Swapping-based Annealed Particle Filter with Occlusion Handling for 3D Human Body Tracking

Xuan Son Nguyen
INRIA, Loria, Campus Scientifique, Vandœuvre-Lès-Nancy, France

Keywords: Particle Filter, Human Body Tracking, Bayesian Network.

Abstract: In this paper, we propose a new approach for 3D human body tracking. We first extend the idea of Swapping-based Partitioned Sampling (SBPS), which was introduced by Dubuisson et al. for solving the articulated object tracking problem in high dimensional state spaces. This extension aims to deal with self-occlusion and constraints between parts of the human body, which are not taken into account in SBPS. We prove that, under the same assumptions required by SBPS, the posterior distribution are correctly estimated in our framework. We then introduce a new approach for 3D human body tracking, based on this new framework and Annealed Particle Filter (APF). Experiments with multi-camera walking sequences from the HumanEva I dataset show the efficiency of the proposed approach in terms of both accuracy and computation time.

1 INTRODUCTION

Tracking the human body with accuracy and within a reasonable time is challenging due to the high complexity of the problem to solve. Various approaches have been proposed for this problem. One class of approaches is known as optimization-based methods (Deutscher and Reid, 2005; Zhang et al., 2010). Typically, these methods are based on the optimization of an objective function corresponding to the matching function between the model and the observed image features.

Another way of reducing the dimensionality of the configuration space is to use decomposition methods within the particle filter framework (MacCormick and Isard, 2000; Rose et al., 2008). Particle Filter (PF) (Gordon et al., 1993) has been shown to be an effective method for solving visual tracking problems. This is due to its ability to deal with non-linear, non-Gaussian and multimodal distributions encountered in such problems. The key idea behind decomposition methods is similar: decompose the state space of the target object into a set of subspaces where particle filter can be applied. Since the dimension of these subspaces is smaller than that of the original state space, sampling in these subspaces will be more efficient than sampling in the original space and therefore, fewer particles are needed to achieve a good performance. Recently, a decomposition approach called Swapping-based Partitioned Sampling (SBPS), which is based on the state-of-the-art algorithm Partitioned Sampling (PS) (MacCormick and Isard, 2000) for tracking in high dimensional state spaces, has been introduced in (Dubuisson et al., 2011; Dubuisson et al., 2013). Under some assumptions, SBPS is guaranteed to produce a correct estimation of the posterior distribution. However, one of the important assumptions required by SBPS is that no self-occlusion occurs during tracking. This assumption is often violated in real-world problems, and the posterior distribution can be poorly approximated by SBPS in such cases. Another disadvantage of SBPS is that it does not take into account constraints between different parts of the articulated object, which have been shown to be very important in human tracking (Sigal et al., 2010). To address these problems, we first introduce an extension of SBPS, which is more flexible than SBPS and allows us to better estimate the posterior distribution when self-occlusion is present. We then introduce a new approach for 3D human body tracking, which is a combination of the new framework and Annealed Particle Filter (APF) (Deutscher and Reid, 2005).

The paper is organized as follows. In Section 2, we give a brief introduction to PF, PS and SBPS. Section 3 presents the proposed approach. Section 4 reports the results of our experimental evaluation. Finally, Section 5 offers some conclusions and ideas for future work.

Nguyen X..
Swapping-based Annealed Particle Filter with Occlusion Handling for 3D Human Body Tracking.
DOI: 10.5220/0004686905760583
In *Proceedings of the 9th International Conference on Computer Vision Theory and Applications* (VISAPP-2014), pages 576-583
ISBN: 978-989-758-009-3

2 RELATED WORK

2.1 Particle Filter

In this paper, human tracking consists of estimating a state sequence $\{\mathbf{x}_t\}_{t=1,\ldots,T}$, whose evolution is given by equation $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{n}_t^{\mathbf{x}})$, from observations $\{\mathbf{y}_t\}_{t=1,\ldots,T}$ related to the states by $\mathbf{y}_t = \mathbf{h}_t(\mathbf{x}_t, \mathbf{n}_t^{\mathbf{y}})$. Usually, $\mathbf{f}_t$ and $\mathbf{h}_t$ are nonlinear functions, and $\mathbf{n}_t^{\mathbf{x}}$ and $\mathbf{n}_t^{\mathbf{y}}$ are i.i.d. noise sequences. From a probabilistic viewpoint, it amounts to estimate, for any t, $p(\mathbf{x}_{1:t}|\mathbf{y}_{1:t})$ where $\mathbf{x}_{1:t}$ denotes the tuple $(\mathbf{x}_1, \ldots, \mathbf{x}_t)$.

The PF framework (Gordon et al., 1993) approximates the posterior densities using weighted samples $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}$, $i = 1, \ldots, N$, where each $\mathbf{x}_t^{(i)}$ is a possible realization of state $\mathbf{x}_t$ called a *particle*. In its *prediction* step, PF propagates the particle set $\{\mathbf{x}_{t-1}^{(i)}, w_{t-1}^{(i)}\}$ using a proposal function $q(\mathbf{x}_t|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_t)$ which may differ from $p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})$ (but, for simplicity, we will assume they do not); in its *correction* step, PF weights the particles using a likelihood function, so that $w_t^{(i)} \propto w_{t-1}^{(i)} p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) \frac{p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t^{(i)}|\mathbf{x}_{1:t-1}^{(i)}, \mathbf{y}_t)}$, with $\sum_{i=1}^N w_t^{(i)} = 1$. The particles can then be resampled: those with the highest weights are duplicated while the others are eliminated. The estimation of the posterior density $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ is then given by $\sum_{i=1}^N w_t^{(i)} \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t)$, where $\delta_{\mathbf{x}_t^{(i)}}$ are Dirac masses centered on particles $\mathbf{x}_t^{(i)}$.

