An Efficient Stochastic Framework For 3D Human Motion Tracking

Bingbing Ni, Stefan Winkler and Ashraf Ali Kassim

Department of Electrical and Computer Engineering
National University of Singapore
4 Engineering Drive 3, Singapore 117576

ABSTRACT

In this paper, we present a stochastic framework for articulated 3D human motion tracking. Tracking full body human motion is a challenging task, because the tracking performance normally suffers from several issues such as self-occlusion, foreground segmentation noise and high computational cost. In our work, we use explicit 3D reconstructions of the human body based on a visual hull algorithm as our system input, which effectively eliminates self-occlusion. To improve tracking efficiency as well as robustness, we use a Kalman particle filter framework based on an interacting multiple model (IMM). The posterior density is approximated by a set of weighted particles, which include both sample means and covariances. Therefore, tracking is equivalent to searching the maximum a posteriori (MAP) of the probability distribution. During Kalman filtering, several dynamical models of human motion (e.g., zero order, first order) are assumed which interact with each other for more robust tracking results. Our measurement step is performed by a local optimization method using simulated physical force/moment for 3D registration. The likelihood function is designed to be the fitting score between the reconstructed human body and our 3D human model, which is composed of a set of cylinders. This proposed tracking framework is tested on a real motion sequence. Our experimental results show that the proposed method improves the sampling efficiency compared with most particle filter based methods and achieves high tracking accuracy.

Keywords: Articulated 3D human motion tracking, IMM Kalman particle filter, simulated physical force/moment

1. INTRODUCTION

Multiple view based, marker-less articulated human motion tracking has attracted a growing research interest in recent years, primarily because of the large number of potential applications such as motion capture, human computer interaction, virtual reality, smart surveillance systems etc. Due to the high dimensionality of human body motion, 3D tracking is inherently a difficult problem, and various different methods have been proposed.[1]

A large number of methods based on deterministic search have been developed; they include space decomposition,[2] incremental tracking,[3] exponential maps,[4] and image forces.[5–7] Taking as inputs the segmented human figure or reconstructed 3D human body (e.g., surface points or voxels), these methods search for the best configuration of the 3D human pose which is consistent with the observation data. Their objective functions to minimize are therefore usually defined either as the matching score between the human contours with the projected edges of the human model, or the fitting score between the 3D human model with the reconstructed human body.

However, due to an inherent limitation (only local optima are guaranteed), these deterministic methods normally lack robustness. They are sensitive to image noise, foreground segmentation errors, self-occlusion, and may easily lose track due to the problem of error accumulation.[3] Therefore, many algorithms based on stochastic sampling have been proposed to address these problems, including particle filter,[8–10] unscented Kalman filter,[11,12] belief propagation,[13] Markov network[14] etc.

The particle filter technique,[15] which is designed for high dimensional, nonlinear and non-Gaussian tracking problems, approximates the underlying probability distribution via a number of weighted particles and thus

Corresponding author: B. Ni (g0501096@nus.edu.sg).
S. Winkler is now with Symmetricom, San Jose, CA 95131.

avoids the analytical inference of the posterior density. Theoretically, the number of particles required increases exponentially when the dimensionality of the tracking problem increases. Therefore, efficient sampling schemes are required when dealing with a high dimensional tracking problem such as human motion tracking. Previous works use techniques based on simulated annealing,[8] analytical inference,[10] sampling space adaption,[13] sample-and-refine[16] and others to reduce the number of particles required. Also, prior information such as human motion dynamical models[9,17] or human appearance models[18] can be used to further improve sampling efficiency.

In our work, the input data are the surface points and surface normals reconstructed from multiple-view images of the human body. Our likelihood function encodes the fitting score between the reconstructed surface points and the 3D model as well as the surface direction alignment between the model and the observation. Given that human motion dynamics are relatively complex, a single motion model sometimes fails to capture the true motion state. Therefore we use a Kalman particle filter based on an interacting multiple model (IMM),[19] which incorporates different motion models, instead of using a standard particle filter algorithm such as CONDENSATION.[15]

The probability distribution of the human motion state is represented by a set of weighted particles, each of which has a sample mean and a covariance matrix associated with each model to represent the pose and velocity vector and its uncertainty. The sample means and covariances are mixed and updated adaptively according to different system models (e.g., zero order dynamics, linear dynamics) in the Kalman particle filtering framework. In particular, the Kalman prediction step is carried out by assuming zero-order and first-order (linear) dynamics, and the estimated motion state can frequently change modes according to the interaction of these models. Based on the prediction, the measurement step during Kalman filtering is accomplished by a 3D registration method based on simulated physical force/moment that we described elsewhere.[20] We show that this stochastic tracking framework can achieve both high accuracy and robustness.

