Simultaneous Partitioned Sampling for articulated object tracking

Christophe Gonzales, Séverine Dubuisson, Xuan Son Nguyen

Laboratoire d'Informatique de Paris 6 (LIP6/UPMC)
4 place Jussieu, 75005 Paris, FRANCE
firstname.name@lip6.fr

Abstract. In this paper, we improve the Partitioned Sampling (PS) scheme to better handle high-dimensional state spaces. PS can be explained in terms of conditional independences between random variables of states and observations. These can be modeled by Dynamic Bayesian Networks. We propose to exploit these networks to determine conditionally independent subspaces of the state space. This allows us to simultaneously perform propagations and corrections over smaller spaces. This results in reducing the number of necessary resampling steps and, in addition, in focusing particles into high-likelihood areas. This new methodology, called *Simultaneous Partitioned Sampling*, is successfully tested and validated for articulated object tracking.

1 Introduction

Articulated object tracking is an important computer vision task for a wide variety of applications including gesture recognition, human tracking and event detection. However, tracking articulated structures with accuracy and within a reasonable time is challenging due to the high dimensionality of the state and observation spaces. In the optimal filtering context, the goal of tracking is to estimate a state sequence $\{\mathbf{x}_t\}_{t=1,\dots,T}$ whose evolution is specified by a dynamic equation $\mathbf{x}_t = \mathbf{f}_t(\mathbf{x}_{t-1}, \mathbf{n}_t^{\mathbf{x}})$ given a set of observations. These observations $\{\mathbf{y}_t\}_{t=1,\dots,T}$, are 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 vector-valued and time-varying transition functions, and $\mathbf{n}_t^{\mathbf{x}}$ and $\mathbf{n}_t^{\mathbf{y}}$ are Gaussian noise sequences, independent and identically distributed. All these equations are usually considered in a probabilistic way and their computation is decomposed in two main steps. First the prediction of the density function $p(\mathbf{x}_t|\mathbf{y}_{1:t-1}) = \int_{\mathbf{x}_{t-1}} p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1})d\mathbf{x}_{t-1}$ with $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ the prior density related to transition function $\mathbf{f}_t$, and then a filtering step $p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{y}_{1:t-1})$ with $p(\mathbf{y}_t|\mathbf{x}_t)$ the likelihood density related to the measurement function $\mathbf{h}_t$.

When functions $\mathbf{f}_t$ and $\mathbf{h}_t$ are linear, or linearizable, and when distributions are Gaussian or mixtures of Gaussians, the sequence $\{\mathbf{x}_t\}_{t=1,\dots,T}$ can be computed analytically by Kalman, Extended Kalman or Unscented Kalman Filters [4]. Unfortunately, most vision tracking problems involve nonlinear functions and non-Gaussian distributions. In such cases, tracking methods based on

particle filters [4, 6], also called Sequential Monte Carlo Methods (SMC), can be applied under very weak hypotheses: their principle is not to compute the parameters of the distributions, but to approximate these distributions by a set of N weighted samples $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}$, also called particles, corresponding to hypothetical state realizations. As optimal filtering approaches do, they consist of two main steps: (i) a prediction of the object state in the scene (using previous observations), that consists in propagating the set of particles $\{\mathbf{x}_t^{(i)}, w_t^{(i)}\}$ according to a proposal function $q(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_t)$, followed by (ii) a correction of this prediction (using a new available observation), that consists in weighting the particles according to a likelihood function, so that $w_t^{(i)} \propto w_{t-1}^{(i)} p(\mathbf{y}_t|\mathbf{x}_t^{(i)}) \frac{p(\mathbf{x}_t^{(i)}|\mathbf{x}_{t-1}^{(i)})}{q(\mathbf{x}_t|\mathbf{x}_{0:t-1}^{(i)}, \mathbf{y}_t)}$, with $\sum_{i=1}^{N} w_t^{(i)} = 1$. Particles can then be resampled, so that those with highest weights are duplicated, and those with lowest weights are suppressed. There exist many models of particle filters, each one having its own advantages. For example, the Condensation algorithm [9] has proved to be robust to clutter and occlusion due to its multiple hypotheses. Unfortunately, the computational cost of particle filters highly depends on the number of dimensions of the state space and, for large state spaces, costs may be unrealistically high due to the large number of particles needed to approximate the distributions and to the costs of computing weights $w_t^{(i)}$. In this paper, we propose a new way to reduce the number of particles necessary to treat high-dimensional state spaces, by reducing the number of resampling steps and the variance of the particle set. This paper is organized as follows. Section 2 gives a short overview of the existing approaches that try to solve this high-dimensionality problem. Section 3 recalls the Partitioned Sampling approach and its limitations, then details our approach. Section 4 gives comparative tracking results on challenging synthetic video sequences. Finally, concluding remarks and perspectives are given in Section 5.