2.2 Partitioned Sampling and Swapping-based Partitioned Sampling

Partitioned Sampling (PS) has been introduced by MacCormick (MacCormick and Isard, 2000). PS's key idea is to exploit some natural decomposition of the system dynamics w.r.t. subspaces of the state space in order to apply PF only on those subspaces. This leads to a significant reduction in the number of particles required for tracking. So, assume that state space $\mathcal{X}$ and observation space $\mathcal{Y}$ can be partitioned as $\mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^P$ and $\mathcal{Y} = \mathcal{Y}^1 \times \cdots \times \mathcal{Y}^P$ respectively. For instance, a system representing a hand could be defined as $\mathcal{X}^{\text{hand}} = \mathcal{X}^{\text{palm}} \times \mathcal{X}^{\text{thumb}} \times \mathcal{X}^{\text{index}} \times \mathcal{X}^{\text{middle}} \times \mathcal{X}^{\text{ring}} \times \mathcal{X}^{\text{little}}$. Assume in addition that the dynamics of the system follows this decomposition, i.e., that:

$$f_t(\mathbf{x}_{t-1}, n_t^{\mathbf{x}}) = f_t^P \circ f_t^{P-1} \circ \cdots \circ f_t^2 \circ f_t^1(\mathbf{x}_{t-1}), \quad (1)$$

where $\circ$ is the usual function composition operator and where each function $f_t^i : \mathcal{X} \mapsto \mathcal{X}$ modifies the particles' states only on subspace $\mathcal{X}^i$ [1].

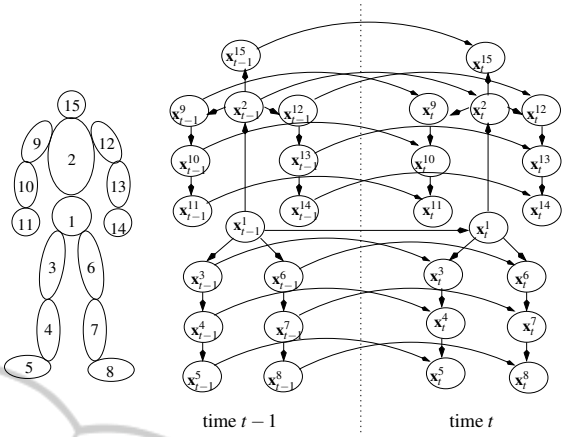

Figure 1: Full body tracking.

In PS, one assumes that the global image likelihood can be factorized as product of local likelihoods:

$$p(\mathbf{y}_t|\mathbf{x}_t) = \prod_{i=1}^{P} p^i(\mathbf{y}_t^i|\mathbf{x}_t^i), \quad (2)$$

where $\mathbf{y}_t^i$ and $\mathbf{x}_t^i$ are the projections of $\mathbf{y}_t$ and $\mathbf{x}_t$ on $\mathcal{Y}^i$ and $\mathbf{X}^i$ respectively.

Swapping-based Partitioned Sampling (SBPS) has been introduced in (Dubuisson et al., 2011). It exploits conditional independences encoded in Dynamic Bayesian Networks (DBNs) (Murphy, 2002) to improve the tracking accuracy and reduce the computational cost of PS. Figure 1 show an example of human body tracking and a DBN used by SBPS for modeling this problem, where $\mathbf{x}_t^k, k = 1, \ldots 15$ represent the parts of the human body. Denote $\mathbf{pa}(\mathbf{x}_t^k)$ and $\mathbf{pa}_t(\mathbf{x}_t^k)$ the parent of node $\mathbf{x}_t^k$ in all time instants and in time instant t, respectively. The assumptions required by SBPS is that the proposal transition function of a given part $\mathbf{x}_t^i$ at time t depends only on that part in time $t-1$ (and possibly on other parts in time t) and the observations depend only on their corresponding state. The set $\{1, \ldots, P\}$ of parts of the target objects can be partitioned into some sets $\{P_1, \ldots, P_K\}$ such that those parts in each P_j are all independent conditionally to $\cup_{h<j} P_h$. For instance, in Figure 1, $P = 15$ and $K = 5$, $P_1 = \{1\}$ corresponds to the pelvis, $P_2 = \{2, 3, 6\}$ to the torso, the right and left thighs, $P_3 = \{4, 7, 9, 12, 15\}$ to the right and left calves, the right and left upper arms, the head, $P_4 = \{5, 8, 10, 13\}$ to the right and left foots, the right and left lower arms, $P_5 = \{11, 14\}$ to the right and left hands. At

[1] Note that, in (MacCormick and Isard, 2000), functions f_t^i are more general since they can modify states on $\mathcal{X}^i \times \cdots \times \mathcal{X}^M$. However, in practice, particles are often propagated only one $\mathcal{X}^j$ at a time.

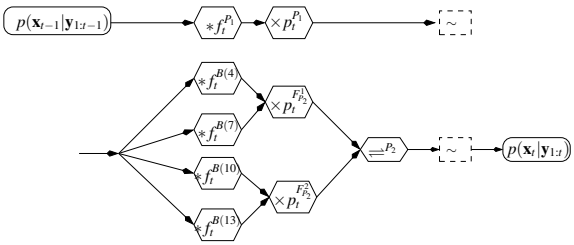

Figure 2: New diagram for SBPS.

the jth stage, SBPS performs the prediction/correction steps for the parts in P_j in parallel instead of part after part as in PS. This enables to produce better particles by swapping their subparts. This is achieved by an operation called *swapping* or *permutation*, which guarantees that the target distribution is correctly estimated. More precisely, whenever two particles are such that they have the same states on some nodes $\mathbf{pa}_t(\mathbf{x}_t^k)$, then swapping their states on $\mathbf{x}_t^k$ and its descendants cannot alter the density estimated by the particle set. For instance, on Figure 1, if two particles have the same value for node $\mathbf{pa}_t(\mathbf{x}_t^1)$, their values on node $\mathbf{x}_t^3$, $\mathbf{x}_t^4$ and $\mathbf{x}_t^5$ can be safely swapped. The set of subparts to permute similarly to $\mathbf{x}_t^k$ in called a *swapping set*. In addition, subpart permutation between two particles can only be performed if they have the same states on some nodes $\mathbf{pa}_t(\mathbf{x}_t^k)$, which leads to the definition of what is *admissible permutation*. In SBPS, the definitions of swapping set and admissible permutation guarantee that densities are correctly estimated.