The paper is organized as follows: Section 2 describes our 3D human model. Section 3 gives a brief introduction of the 3D reconstruction method. Section 4 introduces our IMM based Kalman particle filtering framework and the simulated physical force/moment based measurement method. Section 5 presents our experimental results for a human motion sequences, and Section 6 concludes the paper.

2. 3D HUMAN MODEL

Our human body model consists of a combination of 10 cylinders (the torso can be regarded as a degenerate cylinder since it has an elliptical cross-section), as illustrated in Figure 1. The global coordinate system originates at the center of the torso. Local coordinate frames are defined for each joint of the adjacent body parts.

We further define kinematic constraints for each joint, i.e., movement is restricted to 25 degrees of freedom (DoF), the joint angles are limited to certain ranges. We consider 6 DoF for the torso (global translation and rotation), 3 DoF for upper arms, legs and head (rotating about their X, Y and Z axes), and 1 DoF for lower arms and legs (they are only allowed to rotate about their X axes).

The entire 3D pose of the body is determined by a 25-dimensional pose vector:

$$\mathbf{x}_p = (t_{0x}, t_{0y}, t_{0z}, \theta_{0x}, \theta_{0y}, \theta_{0z}, \theta_{1x}, \theta_{1y}, \theta_{1z}, \theta_{2x}, ...)^T, \tag{1}$$

which contains the joint angles of shoulders, elbows, hips, and knees, plus the global position and orientation of the torso.

3. 3D RECONSTRUCTION OF THE HUMAN BODY

The inputs to our tracking framework are the sparsely reconstructed human surface points and surface normals. These reconstruction data can be obtained via 3D reconstruction algorithms given multiple synchronized images and camera calibration parameters. Segmented human silhouettes can be computed by the foreground detection method[21] provided with background statistics.

We adopt the well known visual hull method[22–24] to reconstruct the 3D scene points as well as their surface normal vectors. These surface reconstruction points are obtained by intersecting the viewing cones from each view, and the corresponding normals are given by the cross product between the viewing lines and the tangent

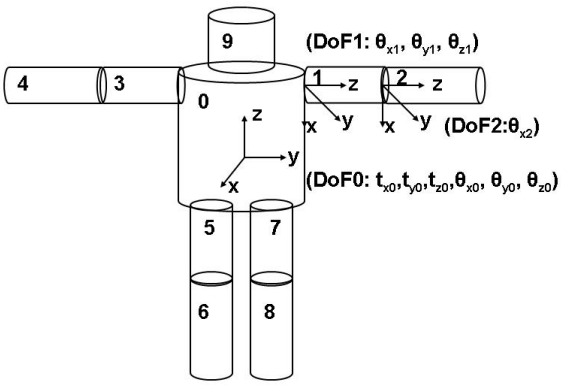


Figure 1. Articulated 3D model of the human body. The global coordinate system originates at the center of the torso; local coordinate frames are defined for body part. Each joint is subject kinematic constraints (see text for details).

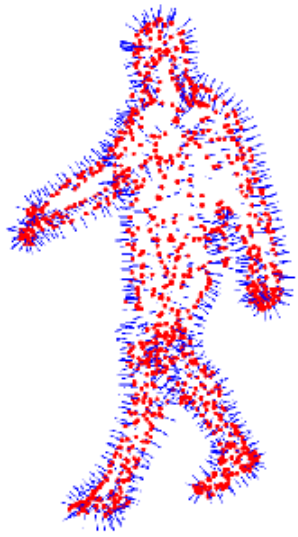


Figure 2. Visual hull reconstruction of the human body, showing surface points (red dots) and normals (blue arrows).

to the image silhouette. Figure 2 shows an example of a 3D reconstruction of the human body. For more details about the reconstruction method, kindly refer to the references.[23,25]

4. HUMAN MOTION TRACKING FRAMEWORK

In the Bayesian framework, the task of human motion tracking can be formulated as inferring the maximum a posteriori (MAP) of the probability density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ given the image observation sequence $\mathbf{y}_{1:t} = (\mathbf{y}_1, \mathbf{y}_2, ...\mathbf{y}_t)$. Here the state vector $\mathbf{x}_t$ denotes the pose vector $\mathbf{x}_{pt}$ augmented with the velocity vector $\mathbf{x}_{vt}$, i.e., $\mathbf{x}_t = (\mathbf{x}_{pt}^T, \mathbf{x}_{vt}^T)^T$. Provided with the previous estimation of the density $p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})$, inferring the posterior density of the current frame is expressed as:

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) = \kappa\ p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1}) p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}, \tag{2}$$

where κ is a normalization constant. $p(\mathbf{y}_t|\mathbf{x}_t)$ is a likelihood term, which measures the probability of observing $\mathbf{y}_t$ given the motion state vector $\mathbf{x}_t$. $p(\mathbf{x}_t|\mathbf{x}_{t-1}, \mathbf{y}_{1:t-1})$ models the transition probability of the motion dynamics.

