Human Body Model Acquisition and Motion Capture Using Voxel Data

Ivana Mikić[2], Mohan Trivedi[1], Edward Hunter[2], Pamela Cosman[1]

[1]Department of Electrical and Computer Engineering, 9500 Gilman Drive, La Jolla, California 92093-0407, phone: 858 822-0002, fax: 858 822-5336
trivedi@ece.ucsd.edu, pcosman@code.ucsd.edu

[2]Q3DM, Inc.
{imikic,ehunter}@q3dm.com

Abstract. In this paper we present a system for human body model acquisition and tracking of its parameters from voxel data. 3D voxel reconstruction of the body in each frame is computed from silhouettes extracted from multiple cameras. The system performs automatic model acquisition using a template based initialization procedure and a Bayesian network for refinement of body part size estimates. The twist-based human body model leads to a simple formulation of the extended Kalman filter that performs the tracking and with joint angle limits guarantees physically valid posture estimates. Evaluation of the approach was performed on several sequences with different types of motion captured with six cameras.

1 Introduction

Motion capture is of interest in many applications such as advanced user interfaces, entertainment, surveillance systems, or motion analysis for sports and medical purposes. In the past few years, the problem of markerless, unconstrained motion capture has received much attention from computer vision researchers [1, 2, 3, 4]. Many systems require manual initialization of the model and then perform the tracking. The systems that use multiple camera images as inputs, most often analyze the data in the image plane, comparing it with the appropriate features of the model projection [5, 6]. Promising results have been reported in using the depth data obtained from stereo [7, 8] for pose estimation. However, only recently the first attempts at using voxel data obtained from multiple cameras to estimate body pose have been reported [9]. This system used a very simple initialization and tracking procedure that did not guarantee a valid articulated body model.

In [10] we have introduced the framework for articulated body model acquisition and tracking from voxel data: video from multiple cameras is segmented and a voxel reconstruction of the person's body is computed from the 2D silhouettes; in the first frame, an automatic model initialization is performed – the head is found first by template matching and the other body parts by a sequential template growing procedure. An extended Kalman filter is then used for tracking. In this system, a body pose where all four limbs were visible as separate was expected in the first frame for

successful initialization. The orientation of each body part was modeled independently and therefore the physically invalid joint rotations were possible. Also, the voxel labeling approach used to compute the measurements for the tracker was based on Mahalanobis distance and worked well only for small frame-to-frame displacements.

We have, therefore, designed a new system with a greatly improved performance (Fig. 1). During model initialization, we have introduced a model refinement phase where a Bayesian network that incorporates the knowledge of human body proportions improves the estimates of body part sizes. A twist-based human body model is now used, which leads to a simple extended Kalman filter formulation and guarantees physically valid posture estimates. We have also designed a voxel labeling approach that takes advantage of the unique qualities of voxel data and of the Kalman filter predictions to obtain quality measurements for tracking. Section 2 contains the description of the twist-based human body model and the formulation of the extended Kalman filter. In Section 3 the voxel labeling and tracking are presented. The model acquisition algorithm is described in Section 4. Experimental evaluation is presented in Section 5. Concluding remarks follow in Section 6.

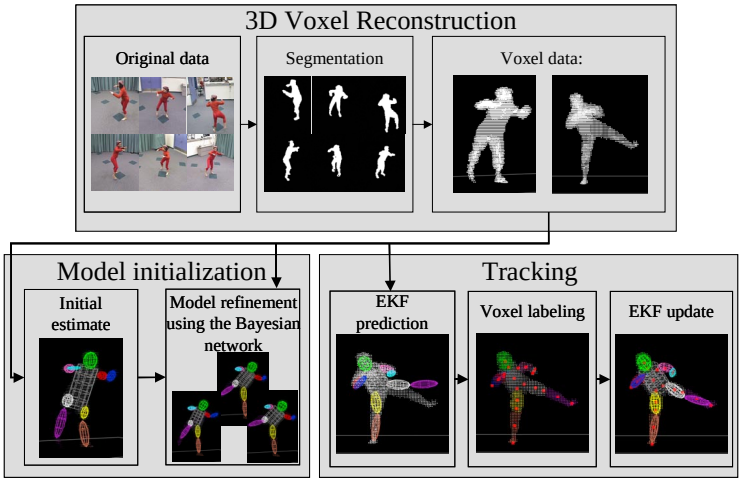

Fig. 1. System components.

2 Human Body Model and the Kalman Filter Formulation

The articulated body model we use is shown in Fig. 2. Sizes of body parts are denoted as $2\lambda_i^{(j)}$, where i is the body part index, and j is the dimension order – smallest dimension is 0 and largest is 2. For all parts except torso, the two smaller dimensions are set to be equal to the average of the two dimensions estimated during initialization. The positions of joints are fixed relative to the body part dimensions in

the torso coordinate system (for example, the hip is at $\begin{bmatrix} 0 & d_H \lambda_0^{(1)} & -\lambda_0^{(2)} \end{bmatrix}^T$ - see Fig. 2). Sixteen axes of rotation are modeled in different joints. Two in the neck, three in each shoulder, two in each hip and one in each elbow and knee. The range of allowed values is set for each angle. For example, the rotation in the knee can go from 0 to 180 degrees - the knee cannot bend forward. The rotations about these axes (relative to the torso) are modeled using twists [11, 12].