The soundness of SBPS can be justified by d-separation analysis on DBNs (see (Dubuisson et al., 2013)).

3 PROPOSED APPROACH

SBPS is based on an important assumption that the global image likelihood can be expressed as a product of individual local likelihoods. The likelihood for each part of the target object is then evaluated independently of other parts without taking into consideration self-occlusion. When self-occlusions are present however, this assumption is violated and the product of local likelihoods gives a poor approximation to the global likelihood. As a result, SBPS tracker does not give good performances when tracking under self-occlusions. To cope with this problem, we propose a new diagram for SBPS. We first introduce a theoretical framework for this diagram in the next section. Then, in section 3.2, we propose a new approach for 3D human body tracking, based on this diagram.

3.1 Theoretical Framework

For computational reason, we will assume from this point on that, within each time slice, the DBN structure is a directed tree.

For any $u, u = 1,\ldots,P$, denote $T(u)$ the set of indices of nodes in the subtree whose root is $\mathbf{x}_t^u$ (the indice of node $\mathbf{x}_t^u$ for any t is u). Denote P_1 any subset of $\{1,\ldots,P\}$ such that: if $u_1 \in P_1$ then $u_2 \in P_1$ where $\mathbf{x}_t^{u_2} = \mathbf{pa}_t(\mathbf{x}_t^{u_1})$. For any $j \geq 2$, assume that $P_1,\ldots,P_{j-1}$ have been defined. Let S_j be any subset of the set of nodes that have not been selected in any set $P_1,\ldots,P_{j-1}$ and any node in S_j is some child node of a node in $P_1 \cup \ldots \cup P_{j-1}$: $S_j \subseteq \{u \in \{1,\ldots,P\} \setminus \cup_{h=1}^{j-1} P_h : \mathbf{pa}_t(\mathbf{x}_t^u) \subseteq \bigcup_{h=1}^{j-1} \bigcup_{v \in P_h} \{\mathbf{x}_t^v\}\}$. For all $u \in S_j$, denote $B(u)$ any subset of $T(u)$ such that: if $u_1 \neq u$ and $u_1 \in B(u)$ then $u_2 \in B(u)$ where $\mathbf{x}_t^{u_2} = \mathbf{pa}_t(\mathbf{x}_t^{u_1})$. Then P_j is defined as: $P_j = \bigcup_{u \in S_j} B(u)$. In order to determine interesting permutations, denote $F_{P_j}^1,\ldots,F_{P_j}^{n_j}$ a partition of P_j such that any $F_{P_j}^h, h = 1,\ldots,n_j$ is a union of some sets $B(u), u \in S_j$. We call $F_{P_j}^1,\ldots,F_{P_j}^{n_j}$ an *interesting partition* of P_j.

As an example, consider the human tracking problem in Figure 1. Here we have $P = \{1,\ldots,15\}$. Assume that $P_1 = \{1,2,3,6,9,12,15\}$ and now we want to define P_2. We can define $S_2 = \{4,7,10,13\}$ since any node in S_2 is a child node of a node in P_1. Then we can define the following sets: $B(4) = \{4,5\}$, $B(7) = \{7,8\}$, $B(10) = \{10,11\}$, $B(13) = \{13,14\}$ and hence $P_2 = B(4) \cup B(7) \cup B(10) \cup B(13) = \{4,5,7,8,10,11,13,14\}$. We partition P_2 into $F_{P_2}^1, F_{P_2}^2$ where: $F_{P_2}^1 = B(4) \cup B(7)$, $F_{P_2}^2 = B(10) \cup B(13)$. In this case, a new diagram for SBPS can be described in Figure 2, where $*f_t^{P_1}, *f_t^{B(4)}, *f_t^{B(7)}, *f_t^{B(10)}, *f_t^{B(13)}$ refer to propagation of nodes in $P_1, B(4), B(7), B(10), B(13)$ in topological orders, respectively. $\times p_t^{P_1}, \times p_t^{F_{P_2}^1}, \times p_t^{F_{P_2}^2}$ refer to the correction step where particle weights are multiplied by $p_t^{P_1}, p_t^{F_{P_2}^1}, p_t^{F_{P_2}^2}$, respectively. $\rightleftharpoons^{P_2}$ refers to the particle subpart swappings, where we permute the substates in $F_{P_2}^1$ among particles having the same states for $\mathbf{x}_t^3 \cup \mathbf{x}_t^6$, and where we permute the substates in $F_{P_2}^2$ among particles having the same states for $\mathbf{x}_t^9 \cup \mathbf{x}_t^{12}$. In general case, the swapping set and admissible permutation are defined as follows:

Definition 1. *Let F be a set in an interesting partition of P_j. Let $u_1,\ldots,u_l \in S_j$ are such that $F = \bigcup_{i=1}^{l} B(u_i)$. The set $\mathbf{x}_t^{\cup_{i=1}^{l} B(u_i)} \bigcup \mathbf{x}_{t-1}^{\cup_{i=1}^{l} (T(u_i) \setminus B(u_i))}$ is called a swapping set. A permutation $\sigma : \{1,\ldots,N\} \mapsto \{1,\ldots,N\}$*

is said to be admissible if and only if $\mathbf{x}_s^{(i),h} = \mathbf{x}_s^{(\sigma(i)),h}$ *for all* $i \in \{1,\ldots,N\}$ *and for all nodes* $\mathbf{x}_s^h \in \cup_{s=1}^{t} \cup_{i=1}^{l} \mathbf{pa}_s(\mathbf{x}_s^{u_i})$.