2 Reducing particle filter's complexity: the problem of high-dimentional state spaces

Dealing with high-dimensional state and observation spaces is a major concern of the vision tracking research community, especially when using the particle filter framework for articulated object tracking. There exist essentially two ways to tackle high-dimensional problems: either reduce the dimension of the state space/search space or exploit conditional independences naturally arising in the state space to partition the latter into low-dimensional spaces where few particles are needed.

Among those algorithms that follow the first way, some exploit tailored proposal functions to better guide particles during the prediction step. For instance, in [3], attractors corresponding to specific known state vectors are used to better diffuse particles and then to explore more efficiently the state space. For articulated body tracking, many model-driven approaches using prior knowledge [7] on the movement of articulated parts derived from physical models, have been successfully applied. In [8], the environment is assumed to influence the movement

of body parts and environmental constraints are thus integrated into the tracking scheme. Search into the state space may also be improved using optimization techniques. In [18], a population-based metaheuristic is for instance exploited: a path relinking scheme is used to resample particles so as to avoid missing modes of the probability distribution to estimate. Deutscher *et al.* [5] also proposed the Annealed Particle Filter that consists of adding to the resampling step simulated annealing iterations to diffuse particles into high-likelihood areas. In [2] a new optimization technique is also considered, which is more efficient than previous classical gradient methods because it incorporates constraints.

The second family of approaches consists of reducing the number of necessary particles by exploiting conditional independences in the state space to divide it into small parts. For instance, in [16], graphical models are used to derive conditional density propagation rules and to model interpart interactions between articulated parts of the object. A belief inference is also used in [19], where the articulated body is modeled by a Dynamical Bayesian Network in which inference is computed using both Belief Propagation and Mean Field algorithms. In [1] a body is modeled by a factor graph, and the marginal of each part of the body is computed using Belief Propagations and a specific particle filter. Then, the global estimation consists of recomputing all the weights by taking into account the links between parts of the body. In [17] Bayesian Networks are exploited to factor the representation of the state space, hence reducing the complexity of the particle filter framework. Partitioned Sampling (PS) [11, 12] was proposed by MacCormick and Isard in 2000 and is one of the most popular frameworks. The key idea, that will be described in Section 3.1, is to divide the joint state space $\mathbf{x}_t$ into a partition of P elements, i.e. one element per object part, and for each one, to apply the transition function and to perform a weighted resampling operation. PS was first applied in multiple object tracking. The order of treatment of the objects was fixed over time, which made it fail when there were occlusions. In Dynamic Partition Sampling [20], the posterior distribution is represented by a mixture model, whose mixture components represent a specific order of treatment of the objects. However, when the set of configurations is large, this approach becomes intractable. The Ranked Partition Sampling [21] proposes to simultaneously estimate the order of treatment of the objects and their distributions. For the articulated object tracking purpose, PS suffers from numerous resampling steps that increase noise as well as decrease the tracking accuracy over time. We propose in this paper an adaptation of PS to efficiently deal with an articulated object by exploiting independence between its parts to reduce the number of resampling steps. We call this new methodology *Simultaneous Partitioned Sampling* (SPS) and derive its modeling in the next section.

3 Proposed approach

3.1 Partitioned Sampling (PS)

Partitioned Sampling (PS) is a very effective Particle Filter that exploits some decomposition of the system dynamics in order to reduce the number of par-

ticles needed to track objects when the state space dimensions are large. The basic idea is to divide the state space into an appropriate set of partitions and to apply sequentially a Particle Filter on each partition, followed by a specific resampling ensuring that the sets of particles computed actually represent the joint distributions of the whole state space.