Although it would be impossible to develop an analytical expression of the posterior density, in a particle filter framework we can approximate it by a set of weighted state vectors called *particles*. Each particle i is denoted as $(\mathbf{x}_t^{(i)}, \omega_t^{(i)})$, representing the motion state and the associated likelihood weight, and these particles are propagated throughout consecutive frames via sequential importance sampling.[26] More specifically, in an IMM Kalman particle filtering framework, each particle is represented by a set of sample means and covariance matrices associated with each model, as well as the likelihood weight, i.e., $(\mathbf{x}_{1,t}^{(i)}, P_{1,t}^{(i)}, ..., \mathbf{x}_{j,t}^{(i)}, P_{j,t}^{(i)}, ..., \mathbf{x}_{M,t}^{(i)}, P_{M,t}^{(i)}, u_{1,t}^{(i)}, ..., u_{j,t}^{(i)}, ...u_{M,t}^{(i)}, \omega_t^{(i)})$, where the subscript j denotes the model index, and $u_j^{(i)}$ denotes the probability of being in model j. Each covariance matrix adaptively controls the sampling step and is recursively updated via Kalman filtering.

We now give detailed descriptions of our model interaction, Kalman filtering, model combination, measurement step and likelihood function.

4.1 Model Interaction

Before the inference of the current frame t, all the particles sampled from the last frame $t-1$ are first mixed according to their model probabilities as follows:

$$u_{k|j,t-1}^{(i)} = \frac{p_{kj} u_{k,t-1}^{(i)}}{c_j^{(i)}}, \tag{3}$$

where $u_{k|j,t-1}^{(i)}$ denotes the mixture probability from model j to k for particle i, p_{kj} is the model transition probability, and the normalization factor $c_j^{(i)}$ is defined as:

$$c_j^{(i)} = \sum_{k=1}^{M} p_{kj} u_{k,t-1}^{(i)}. \tag{4}$$

The mixed mean $\mathbf{x}_{0j,t-1}^{(i)}$ and covariance $P_{0j,t-1}^{(i)}$ are expressed as:

$$\mathbf{x}_{0j,t-1}^{(i)} = \sum_{k=1}^{M} u_{k|j,t-1}^{(i)} \mathbf{x}_{k,t-1}^{(i)}, \tag{5}$$

$$P_{0j,t-1}^{(i)} = \sum_{k=1}^{M} u_{k|j,t-1}^{(i)} (\mathbf{P}_{k,t-1}^{(i)} + (\mathbf{x}_{k,t-1}^{(i)} - \mathbf{x}_{0j,t-1}^{(i)})(\mathbf{x}_{k,t-1}^{(i)} - \mathbf{x}_{0j,t-1}^{(i)})^T). \tag{6}$$

4.2 Kalman Filtering

After we get the set of mixed means and covariances for each particle and each model, we perform Kalman filtering according to each model's dynamical equation. Specifically, the process and measurement are updated via the following well-known Kalman equations:[27]

$$\mathbf{x}_t = F\mathbf{x}_{t-1} + \mathbf{v}_{t-1}, \tag{7}$$

$$\mathbf{z}_t = H\mathbf{x}_t + \mathbf{n}_t. \tag{8}$$

Here $\mathbf{v}_{t-1}$ and $\mathbf{n}_t$ are process and measurement noise, which have zero mean, and their associated covariances are Q_{t-1} and R_t, respectively. F denotes the state transition matrix, which in our case can take two possible forms. In the first-order (i.e., linear) dynamics model,

$$F = \begin{pmatrix} I & I \\ 0 & I \end{pmatrix},$$

and in the zero-order dynamics (i.e., static) model,

$$F = \begin{pmatrix} I & 0 \\ 0 & I \end{pmatrix}.$$

Each I is a 25×25 identity matrix. $H = (I \;\; 0)$ denotes the measurement model for both cases, which means that only the pose component of the motion state vector is measured.