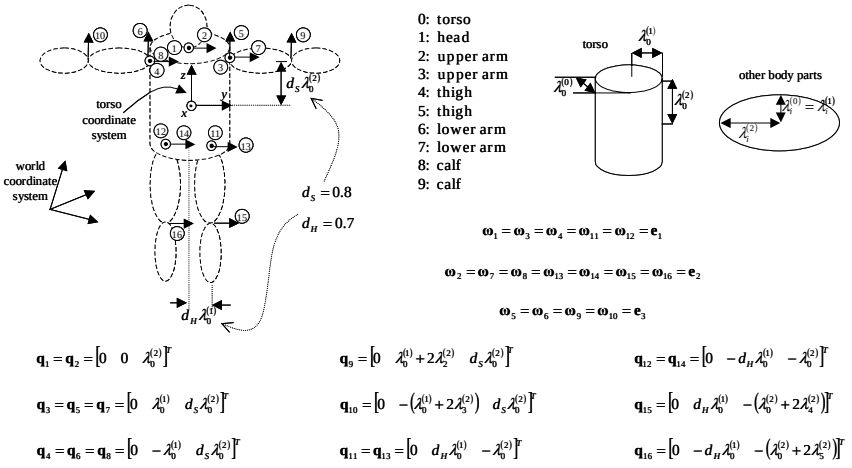

Fig. 2. Articulated body model. Sixteen axes of rotation (marked by circled numbers) in body joints are modeled using twists relative to the torso-centered coordinate system. To describe an axis of rotation, a unit vector along the axis and a coordinate of a point on the axis in the "initial" position of the body are needed. As initial position, we chose the one where legs and arms are straight and arms are pointing away from the body as shown in the figure. Dimensions of body parts are determined in the initialization procedure and are held fixed thereafter. Body part dimensions are denoted by λ; subscript refers to the body part number and superscript to dimension order – 0 is for the smallest and 2 for the largest of the three. For all body parts except the torso, the two smaller dimensions are set to be equal.

Rotation about an axis is described by a unit vector along the axis (**ω**), an arbitrary point on the axis (**q**) and the angle of rotation (θ_i). A twist associated with an axis of rotation is defined as:

$$\boldsymbol{\xi} = \begin{bmatrix} \boldsymbol{\omega} \\ \mathbf{v} \end{bmatrix}, \text{ where } \mathbf{v} = -\boldsymbol{\omega} \times \mathbf{q} \tag{1}$$

An exponential map, $e^{\hat{\xi}\theta}$ maps the homogeneous coordinates of a point from its initial values to the coordinates after the rotation is applied [12]:

$$\mathbf{p}(\theta) = e^{\hat{\xi}\theta}\mathbf{p}(0) = \mathbf{T}\mathbf{p}(0) \tag{2}$$

where

$$\mathbf{p}(\theta) = \begin{bmatrix}\mathbf{x}^T(\theta) & 1\end{bmatrix}^T = \begin{bmatrix}x(\theta) & y(\theta) & z(\theta) & 1\end{bmatrix}^T, \quad e^{\hat{\omega}\theta} = \mathbf{I} + \hat{\boldsymbol{\omega}}\sin\theta + \hat{\boldsymbol{\omega}}^2(1-\cos\theta)$$

$$\mathbf{T} = e^{\hat{\xi}\theta} = \begin{bmatrix} e^{\hat{\omega}\theta} & (\mathbf{I} - e^{\hat{\omega}\theta})(\boldsymbol{\omega}\times\mathbf{v}) + \boldsymbol{\omega}\boldsymbol{\omega}^T\mathbf{v}\theta \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} \mathbf{R} & \mathbf{t} \\ 0 & 1 \end{bmatrix}, \quad \hat{\boldsymbol{\omega}} = \begin{bmatrix} 0 & -\omega_2 & \omega_1 \\ \omega_2 & 0 & -\omega_0 \\ -\omega_1 & \omega_0 & 0 \end{bmatrix} \qquad (3)$$

Even though the axes of rotation change as the body moves, in the twists formulation the descriptions of the axes stay fixed and are determined in the initial body configuration. We chose the configuration with extended arms and legs and arms pointing to the side of the body (as shown in Fig. 2). In this configuration, all angles θ_i are zero. The figure also gives values for the vectors $\boldsymbol{\omega}_i$ and $\mathbf{q}_i$ for each axis.

Knowing the dimensions of body parts and using the body model shown in Fig. 2, the configuration of the body is completely captured with angles of rotation about each of the axes (θ_1 - θ_{16}) and the centroid and orientation of the torso. Orientation (rotation matrix) of the torso is parameterized with a quaternion, which is equivalent to the unit vector $\boldsymbol{\omega}_0$ and the angle θ_0. Therefore, the position and orientation of the torso are captured using seven parameters – three coordinates for centroid location and four for the orientation. The configuration of the described model is fully captured by 23 parameters, which we include into the Kalman filter state, $\mathbf{x}_k$. For the measurements of the Kalman filter (contained in the vector $\mathbf{z}_k$) we chose 23 points on the human body: centroids and endpoints of each of the ten body parts, neck, shoulders, elbows, hips, knees, feet and hands (Fig. 3).