This specific resampling is called a "Weighted Resampling". Let $g : \mathcal{X} \mapsto \mathbb{R}$ be a strictly positive continuous function on $\mathcal{X}$ called a *weighting function*. Given a set of particles $\mathcal{P}_t = \{\mathbf{x}_t^{(i)}, w_t^{(i)}\}_{i=1}^N$ with weights $w_t^{(i)}$, weighted resampling produces a new set of particles $\mathcal{P}'_t = \{\mathbf{x'}_t^{(i)}, w'_t^{(i)}\}_{i=1}^N$ representing the same distribution as $\mathcal{P}_t$ while located at the peaks of function g. To achieve this, let an "*importance distribution*" ρ_t be defined on $\{1, \ldots, N\}$ by $\rho_t(i) = g(\mathbf{x}_t^{(i)}) / \sum_{j=1}^N g(\mathbf{x}_t^{(j)})$ for $i = 1, \ldots, N$. Select independently indices $k_1, \ldots, k_N$ according to probability ρ_t. Finally, construct a set of particles $\mathcal{P}'_t = \{\mathbf{x'}_t^{(i)}, w'_t^{(i)}\}_{i=1}^N$ defined by $\mathbf{x'}_t^{(i)} = \mathbf{x}_t^{(k_i)}$ and $w'_t^{(i)} = w_t^{(k_i)} / \rho_t(k_i)$. MacCormick [10] shows that $\mathcal{P}'_t$ represents the same probability distribution as $\mathcal{P}_t$ while focusing on the peaks of g.

The basic idea underlying Partitioned Sampling is to exploit a "natural" decomposition of the system dynamics w.r.t. subspaces of the state space in order to apply Particle Filtering only on those subspaces. This allows for a significant reduction in the number of particles needed to track complex objects. More precisely, assume that state space $\mathcal{X}$ can be partitioned as $\mathcal{X} = \mathcal{X}^1 \times \cdots \times \mathcal{X}^P$, i.e., the system is viewed as being composed of P parts. For instance, a system representing a hand could be decomposed as $\mathcal{X}^{\text{hand}} = \mathcal{X}^{\text{palm}} \times \mathcal{X}^{\text{thumb}} \times \mathcal{X}^{\text{index}} \times \mathcal{X}^{\text{middle}} \times \mathcal{X}^{\text{ring}} \times \mathcal{X}^{\text{little}}$. In addition, assume that the dynamics of the whole system follows this decomposition, i.e., that there exist functions $f_t^i : \mathcal{X} \mapsto \mathcal{X}$ satisfying that the projection of $\mathbf{x}'$ over $\mathcal{X}^1 \times \cdots \times \mathcal{X}^{i-1}$ equals that of $\mathbf{x}$ whenever $\mathbf{x}' = f_t^i(\mathbf{x})$ and such that:

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

where $\circ$ is the usual function composition operator. By definition, each function f_t^i can propagate the particles over subspace $\mathcal{X}^i \times \cdots \times \mathcal{X}^P$, i.e., it can only modify the substates of the particles defined on $\mathcal{X}^i \times \cdots \times \mathcal{X}^P$. However, in practice, function f_t^i usually just modifies the substate defined on $\mathcal{X}^i$.

One step of a "standard" Particle Filter would resample particles, propagate them using proposal function f_t and finally update the particle weights using the observations at hand. Here, exploiting the features of weighted resampling, Partitioned Sampling achieves the same result by substituting the f_t propagation by a sequence of applications of the f_t^i followed by weighted resamplings, as shown in Fig. 1. In this figure, operations "$*f_t^i$" refer to propagations of particles using proposition function f_t^i as defined above and operations "$\sim g_t^i$" refer to weighted resamplings w.r.t. importance function g_t^i. Of course, to be effective, PS needs g_t^i to be peaked with the same region as the posterior distribution restricted to $\mathcal{X}^i$. As an example, assume that $\mathcal{X} = \mathcal{X}^1 \times \mathcal{X}^2$ and consider that the large square on Fig. 2.a represents the whole of $\mathcal{X}$. Then, the effect of propagating particles according to f_t^1 and resampling w.r.t. g_t^1 corresponds to direct the set of particles into the vertical shaded rectangle (where the peaks of g_t^1 are

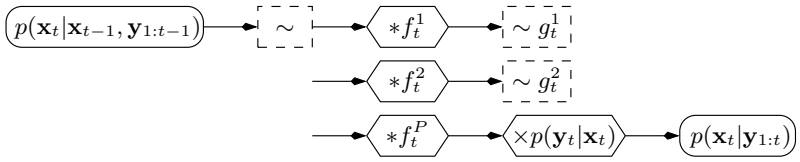
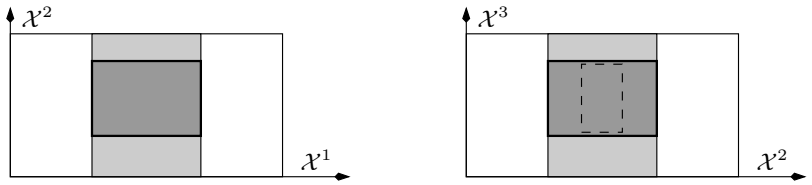


