Particle Filter with Analytical Inference for Human Body Tracking

Mun Wai Lee, Isaac Cohen and Soon Ki Jung[1]
Institute for Robotics and Intelligent Systems
Integrated Media Systems Center
University of Southern California
Los Angeles, CA 90089-0273
{munlee|icohen|soonki}@usc.edu

Abstract

This paper introduces a framework that integrates analytical inference into the particle filtering scheme for human body tracking. The analytical inference is provided by body parts detection, and is used to update subsets of state parameters representing the human pose. This reduces the degree of randomness and decreases the required number of particles. This new technique is a significant improvement over the standard particle filtering with the advantages of performing automatic track initialization, recovering from tracking failures, and reducing the computational load.

1. Introduction

Human body motion tracking and analysis has received a significant amount of attention in the computer vision research community in the past decade. This has been motivated mainly by the desire of understanding human pose and gestures for building the next generation user interface. Inspired from human to human interactions, such an interface will go beyond the mouse-keyboard interaction, defining a system that responds naturally to user gestures. Naturally other applications related to a marker-less capture of the human body motion can be considered within the presented framework (refer to [1] for a survey).

We present in this paper the first step towards such an interface by providing a robust human body tracking from set of synchronized and geometrically registered video streams. The method relies on inferring an articulated body model from: the observed silhouettes and an analytical inference of human body parts.

An articulated human body model (stick-figure) is often used for detailed motion capture. Indeed, it provides an effective representation of the physical structure and constraints of the human body. Fitting and tracking the articulated body model becomes a problem of estimating the state vector describing the human pose, where each state parameter represents one degree of freedom (e.g. joint angle) of the human model. However, the large number of degrees of freedom associated with the model cannot be all analytically inferred from image features.

Various methods have been proposed to address this problem, which rely on introducing additional constraints to reduce the state space, using learned dynamic models [2] or PCA-based dimensional reduction [3]. However, these methods restrict the pose space and are not suitable for a general motion capture application.

The particle filter technique [4][5][6] is a promising method for human body tracking [7] because it avoids complex analytical computations. Based on the Monte-Carlo simulation, particle filter provides a suitable framework for state estimation in a nonlinear, non-Gaussian system [6].

However particle filter requires an impractically large number of particles to sample the high dimensional state space effectively; otherwise, it is easy to lose track and difficult to recover tracking failure because of sample depletion in the state space. In addition, particle filter requires an accurate model initialization. Often, initialization is done manually, which is undesirable in many applications. Recent works that used particle filter for human tracking have focused on improving efficiency using variance analysis [8] and simulated annealing

[1] Department of Computer Engineering Kyungpook National University 1370 Sankyuk-dong Buk-gu Daegu 702-701 South Korea skjung@knu.ac.kr

0-7695-1860-5/02 $17.00 © 2002 **IEEE**

approach [9]. Another work [10] uses 2D image motion as features to improve the likelihood measure.

In this paper, we propose a novel method for human body tracking which combines particle filter with analytical inference techniques. We propose methods for detecting body parts such as the head, the hands and torso. The results of these detections are used to make inference on a subset of state parameters corresponding to the observed human body pose. This additional inference is used to improve the state estimation within the particle filtering framework. The basic modules of this proposed framework are illustrated in Figure 1, and will be described in details in Section 4.

There are several advantages of combining particle filter with analytical inference. Firstly, the inference helps to reduce the degree of freedom that is dependent on Monte-Carlo simulation during state estimation. This allows the use of a smaller number of particles and henceforth reduces the computational complexity. Secondly, the analysis is useful for automatic model initialization and recovery of lost tracks.