For each particle $\mathbf{x}_{j,t-1}^{(i)}$, the predicted means and covariances are computed, according to different motion dynamical models:

$$\mathbf{x}_{j,t|t-1}^{(i)} = F_j \mathbf{x}_{j,t-1}^{(i)}, \tag{9}$$

$$P_{j,t|t-1}^{(i)} = Q_{j,t-1}^{(i)} + F_j P_{j,t-1}^{(i)} F_j^T. \tag{10}$$

The Kalman updates of the means and covariances are made for each particle based on the measurement $\mathbf{z}_t^{(i)}$ according to:

$$\mathbf{x}_{j,t}^{(i)} = \mathbf{x}_{j,t|t-1}^{(i)} + K_{j,t}^{(i)}(\mathbf{z}_{j,t}^{(i)} - H_j\mathbf{x}_{j,t|t-1}^{(i)}), \tag{11}$$

$$P_{j,t}^{(i)} = P_{j,t|t-1}^{(i)} - K_{j,t}^{(i)}H_jP_{j,t|t-1}^{(i)}, \tag{12}$$

$$K_{j,t}^{(i)} = P_{j,t|t-1}^{(i)}H_j^T(H_jP_{j,t|t-1}^{(i)}H_j^T + R_{j,t}^{(i)})^{-1} \tag{13}$$

The new particle $\mathbf{x}_{j,t}^{(i)}$ is drawn according to the Gaussian proposal distribution based on the updated means and covariances:

$$\mathbf{x}_{j,t}^{(i)} \sim \mathcal{N}(\mathbf{x}_{j,t}^{(i)}, P_{j,t|t-1}^{(i)}). \tag{14}$$

4.3 Model Combination

For each newly generated particle $\mathbf{x}_{j,t}^{(i)}$, their models are combined to yield the mixture probability $u_{j,t}^{(i)}$ of the current frame, as well as the mean $\overline{\mathbf{x}_t^{(i)}}$ and covariance $\overline{P_t^{(i)}}$:

$$u_{j,t}^{(i)} = \frac{1}{c}\Lambda_{j,t}^{(i)}\sum_{k=1}^{M} p_{kj}u_{k,t-1}, \tag{15}$$

$$\overline{\mathbf{x}_t^{(i)}} = \sum_{k=1}^{M} u_{k|j,t}\mathbf{x}_{k,t}^{(i)}, \tag{16}$$

$$\overline{P_t^{(i)}} = \sum_{k=1}^{M} u_{k|j,t}(\mathbf{P}_{k,t}^{(i)} + (\mathbf{x}_{k,t}^{(i)} - \mathbf{x}_{0j,t}^{(i)})(\mathbf{x}_{k,t}^{(i)} - \mathbf{x}_{0j,t}^{(i)})^T. \tag{17}$$

Here, c is a normalization factor which can be expressed as:

$$c = \sum_{j=1}^{M}\Lambda_{j,t}^{(i)}c_j^{(i)} \tag{18}$$

where $c_j^{(i)}$ is defined in Eq. (4).

The innovation is assumed to be a Gaussian distribution with zero mean and covariance $S_{j,t}^{(i)}$, i.e., $\mathcal{N}(0, S_{j,t}^{(i)})$, and the model likelihood $\Lambda_{j,t}^{(i)}$ is therefore denoted as:

$$\Lambda_{j,t}^{(i)} = \mathcal{N}(r_{j,t}^{(i)}; 0, S_{j,t}^{(i)}), \tag{19}$$

where the innovation $r_{j,t}^{(i)}$ and the residual covariance $S_{j,t}^{(i)}$ are defined as:

$$r_{j,t}^{(i)} = \mathbf{z}_{j,t}^{(i)} - H_jF_j\mathbf{x}_{j,t-1}^{(i)}, \tag{20}$$

$$S_{j,t}^{(i)} = H_jP_{j,t|t-1}^{(i)}H_j^T + R_{j,t}^{(i)}. \tag{21}$$

4.4 Measurement Step

Starting from the prediction $\mathbf{x}_{j,t|t-1}^{(i)}$, we obtain the measurement vector $\mathbf{z}_{j,t}^{(i)}$ by applying a 3D registration method based on simulated physical force/moment, which was proposed in our previous work.[20] Our registration method is based on the well-known iterative closest points (ICP) concept,[28] which can align a model with scene points in an iterative manner.

Suppose a scene point is denoted by $\mathbf{p}$, and the corresponding point (i.e., with the closest distance) on the model is denoted as $\mathbf{p}'$. We create a force that is a function of their displacement:

$$\overrightarrow{F} = \rho\left(m(\overrightarrow{\mathbf{n_p}}, \overrightarrow{\mathbf{n_{p'}}})\,|\mathbf{pp}'|\right)\overrightarrow{\mathbf{a_{pp'}}}. \tag{22}$$

Here $|\overrightarrow{\mathbf{pp}'}|$ denotes the Euclidean distance between the scene point $\mathbf{p}$ and the point on the model $\mathbf{p}'$, $\overrightarrow{\mathbf{a_{pp'}}}$ is the unit vector pointing to $\overrightarrow{\mathbf{pp}'}$, and $m(\overrightarrow{\mathbf{n_p}}, \overrightarrow{\mathbf{n_{p'}}})$ is a modulation factor of the form $m(\overrightarrow{\mathbf{n}_1}, \overrightarrow{\mathbf{n}_2}) = \cos(\overrightarrow{\mathbf{n}_1}, \overrightarrow{\mathbf{n}_2})$ that accounts for the alignment of the surface normals between the model and the scene. The magnitude term is further embedded into a robust function $\rho(x)$ to suppress the effect of outliers. We have chosen a truncated quadratic function, $\rho(x) = \min(x^2, \beta)$, where β is some upper bound.