Fig. 1. Partitioned Sampling condensation diagram.



a) Partition Sampling b) Simultaneous Partition Sampling

Fig. 2. Interpretation of Partitioned Sampling ans Simultaneous Partition Sampling.

located). Further propagating these particles using f_t^2 and resampling w.r.t. g_t^2 head them toward the small shaded rectangle that precisely corresponds to the peaks of $p(\mathbf{x}_t|\mathbf{y}_{1:t})$.

This scheme can be significantly improved when the likelihood function decomposes on subsets $\mathcal{X}^i$, i.e., when:

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

where $\mathbf{y}_t^i$ and $\mathbf{x}_t^i$ are the projections of $\mathbf{y}_t$ and $\mathbf{x}_t$ on $\mathcal{X}^i$ respectively. Such a decomposition naturally arises when tracking articulated objects. In these cases, PS condensation diagram can be substituted by that of Fig. 3. MacCormick and Isard show that this new diagram produces mathematically correct results [12].

3.2 Our contribution: Simultaneous Partitioned Sampling (SPS)

In a sense, the hypotheses used by Partition Sampling can best be explained on a dynamic Bayesian Network (DBN) representing the conditional independences

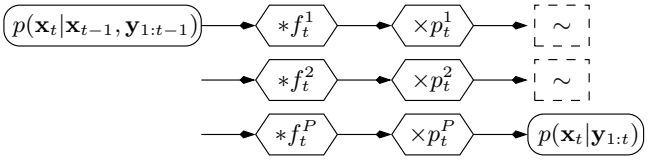


Fig. 3. Improved Partitioned Sampling condensation diagram.

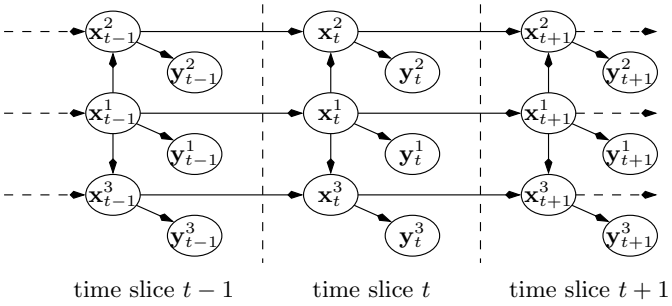


Fig. 4. A Dynamic Bayesian network.

between random variables of states and observations [13]. Assume for instance that an object to be tracked is composed of 3 parts: a torso, a left arm and a right arm. Let $\mathbf{x}_t^1, \mathbf{x}_t^2, \mathbf{x}_t^3$ represent these parts respectively. Then, the probabilistic dependences between these variables and their observations $\mathbf{y}_t^1, \mathbf{y}_t^2, \mathbf{y}_t^3$, can be represented by the DBN of Fig. 4.

In this figure, Eq. (2) implicitly holds because, conditionally to states $\mathbf{x}_t^i$, observations $\mathbf{y}_t^i$ are independent of the other random variables. In addition, the probabilistic dependences between substates $\mathbf{x}_t^1, \mathbf{x}_t^2, \mathbf{x}_t^3$ suggest that the dynamics of the system is decomposable on $\mathcal{X}^1 \times \mathcal{X}^2 \times \mathcal{X}^3$. As a consequence, the condensation diagram of Fig. 3 can be exploited to track the object.

Through the d-separation criterion [14], DBNs offer a strong framework for analyzing probabilistic dependences among sets of random variables. By this criterion, it can be remarked that, on Fig. 4, $\mathbf{x}_t^3$ is independent of $\mathbf{x}_t^2$ conditionally to $\mathbf{x}_t^1$ and $\mathbf{x}_{t-1}^3$. Similarly, $\mathbf{x}_t^2$ can be shown to be independent of $\mathbf{x}_t^3$ conditionally to $\mathbf{x}_t^1$ and $\mathbf{x}_{t-1}^2$. As a consequence, propagations/corrections over subspaces $\mathcal{X}^2$ and $\mathcal{X}^3$ can be performed simultaneously (since they are independent). This suggests the condensation diagram of Fig. 5, which we call a *Simultaneous Partitioned Sampling* (SPS). It is easily seen that, as for PS, the set of particles resulting from SPS represents probability distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$.

