Human Body Tracking with Auxiliary Measurements

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

Abstract

This paper presents two techniques for improving human body tracking within the particle filtering scheme. Both techniques explore the use of auxiliary measurements. The first technique uses optical flow cues to improve the sampling distribution. The second technique involves the detection of individual body parts, namely the hand, head and torso; and using these detection results to provide additional inference on subsets of state parameters. This method enables the automatic initialization of state vector and allows recovering from tracking failures. These two methods improve the overall accuracy, efficiency and robustness of human body tracking as illustrated by the experimental results.

1 Introduction

Human body tracking is an active research area with many applications in interactive virtual environment, human computer interaction, motion capture for human animation and video surveillance (see [13] for a survey). There are many methods for human body tracking, including blob tracking [22], gradient methods [4][2] and voxel reconstruction [12]. Recently, the particle filter (a.k.a. Condensation [7], sequential Monte Carlo) has been widely used [5][6] [10][18][19]. Based on sampling approximation and likelihood computation, particle filter is able to handle pose ambiguities due to motion singularities and occlusion [5].

There are however some problems in the use of particle filter. Firstly, in human body tracking, the state vector is of a large dimension (typically ≈ 30) resulting in inefficient sampling and high computation complexity. Secondly, an accurate pose initialization is required prior to the tracking. Often, initialization is done manually and therefore reduces the practicality of the method in many applications.

1.1 Related Work

Recent works using particle filter for human body tracking have focused on improving efficiency using various methods such as covariance scaled sampling [19], simulated annealing approach [6], partitioned sampling [10] and hybrid Monte Carlo approach [3]. Other works include combining particle filter with deterministic gradient descent search [21], and using dynamic motion model [18][18]. However, the large computation complexity remains an important issue for body tracking using particle filter.

Detections of head, face, hand and torso have been used in surveillance and hand gesture applications. While there are problems with view-point variation and occlusion, the detection results are generally reliable in majority of viewing conditions. These body part detection methods have been used to provide coarse estimates of the human pose [22][12] and for human detection in static images [14]. However these methods have not been used for detailed human body tracking because of the difficulty in detecting immediate joints such as elbows, shoulders, knees and hips.

Motion features can also be used for performing a better tracking and reducing the numerical complexity of the particle filter. Optical flow has been used for detection and tracking problems [23][18]. It provides a dense field that is useful for inferring movement of the subject and estimating the body joints displacements.

In our work, we consider body part detection and motion features as auxiliary measurements that are useful for inferring some aspects of the human pose and motion. Our contributions are new techniques for integrating these measurements to the sequential Monte Carlo scheme for a more robust and efficient tracking. In this work, we are using three calibrated stationary cameras.

1.2 Our Approach

Our approach is based on the use of auxiliary measurement to improve the state estimation of particle

filter. These auxiliary measurements include motion cue and body part detection. Here, the motion cue is used to improve the sampling distribution by providing better estimations of the current positions of the upper and lower arms. The second type of auxiliary measurement is the detection of hands, head 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. The use of auxiliary measurements reduces the degree of randomness during state vector propagation and also reduces the reliance on Monte Carlo simulation, thereby improving efficiency. The use of auxiliary measurement also generates reliable hypotheses on joint positions that help in initializing the pose and recovering from tracking failure. The proposed approach is fully automatic not requiring the initialization of the articulated body model and moreover, the algorithm bootstraps itself allowing the recovery of body joints after loss of tracking.

The paper is organized as follows: Section 2 briefly describes the human model and the likelihood measure considered. Section 3 describes the use of motion cues. Section 4 describes the use of body parts detection. Experimental results are presented and discussed in Section 5.

2 Particle Filter for Articulated Human Body Model Fitting

The use of particle filter for human body tracking has been well described in previous works [5][6][10][18]. Among these works, there are some variations in the implementation, such as the sampling strategies and the likelihood functions.

In this section, we will briefly describe the articulated human body model we use and the likelihood function considered. The techniques proposed in this paper are generally independent of the likelihood computation and are also applicable in conjunction with other sampling strategies such as annealing particle filter and partitioned sampling.