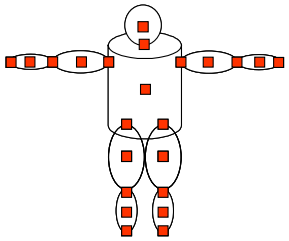

Fig. 3. 23 points on the human body, chosen to form a set of measurements for the Kalman filter

To design the Kalman filter, the relationship between the measurement and the state needs to be described. For a point $\mathbf{p}$, we define the "set of significant rotations" that contains the rotations that affect the position of the point – if $\mathbf{p}$ is the fingertip, there would be four: three in the shoulder and one in the elbow. Set of angles $\boldsymbol{\theta}_\mathbf{p}$ contains the angles associated with the set of significant rotations. The position of a point $\mathbf{p}_t(\boldsymbol{\theta}_\mathbf{p})$ with respect to the torso is given by the product of exponential maps that correspond to the set of significant rotations and of the position of the point in the initial configuration $\mathbf{p}_t(\mathbf{0})$ [12]:

$$\mathbf{p}_t(\boldsymbol{\theta}_\mathbf{p}) = \mathbf{T}_{i1}\mathbf{T}_{i2}\ldots\mathbf{T}_{in}\mathbf{p}_t(\mathbf{0}), \text{ where } \boldsymbol{\theta}_\mathbf{p} = \{\theta_{i1}, \theta_{i2}, \ldots\theta_{in}\} \qquad (4)$$

We denote with $\mathbf{T}_0$ the mapping that corresponds to the torso position and orientation:

$$\mathbf{T}_0 = \begin{bmatrix} \mathbf{R}_0 & \mathbf{t}_0 \\ 0 & 1 \end{bmatrix} = \begin{bmatrix} e^{\hat{\omega}_0 \theta_0} & \mathbf{t}_0 \\ 0 & 1 \end{bmatrix} \quad (5)$$

where $\mathbf{R}_0 = e^{\hat{\omega}_0 \theta_0}$ and $\mathbf{t}_0$ is the torso centroid. The homogeneous coordinates of a point with respect to the world coordinate system can now be expressed as:

$$\mathbf{p}_0(\boldsymbol{\theta}_\mathbf{p}, \boldsymbol{\omega}_0, \mathbf{t}_0) = \mathbf{T}_0 \mathbf{p}_t(\boldsymbol{\theta}_\mathbf{p}) \quad (6)$$

It follows that Cartesian coordinate of this point in the world coordinate system is:

$$\mathbf{x}_0(\boldsymbol{\theta}_\mathbf{p}) = \mathbf{R}_0(\mathbf{R}_{i1}(\mathbf{R}_{i2}(...(\mathbf{R}_{in}\mathbf{x}_\mathbf{t}(\mathbf{0}) + \mathbf{t}_{in}) + ...) + \mathbf{t}_{i2}) + \mathbf{t}_{i1}) + \mathbf{t}_0 \quad (7)$$

For any of the chosen 23 measurement points, its location relative to the torso centroid in the initial configuration $\mathbf{x}_\mathbf{t}(\mathbf{0})$ is a very simple function of the body part dimensions and the joint locations (which are also defined relative to the body part dimensions). For example, left foot is at $\left[0 \quad d_H \lambda\binom{1}{0} \quad -\left(\lambda_0^{(2)} + 2\lambda_4^{(2)} + 2\lambda_8^{(2)}\right)\right]^T$.

In the Kalman filter equations [13]:

$$\begin{aligned} \mathbf{x}_{k+1} &= \mathbf{F}\mathbf{x}_k + \mathbf{u}_k \\ \mathbf{z}_k &= \mathbf{H}(\mathbf{x}_k) + \mathbf{w}_k \end{aligned} \quad (8)$$

the relationship between the measurements and the state is nonlinear, except for the torso centroid. It is, therefore, linearized around the predicted state in the extended Kalman filter, i.e. the Jacobian of $H(x_k)$ is computed at the predicted state. This Jacobian consists of partial derivatives of coordinates of each of the 23 measurement points with respect to the torso centroid, torso quaternion and the 16 angles of rotation. All these derivatives are straightforward to compute from the Equation 7, since only one rotation matrix and one translation vector depend on any given angle θ_i. Rotation matrices associated with axes of rotation are very simple since all axes coincide with one of the coordinate axes. The rotation matrix that describes the torso orientation is arbitrary, but its derivatives with respect to ω_0 and θ_0 are also easy to compute.

To ensure that the model configuration represents a valid posture of the human body, the angles in different joints need to be limited. We impose these constraints on the updated Kalman filter state that contains the estimate of these angles by setting the value of the angle to the interval limit if that limit has been exceeded.

3. Voxel Labeling and Tracking

The algorithm for model acquisition, which estimates body part sizes and their locations in the beginning of the sequence, will be presented in the next section. For now, we will assume that the dimensions of all body parts and their approximate locations in the beginning of the sequence are known. For every new frame, the tracker updates model position and configuration to reflect the motion of the tracked person. The labeling of the voxel data is necessary for obtaining the measurements

used for tracking. From the labeled voxels, it is easy to compute the locations of the 23 points shown in Fig. 3, since those points are either centroids or endpoints of different body parts.

Initially, we labeled the voxels based on the Mahalanobis distance from the predicted positions of body parts. However, in many cases, this led to loss of track. This was due to the fact that labeling based purely on distance cannot produce a good result when the prediction is not very close to the true positions of body parts. We have, therefore, designed an algorithm that takes advantage of the unique qualities of voxel data to perform reliable labeling even for very large frame-to-frame displacements.