The correctness of the diagram in Figure 2 follows from Proposition 1 (see Appendix for the proof).

Proposition 1. *Under the same assumptions required by SBPS, the set of particles resulting from the diagram of Figure 2 represents probability distribution* $p(\mathbf{x}_t|\mathbf{y}_{1:t})$.

It should be noted that the application of our theoretical framework is not limited for DBNs whose structures are directed trees. However, since the swapping sets and admissible permutations can be identified more easily in directed tree cases, an efficient implementation of the tracking algorithm can be obtained for these cases.

The main advantage of the diagram in Figure 2 compared to that of SBPS is that self-occlusion or constraints between parts of the articulated object can be efficiently taken into account. Let us consider the case of self-occlusion. Actually, there are two interesting cases where the diagram in Figure 2 are particularly useful when dealing with self-occlusion. The first case is when we do not have observation for each part but only for a group of parts. In this case, the evaluation of the likelihood for each body part, as required by SBPS, often gives poor approximations, resulting in bad tracking results due to poor approximations of the global likelihood in Equation 2. Now, supposing that P_1, $F_{P_2}^1$ and $F_{P_2}^2$ are groups of parts where we can obtain good observation for each group. In other words, we can obtain good approximations for $p_t^{P_1}$, $p_t^{F_{P_2}^1}$ and $p_t^{F_{P_2}^2}$. Then, the global likelihood can be better approximated as follows:

$$p(\mathbf{y}_t|\mathbf{x}_t) = p_t^{P_1} \times p_t^{F_{P_2}^1} \times p_t^{F_{P_2}^2}$$

The second case where the diagram in Figure 2 is useful for dealing with self-occlusion is when the observation on some parts (*guiding parts*) provides search constraints for some other parts (*primary parts*) that are estimated just before the guiding parts when tracking the object. In this case, it could be more efficient to combine the guiding parts and the primary parts into some set F (for instance, $F_{P_2}^1$ or $F_{P_2}^2$ in the diagram of Figure 2) in order to improve the estimates for the primary parts and thus for the guiding parts.

The diagram in Figure 2 also allows us to apply some types of hard prior which SBPS finds it difficult to deal with. These hard priors eliminate any particle that violates some constraint between parts of the articulated object, thus reduce the search space. Let us consider the case when one wants to impose a pairwise constraint between two parts belonging to two different branches of the tree representing the articulated object. In the case their parent parts are poorly estimated, then at the prediction step of these parts, it could happen that the process of generating new particles for them must be repeated a lot of times to obtain the required number of particles satisfying the constraint. This process may increase the computational time by unpredictable amounts. In such situation, one might want to backtrack and regenerate new configurations for their parent parts. Such a task can not be achieved in SBPS since any part in SBPS is always processed after its parent parts and before its child parts. In the diagram of Figure 2, the capability of processing in parallel a group of parts belonging to the same branch of the tree enables an efficient way of dealing with some types of hard prior.

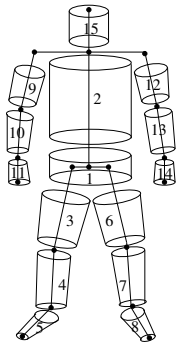

Figure 3: Human body model.

3.2 Our Approach for Human Tracking

3.2.1 Body Model and Likelihood Model

We use the human body model in (Sigal et al., 2010), which is shown in Figure 3. The model consists of 15 parts where body parts are connected by joints. Tracking consists of estimating at each time instant a vector of 34 parameters comprising the global position and orientation of the pelvis and the relative joint angles between neighboring limbs.

The likelihood model consists of a edge-based part and a silhouette-based part, which is commonly-used likelihood model for human body tracking. We refer the reader to (Deutscher and Reid, 2005; Sigal et al., 2010). Our human body tracking problem can be modeled by the DBN in Figure 1.

3.2.2 Self-occlusion Handling

Like SBPS, our idea is to partition P into some subsets. Body parts in these subsets are estimated se-

quentially, while within each subset, body parts or groups of body parts are processed in parallel. Here, we partition P into 2 subsets: $P_1 = \{1,2\}$, $P_2 = \{3,4,5,6,7,8,9,10,11,12,13,14,15\}$. P_2 is then partitioned into $F^1_{P_2}$ and $F^2_{P_2}$ where: $F^1_{P_2} = \{3,4,5,6,7,8\}$, $F^2_{P_2} = \{9,10,11,12,13,14,15\}$. It is easy to see that the sets $P_1, P_2, F^1_{P_2}, F^2_{P_2}$ follow the definitions in section 3.1. This partition is motivated by the fact that during normal human activities like walking or jogging, the left and right legs often occlude each other and should be processed in parallel. Also, the left and right arms often occlude each other and should be processed in parallel.

3.2.3 Dealing with Hard Prior

In (Sigal et al., 2010), to reduce the search space, a hard prior is applied that eliminates any particle that corresponds to implausible body poses such as having angles exceeding anatomical joint limits or interpenetrating limbs. This has been shown to improve significantly the performance of the tracking algorithms. By partitioning P into P_1, P_2, such a hard prior can naturally be applied in our approach. More precisely, when processing P_2, we can test for intersections between the left and right calves, since P_2 contains all parameters describing the configuration of the legs. Similarly, test for intersections between the lower arms and the torso can be performed when processing P_2, since P_2 contains all parameters describing the configuration of the arms, and at the time we process P_2, the configuration of the torso has been previously determined in P_1. Finally, constraints on anatomical joint limits can be imposed at any time we process P_1 and P_2.