Human Body Model

The articulated human body model consists of 10 joints and 14 segments, representing the head, torso and limbs. Each segment is represented by a tapered 3D cone with an elliptical cross-section (Figure 1). The model has 32 degrees of freedom that include the global scale, translation, rotation and local joint rotations (Table 1). These degrees of freedom are represented in the state vector x_t .

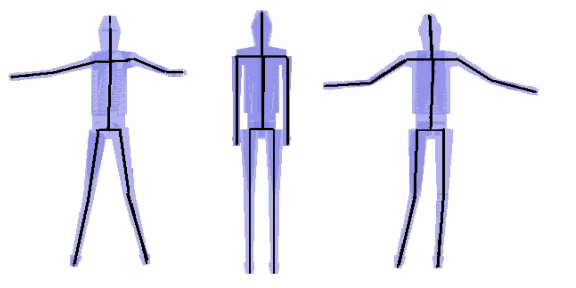


Figure 1. The articulated human body model considered in this work consists of 10 joints and 14 segments.

Parameters Name	Number
Global Scale	1
Global Translation	3
Global Rotation	3
Joint Angles	25

Table 1. Model Parameters

Likelihood Computation

Fitting and tracking the articulated model to the detected human in the video streams requires the definition of a likelihood function allowing the mapping of the degree of freedom of the model onto image properties. The likelihood computation is based on two types of features, the silhouette boundary and silhouette regions.

Given a predicted pose (described by a state vector of a particle), we synthesize a silhouette onto each camera view. This synthesized silhouette allows the matching of the estimated pose to the perceived silhouette derived by a foreground detection method. The likelihood measure is computed by matching the boundaries and regions with the extracted foreground boundary. Further details can be found in [1].

3 Motion Cue for Improved Sampling Distribution

The 3D dynamic state vector x_t consists of the global scale, position and orientation of the human and the joint angles. The state vector characterizes the pose of the person at time *t*. The state transition model is defined by:

$$x_t = x_{t-1} + \eta, \quad (1)$$

where η is a vector of standard normal random variables representing the process noise. This assume a zero order motion model and that the human is generally stationary but with some random perturbations. This is an

engineering trick as it is difficult to model the human dynamic characteristics accurately in an uncontrolled environment. At the prediction stage of the particle filter, the sampling of the *i*th particle, $x_t^{(i)}$, employs a proposal distribution that is a Gaussian centered on the previous state $\tilde{x}_{t-1}^{(i)}$, and defined by:

$$x_t^{(i)} \sim \mathcal{N}\left(\tilde{x}_{t-1}^{(i)}, \Sigma\right), \tag{2}$$

where Σ is the covariance matrix of the process noise.

When the person is moving, the use of this proposal distribution is not suitable since the process noise will have a large variance. However, without additional information, the previous state is the optimal unbiased estimation.

An alternative approach consists in using a first or higher order motion model. For example, first order models are commonly used for tracking walking person. However, the application we are focusing is where the person is gesturing freely and we do not expect body joint angles to change in a consistent way over a reasonable period of time. In addition, increasing the order of motion model introduces additional parameters and will affect tracking stability and increase the computation time.

Another alternative is to model human dynamics and apply dimensionality reduction technique to reduce complexity. This is useful for tracking a restrictive set of human motion (e.g. walking). For applications where there is higher variability in human movement, the benefit of this approach is unclear, and a large set of training data would be required for learning the dynamic model.

In the basic particle filter method, the particles propagate using the transition prior and do not take into account current observation data, until at a later stage when the likelihood measure is computed. This leads to inefficient sampling. Various approaches have been suggested to address this issue. The unscented particle filter (UPF) [11], uses the current observation to derive an improved proposal distribution. Similar strategies are used in [8][17][15] for object tracking. However, for human body tracking, it is difficult to apply UPF directly because it requires good grouping and labeling of edge features which usually involves searching and matching of extracted features with synthesized features obtained by projecting the estimated body model onto the all camera views.

