] N. Werghi, "Segmentation and modeling of full human body shape from 3-D scan data: A survey," *IEEE trans. Systems, Man, and Cybernetics, Part C*, vol. 37, no. 6, pp. 1122–1136, 2007.

[5] J. Davis, S. R. Marschner, M. Garr, and M. Levoy, "Filling holes in complex surfaces using volumetric diffusion," In *Proceedings of International Symposium on 3D Data Processing Visualization and Transmission*, 2002.

[6] S. Rusinkiewicz and M. Levoy, "Efficient variants of the ICP algorithm," In *Proceedings of 3-D Digital Imaging and Modeling*, pp. 145-152, 2001.

[7] K. Yamauchi and Y. Sato, "3D human body measurement by multiple range images," In *Proceedings of International Conference on Pattern Recognition*, vol. 4, pp. 833–836, 2006.