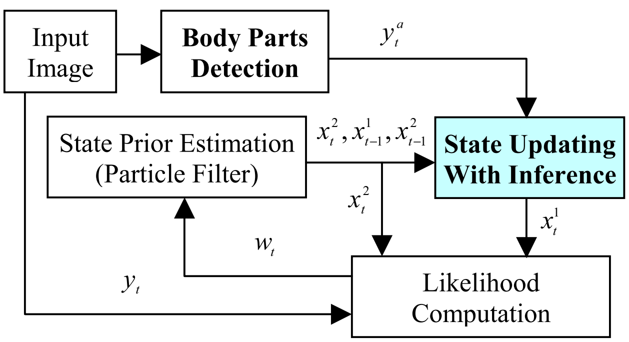


Figure 1: Particle filtering with inference. The notations are described in Section 4.

2. Particle Filter Articulated Model Fitting

2.1. Particle Filter

Particle filter, also known as the Condensation algorithm [5] is a robust online filtering technique, based on the Bayesian framework. This technique provides a suitable basic framework for estimating the degrees of freedom of an articulated body model: Particle filter estimates the states by recursively updating sample approximations of posterior distribution.

The posterior distribution at time *t* is represented by a set of *N* particles denoted by $\{x_t^{(i)}\}_{i=1,\ldots,N}$, with weights $\{w_t^{(i)}\}_{i=1,\ldots,N}$. There are 3 basic steps: *selection, prediction* and *updating.*

Selection. Resample with replacement to produce the *N* particles $\{\tilde{x}_{t-1}^{(i)}\}_{i=1,\ldots,N}$, from the set $\{x_{t-1}^{(i)}\}_{i=1,\ldots,N}$. The probability of selecting a particle $x_{t-1}^{(i)}$ is proportional to its normalized weight $w_{t-1}^{(i)}$.

Prediction. The samples are updated according to a stochastic diffusion model,

$$x_t^{(i)} = \tilde{x}_{t-1}^{(i)} + \eta, \tag{1}$$

where η is a vector of standard normal random variables.

Updating. Given an observation y_t, the weights are updated by the likelihood estimates, $w_t^{(i)} \propto p(y_t | x_t^{(i)})$, and are normalized. The main problem in particle filtering is to define the appropriate likelihood estimate based on the available observations and priors.

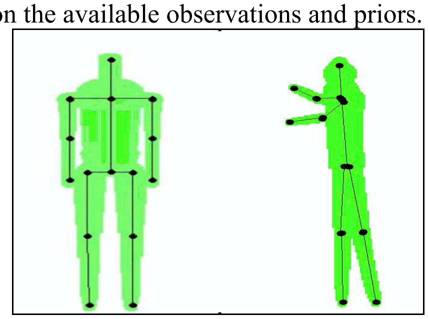


Figure 2: Articulated human body model.

2.2. Human Body Model

Various articulated body models have been proposed in the literature according to the targeted application. Some have a very small number of degrees of freedom and focus on the limbs [11] while others have proposed a model that contains hand and fingers joints.

The articulated human body model we use consists of 10 joints and 14 segments, representing the head, torso and limbs (Figure 2). A tapered 3D cone with an elliptical cross-section represents each segment. The model has 32 degrees of freedom that include the global translation, rotation and scale, and local joint rotations.

Fitting and tracking the articulated model to the detected humans in the video streams requires the definition of a likelihood function allowing to map the degree of freedom of the model onto image properties.

The likelihood computation is based on two components: the foreground boundary of the moving person and the detected silhouette regions.

Matching of Foreground Boundary. This involves matching the boundary of the foreground in the input images to the predicted silhouette boundary of the human body model.

For each particle $x_t^{(i)}$, we compute the position of each segment of the human model, $S_t^{(i)} = \{s_1, \ldots, s_{N_S}\}$, where N_s =14 is the number of segments of the human body model. For clarity, we omit the superscript (*i*) and subscript *t*.