Due to its unique shape and size, the head is easiest to find and is located first (Fig. 4). We create a spherical crust template whose inner and outer diameters correspond to the smallest and largest head dimensions. For the head center we choose the location of the template center that maximizes the number of surface voxels that are inside the crust. Then, the voxels that are inside the sphere of the larger diameter, centered at the chosen head center are labeled as belonging to the head, and the true center is recomputed from those voxels. The location of the neck is found as an average over head voxels with at least one neighbor a non-head body voxel. The prediction of the head center location is available and we therefore search for it only in the neighborhood of the prediction. This speeds up the search and also decreases the likelihood of error.

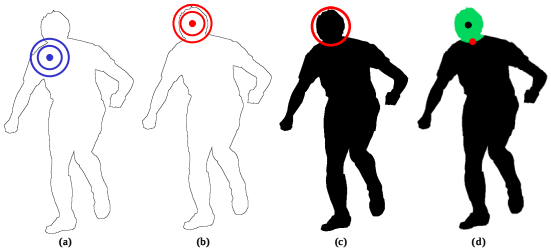

Fig. 4. Head location procedure illustrated in a 2D cross-section. (a) search for the location of the center of a spherical crust template that contains the maximum number of surface voxels (b) the best location is found (c) voxels that are inside the sphere of a larger diameter are labeled as belonging to the head (d) head voxels (green), the head center (black) and the neck (red)

Next, the torso voxels are labeled. The template of the size of the torso (with circular cross-section whose radius is the larger of the two torso base dimansions) is then placed with its base at the neck and with its axis going through the centroid of non-head voxels. The voxels inside this template are then used to recompute a new centroid, and the template is rotated so that its axis passes through it (torso is anchored to the neck at the center of its base at all times). This procedure is repeated until the template stops moving, which is accomplished when the template is entirely inside the torso or is well centered over it. Even with an initial centroid that is completely outside the body, this procedure converges, since in the area close to the neck, the template always contains some torso voxels that help steer the template in the right direction (see Fig. 5). The voxels inside the template are labeled as belonging to the torso.

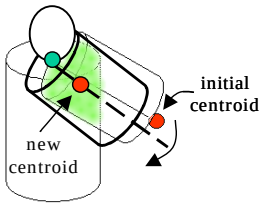

Fig. 5. Fitting the torso. Initial torso template is placed so that its base is at the neck and its main axis passes through the centroid of non-head voxels. Voxels that are inside the template are used to calculate new centroid and the template is rotated to align the main axis with the new centroid. The process is repeated until the template stops moving which happens when it is entirely inside the torso or is well centered over it.

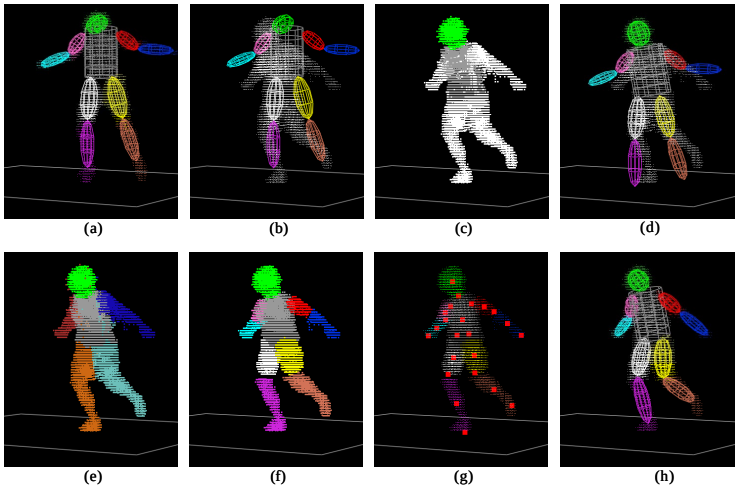

Fig. 6. Voxel labeling and tracking. (a) tracking result in the previous frame (b) model prediction in the new frame; (c) head and torso located; (d) limbs moved to preserve the predicted hip and joint angles for the new torso position and orientation; (e) four limbs are labeled by minimizing the Mahalanobis distance from to the limb positions shown in (d); (f) upper arms and thighs are labeled by fitting them inside the limbs, anchored at the shoulder/hip joints. The remaining limb voxels are labeled as lower arms and thighs; (g) the measurement points are easily computed from the labeled voxels (h) tracker adjusts the body model to fit the data in the new frame

Then, the predictions for the four limbs are modified to maintain the predicted hip and shoulder angles with the new torso position. The remaining voxels are then assigned to the four limbs based on Mahalanobis distance from these modified limb positions. To locate upper arms and thighs inside the appropriate limb voxel blobs, the same fitting procedure used for torso is repeated, with templates anchored at the shoulders/hips. When the voxels belonging to upper arms and thighs are labeled, the remaining voxels in each of the limbs are labeled as lower arms or calves. Using

modified predictions of the limb locations enables the system to handle large frame to frame displacements. Once all the voxels are labeled, the 23 measurement points are easily computed as centroids or endpoints of appropriate blobs. The extended Kalman described in the previous section is then used to adjust the model to the measurements in the new frame and to produce the prediction for the next frame. Fig. 6 illustrates the voxel labeling and tracking.