3.2.4 Algorithm

Our approach is based on the approach in (Dubuisson and Gonzales, 2012) (we call it SB-PSAPF), which combines the idea of SBPS and Partitioned Sampling Annealed Particle Filter (PSAPF) (Bandouch et al., 2008). In this approach, an annealing run consists of parallel propagation/correction steps for a set of parts, followed by a swapping operation over this set, and finally by a resampling. By testing on a human tracking problem, this approach has been shown to be effective. However, the tracking conditions in this problem is quite simple, where no cluttered background is present and no self-occlusion occurs during tracking. Actually, these conditions are necessary for the assumptions in Equation 2 to be satisfied. In real world problems, when these assumptions do not hold, the global likelihood might be poorly approximated, resulting in poor tracking results. Furthermore, the use of the swapping operation in multiple layers framework, as proposed in SB-PSAPF, creates another issue. Although this operation generates more particles near the modes of the target distribution, it creates the well-known *sample impoverishment* problem. The reason is that after applying it on a particle set, it increases the differences between particle weights. After resampling, only particles with highest weight are multiplied, while the remaining particles (those with medium weights and low weights) have little chance to survive. In SB-PSAPF, the sample impoverishment problem is worst since at each layer, the swapping operation is performed once and the diversity of the particle set decreases as the number of layers increases.

Another drawback of SB-PSAPF is that it estimates the first set of body parts, without taking into account their relation with the remaining body parts. In our case, it estimates the pelvis and the torso without looking at the head and the limbs. In practice, however, some body parts, such as and the legs and the head often provide important constraints for finding the pelvis and the torso, and therefore it is not always possible to localize the pelvis and the torso separately from other body parts. When the pelvis and the torso are poorly estimated, this affects the estimates of the head and the limbs and the performance of the tracking algorithm degrades.

To address the problems discussed above, we propose a two-stages tracking strategy, where in the first stage, human body parts are tracked using APF and in the second stage, the estimates of the head and the limbs are refined using SB-PSAPF, except that we omit the optimization step for the pelvis and the torso. At each time instant, an annealing run in the second stage of our algorithm consists of the following steps:

Step 1: The particle set is resampled

Step 2: The parts in P_2 are propagated using their dynamic functions. The hard prior, which is discussed in this section, is applied. This step is repeated until obtaining the required number of particles.

Step 3: For each particle, the likelihoods of the substates corresponding to the body parts in $F^1_{P_2}$ and $F^2_{P_2}$ are evaluated (for the sake of convenience, we call them the likelihoods of $F^1_{P_2}$ and $F^2_{P_2}$, respectively). For $F^1_{P_2}$, the legs of the human body model are first projected into edge images and silhouette images and then the likelihood of $F^1_{P_2}$ is computed as in (Sigal et al., 2010). In this way, one evaluation of the likelihood function in this step requires less computation time than that in APF, since only body parts related to the evaluation are projected.

Step 4: The substates corresponding to the body parts in $F^1_{P_2}$ can be permuted among particles having the same value for the pelvis. Also, the substates cor-

Table 1: Tracking errors and standard deviations (mm) of APF, SB-PSAPF and our approach from tracking the walking sequences of S1, S2, S3 with 590, 438 and 448 frames, respectively. The processing times shown below each method are computed for one frame and one camera.

Sequence	S1	S2	S3
SB-PSAPF (12s)	242±40	230±34	285±45
SB-PSAPF2 (11s)	220±32	211±25	265±35
APF (6s)	125±37	105±20	199±41
Our approach (4s)	120±26	95±12	182±30

responding to the body parts in $F^2_{P_2}$ can be permuted among particles having the same value for the torso. The new weight of each particle is computed by taking the product of the likelihoods of $F^1_{P_2}$ and $F^2_{P_2}$. In order to generate a better particle set, at each step, the particle with highest weight resulting from all possible permutations is constructed and added into the new particle set. This particle is then eliminated from the current particle set and the process is repeated.

Step 5: The particle weights are annealed as in (Deutscher and Reid, 2005) so that about a half of the particle set will survive the next resampling step.

4 EXPERIMENTAL RESULTS

In this section, we quantitatively compare our approach with APF, SB-PSAPF and SB-PSAPF2. For SB-PSAPF, we estimate the pelvis and the torso at the first stage, the head, the left and right upper arms, the left and right thighs at the second stage, the left and right lower arms, the left and right calves at the third stage, and finally the left and right foots, the left and right hands at the forth stage. SB-PSAPF2 is a variant of SB-PSAPF where we use the same partition as our approach. For both SB-PSAPF and SB-PSAPF2, the same hard prior used in our approach is applied to improve their performance. The comparison between SB-PSAPF and SB-PSAPF2 will show the interest of the diagram in Figure 2 when dealing with self-occlusion. In order to ensure a fair comparison, we use the HumanEva dataset (Sigal et al., 2010), which provides an implementation of APF. We implement our algorithm, SB-PSAPF and SB-PSAPF2 within their framework, while keeping the remaining codes unchanged. We use the error measure in (Sigal et al., 2010), which averages the Euclidean distance between 15 markers on the true pose and the corresponding points computed from the estimated pose. The likelihood model is constructed from edges and silhouettes (Sigal et al., 2010). The image sequences from 5 camera views: C1, C2, C3, BW1, BW2 are used. First-order temporal dynamics are used where the sampling covariance is learnt from all training data in the dataset.

For our approach, we use 400 evaluations of the likelihood function per frame, which is equal to 5 layers with 50 particles per layer for the first stage, and 3 layers with 50 particles per layer for the second stage. For APF, we use 500 evaluations of the likelihood function per frame, which is equal to 5 layers with 100 particles per layer. For SB-PSAPF and SB-PSAPF2, we use 1000 evaluations of the likelihood function per frame, which is equal to 5 layers with 50 particles per layer in 4 stages for SB-PSAPF, and 5 layers with 100 particles per layer in both stages for SB-PSAPF2.