Given a foreground segmentation, a set of contour points along the boundary of foreground is extracted, denoted by $C = \{c_j\}_{j=1,\ldots,N_c}$, where N_c is the number of contour points. Each contour point c_j is matched to the closest segment s_k, such that

$$k = \arg\min_l d(c_j, s_l),$$

where $d(c_j, s_l)$ is the distance of point c_j to the edge of segment s_l projected on the image. The similarity measure is given by,

$$p(c_j \mid s_k) = \exp[-d^2(c_j, s_k)/\sigma^2],$$

where σ^2 is the variance of model and input edge disparity. As the segmentation of moving objects characterizing the foreground regions is not error-free we have to accommodate for errors in the segmentation and account for outliers among the contour points *C*. Denoting $P_{outlier}$ as the probability of the point c_j being an outlier, the likelihood measure for contour point c_j given S is

$$p(c_j \mid S) = (1 - P_{outlier})\, p(c_j \mid s_k) + P_{outlier}.$$

The value of $P_{outlier}$ depends on the quality of foreground segmentation and is derived empirically. Combining the likelihood of all contour points, the likelihood measure for boundary matching is defined by:

$$L_{\text{Boundary}} = p(C \mid S) = \prod_{j=1}^{N_c} p(c_j \mid S).$$

The above likelihood measure is insufficient because it often leads to over-estimation of the human size. We augment this with the second likelihood component described as follows.

Matching of Predicted Silhouette Region. The human body model, when projected onto the 2D images, should lie inside the extracted foreground silhouette. The second likelihood component penalizes any part of the body model that lies outside the silhouette.

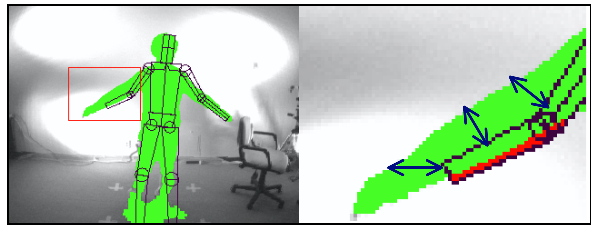
Given a predicted body pose, we compute the projection of the human body model on the 2D images and count the number of pixels, n_α, that are inside the projection but lie outside the detected foreground region. The likelihood is expressed as:

$$L_{\text{Region}} = (P_\alpha)^{n_\alpha},$$

where P_α is a probability of false negative errors in foreground extraction.

The combined likelihood similarity for matching the articulated body model to the detected silhouettes is given by:

$$L = L_{\text{Boundary}} \times L_{\text{Region}}.$$



(a) (b)

Figure 3: Likelihood computation. In (a), the foreground extraction is shown in light gray and a slightly misaligned human model is overlaid on the image for illustration. The enclosed box is enlarged and shown in (b), where the arrows illustrate boundary errors (1st likelihood component), and the dark shaded area at the bottom edge of the model's lower arm indicates region matching errors (2nd likelihood component).

3. Detection of Human Body Parts

The likelihood measure, defined in the previous section, provides a good estimation of the state vector characterizing the articulated model when provided with a correct initialization and there are sufficient particles sampling the state vector distribution. However, requiring a good initialization limits the use of such human body tracking to a set of specified body motion or postures and prevents us from using this system for human body motion capture. We would like to extend the system capabilities such that it does not require a manual initialization step and can overcome the problem of samples depletion during the tracking process.

We propose an approach that incorporates additional cues in order to perform automatic initialization and recover from body-parts tracking failures. In the following, we describe the use of hands, head and torso location for defining additional cues in the particle filter based tracking.

3.1. Hands Detection

Detection. The hands are detected along the outlines of the foreground. Peaks of convex curvature are extracted along the silhouette boundary. We match these curvature peaks in different images using epipolar constraints, and reconstruct their 3D positions. Using prior estimation of the hand positions, based on human body structure and tracking information, we can further eliminate unlikely hand positions.

This requires that the hand is visible in at least two camera views. As this is not always the case, we assign a prior probability P_n that there is an occlusion and therefore no correct hand measurement can be obtained.