We have adopted a different strategy based on using motion features. This is achieved by considering optical flow field to estimate the motion of individual body parts and generating a better proposal distribution for the dynamic state parameters. Since optical flow is a dense field, motion inference can be readily obtained without the need of searching and labeling. We define an auxiliary parameter vector (that we will refer to as motion parameters) as the change in state vector

$$v_t = x_t - x_{t-1} \tag{3}$$

and the state transition model becomes

$$x_t = x_{t-1} + v_t \tag{4}$$

The change in state parameter will induce optical flow in the images that can be measured. We define z_t as a measurement of motion parameters, so that

$$z_t = v_t - \eta', \tag{5}$$

where η' is the measurement noise, and the state transition becomes

$$x_t = x_{t-1} + z_t + \eta'. \tag{6}$$

If the measurement, z_t, (which will be described later) is reasonably accurate, then the measurement noise η' will have a small variance compared to the process noise η. We approximate the measurement noise η' by a vector of standard normal random variables:

$$\eta' \sim \mathcal{N}(0, \Sigma').$$

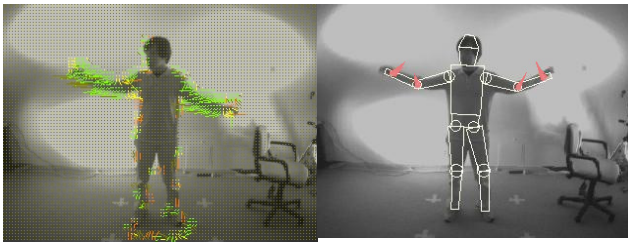
We then modify the proposal distribution as,

$$x_t^{(i)} \sim \mathcal{N}\left(\tilde{x}_{t-1}^{(i)} + z_t^{(i)}, \Sigma'\right). \tag{7}$$

We note that the measurement $z_t^{(i)}$ is conditioned on the previous state $\tilde{x}_{t-1}^{(i)}$ because the positions of the body parts are required before we can extract the relevant motion parameters from the optical field. In the following, we describe this in more details.

For the present discussion, we focus on the extraction of motion parameters for the lower arms in order to extract the change of joint angle at the elbow. For each particle, given its previous state $\tilde{x}_{t-1}^{(i)}$, we can determine the position of the lower arm on the image. We integrate the optical flow along the upper and lower edges of the lower arm and estimate the 2D image motion of the hand with respect to the elbow joint. We only consider motion in the direction perpendicular to the lower arm in the image.

From each image, we can only obtain an estimate of the motion projected onto the plane orthogonal to the optical ray. This introduces two constraints: combining the estimates from three different images results in an over-constraint problem. The 3D motion of the lower arm is then estimated using a Least Mean Square method. The change in the joint angle of the elbow is then estimated using the geometry of the articulated model.



(a) (b)

Figure 2 Measurement of motion parameters at lower and upper arm. (a) Optical flow field, (b) integration of flow along predicted edges of body parts and estimation of motion along the normal direction.

In our implementation, we use the approach described above to estimate the 3D motion of each of the lower and upper arms and update the joint angles at the elbows and shoulders.

This approach is particularly useful for inferring the arm movement, because of the following reasons:

1. In indoor environment, while interacting with a system the movements of the arms are usually faster and more varying compared to the rest of the body. Therefore the joint angles at the elbow and shoulder have larger noise variances in the transition model, reducing the efficiency and the robustness during the tracking.
2. The movements of the arms are usually observable from at least two of the cameras, so that reasonably good estimates of the arm movement can be extracted from the optical flow field.

4 Parts Detection for Partial Updating of State Vector

The second technique involves the detection of body parts to improve tracking by updating subsets of the state parameters, within the particle filtering scheme.

In our implementation, this scheme consists of three stages, where each stage involves the detection of a specific body part, namely the hand, head and torso respectively. The detection results are used to update subsets of state vector using inverse kinematics. The state parameters that are updated at each stage are given in Table 2. The additional inference provided by the detection of body part improves the state estimation. The state estimation process for a particle is shown schematically in Figure 3.

Detection y_t^a	Number of Constraints	Updated parameters x_t^1
Hand	3	Elbow joint (1dof) Shoulder joint (2dof).
Head	3	Neck joint (2dofs) Neck position (1dof)
Torso	4	Body position (2dofs) Body orientation (2dofs)

Table 2. Subset of the state vector updated with the body part detection.