Similarly, the moment we create can be denoted as:

$$\overrightarrow{M} = \overrightarrow{F}L, \tag{23}$$

where L is the vertical distance from force $\overrightarrow{F}$ to the rotation center of each joint.

As in the ICP, we iteratively compute the closest points and then update the pose measurement vector according to the estimated transform. During each iteration step, the displacements between all 3D scene points and the model are calculated, and all forces and moments are summed up, resulting in a translation and a rotation vector, and we transform the pose measurement vector in a hierarchical manner.[20]

Furthermore, we also limit the joint angles in the updating step according to the kinematic constraints. Given the pose measurement vector $\widehat{\mathbf{z}_{j,t}^{(i)}}$ estimated from the previous iteration and the updating vector $\delta\mathbf{z}$ generated by our registration algorithm, we clamp the new pose vector to avoid the violation of any constraint by the following inequality:

$$\mathbf{z}_{lb} \preceq \widehat{\mathbf{z}_{j,t}^{(i)}} + \delta\mathbf{z} \preceq \mathbf{z}_{ub}, \tag{24}$$

where $\mathbf{z}_{lb}$ and $\mathbf{z}_{ub}$ are the lower and upper bounds of the joint angles.

For each prediction $\mathbf{x}_{j,t|t-1}^{(i)}$, we generate the measurement $\mathbf{z}_{j,t}^{(i)}$ by the above method. It normally takes less than 5 iterations to get a measurement $\mathbf{z}_{j,t}^{(i)}$, which is efficient.

4.5 Likelihood Function

Given a set of 3D reconstruction points with surface normals for each frame and a 3D human model composed of a set of connected body parts, our likelihood function is designed as:

$$p(\mathbf{y}_t|\mathbf{x}_t) = \kappa e^{-D(M,S)/\sigma^2}, \tag{25}$$

where κ is some normalization constant. M and S denote the model and the set of scene points, respectively; σ is the variance. The distance term $D(M, S)$ is computed as:

$$D(M, S) = \frac{1}{N}\sum_i d(\mathbf{p}_i, \mathbf{p}_i'), \tag{26}$$

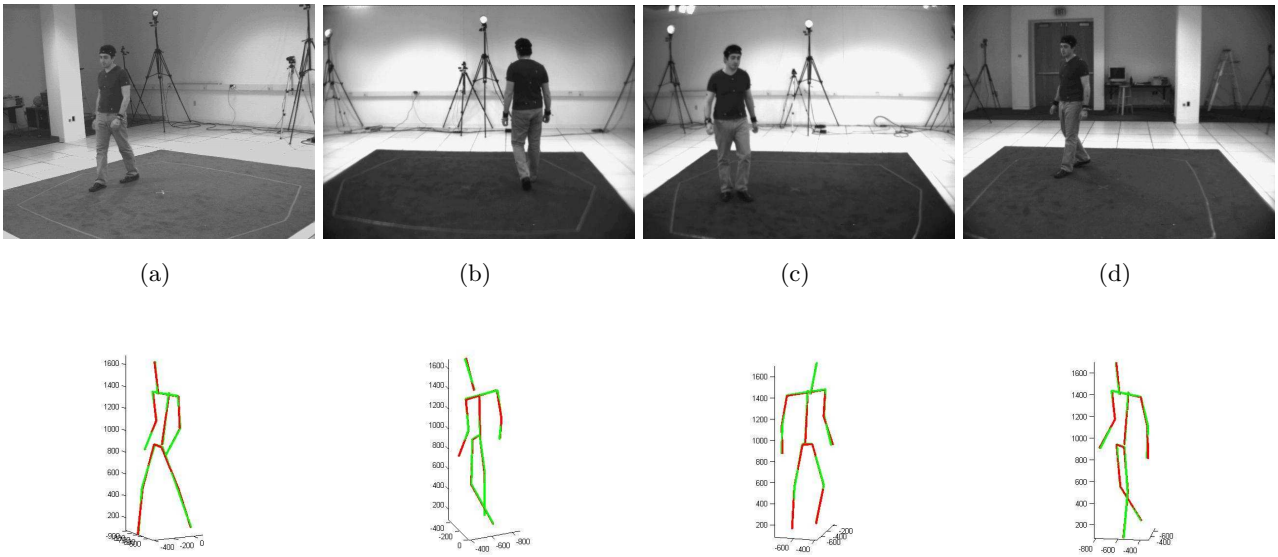
which can be regarded as the average distance between a 3D scene point $\mathbf{p}_i$ and the corresponding point on the 3D model $\mathbf{p}_i'$. $d(\mathbf{p}_i, \mathbf{p}_i')$ can be expressed as:

$$d(\mathbf{p}_i, \mathbf{p}_i') = \rho\left(\left(s(\overrightarrow{\mathbf{n_{p_i}}}, \overrightarrow{\mathbf{n_{p_i'}}})\left|\overrightarrow{\mathbf{p}_i\mathbf{p}_i'}\right|\right)^2\right). \tag{27}$$

The function s is used to penalize the misalignment of surface normals between the model and the scene:

$$s(\overrightarrow{\mathbf{n}_1}, \overrightarrow{\mathbf{n}_2}) = 1 - \epsilon\cos(\overrightarrow{\mathbf{n}_1}, \overrightarrow{\mathbf{n}_2}) \tag{28}$$

with $0 < \epsilon < 1$.



(a) (b) (c) (d)

Figure 3. Examples of tracking results for the walking sequence. The top row shows the captured views; the bottom row shows the corresponding tracking results. Our estimated human pose is shown in green and the ground truth pose in red.

5. RESULTS AND DISCUSSION

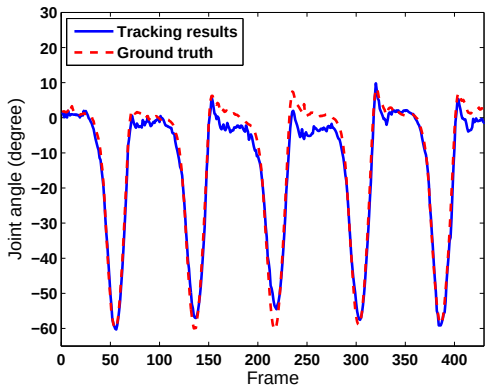
We validate our proposed method using the real human motion dataset HumanEva,[29] which also provides the ground truth of the human motion. We choose the walking sequence of subject 2 with 430 frames for our validation. The videos were captured by 7 synchronized digital cameras surrounding the scene with a resolution of 640×480 pixels. The 3D human surface reconstruction points as well as the corresponding surface normals are computed via the method presented here. We assume two motion models, namely zero-order and first-order dynamics, in the IMM particle filter framework.

Examples of the tracking results are shown in Figure 3. It can be observed that our human tracking method performs well; the estimated poses closely follow the ground truth.

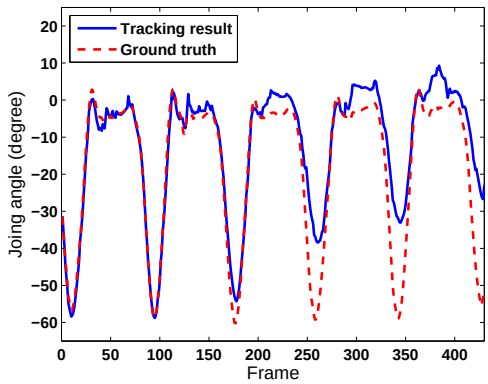
Figure 4 compares the ground truth and the estimated joint angles for both knees and elbows in the walking sequence. It can be observed that in the walking sequence, our estimated joint angles follow the periodical motion very accurately.

Figure 5 shows the combined root mean squared error (RMSE) of the estimated joint angles for the walking sequence. There is still some error accumulation over time, which we are currently working to resolve. The remaining variations in the plot are due to ambiguities of the lower arms in certain poses, which lead to increased misalignments. Despite these issues, the average RMSE of the estimated joint angles is 4.83 degrees, which shows the high accuracy of our tracking method compared with the quantitative results reported in previous works,[9,12,30] where the estimated RMSE for the joint angles are over 10 degrees.

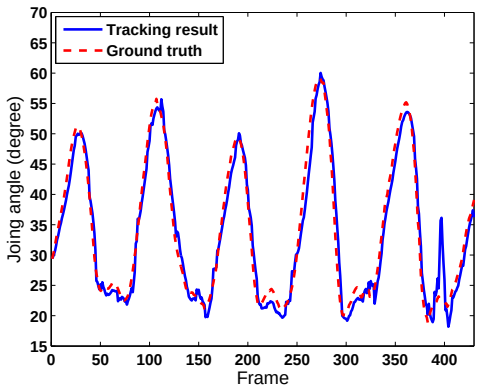
Our IMM particle filter based tracking framework is also efficient. Due to the application of IMM based particle filter together with our 3D registration method, only 20 particles are enough to track all the motion sequences successfully. Previous works based on particle filters[8–10,13,14,17,31] require a much larger number of particles (hundreds or even thousands).