We present in this section the tracking results for the walking sequences of S1, S2 and S3. Table 1 shows the tracking errors and standard deviations of 4 approaches from tracking S1, S2 and S3. As can be observed, SB-PSAPF has the worst performance. Due to self-occlusions between the legs and between the torso and the arms, the evaluation of the likelihood for each body part, as required by SB-PSAPF, gives poor approximations. This leads to poor performances of SB-PSAPF. By taking into account self-occlusion, SB-PSAPF2 improves the performance of SB-PSAPF. For 3 sequences, our approach achieves better performances than the other approaches in term of tracking error, standard deviation and computation time. We can observe that the tracking errors obtained from tracking S3 are higher than those obtained from tracking S1 and S2. This is due to larger movements of the legs of S3 compared to S1 and S2. We can also observe that the difference between the tracking errors of our approach and APF from tracking S3 are larger that those obtained from tracking S1 and S2. This result suggests that our approach is robust in tracking people with strong motions.

Figure 4 shows some frames from tracking S3, using APF and our approach. At frame 100, our approach fails to track the right leg, as APF does (see the 2nd column). However, at the next frames, our approach can recover from tracking failure and tracks the legs quite well, while APF fails to track one of the two legs correctly. This highlights the effectiveness of the second stage of our approach in refining the estimate for the limbs. Figure 5 shows the track-

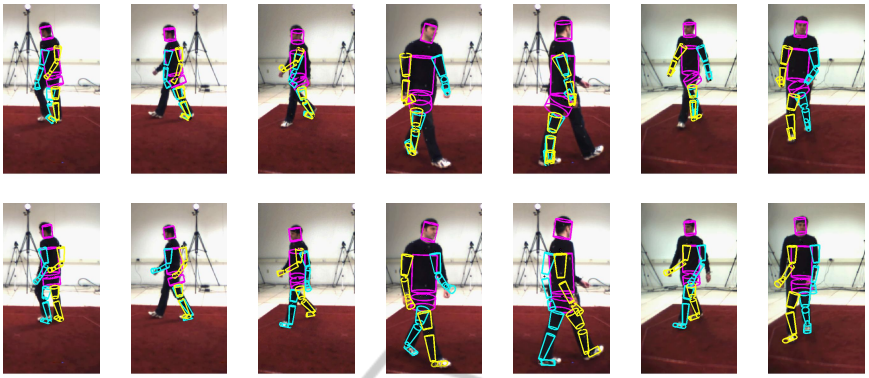

Figure 4: Tracking results of the sequence S3 walking for frames 51,100, 240 303,340, 382. APF (*1*th row), our approach (*2*nd row). The images in the first 5 frames are obtained from one view and those in the last 2 frames are obtained from another view to better visualize the tracking results for the legs of 2 approaches.

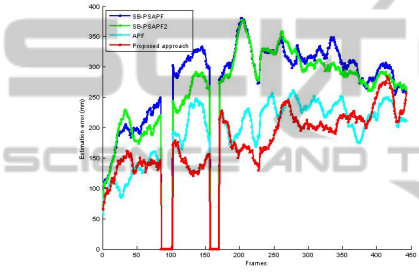

Figure 5: Tracking errors (mean over 10 runs) of SB-PSAPF, SB-PSAPF2, APF and our approach for the walking sequence of S3.

ing error curves of 4 approaches from tracking S3. There are two intervals where we can not obtain the groundtruth, due to bad data. Hence, we set the tracking errors of 4 approaches to zero for these intervals. As can be observed, our approach is clearly better than the other approaches at most of the frames.

5 CONCLUSIONS

We have presented a new particle filter based approach for 3D human body tracking. Our approach consists of 2 stages. At the first stage, the human body is tracked using Annealed Particle Filter. Then, the limbs are refined using another stage, where more particles near the modes of the likelihood function are generated by swapping the substates of the particles. This second stage is also based on the framework of Annealed Particle Filter. Our quantitative evaluation on real sequences has shown the robustness of the proposed approach. It should be noted that our theoretical framework proposed in section 3.1, can be applied not only for tracking articulated objects but also for tracking multiple objects interacting with each others. Here, the advantage compared to SBPS is that occlusions and constraints between objects can be effectively taken into account. Although our approach improves the estimate of the limbs thanks to its second stages, the obtained tracking results are not satisfactory due to self-occlusion. Our current work focus on developing more efficient methods for dealing with these problems within our framework.

REFERENCES

Bandouch, J., Engstler, F., and Beetz, M. (2008). Evaluation of Hierarchical Sampling Strategies in 3D Human Pose Estimation. In *BMVC*, pages 925–934.

Deutscher, J. and Reid, I. (2005). Articulated body motion capture by stochastic search. *International Journal of Computer Vision*, 61(2):185–205.

Dubuisson, S. and Gonzales, C. (2012). An optimized dbn-based mode-focussing particle filter. In *Proceedings of the 2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society.

Dubuisson, S., Gonzales, C., and Nguyen, X. S. (2011). Swapping-based partitioned sampling for better complex density estimation: Application to articulated object tracking. In *International Conference on Scalable Uncertainty Management*, pages 525–538.

Dubuisson, S., Gonzales, C., and Nguyen, X. S. (2013). Sub-sample swapping for sequential monte carlo approximation of high-dimensional densities in the context of complex object tracking. *International Journal of Approximate Reasoning*, 54(7):934 – 953.

Gordon, N., Salmond, D. J., and Smith, A. (1993). Novel approach to nonlinear/non-Gaussian Bayesian state estimation. *IEE Proceedings of Radar and Signal Processing*.

MacCormick, J. and Isard, M. (2000). Partitioned sampling, articulated objects, and interface-quality hand tracking. In *ECCV*, pages 3–19.

Murphy, K. P. (2002). *Dynamic Bayesian Networks : Rep-*

resentation, Inference and Learning. PhD thesis, University of California, Berkeley.

Rose, C., Saboune, J., and Charpillet, F. (2008). Reducing particle filtering complexity for 3D motion capture using dynamic Bayesian networks. *ICAI*, pages 1396–1401.

Sigal, L., Balan, A., and Black, M. (2010). Humaneva: Synchronized video and motion capture dataset and baseline algorithm for evaluation of articulated humanmotion. *International Journal of Computer Vision*, 87(1-2):4–27.