The proposed technique is related to Rao-Blackwellised particle filtering (RBPF) [24], which improves efficiency by marginalizing some of the state parameters analytically. Most previous work in human body tracking, uses edge and region features as measurement, from which it is difficult to compute any particular subset of the state vector directly. Therefore, RBPF is not readily applicable to this problem without using different types of measurement. Our proposed technique, adopting the RBPF approach, introduces the use of body parts detection to measure the positions of specific body parts and therefore is able to marginalize subsets of the state vector.

In the following sections, we first describe the methods for detecting the body parts and how they are used to update subsets of state vector. In the later part of this section, we describe how this inference is integrated into the overall state estimation framework.

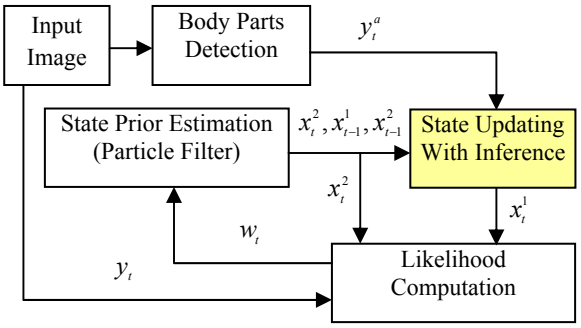


Figure 3. Body parts detection and state updating.

Hand Detection

For hand detection, we locate candidate positions as the peaks of convex curvature along the outlines of the silhouette (Figure 4a). We match these curvature peaks in different images using epipolar constraints, and reconstruct their 3D positions.

The measurement (giving 3 constraints) 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 previous elbow position. The joint angles at the elbow and shoulder are then updated accordingly.

If a candidate hand position violates the physical constraints of the body (e.g. too far away from the shoulder), this candidate will be regarded as an outlier and ignored.

Head Detection

The head detection is performed using a reference chain code representation of a head-shoulder contour (Figure 4b) as a template for the head. We match this template along the contour boundary of the extracted silhouette to detect the location of 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. Positions along the silhouette boundary that match well are candidate for head positions. To obtain the head position in 3D space, the head must be detected in at least two input images. Epipolar constraint is used to remove the false measurements.

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 related to the head position: the orientation of the head (2 dofs) and the position of the neck along the body main axis (1 dof). Change in the 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.

Torso 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 are computed. A line is then fitted to these 3D points using PCA and RANSAC methods, based on least square criterion. This extracted line, illustrated in Figure 4d, provides a measurement of the torso orientation, and a constraint that the torso must lay along the line. This gives four constraints, which are used to update the state parameters.

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 modification, we apply the same strategy as for head detection to keep the positions of hands and legs constant and update the angles at intermediate joints.

Detection Errors

The body part detections have problems with occlusions, as well as outlier errors, which will lead to negative or false detection. To overcome this, not all particles are updated with the detection results. Instead, we use the likelihood measure of the detection (based on matching error during the measurement stages) to determine the confidence of the detection result. Monte Carlo method is then used to decide whether to use the detection result for each particle propagation. When the detection result is not used, the state vector will be propagated according to the stochastic state transition model used in the standard particle filter algorithm.

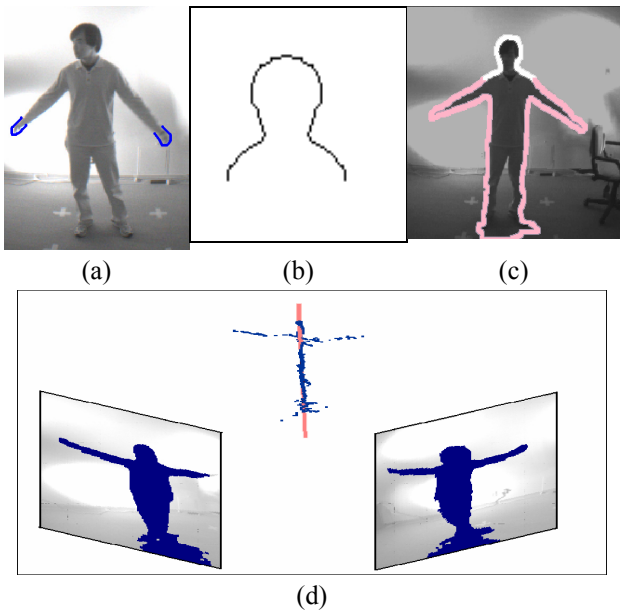


Figure 4. Body parts detection. (a) Hand detection, (b) Template of the head. (c) detected head, (d) torso detection, the medial axis of the silhouette in two views is extracted, matched and reconstructed in 3D. The torso axis is found by fitting a line to the medial axis points in 3D.