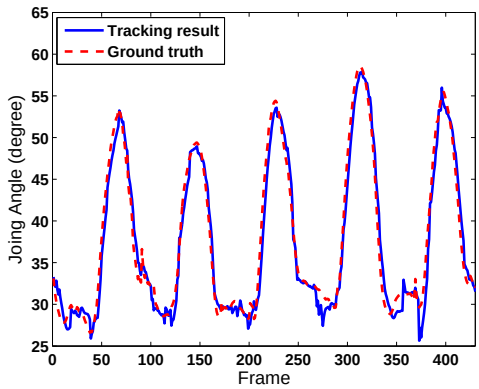
(a) Left knee



(b) Right knee

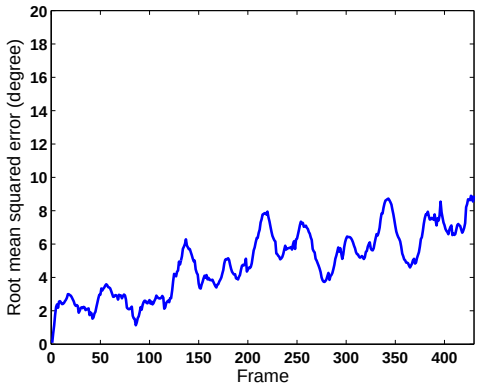


(c) Left elbow

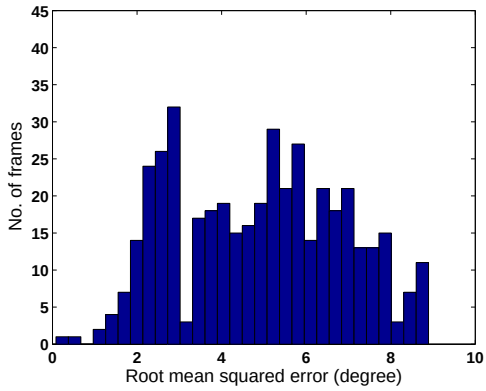


(d) Right elbow

Figure 4. Comparison between the ground truth and the estimated joint angles for the walking sequence.



(a) Development over time



(b) Histogram

Figure 5. RMSE for the walking sequence.

6. CONCLUSIONS

Tracking full body human motion is a challenging task given its high dimensionality. We proposed a novel tracking framework using Kalman particle filtering based on interacting multiple models. By assuming different dynamical models associated with the human motion, the tracking results for a real walking sequence have shown the accuracy of our proposed method. The application of our simulated physical force/moment based registration algorithm further reduces the number of samples required, since each particle is guaranteed to approach its local peak by this iterative registration method. Future work will concentrate on tests with additional motion capture sequences as well as the modeling of the probability of human motion dynamics, which can be used as a prior information for more efficient tracking.