4. Model Acquisition

The human body model is chosen a priori and is the same for all humans. However, the actual sizes of body parts vary from person to person. Obviously, for each captured sequence, the initial locations of different body parts will vary also. Model acquisition, therefore, involves both locating the body parts and estimating their true sizes from the data in the beginning of a sequence. It is performed in two stages. First, the rough estimates of body part locations and sizes in the first frame are generated using a template fitting and growing algorithm. In the second stage, this estimate is refined over several subsequent frames using a Bayesian network that takes into account both the measured body dimensions and the known proportions of the human body. During this refinement process, the Bayesian network is inserted into the tracking loop, using the body part size measurements produced by the voxel labeling to modify the model, which is then adjusted to best fit the data using the extended Kalman filter. When the body part sizes stop changing, the Bayesian network is "turned off" and the regular tracking continues.

4.1 Initial Estimation of Body Part Locations and Sizes

This procedure is similar to the voxel labeling described in Section 6.2. However, the prediction from the previous frame does not exist (this is the first frame) and the sizes of body parts are not known. Therefore, several modifications and additional steps are needed.

The algorithm illustrated in Fig. 4 is still used to locate the head, however, the inner and outer diameters of the spherical crust template are now set to the smallest and largest head diameters we expect to see. Also, the whole volume has to be searched. The errors are more likely than during voxel labeling for tracking, but are still quite rare: in our experiments on 600 frames, this version located the head correctly in 95% of the frames.

To locate the torso, the same fitting procedure described for voxel labeling is used (Fig. 5), but with the template of an average sized torso. Then, the torso template is shrunk to a small, predetermined size in its new location and grown in all dimensions until further growth starts including empty voxels. At every step of the growing, the torso is reoriented as shown in Fig. 5 to ensure that it is well centered during growth. In the direction of the legs, the growing will stop at the place where legs part. The voxels inside this new template are labeled as belonging to the torso (Fig. 7).

Next, the four regions belonging to the limbs are found as the four largest

connected regions of remaining voxels. The hip and shoulder joints are located as the centroids for voxels at the border of the torso and each of the limbs. Then, the same fitting and growing procedure described for the torso is repeated for thighs and upper arms. The lower arms and calves are found by locating connected components closest to the identified upper arms and thighs. Fig. 8. shows the described initial body part localization on real voxel data.

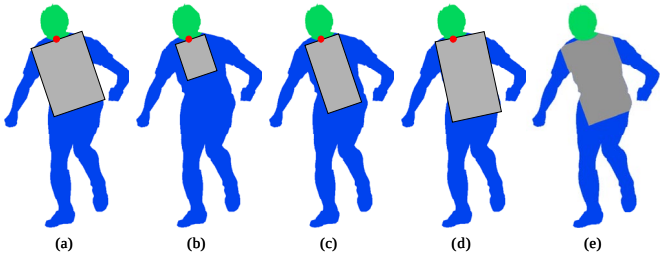

Fig. 7. Torso locating procedure illustrated in a 2D cross-section. (a) Initial torso template is fitted to the data; (b) It is then replaced by a small template of predetermined size which is anchored at the same neck point and oriented the same way; (c) the template is then grown and reoriented at every step of growing to ensure the growth does not go in the wrong direction; (d) the growing is stopped when it starts including empty voxels; (e) voxels inside the final template are labeled as belonging to the torso

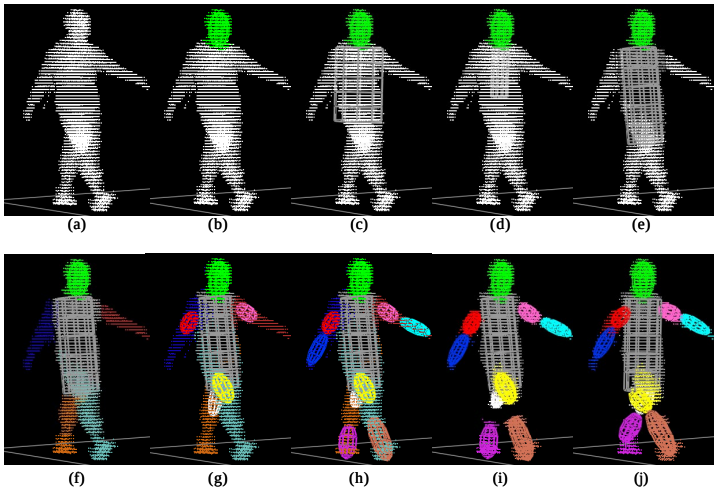

Fig. 8. Initial body part localization. (a) 3D voxel reconstruction; (b) head located; (c) initial torso template anchored at the neck centered over the non-head voxels; (d) start of the torso growing; (e) final result of torso growing with torso voxels labeled; (f) four limbs labeled as four largest remaining connected components; (g) upper arms and thighs are grown anchored at the shoulders/hips with the same procedure used for torso; (h) lower arms and calves are fitted to the remaining voxels; (i) all voxels are labeled; (j) current model adjusted to the data using the EKF to ensure a kinematically valid posture estimate.

4.2 Model Refinement

The estimates of body part sizes and locations in the first frame produced by the algorithm described in the previous section performs robustly, but the sizes of the torso and the limbs are often very inaccurate and depend on the body pose in the first frame. For example, if the person is standing with legs straight and close together, the initial torso will be very long and include much of the legs. The estimates of the thigh and calf sizes will be very small. Obviously, an additional mechanism for estimating true body part sizes is needed.