Integration with Particle Filter

In the following, we describe the framework on how the inference provided by body parts detection is

integrated into the particle-filtering scheme. Denoting x as the hidden state and y as the observation. At any time t, the state estimation is given by:

$$p(x_t \mid x_{t-1}, y_t) \propto p(y_t \mid x_t) p(x_t \mid x_{t-1}), \quad (8)$$

where $p(x_t \mid x_{t-1})$ is the transition probability distribution. In this section, we consider only the state propagation for one particle.

Suppose that the state vector x_t is decomposed into two parts (x_t^1, x_t^2) and denoting y_t^a as an auxiliary measurement (from body parts detection) that can be used to estimate x_t^1 analytically. We can rewrite the estimation expression as,

$$\begin{aligned} & p(x_t^1, x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a, y_t) \\ & \propto p(y_t \mid x_t^1, x_t^2) p(x_t^1, x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a) \\ & \propto p(y_t \mid x_t^1, x_t^2) p(x_t^1 \mid y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2) p(x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a) \end{aligned} \quad (9)$$

For example, during the hand detection stage, the decomposition of the state vector is such that x_t^1 consists of the elbow joint angle and two joint angles at the shoulder, which can be inferred using inverse kinematics as described earlier. The other subset x_t^2 consists of the remaining 29 state parameters.

Since y_t^a is a measurement on a local body part (e.g. detection of left hand), the rest of the state parameters in x_t^2 (which relate to other parts of the body, e.g. right arms, legs etc) are relatively independent of y_t^a. We can approximate the estimation expression as:

$$\begin{aligned} & p(x_t^1, x_t^2 \mid x_{t-1}^1, x_{t-1}^2, y_t^a, y_t) \\ & \propto p(y_t \mid x_t^1, x_t^2) p(x_t^1 \mid y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2) p(x_t^2 \mid x_{t-1}^1, x_{t-1}^2) \end{aligned} \quad (10)$$

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

$$p\left(x_t^1 \middle| y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2\right) = \delta(x_t^1 - f(y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2)) \quad (11)$$

where $\delta(\cdot)$ is the Dirac delta function. Then the estimation becomes:

$$\begin{aligned} & p\left(x_t^1, x_t^2 \middle| x_{t-1}^1, x_{t-1}^2, y_t^a, y_t\right) \\ & \propto p(y_t \middle| x_t^1, x_t^2) p(x_t^2 \middle| x_{t-1}^1, x_{t-1}^2) \delta(x_t^1 - f(y_t^a, x_t^2, x_{t-1}^1, x_{t-1}^2)) \end{aligned} \quad (12)$$

This allows Monte Carlo sampling to be applied only to the reduced state vector x_t^2, therefore improving the efficiency.

Multiple Hand Detection Candidates

For the hand detection, we may have detected a number of candidates for hand position. In this case, 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 by Monte Carlo method, using the confidence measures of each candidate position (based on matching error). In this scenario, the 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 been reduced considerably.

In this work, we are assuming that the measurement noises are negligible. This is a reasonable assumption for head and hand detections that generally have good localization properties. In our current work, we are addressing the issue of measurement noise and are reformulating the inference in a probabilistic framework, taking into account the measurement uncertainties.

5 Experimental Results

5.1 Experiment Setup and Tracking Initialization

Three calibrated cameras are set up to capture video sequences in a room. A background of the empty scene is first learned for detection purposes. As the person enters the views, the silhouettes in the three views are extracted using a background subtraction method. Once the person has fully entered the scene, the particle filter is initialized using the body parts detection methods and additional heuristic rules described in [1]. The initialization is fully automatic and does not require the person to stand in a particular posture. The first initialization may not be accurate due to self-occlusions and pose singularities. But the use of auxiliary measurements is able to generate additional good state hypotheses to help in recovering more accurate pose at subsequent frames. Details of this initialization step can be found in [1].