The major difference with PS is that, by resampling only after both $\mathbf{x}_t^2$ and $\mathbf{x}_t^3$ have been processed, we can gain in accuracy. Actually, consider Fig. 2.b in which the shaded rectangles explain how PS achieves concentrating iteratively on the peaks of $p(\mathbf{x}_t|\mathbf{y}_{1:t})$. After processing subspace $\mathcal{X}^2$, PS focuses on the light

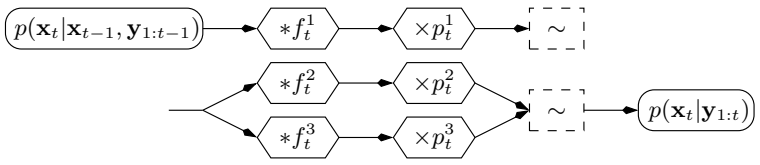


Fig. 5. Basic Simultaneous Partitioned Sampling condensation diagram.

gray rectangle. Therefore, this rectangle is determined only using observation $\mathbf{y}_t^2$. If, for a given particle, substate $\mathbf{x}_t^1$ was near the edge of a rectangle (i.e., not too close to a peak), then one observation $\mathbf{y}_t^2$ may be insufficient to discard this value of $\mathbf{x}_t^1$ whereas two observations $\mathbf{y}_t^2$ and $\mathbf{y}_t^3$ may well be sufficient. In other words, taking into account multiple independent observations can focus the particles on smaller peaked regions of the state space. For instance, on Fig. 2.b, it may well be the case that, instead of ending up with particles located in the dark shaded area, SPS focus them on the smaller dashed rectangle.

Of course, to be effective, SPS needs that propagations/corrections on all the subspaces processed simultaneously be "good" in the sense that they end up with high weights. A naive approach to SPS would not guarantee this property. For instance, on the example of Fig. 4, a particle may be close to the true state of the left arm and far from that of the right arm. In such a case, the overall weight of the particle would be low. If numerous particles had this feature, SPS would perform poorly. Fortunately, the conditional independences exploited by SPS also enable a *substate swapping* operation that improves significantly the concentration of the particles in the high likelihood areas. The idea is that if two particles, say $(\mathbf{x}_t^{1,(i)}, \mathbf{x}_t^{2,(i)}, \mathbf{x}_t^{3,(i)})$ and $(\mathbf{x}_t^{1,(j)}, \mathbf{x}_t^{2,(j)}, \mathbf{x}_t^{3,(j)})$, have the same substate on $\mathcal{X}^1$, i.e., $\mathbf{x}_t^{1,(i)} = \mathbf{x}_t^{1,(j)}$, then we can swap their substates on $\mathcal{X}^2$ or $\mathcal{X}^3$ (thus creating new particles $(\mathbf{x}_t^{1,(i)}, \mathbf{x}_t^{2,(j)}, \mathbf{x}_t^{3,(i)})$ and $(\mathbf{x}_t^{1,(j)}, \mathbf{x}_t^{2,(i)}, \mathbf{x}_t^{3,(j)})$) without altering the probability distribution represented by the set of particles. This feature is actually guaranteed by the d-separation criterion. Consequently, if one particle is close to the true state of the left arm and far from that of the right arm while another is close to the true state of the right arm and far from that of the left arm, provided they have the same value on $\mathcal{X}^1$, we can substitute them by a new particle close to the true states of both arms and a new particle far from those true states. Of course, after swapping, resampling will essentially take the best particle into account. By having the best values on both $\mathcal{X}^2$ and $\mathcal{X}^3$, this particle allows SPS to concentrate on smaller high-peaked regions than PS. This leads to the new condensation diagram of Fig. 6, where $\rightleftharpoons^{2,3}$ represents substate swapping on $\mathcal{X}^2$ and $\mathcal{X}^3$.