Zhang, X., Hu, W., Wang, X., Kong, Y., Xie, N., Wang, H., Ling, H., and Maybank, S. (2010). A swarm intelligence based searching strategy for articulated 3D human body tracking. In *CVPR*, pages 45–50.

APPENDIX

Proposition 1. We denote by $Q_j = \cup_{h=1}^{j} P_h$ and $R_j = \cup_{h=j+1}^{K} P_h$ the set of parts already processed after the jth step of the diagram in Figure 2 (the jth step consists of processing parts in P_j) and those still to be processed, respectively. By convention, we set $Q_0 = R_K = \emptyset$. We shall show that 1) after the propagations of parts in P_j (prediction step), the particle set represents the density $p(\mathbf{x}_t^{Q_j}, \mathbf{x}_{t-1}^{R_j} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_{j-1}})$; 2) after the corrections of parts in P_j, the particle set represents the density $p(\mathbf{x}_t^{Q_j}, \mathbf{x}_{t-1}^{R_j} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_j})$; 3) after the swapping of parts in P_j, the particle set represents the density $p(\mathbf{x}_t^{Q_j}, \mathbf{x}_{t-1}^{R_j} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_j})$; Since the resampling step does not alter the distribution represented by the particle set, at the last step (Kth step) of the diagram in Figure 2, the particle set represents the density $p(\mathbf{x}_t^{Q_K}, \mathbf{x}_{t-1}^{R_K} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_K}) = p(\mathbf{x}_t^{P} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{P}) = p(\mathbf{x}_t | \mathbf{y}_{1:t})$.

We prove 1), 2), 3) by introduction on j. Assume that after the *(j-1)*th step, the particle set represents the density $p(\mathbf{x}_t^{Q_{j-1}}, \mathbf{x}_{t-1}^{R_{j-1}} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_{j-1}})$. For the proof of 1) and 2), we refer the reader to (Dubuisson et al., 2013). We now prove 3). Let F be a set in an interesting partition of P_j and let $u_1, \ldots, u_l \in S_j$ are such that $F = \bigcup_{i=1}^{l} B(u_i)$. Denote $\mathbf{x}_t^C = \mathbf{x}_t^{\cup_{i=1}^{l} B(u_i)}$, $\mathbf{x}_{t-1}^D = \mathbf{x}_{t-1}^{\cup_{i=1}^{l} (T(u_i) \setminus B(u_i))}$, $\mathbf{x}_t^A = \cup_{i=1}^{l} \mathbf{pa}_t(\mathbf{x}_t^{u_i})$, $\mathbf{x}_t^V = \mathbf{x}_t^{Q_j \setminus (C \cup A)}$, $\mathbf{x}_{t-1}^W = \mathbf{x}_{t-1}^{R_j} \setminus \mathbf{x}_{t-1}^D$. From 2), after the corrections of parts in P_j, the particle set estimates the distribution $p(\mathbf{x}_t^{Q_j}, \mathbf{x}_{t-1}^{R_j} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_j})$. We have:

$$
\begin{aligned}
&p(\mathbf{x}_t^{Q_j}, \mathbf{x}_{t-1}^{R_j} | \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_j}) \propto p(\mathbf{x}_t^{Q_j}, \mathbf{x}_{t-1}^{R_j}, \mathbf{y}_{1:t-1}, \mathbf{y}_t^{Q_j}) \\
&= p(\mathbf{x}_t^{C \cup A \cup V}, \mathbf{x}_{t-1}^{D \cup W}, \mathbf{y}_{1:t}^{C \cup A \cup V}, \mathbf{y}_{1:t-1}^{D \cup W}) \\
&= \int p(\mathbf{x}_t^{C \cup V}, \mathbf{x}_{1:t}^{A}, \mathbf{x}_{t-1}^{D \cup W}, \mathbf{y}_{1:t}^{C \cup A \cup V}, \mathbf{y}_{1:t-1}^{D \cup W}) d\mathbf{x}_{1:t-1}^{A} \\
&= \int p(\mathbf{x}_{1:t}^{A}, \mathbf{y}_{1:t}^{A}, \mathbf{x}_t^{C}, \mathbf{x}_{t-1}^{D}, \mathbf{y}_{1:t}^{C}, \mathbf{y}_{1:t-1}^{D}, \mathbf{x}_t^{V}, \mathbf{x}_{t-1}^{W}, \mathbf{y}_{1:t}^{V}, \mathbf{y}_{1:t-1}^{W}) \\
&d\mathbf{x}_{1:t-1}^{A} \\
&= \int p(\mathbf{x}_{1:t}^{A}, \mathbf{y}_{1:t}^{A}) . p(\mathbf{x}_t^{C}, \mathbf{x}_{t-1}^{D}, \mathbf{y}_{1:t}^{C}, \mathbf{y}_{1:t-1}^{D} | \mathbf{x}_{1:t}^{A}, \mathbf{y}_{1:t}^{A}) \\
&. p(\mathbf{x}_t^{V}, \mathbf{x}_{t-1}^{W}, \mathbf{y}_{1:t}^{V}, \mathbf{y}_{1:t-1}^{W} | \mathbf{x}_{1:t}^{A}, \mathbf{y}_{1:t}^{A}, \mathbf{x}_t^{C}, \mathbf{x}_{t-1}^{D}, \mathbf{y}_{1:t}^{C}, \mathbf{y}_{1:t-1}^{D}) d\mathbf{x}_{1:t-1}^{A} \\
&= \int p(\mathbf{x}_{1:t}^{A}, \mathbf{y}_{1:t}^{A}) . p(\mathbf{x}_t^{C}, \mathbf{x}_{t-1}^{D}, \mathbf{y}_{1:t}^{C}, \mathbf{y}_{1:t-1}^{D} | \mathbf{x}_{1:t}^{A}) \\
&. p(\mathbf{x}_t^{V}, \mathbf{x}_{t-1}^{W}, \mathbf{y}_{1:t}^{V}, \mathbf{y}_{1:t-1}^{W} | \mathbf{x}_{1:t}^{A}, \mathbf{x}_t^{C}, \mathbf{x}_{t-1}^{D}, \mathbf{y}_{1:t}^{C}, \mathbf{y}_{1:t-1}^{D}) \\
&d\mathbf{x}_{1:t-1}^{A} \qquad (4)
\end{aligned}
$$