5.2 Improvement from Motion Cue

In Section 3, we describe the use of motion cue to modify the proposal distribution for particle sampling, so that the distribution is centered on $\tilde{x}_{t-1}^{(i)} + z_t^{(i)}$, where $\tilde{x}_{t-1}^{(i)}$ is the previous state after re-sampling, and $z_t^{(i)}$ is the estimated change in parameters from motion cue. Without the use of motion cue, the distribution will be centered on $\tilde{x}_{t-1}^{(i)}$. In this experiment, for each time frame, we evaluate the likelihood measures at the two centers of distributions using the current image, i.e. we compute and compare $p(y_t \middle| \tilde{x}_{t-1}^{(i)} + z_t^{(i)})$ and $p(y_t \middle| \tilde{x}_{t-1}^{(i)})$, where $\tilde{x}_{t-1}^{(i)}$ is the previous

state parameters of the best matched particle at time *t*-1. The result is shown in Figure 5, which shows that the likelihood measure is generally higher with the improved proposal distribution generated using the motion cue.

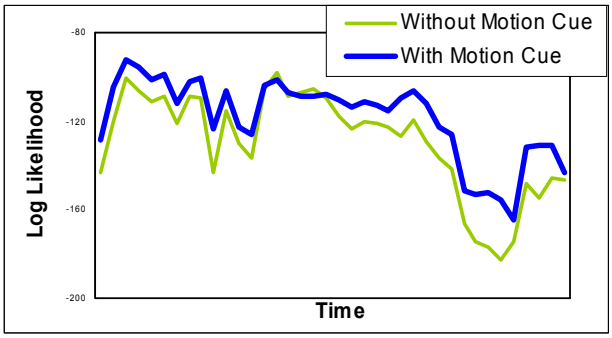


Figure 5. Likelihood measures with and without the use of motion cue. Note that the likelihood is expressed in the logarithmic scale.

5.3 Recovering Tracking Failure and Robust Tracking

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 loose track and recovering from 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 for maintaining good tracking and recovering from 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 6). After the hands reappear, the standard particle filter is unable to recover from the lost track. In the proposed method, the articulated body model is 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 proposed method, the reduction in computation due to the smaller number of particles has more than offset the additional computation in body parts detection. Body parts detection is performed only once for each frame. The main computationally intensive processing is the likelihood computation which is performed for every particle at each frame. Both methods use the same likelihood computation.

Figure 7 shows a result from another video of a person who is turning her body and waving her hands. The rotation of the torso is not easily detected from 2D silhouette edges, but using the cues provided by the limbs, the tracker is able to estimate the pose correctly as shown by the rendered human model.

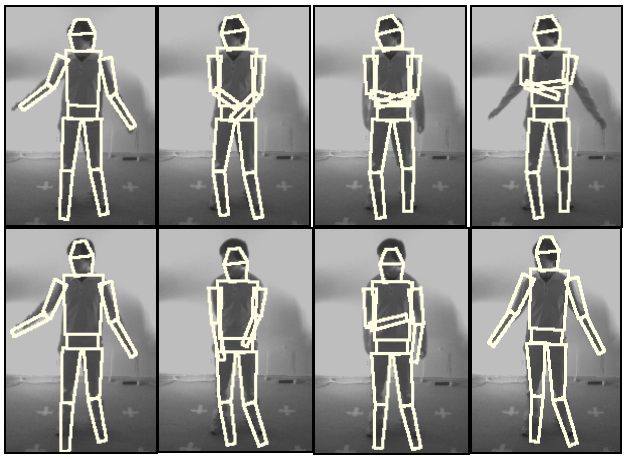


Figure 6: Track recovery after self-occlusion. First row shows the result from standard particle filter without inference. Second row shows the result from improved particle filter with inference. For each row: (a) before occlusion, (b) both hands are occluded behind the person body, (c) one hand reappears, (d) both hands are visible.

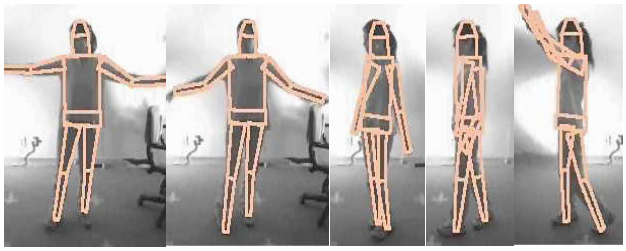


Figure 7. Tracking a turning person: The person is turning while waiving her hands.