The above method generates a number of hypotheses of hand position $\{y_{hand}\}$. These hypotheses are rated using prior information obtained from tracking. The prior probability distribution of the hand position is approximated by a multivariate Gaussian function:

$$\Pr(y_{hand}) \propto N(y_{hand}, \mu_{hand}, \Sigma_{hand}),$$

where μ_{hand} and Σ_{hand} are the mean and covariance of the hand position based on weighted particle samples. The probabilities are normalized so that the sum of probability is equal to (1- P_n), where P_n is the probability of no correct hand measurement.

Updating. In the following we show how to integrate this prior into the updating process of the particle states. For each given particle of state vector, we use a Monte Carlo method to select the hand position measurement y_{hand}, based on the probability $P(y_{hand})$. There is a probability P_n that none of the measurements is used, and the state parameters are propagated using stochastic dynamic model as in the standard particle filter, as expressed in (1) in Section 2.1.

The measurement is used to update the positions of hand and elbow, while keeping the rest of the body position unchanged. Given the positions of the hand and shoulder, the elbow lies along a circle. We choose the point on this circle that is closest to the prior estimate of the elbow position. This is equivalent to keeping the "azimuth" angle at the shoulder unchanged and updating three other joint angles (one at the elbow, two at the shoulder).

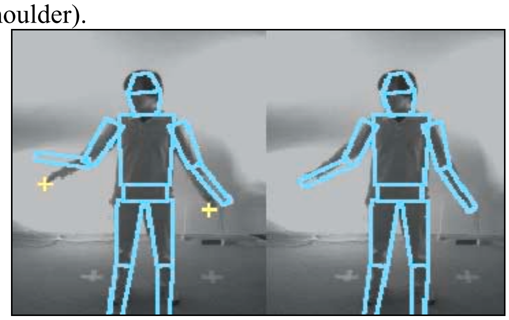


Figure 4: Hand detection. Left: Prior estimated pose with detected hands shown as crosses. Right: Updated pose.

3.2. Head Detection

Detection. The head detection is performed using a reference chain code representation of a head-shoulder contour (Figure 5) as a template for head. We match this template along the contour boundary of the extracted silhouette to detect the head. To achieve scale invariance, the contours are rescaled with respect to the estimated human height. The chain code features are normalized before comparison to achieve rotation invariance. Matching error is based on chain code differencing.

To obtain head position in 3D space, the head must be detected in at least two input images. Epipolar constraint is used to remove false measurements and to achieve more accurate localization.

In most cases, the head is accurately located. However, occlusions can occur and incomplete edge description of the head can be observed. We express the probability of false measurement as a zero mean Gaussian function of the matching error:

$$P_{head} = \frac{1}{\sqrt{2\pi}\sigma_c} \exp\left(\frac{e^2}{\sigma_c^2}\right)$$

where e is the error in chain code differencing. The variance σ_c^2 is obtained empirically. With this probability, a Monte Carlo method is used to decide whether to use the head measurement during updating, using the same approach as in the hand detection described previously.

Updating. A measurement of the head position in the 3D space provides three constraints for updating the state parameters. We choose to update the three degrees of freedom that are most related to the head position: the orientation of the head (2 dofs) and the position of neck along the body main axis (1 dof).

Change in neck position necessarily generates a change of the positions of torso and other body parts such as the shoulders and hips. Our strategy is to minimize changes in the positions of the "end-effectors" such as the hands and legs. We shift the torso only along its axis and compute the joint angles at the intermediate joints (shoulders, elbows, hips and knees) while keeping the hands and legs fixed.

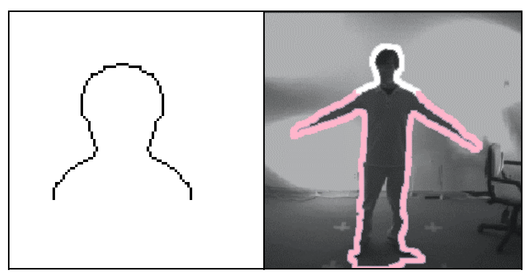


Figure 5: Head detection. Left: Template of head. Right: extracted silhouette boundary, and detected head.