In addition to the initial estimate of the body part sizes and of the person's height, a general knowledge of human body proportions is available. To take that important knowledge into account when reasoning about body part sizes, we are using Bayesian networks (BNs). A BN is inserted into the tracking loop (Fig. 9), modifying the estimates of body part lengths at each new frame. The EKF tracker adjusts the new model position and configuration to the data, the voxel labeling procedure provides the measurements in the following frame, which are then used by the BN to update the estimates of body part lengths. This procedure is repeated until the body part lengths stop changing.

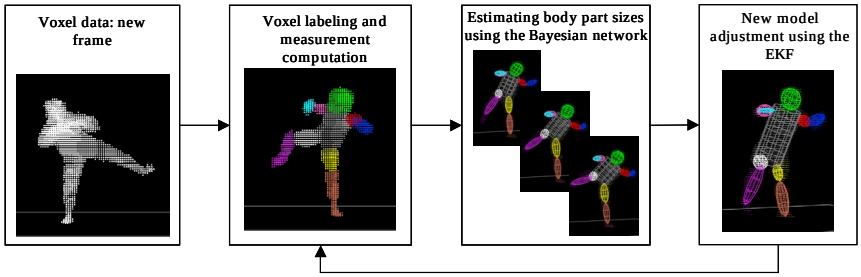

Fig. 9. Body part size estimation

The domain knowledge that is useful for designing the Bayesian network is: the human body is symmetric, i.e., the corresponding body parts on the left and the right sides are of the same dimensions; the lengths of the head, the torso, the thigh and the calf add up to person's height; the proportions of the human body are known.

The measurements that can be made from the data are the sizes of all body parts and the person's height. The height of the person, the dimensions of the head and the two width dimensions for all other body parts are measured quite accurately. The lengths of different body parts are the ones that are inaccurately measured. This is due to the fact that the measured lengths depend on the borders between body parts, which are hard to locate accurately. For example, if the leg is extended, it is very hard to determine where the thigh ends and the calf begins, but the two width dimensions can be very accurately determined from the data.

Taking into account what is known about the human body and what can be measured from the data, we can conclude that there is no need to refine our estimates of the head dimensions or the width dimensions of other body parts since they can be

accurately estimated from the data, and our knowledge of body proportions would not be of much help in these cases anyway. However, for body part lengths, the refinement is necessary and the available prior knowledge is very useful. Therefore, we have built a Bayesian network shown in Fig. 10 that estimates the lengths of body parts and that takes into account what is known and what can be measured.

Each node represents a continuous random variable. Leaf nodes Thm, Cm, UAm and LAm are the measurements of the lengths of the thigh, calf, upper and lower arm in the current frame. Leaf node Height is the measurement of the person's height (minus head length) computed in the first frame. If the person's height is significantly smaller than the sum of measured lengths of appropriate body parts, we take that sum as the true height – in case the person is not standing up. Leaf nodes Thm0, Cm0, UAm0 and LAm0 are used to increase the influence of past measurements and speed up the convergence. Each of these nodes is updated with the mean of the marginal distribution of its parent from the previous frame. Other nodes (Torso, Thigh, Calf, UpperArm and LowerArm) are random variables that represent true body part lengths. Due to the body symmetry, we include only one node for each of the lengths of the limb body parts and update the corresponding measurement node with the average of the measurements from the left and right sides. The measurement of the torso length is not used because the voxel labeling procedure just fits the known torso to the data, therefore the torso length measurement is essentially the same as the torso length in the model from the previous frame.

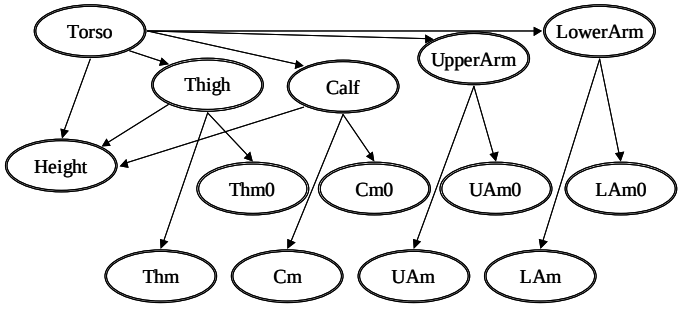

Fig. 10. The Bayesian network for estimating body part lengths. Each node represents a length. The leaf nodes are measurements (Thm represents the new thigh measurement, Thm0 reflects the past measurements etc.). Nodes Torso, Thigh, Calf, UpperArm and LowerArm are random variables that represent true body part lengths.

All variables are Gaussian and the distribution of a node Y with continuous parents **Z** is of the form:

$$p(Y/\mathbf{Z}=\mathbf{z}) = \mathcal{N}\left(\alpha + \boldsymbol{\beta}^T \mathbf{z}, \sigma^2\right) \quad (\mathbf{9})$$

Therefore, for each node with n parents, a set of n weights $\boldsymbol{\beta} = [\beta_1 \quad \ldots \quad \beta_n]^T$, a standard deviation σ and possibly a constant α are the parameters that need to be chosen. These parameters have clear physical interpretation (body proportions) and are quite easy to select.

5. Experimental Evaluation

We have evaluated the described system on several sequences with different types of motion such as walking, running, jumping, sitting and dancing, captured with six cameras. Data was captured at frame rates between 5 and 10 frames/second. Each camera was captured at resolution of 640×480 pixels. We illustrate the tracking results on two sequences - dance and walking.