6 Conclusion

We have proposed two improvements to the particle filter technique for human body tracking by using two types of auxiliary measurements. Firstly, motion cue is used to improve the sampling distribution by providing

better estimations of the current positions of the upper and lower arms. Secondly, auxiliary measurement from the detection of hands, head and torso, are used to infer subsets of state parameters and improve the estimation of human pose within the particle-filtering scheme. This has the advantages of performing automatic track initialization, recovering lost track, and reducing the computational complexity of the particle filtering method.

Acknowledgement

The research has been funded in part by the Integrated Media Systems Center, a National Science Foundation Engineering Research Center, and Cooperative Agreement No. EEC-9529152.

7 References

[1] Mun Wai Lee, Isaac Cohen, Soon Ki Jung, "Particle Filter with Analytical Inference for Human Body Tracking", *IEEE Workshop on Motion and Video Computing*, 2002.

[2] C. Bregler, J. Malik, "Tracking people with twists and exponential maps," *CVPR* 1998, pp.8 -15.

[3] K. Choo, and D.J. Fleet, "People tracking with hybrid Monte Carlo." *ICCV* 2001, vol II, pp. 321-328.

[4] Q. Delamarre, O. Faugeras, "3D articulated models and multi-view tracking with physical forces," *CVIU*, vol 81, no. 3, March 2001, pp. 328-357.

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

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

[7] M. Isard, A. Blake, "CONDENSATION – conditional density propagation for visual tracking," *IJCV,* 1998.

[8] M. Isard and A. Blake, "ICondensation: Unifying low-level and high-level tracking in a stochastic framework", *ECCV* 1998. vol. 1, pp 893-908.

[9] I. A. Kakadiaris, D. Metaxas, "3D human body model acquisition from multiple views," *ICCV* 1995, pp. 618 -623.

[10] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking," *ECCV* 2000, vol 2, pp. 3-19.

[11] R. van der Merwe, A. Doucet, N. de Freitas and E. Wan, "The Unscented Particle Filter", in *Advances in Neural Information Processing Systems (NIPS13)*, *MIT Press*, Eds. T. K. Leen, T. G. Dietterich and V. Tresp, Dec, 2000.

[12] I. Mikic, M. Trivedi, E. Hunter, P. Cosman, "Articulated body posture estimation from multi-camera voxel data," *CVPR* 2001, Vol. 1, pp. 455-460.

[13] T. B. Moeslund, Erik Granum, "A survey of computer vision-based human motion capture", *CVIU*, 2001, 81(3), pp. 231-268.

[14] A. Mohan, C. Papageorgiou, T. Poggio, "Example-based object detection in images by components," *PAMI*, Vol. 23 Issue: 4 , April 2001, pp. 349 –361.

[15] H. Moon, R. Chellappa, A. Rosenfeld, "Tracking of human activities using shape-encoded particle propagation," *ICIP* 2001, vol 1, pp. 357 -360.

[16] R. Plankers, P. Fua, "Articulated soft objects for video-based body modeling,", *ICCV*, 2001, vol 1, pp. 394 -401.

[17] Y. Rui, Y. Chen, "Better Proposal Distributions: Object tracking using unscented particle filter," *CVPR* 2001, vol. 2, pp. 786-793.

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

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

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

[21] J. Sullivan, J. Rittscher**, "**Guiding random particles by deterministic search" *ICCV* 2001, vol.1, pp 323 - 330.

[22] C. R. Wren, A. Azarbayejani, T. Darrell, A.P. Pentland, "Pfinder: real-time tracking of the human body," *PAMI* vol. 19 Issue: 7, July 1997 pp. 780 - 785.

[23] T. Zhao, R. Nevatia, "Stochastic Human Segmentation from a Static Camera," *IEEE Workshop on Motion and Video Computing* 2002.

[24] Doucet, A., de Freitas, J. F. G., Murphy, K., and Russell, S. (2000). "Rao Blackwellised particle filtering for dynamic Bayesian networks," *Uncertainty in Artificial Intelligence.*