3.3. Torso Detection

Detection. A simple method is used to extract the main axis of the torso. We first extract the medial axis of the 2D silhouettes. The medial axis points in different views are matched using epipolar constraint, and the 3D positions computed (refer to [12] for details.) A line is then fitted to these 3D points using PCA and RANSAC method (Figure 6). This extracted line provides a measurement of the torso orientation, and a constraint that the torso must lay along the line. This gives us four constraints, which are used to update the state parameters:

Updating. The position (2 dofs) and orientation (2 dofs) of the human model are updated so that the torso is aligned to the extracted medial axis. (Note that the position of torso along the axis, and the rotation around the axis remain unchanged.) During this update, we apply the same strategy as in head detection to keep the positions of hands and legs constant and update the angles at intermediate joints.

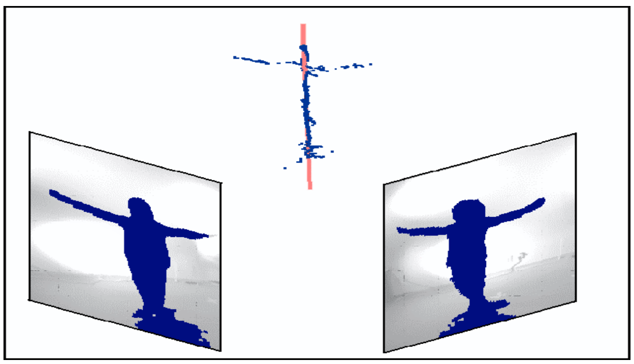


Figure 6: Torso detection. The medial axis of the silhouette in two views is extracted, matched and reconstructed in 3D. The torso axis is found by fitting a line to the medial axis points in 3D.

4. Particle Filter with Inference

As discussed in Section 1, Monte Carlo simulation is not appropriate for estimating high dimensional state parameters. The main idea of this paper is to use analytical computation (inferred from the detection of body parts) to infer a subset of the state parameters. This will reduce the degree of dependence on the Monte Carlo simulation.

In this section, we present a general framework on how the analysis result is incorporated into the particle-filtering scheme. We state the tracking problem as a Markovian state-space model, where x is the hidden state and y is the observation. At any time t, the posterior distribution is given by Bayes' theorem:

$$p(x_{0:t} \mid y_{1:t}) \propto p(y_{1:t} \mid x_{0:t}) p(x_{0:t}) .$$

For first order Markov process, the estimation can be solved recursively [6],

$$p(x_t \mid y_{1:t}) \propto p(y_t \mid x_t) p(x_t \mid y_{1:t-1}) ,$$

where $p(y_t \mid x_t)$ is the likelihood distribution. Under the particle filtering scheme, a prior is constructed from a sample $x_{t-1}^{(i)}$ drawn from the distribution $p(x_{t-1}^{(i)} \mid y_{1:t-1})$. Omitting the superscript (i) for clarity, the state estimation for each particle, after the sample is drawn, becomes:

$$p(x_t \mid x_{t-1}, y_t) \propto p(y_t \mid x_t) p(x_t \mid x_{t-1}) ,$$

where $p(x_t \mid x_{t-1})$ is the transition probability distribution. From here on, we consider only the estimation for one particle.