First, we show the results of the model acquisition. Fig. 11 shows the original camera views and the corresponding acquired models for five people. The models successfully capture the main features of these very different human bodies. All five models were acquired using the same algorithm and the same Bayesian network with fixed parameters. The convergence is achieved in three to four frames.

Fig. 12 and 13 show the tracking results for the dance and walking sequences. Fig. 14 shows some joint angles as functions of time for the walking sequence. The sequence contained fourteen steps, seven by each leg – which are easily correlated with the joint angle plots.

Tracking results look very good. However, the resolution of the model that was chosen limits the types of motions that can be accurately tracked. For example, we do not model the rotation in the waist, i.e. the shoulders and hips are expected to lie in the same plane. This will result in tracking errors when waist rotation is present in the analyzed motion. However, including additional axes of rotation or additional body parts to the model is very simple, following the framework described in this paper.

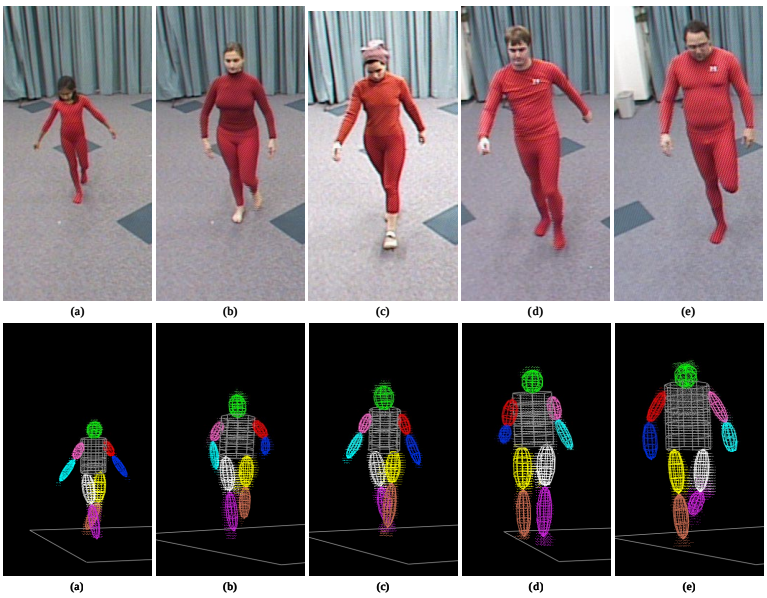

Fig. 11. Original views and estimated models of the five people (a) Aditi (height: 1295.4mm); (b) Natalie (height: 1619.25mm); (c) Ivana (height: 1651mm); (d) Andrew (height: 1816.1mm); (e) Brett (height: 1879mm)

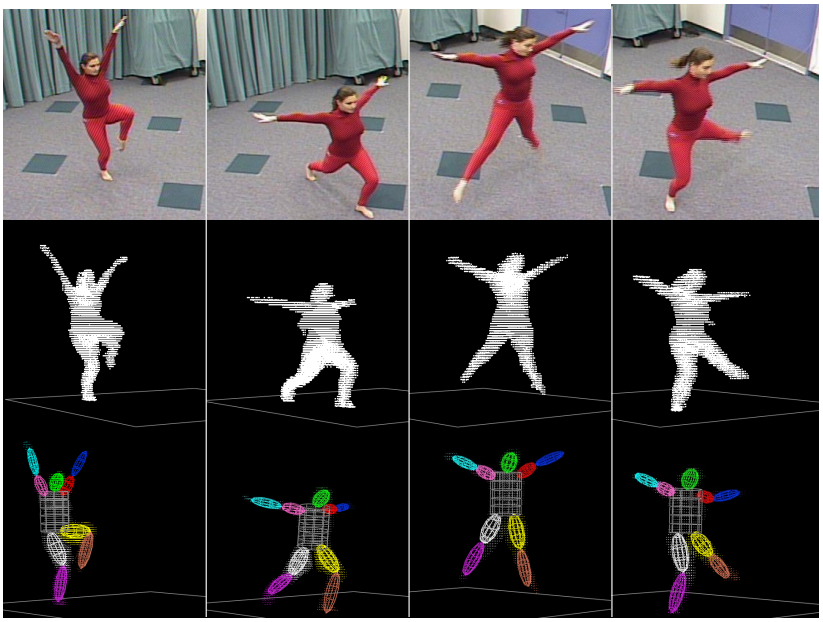

Fig. 12. Tracking results for the dance sequence and one of the six original camera views

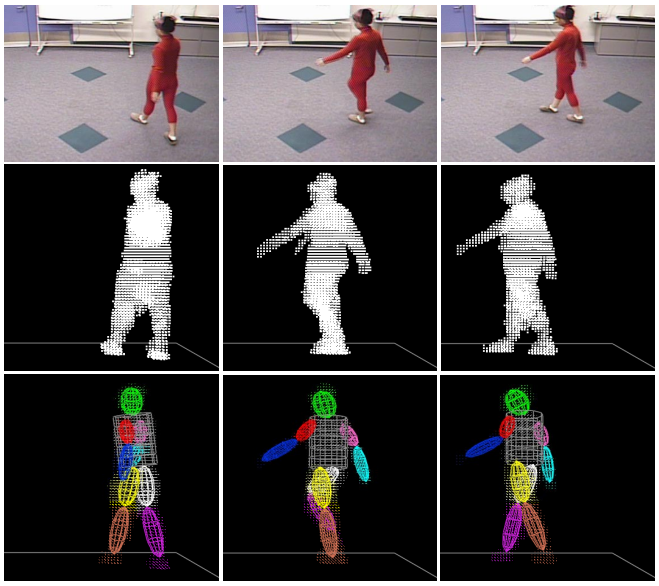

Fig. 13. Tracking results for the walking sequence. First row: one of the six original camera views. Second row: 3D voxel reconstruction viewed from a similar viewpoint. Third row: tracking results