Of course, this scheme can be easily generalized. Assume that $\mathcal{X} = \prod_{i=1}^{P} \mathcal{X}^i$. Partition set $\{\mathcal{X}^i\}_{i=1}^{P}$ into subsets $\mathcal{Y}^j = \{\mathcal{X}^i\}_{i \in I_j}$ such that $\cup I_j = \{1, \ldots, P\}$ and such that each pair $\mathcal{X}^{i_1}, \mathcal{X}^{i_2}$ belonging to the same set $\mathcal{Y}^j$ are independent conditionally to subspaces $\mathcal{X}^i \in \mathcal{Y}^{j'}$ with $j' < j$. Then, for each j, the simultaneous propagation/correction/swapping/resampling scheme described above can

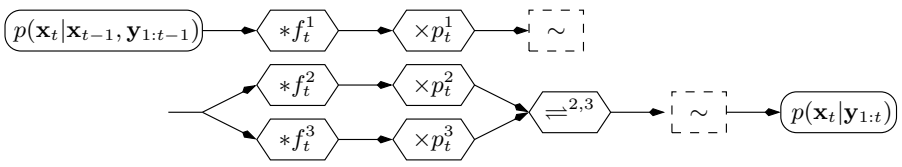


Fig. 6. Complete Simultaneous Partitioned Sampling condensation diagram.

be applied on the sets of $\mathcal{Y}^j$ (note however that, to guarantee that probability distributions remain unchanged, swapping some substate $x_t^i \in \mathcal{Y}^j$ actually requires swapping accordingly all the substates x_t^s that are not d-separated from x_t^i conditionally to $\{x_t^r \in \mathcal{Y}^k : k < j\}$). In the next section, we highlight the advantages of SPS by comparing it to PS on challenging synthetic video sequences.

4 Experimental results

We have chosen to test our method and to compare it with PS on synthetical video sequences because we wanted to highlight its interest in terms of dimensionality reduction and tracking accuracy without having to take into account specific properties of images (noise, *etc.*). Moreover, it is possible to simulate specific motions and then to test and compare with accuracy our method with PS. We have generated our own synthetic video sequence, each one containing 300 frames, showing a P-part articulated object ($P = \{3, 5, 7, 9, 11\}$) translating and distorting over time, see examples of frames in Fig. 7. The goal, here, is then to observe the capacity of PS and SPS to deal with articulated objects composed of a varying number of parts and subject to weak or strong motions.

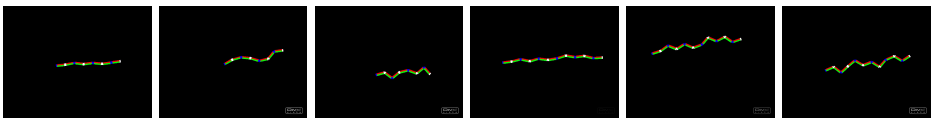


Fig. 7. Some frames of synthetic sequences: an articulated object with 7 and 9 parts.

The tracked articulated object is modeled by a set of P rectangles. The state space contains parameters describing each rectangle, and is defined by $\mathbf{x}_t = \{\mathbf{x}_t^1, \mathbf{x}_t^2, \ldots, \mathbf{x}_t^P\}$, with $\mathbf{x}_t^p = \{x_t^p, y_t^p, \theta_t^p\}$, where (x_t^p, y_t^p) denotes the center of the p^{th} rectangle, and θ_t^p is its orientation. A particle $\mathbf{x}_t^{(i)} = \{\mathbf{x}_t^{1,(i)}, \mathbf{x}_t^{2,(i)}, \ldots, \mathbf{x}_t^{P,(i)}\}$ is then a possible configuration of an articulated object. In the first frame, particles are uniformly generated around the object. During the prediction step, particles are propagated following a random walk whose variance has been manually chosen. The weights of the particles are then computed using the current observation (i.e. the current frame). A classical approach consists in integrating the color distributions given by histograms into particle filtering [15], by measuring the similarity between the distribution of pixels in the region of the estimated parts of the articulated object and of the corresponding reference region. This similarity is determined by computing the Bhattacharyya distance d between the histograms of the target and the reference regions. Finally, the particle's weights are given by $w_t^{(i)} = w_{t-1}^{(i)} p(\mathbf{y}_t | \mathbf{x}_t^{(i)}) \propto w_{t-1}^{(i)} e^{-\lambda d^2}$, with $\lambda = 50$ in our tests. For both approaches, the articulated object global joint distribution is estimated by starting from its center part. PS then propagates and corrects particles part after part to derive a global estimation of the object. SPS considers the left and right

parts as totally independent, and thus propagates and corrects simultaneously in these parts. PS and SPS are compared in terms of tracking accuracy. For that, we measure the tracking error as the distance between the ground truth and the estimated articulated object at each instant. This distance is given by the sum of the Euclidean distances between the corners of the estimated rectangles and their corresponding corners of the ground truth shape. We also measure the variance of the particle set.