REFERENCES

1. T. Weingaertner, S. Hassfeld, and R. Dillmann, "Human motion analysis: A review," in *Proc. IEEE Workshop on Motion of Non-Rigid and Articulated Objects (NAM)*, p. 90, 1997.
2. D. M. Gavrila and L. S. Davis, "3-D model-based tracking of humans in action: A multi-view approach," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 73–80, (San Francisco, CA, USA), June 1996.
3. M. Yamamoto, A. Sato, S. Kawada, T. Kondo, and Y. Osaki, "Incremental tracking of human actions from multiple views," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2–7, (Santa Barbara, CA, USA), 1998.
4. C. Bregler and J. Malik, "Tracking people with twists and exponential maps," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8–15, (Santa Barbara, CA, USA), 1998.
5. Q. Delamarre and O. Faugeras, "3D articulated models and multi-view tracking with silhouettes," in *Proc. International Conference on Computer Vision (ICCV)*, pp. 716–721, (Corfu, Greece), September 1999.
6. Q. Delamarre and O. Faugeras, "3D articulated models and multi-view tracking with physical forces," *Computer Vision and Image Understanding* **81**(3), pp. 328–357, 2001.
7. L. Kakadiaris and D. Metaxas, "Model-based estimation of 3D human motion," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **22**(12), pp. 1453–1459, 2000.
8. J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 126–133, (Hilton Head, SC, USA), June 2000.
9. J. Gall, B. Rosenhahn, T. Brox, and H. P. Seidel, "Learning for multi-view 3D tracking in the context of particle filters," in *Proc. Second International Symposium on Advances in Visual Computing*, pp. 59–69, (Lake Tahoe, NV, USA), November 2006.
10. M. W. Lee, I. Cohen, and S. K. Jung, "Particle filter with analytical inference for human body tracking," in *Proc. Workshop on Motion and Video Computing*, pp. 159–165, (Orlando, FL, USA), 2002.
11. B. Stenger, P. R. S. Mendonca, and R. Cipolla, "Model-based 3D tracking of an articulated hand," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 310–315, (Kauai, HI, USA), 2001.
12. J. Ziegler, K. Nickel, and R. Stiefelhagen, "Tracking of the articulated upper body on multi-view stereo image sequences," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 774–781, (New York, NY), June 2006.
13. T. X. Han and T. S. Huang, "Articulated body tracking using dynamic belief propagation," in *Proc. IEEE International Workshop on Human-Computer Interaction*, pp. 26–35, (Beijing, China), October 2005.
14. Y. Wu, G. Hua, and T. Yu, "Tracking articulated body by dynamic Markov network," in *Proc. International Conference on Computer Vision (ICCV)*, pp. 1094–1101, (Nice, France), October 2003.
15. M. Isard and A. Blake, "CONDENSATION – conditional density propagation for visual tracking," *International Journal of Computer Vision* **29**(1), pp. 5–28, 1998.
16. C. Sminchisescu and B. Triggs, "Covariance scaled sampling for monocular 3D body tracking," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 447, (Kauai, HI, USA), 2001.
17. H. Sidenbladh, M. J. Black, and L. Sigal, "Implicit probabilistic models of human motion for synthesis and tracking," in *Proc. European Conference on Computer Vision (ECCV)*, p. 784, (Copenhagen, Denmark), May 2002.

18. H. Lim, O. I. Camps, M. Sznaier, and V. I. Morariu, "Dynamic appearance modeling for human tracking," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 751–757, (New York, NY), 2006.
19. B. Chen and J. K. Tugnait, "Interacting multiple model fixed-lag smoothing algorithm for Markovian switching systems," *IEEE Transactions on Aerospace and Electronic Systems* **36**(1), pp. 243–250, 2000.
20. B. Ni, S. Winkler, and A. A. Kassim, "Articulated object registration using simulated physical force/moment for 3D human motion tracking," in *2nd Workshop on Human Motion Understanding, Modeling, Capture and Animation (in conjunction with ICCV)*, *Lecture Notes in Computer Science* **4814**, (Rio de Janeiro, Brazil), October 20, 2007.
21. C. Stauffer and W. E. L. Grimson, "Adaptive background mixture models for real-time tracking," in *Proc. Conference on Computer Vision and Pattern Recognition (CVPR)*, p. 2246, (Ft. Collins, CO, USA), 1999.
22. A. Laurentini, "The visual hull concept for silhouette-based image understanding," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **16**(2), pp. 150–162, 1994.
23. J. S. Franco and E. Boyer, "Exact polyhedral visual hulls," in *Proc. British Machine Vision Conference*, 2003.
24. W. Matusik, C. Buehler, R. Raskar, S. J. Gortler, and L. McMillan, "Image-based visual hulls," in *Proc. 27th Annual Conference on Computer Graphics and Interactive Techniques (SIGGRAPH)*, pp. 369–374, (New Orleans, LA, USA), 2000.
25. M. Niskanen, E. Boyer, and R. Horaud, "Articulated motion capture from 3-D points and normals," in *Proc. British Machine Vision Conference*, 2004.
26. A. Doucet, "On sequential Monte Carlo methods for Bayesian filtering," Tech. Rep. 5500, Cambridge University, 1998.
27. R. Signals, R. Estab, and U. K. Malvern, "An introduction to Kalman filters," *Measurement and Control* **19**(2), pp. 69–73, 1986.
28. P. J. Besl and H. D. McKay, "A method for registration of 3-D shapes," *IEEE Transactions on Pattern Analysis and Machine Intelligence* **14**(2), pp. 239–256, 1992.
29. L. Sigal and M. J. Black, "HumanEva: Synchronized video and motion capture dataset for evaluation of articulated human motion," Tech. Rep. CS-06-08, Brown University, 2006.
30. J. P. Luck, C. Debrunner, W. Hoff, Q. He, and D. E. Small, "Development and analysis of a real-time human motion tracking system," in *Proc. IEEE Workshop on Applications of Computer Vision*, (Orlando, FL, USA), 2002.
31. H. Sidenbladh, M. J. Black, and D. J. Fleet, "Stochastic tracking of 3D human figures using 2D image motion," in *Proc. European Conference on Computer Vision (ECCV)*, pp. 702–718, (Dublin, Ireland), 2000.