We will show that: $\{\mathbf{x}_t^C \cup \mathbf{x}_{t-1}^D \cup \mathbf{y}_{1:t}^C \cup \mathbf{y}_{1:t-1}^D\}$ and $\{\mathbf{x}_t^V \cup \mathbf{x}_{t-1}^W \cup \mathbf{y}_{1:t}^V \cup \mathbf{y}_{1:t-1}^W\}$ are independent conditionally to $\{\mathbf{x}_{1:t}^A\}$. In other word, $\{\mathbf{x}_{1:t}^A\}$ d-separate $\{\mathbf{x}_t^C \cup \mathbf{x}_{t-1}^D \cup \mathbf{y}_{1:t}^C \cup \mathbf{y}_{1:t-1}^D\}$ and $\{\mathbf{x}_t^V \cup \mathbf{x}_{t-1}^W \cup \mathbf{y}_{1:t}^V \cup \mathbf{y}_{1:t-1}^W\}$. Since we assume that the observation of a part depends only on this part, all paths that connect $\mathbf{y}_{1:t}^C$ to the rest of the DBN must go through $\mathbf{x}_{1:t}^C$. This is also true for $\mathbf{y}_{1:t-1}^D$, $\mathbf{y}_{1:t}^V$ and $\mathbf{y}_{1:t-1}^W$. Hence, it is enough to show that: $\{\mathbf{x}_{1:t}^A\}$ d-separate $\{\mathbf{x}_{1:t}^C \cup \mathbf{x}_{1:t-1}^D\}$ and $\{\mathbf{x}_{1:t}^V \cup \mathbf{x}_{1:t-1}^W\}$. We will show that: $\{\mathbf{x}_{1:t}^A\}$ d-separate $\{\mathbf{x}_{1:t}^C \cup \mathbf{x}_{1:t}^D\}$ and $\{\mathbf{x}_{1:t}^V \cup \mathbf{x}_{1:t}^W\}$. We have: $\{\mathbf{x}_{1:t}^C \cup \mathbf{x}_{1:t}^D\} = \mathbf{x}_{1:t}^{C \cup D} = \mathbf{x}_{1:t}^{\cup_{i=1}^{l} T(u_i)}$. Consider a path that connects a node in $\mathbf{x}_{1:t}^{\cup_{i=1}^{l} T(u_i)}$ and a node in the rest of the DBN. If this path goes through a node at a time instant $t' > t$, then there must exist a node in the path at a time instant $t'' > t$ such that its two neighbor arcs converge to this node (these arcs point to this node). In this case, neither this node nor its descendants is in $\{\mathbf{x}_{1:t}^A\}$, hence $\{\mathbf{x}_{1:t}^A\}$ d-separate this path. If the path does not go through any node at a time instant $t' > t$, then it must go through a node in $\{\mathbf{x}_{1:t}^A\}$, and in this case the path is blocked at this node. In all cases, we have shown that $\{\mathbf{x}_{1:t}^A\}$ d-separate $\{\mathbf{x}_{1:t}^C \cup \mathbf{x}_{1:t}^D\}$ and the rest of the DBN. Hence $\{\mathbf{x}_{1:t}^A\}$ d-separate $\{\mathbf{x}_{1:t}^C \cup \mathbf{x}_{1:t}^D\}$ and $\{\mathbf{x}_{1:t}^V \cup \mathbf{x}_{1:t}^W\}$.

Now, (4) is equal to:
$\int p(\mathbf{x}_{1:t}^{A}, \mathbf{y}_{1:t}^{A}) . p(\mathbf{x}_t^{C}, \mathbf{x}_{t-1}^{D}, \mathbf{y}_{1:t}^{C}, \mathbf{y}_{1:t-1}^{D} | \mathbf{x}_{1:t}^{A})$
$. p(\mathbf{x}_t^{V}, \mathbf{x}_{t-1}^{W}, \mathbf{y}_{1:t}^{V}, \mathbf{y}_{1:t-1}^{W} | \mathbf{x}_{1:t}^{A}) d\mathbf{x}_{1:t-1}^{A}$

Permuting particles over parts $\mathbf{x}_t^C \cup \mathbf{x}_{t-1}^D$ for fixed value of $\mathbf{x}_{1:t}^A$ cannot change the estimation of density $p(\mathbf{x}_t^C, \mathbf{x}_{t-1}^D, \mathbf{y}_{1:t}^C, \mathbf{y}_{1:t-1}^D | \mathbf{x}_{1:t}^A)$ because estimations by samples are insensitive to the order of the elements in the samples. Moreover, it can neither affect the estimation of density $p(\mathbf{x}_t^V, \mathbf{x}_{t-1}^W, \mathbf{y}_{1:t}^V, \mathbf{y}_{1:t-1}^W | \mathbf{x}_{1:t}^A)$, since $\{\mathbf{x}_t^C \cup \mathbf{x}_{t-1}^D \cup \mathbf{y}_{1:t}^C \cup \mathbf{y}_{1:t-1}^D\}$ and $\{\mathbf{x}_t^V \cup \mathbf{x}_{t-1}^W \cup \mathbf{y}_{1:t}^V \cup \mathbf{y}_{1:t-1}^W\}$ are independents conditionally to $\{\mathbf{x}_{1:t}^A\}$. Hence, permuting particles over the swapping set and within admissible permutations, whose definitions are given in Definition 1, guarantee that the target distribution is correctly estimated.