Suppose that the state vector x_t can be decomposed into two parts (x_t^1, x_t^2) and the observation y_t is augmented by another measurement y_t^a that can be used to estimate x_t^1 analytically. In other words, $p(x_t^1 \mid y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2)$ can be computed analytically. With this decomposition, we can rewrite the estimation expression as,

$$\begin{aligned} & p(x_t^1, x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a, y_t) \\ & \propto p(y_t \mid x_t^1, x_t^2) p(x_t^1, x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a) \\ & \propto p(y_t \mid x_t^1, x_t^2) p(x_t^1 \mid y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2) p(x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a) \end{aligned} \tag{2}$$

If the decomposition of state vector $x_t = (x_t^1, x_t^2)$ is such that x_t^2 (conditioned on x_{t-1}^1, x_{t-1}^2) is relatively independent on y_t^a, then the third term on the right hand side of (2) can be approximated by:

$$p(x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a) \approx k \times p(x_t^2 \mid x_{t-1}^1, x_{t-1}^2) ,$$

where k is a constant.

For example, this approximation is valid when y_t^a is a hand detection measurement, x_t^1 are the state parameters of the hand and elbow, and x_t^2 are the state parameters of other parts of the body except the hand and elbow. In this case, we cannot draw much information on x_t^2 using y_t^a. The estimation expression then becomes:

$$\begin{aligned} & p(x_t^1, x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a, y_t) \\ & \quad \propto p(y_t \mid x_t^1, x_t^2) p(x_t^1 \mid y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2) p(x_t^2 \mid x_{t-1}^1, x_{t-1}^2) \end{aligned}$$

The above expression suggests that analytical inference, as represented by the term $p(x_t^1 \mid y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2)$, can be used for state estimation. We call this inference "analytical" because it involves the estimation of intermediate joint angles using geometry and inverse kinematics.

This framework is helpful because the prediction by analytical inference often has a much lower variance compared to the initial prior probability.

For discussion, we consider the simplest case, where we can compute x_t^1 deterministically by a function $x_t^1 = f(y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2)$, such that:

$$p\left(x_t^1 \middle| y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2\right) = \begin{cases} 1, & x_t^1 = f(y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2), \\ 0, & \text{otherwise.} \end{cases}$$

then the estimation becomes:

$$p\left(x_t^1, x_t^2 \middle| x_{t-1}^1, x_{t-1}^2, y_t^a, y_t\right)$$
$$\propto \begin{cases} p\left(y_t \middle| x_t^1, x_t^2\right) p\left(x_t^2 \middle| x_{t-1}^1, x_{t-1}^2\right), & \text{when } x_t^1 = f\left(y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2\right) \\ 0, & \text{otherwise.} \end{cases}$$

This shows that the Monte Carlo computation of the prior probability is now effectively applied only to the reduced state vector x_t^2. The simulation of $p\left(x_t^1 \middle| x_{t-1}^1, x_{t-1}^2\right)$ is not required. This reduces the required number of particles and the computational load.

For the hand detection, we have a number of hypotheses for hand position. If we assume negligible localization error in these hypotheses, the inference $p\left(x_t^1 \middle| y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2\right)$ is non-zero for a small finite set of discrete x_t^1 values. For each particle, we sample x_t^1 from this finite set using Monte Carlo method. In other words, the prior distribution, which is used as the importance sampling distribution, has collapsed from a continuous x_t^1 space to a few discrete values. While the inference is not totally deterministic, the degree of randomness has greatly reduced.

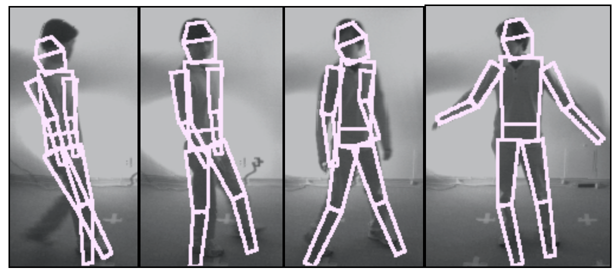
The state estimation process for a particle is shown schematically in Figure 1. In the figure, the shaded box represents the updating of state parameters from body parts detection. In our implementation, it consists of a cascade of three updating stages using inference result from the detection of torso, head and hands respectively.