In the first test, we compare the convergence of PS and SPS. For that, we use synthetic sequences showing the same articulated object. We compute the tracking errors by averaging over 50 different runs, and repeat these 50 runs for different numbers N of particles. Tracking errors of articulated objects composed of 7 and 9 parts are shown in Figure 8.(a-b). As we can see on this figure, SPS always outperforms PS, even when N becomes very high. One can also notice that with a very small number of particles ($N = 2$), SPS shows more robustness. In fact SPS is more stable during periods in the sequence where the motion is stronger, when PS totally looses the object to track. Figure 9 confirms the stability of the proposed approach: the variance of the particle set is lower with SPS. This proves that the particles are more concentrated around high likelihood values than with PS. This is mainly due to the fact that PS does perform twice more resampling steps than SPS, introducing more noise.

To test the stability of our approach, we have generated video sequences in which the motion at the beginning of the sequence is strong. Comparative results of tracking errors of PS and SPS are reported in Figure 10, for different articulated object (i.e. containing 3, 5, 7, 9 and 11 parts). Here again we see the SPS is less disturbed by this strong motion that PS. Moreover, Table 1 shows that the efficiency of SPS over PS is not achieved at the expense of response times (SPS is usually not more than 4% slower than PS).

We have also compared the efficiency of PS and SPS in cases of very strong and erratic movements of parts of the articulated object throughout the sequence. Comparative tracking results for a 5-part articulated object are given in Figure 11. We can see two examples of frames of this sequence and the deformation that undergoes the articulated object. Tracking errors are considerably decreased with SPS, as is the variance.

5 Conclusion

We have presented a new methodology, the *Simultaneous Partitioned Sampling*, that uses independence properties to simultaneous propagate and correct particles in conditionally independent subspaces. As a result, the particle set is more concentrated into high-likelihood areas. Thus, the estimation of the probability density of the tracked object is more accurate. Empirical tests have shown that SPS outperforms PS, especially in cases where the object motion is strong and when the dimension of the state space increases (i.e., the number of parts is large). There still remains to validate this approach on real video sequences. There is still room for improving SPS, especially its swapping method. Cur-