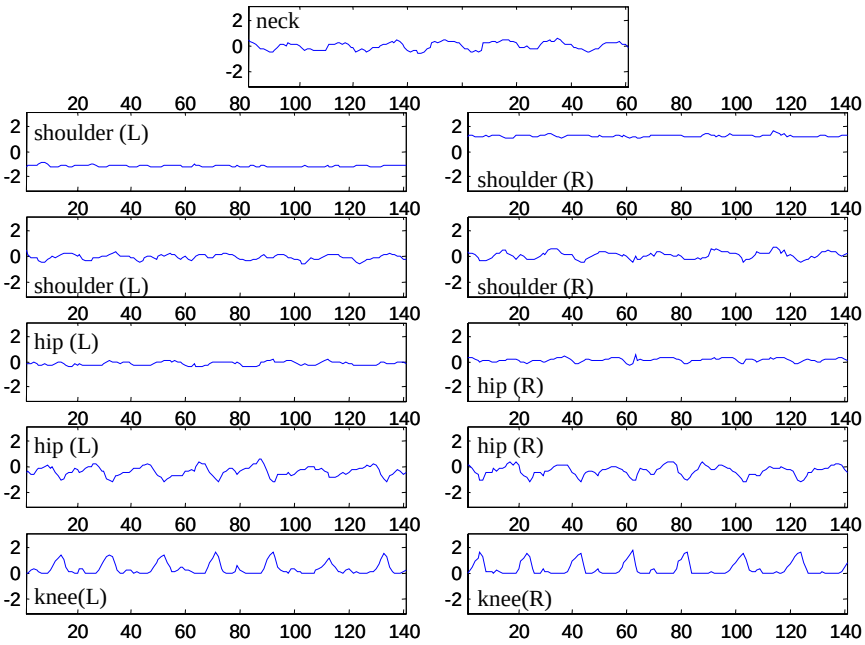

Fig. 14. Joint angles as the functions of the frame number for the walking sequence

6. Concluding Remarks

We have demonstrated that the body posture estimation from voxel data is robust and convenient. Since the voxel data is in the world coordinate system, algorithms that take advantage of the knowledge of average dimensions and shape of some parts of the human body are easily implemented. This leads to effective model acquisition and voxel labeling algorithms. The use of the Bayesian network to impose known human body proportions during model acquisition phase gives excellent results. The twist-based human body model leads to a simple extended Kalman filter formulation and with imposed angle limits guarantees physically valid posture estimates. The framework is easily expandable for more detailed body models, which are sometimes needed.

References

1. Gavrila, D.: Visual Analysis of Human Movement: A Survey. Computer Vision and Image Understanding, vol. 73, no. 1 (1999) 82-98
2. Kakadiaris, I., Metaxas, D.: Model-Based Estimation of 3D Human Motion with Occlusion Based on Active Multi-Viewpoint Selection. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, California (1996)
3. Delamarre, Q., Faugeras, O.: 3D Articulated Models and Multi-View Tracking with Physical Forces. Computer Vision and Image Understanding, Vol. 81, No. 3 (2001) 328-357
4. Bregler, C.: Learning and Recognizing Human Dynamics in Video Sequences. IEEE Conference on Computer Vision and Pattern Recognition, San Juan, Puerto Rico, (1997)
5. Deutscher, J., Blake, A., Reid, I.: Articulated Body Motion Capture by Annealed Particle Filtering. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina (2000)
6. Gavrila, D., Davis, L.: 3D model-based tracking of humans in action: a multi-view approach. IEEE Conference on Computer Vision and Pattern Recognition, San Francisco, California (1996)
7. Covell, M., Rahimi, A., Harville, M., Darrell, T.: Articulated-pose estimation using brightness- and depth-constancy constraints. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina (2000)
8. Jojić, N., Turk, M., Huang, T.: Tracking Self-Occluding Articulated Objects in Dense Disparity Maps. IEEE International Conference on Computer Vision, Corfu, Greece (1999)
9. Cheung, G., Kanade, T., Bouguet, J., Holler, M.: A Real Time System for Robust 3D Voxel Reconstruction of Human Motions. IEEE Conference on Computer Vision and Pattern Recognition, Hilton Head Island, South Carolina (2000)
10. Mikić, I., Trivedi, M., Hunter, E., Cosman, P.: Articulated Body Posture Estimation from Multi-Camera Voxel Data. IEEE Conference on Computer Vision and Pattern Recognition, Kauai, Hawaii (2001)
11. Bregler, C., Malik, J.: Tracking People with Twists and Exponential Maps. IEEE Conference on Computer Vision and Pattern Recognition, Santa Barbara, California (1998)
12. Murray, R. Li, Z., Sastry, S.: A Mathematical Introduction to Robotic Manipulation, CRC Press (1993)
13. Bar-Shalom, Y., Fortmann, T.: Tracking and Data Association. Academic Press (1987)