5. Experimental Results

Experiment Setup and Tracking Initialization:

Three calibrated cameras are set to capture sequences of a single person moving in a room. An empty scene background is first learned for detection purposes. As the person enters the field of view of the cameras, the silhouettes in the 3 views are extracted using a background subtraction. Once the person is within the field of views of the three cameras, the particle filter is initialized using the inference methods described earlier and the following rules: the height of the person is estimated from the height of the silhouette, and the orientation of the shoulders and hips are inferred from the second principal axis of the silhouette. The initialization is fully automatic and does not require the person to stand in a standard posture. Details of this initialization step can be found in [12].

Depending on the pose, the first initialization may not be accurate due to self-occlusions and pose singularities. Using body part detection with particle filter, the tracking algorithm is able to recover the pose at subsequent frames. Figure 7 shows a sequence where the tracking starts when a person is walking towards the center of the room. It demonstrates how the method is able to recover from inaccurate initialization.



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

Figure 7: Initialization and tracking. (a) The initialization of the model when the person enters the scene, the model is not matched properly. (b) As tracking continues, the legs and head are recovered. (c) The hands start to appear as the person turns. (d) The hands are recovered.

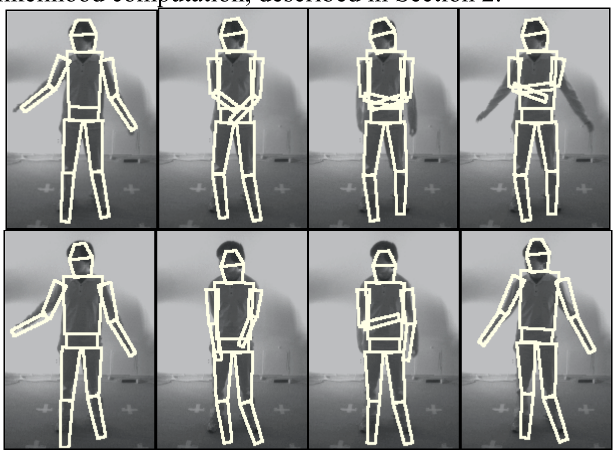
Comparison with Standard Particle Filter:

As only three cameras are used, pose ambiguities will occur while observing person gesturing due to self-occlusion, motion singularities and background clutter. The tracking method should be robust against these problems. While the standard particle filter uses multiple particles to sample the posterior distribution of the state space, it suffers from the problem of high dimensionality, which causes sample depletion in most of the state space. As a result, when an ambiguity occurs, it is easy to lose track and recovering lost tracking is difficult.

Our method uses body part detection to infer some of the state parameters and is able to generate good state hypotheses, even in sample-depleted state space region. This provides an avenue to maintain good tracking and to recover the state parameters after lost tracking.

We compare the performance of our method with the standard particle filter. Both algorithms were tested with a sequence that contains a brief period of about 10 seconds when the hands of the person were hidden behind his back and were occluded from all cameras (Figure 8). After the hands reappear, the standard particle filter is unable to recover from the lost track. In the proposed method, the hands of the human model the articulated body model are able to adjust to the correct hand positions after the person's hands reappear.

In this experiment, 100 particles are used for the new proposed method, while 400 particles were required by the standard particle filter algorithm. In the new method, the reduction in computation due to the smaller number of particles has more than offset the additional computation in body parts detection. Both methods use the same likelihood computation, described in Section 2.



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

Figure 8: Track recovery after self-occlusion. First row shows the result from standard particle filter without inference. Second row shows the result from improved particle filter with inference. For each row: (a) before occlusion, (b) both hands are occluded behind the person body, (c) one hand reappears, (d) both hands are visible.

Experiment with Different Person: Figure 9 presents result with a video of a person who is turning her body and waving her hands. With estimated poses, the human model can be rendered in arbitrary views (Figure 10).

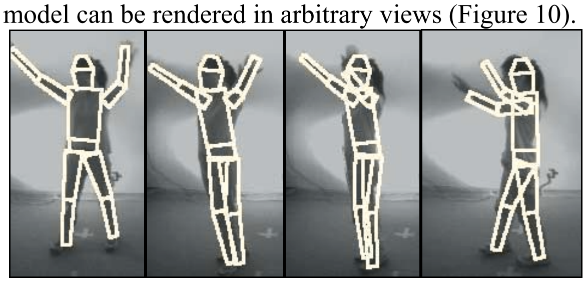


Figure 9. Sequence with different person.

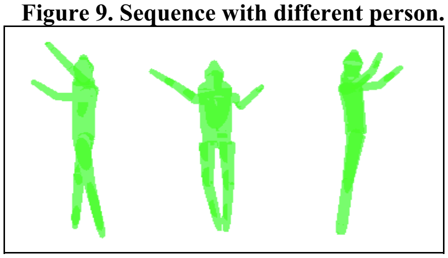


Figure 10. Rendered human models with estimated pose.

6. Conclusion

There are two main approaches to human body tracking: analytical-based methods and synthesis-based methods. This paper introduces a novel technique that combines both approaches by using analytical methods to infer subsets of state parameters and improve the state estimation within the particle-filtering scheme.

This new method has the advantage of handling track initialization, recovering lost track, and reducing the computational load.

7. Acknowledgement

This research was partially funded by the Institute for Creative Technologies. The authors would like to thank Ramakant Nevatia for his guidance, and Hongxia Li, Xuefeng Song and Tao Zhao for their contribution in this work.

8. References

[1] Thomas B. Moeslund, Erik Granum, "A survey of computer vision-based human motion capture", *Computer Vision and Image Understanding* 18 (2001), 231-268.

[2] V. Pavlovic, J. M. Rehg, T. J. cham, and K. P. Murthy, "A dynamic Bayesian network approach to figure tracking using learned dynamic models," *ICCV* 1999, vol. 1, 94-101.

[3] Sidenbladh, H., Black, M. J., and Sigal, L., "Implicit probabilistic models of human motion for synthesis and tracking," *ECCV* 2002. vol. 1, 784-800.

[4] N. Gordon, D. J. Salmond, A. F. M. Smith, "Novel approach to nonlinear/non-Gaussian Bayesian state estimation," *Proc. Inst. Elect. Eng. F*, v140, n2, 1993.

[5] M. Isard, and A. Blake, "Visual tracking by stochastic propagation of conditional density". *ECCV* 1996, 343-356.

[6] M. Sanjeev Arulampalam, S. Maskell, N. Gordon, T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking", *IEEE Trans. Signal Processing,* vol. 50, No. 2, Feb 2002, 174-188.

[7] J. Deutscher, A. Blake, B. North, and B. Bascle, "Tracking through singularities and discontinuities by random sampling," *ICCV* 1999, vol. 2, 1144-1149.

[8] C. Sminchisecu, B. Triggs, "Covariance Scaled Sampling for Monocular 3D Body Tracking", *CVPR* 2001, vol 1, pp 447-454.

[9] J. Deutscher, A. Blake, I. Reid, "Articulated Body Motion Capture by Annealed Particle Filtering," *CVPR* 2000, vol 2, 126-133.

[10] Hedvig Sidenbladh, Michael J. Black, D. J. Fleet, "Stochastic Tracking of 3D Human Figures Using 2D Image Motion" *ECCV* 2000, 702-718.

[11] Tao Zhao, Ram Nevatia, Fengjun Lv, "Segmentation and tracking of multiple humans in complex situations," *CVPR* 2001, vol 2, 194-201.

[12] Isaac Cohen, Mun Wai Lee, "3D Body Reconstruction for Immersive Interaction*", Second International Workshop on Articulated Motion and Deformable Objects* Palma de Mallorca, Spain, 21-23 November, 2002.