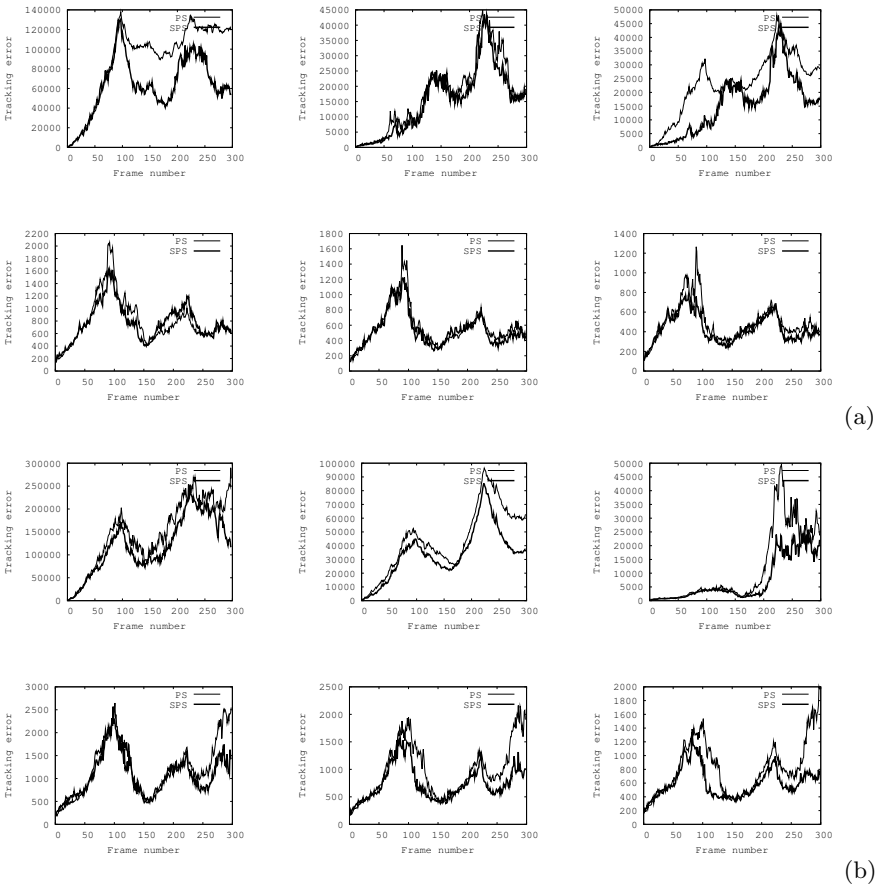


Fig. 8. Convergence study (50 runs): comparison between PS and SPS, from left to right, top to bottom: $N = 2$, $N = 5$, $N = 10$, $N = 20$, $N = 30$ and $N = 40$ for an articulated composed of (a) 7 parts, (b) 9 parts.

rently, we are working on linear programming-based techniques to determine the optimal swappings.

References

1. Bernier, O., Cheungmonchan, P., Bouguet, A.: Fast nonparametric belief propagation for real-time stereo articulated body tracking. Computer Vision and Image Understanding 113(1), 29–47 (jan 2009)
2. Bray, M., Koller-Meier, E., Müller, P., Schraudolph, N.N., Gool, L.V.: Stochastic optimization for high-dimensional tracking in dense range maps. IEE Proceedings Vision, Image and Signal Processing 152(4), 501–512 (2005)

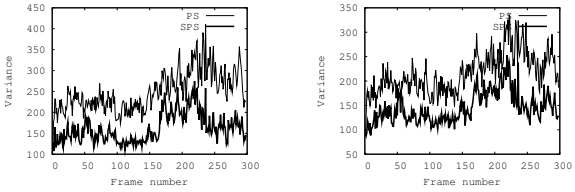


Fig. 9. Comparison of variances (50 runs) obtained for PS and SPS, for a set of $N = 10$ particles for an articulated composed of, from left to right, 7 and 9 parts.

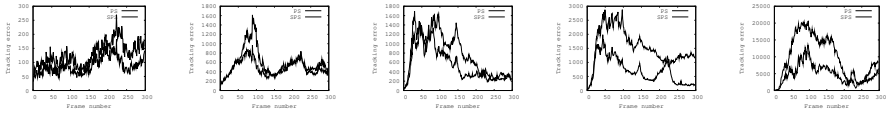


Fig. 10. PS vs. SPS: case of strong motions at the beginning of the sequence, $N = 20$ (50 runs). Tracking errors for an articulated object of, from left to right: 3, 5, 7, 9 and 11 parts.

3. Chang, W.Y., Chen, C.S., Jian, Y.D.: Visual tracking in high-dimensional state space by appearance-guided particle filtering. IEEE Transactions on Image Processing 17(7), 1154–1167 (2008)
4. Chen, Z.: Bayesian filtering: from kalman filters to particle filters, and beyond (2003)
5. Deutscher, J., Davison, A., Reid, I.: Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. In: CVPR. vol. 2, pp. 669–676 (2005)
6. Gordon, N., Salmond, D.J., Smith, A.: Novel approach to nonlinear/non-gaussian bayesian state estimation. IEE Proceedings of Radar and Signal Processing 140(2), 107–113 (1993)

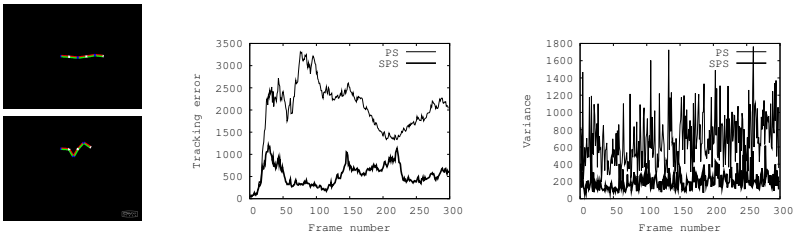


Fig. 11. PS vs. SPS: case of strong motions for a 5-part object, $N = 10$, from left to right: two frames, tracking errors and variance of the particle set (50 runs). SPS outperforms PS because of its capacity to more concentrate particles around high-likelihood areas.

Table 1. Execution times in seconds of PS and SPS, for $N = 10$ to 200 particles (the times reported are averages over 20 runs).

# parts	tracker	10	20	40	60	80	100	200
3	PS	0.54	0.88	1.58	2.99	2.82	3.45	6.80
	SPS	0.56	0.91	1.61	2.33	2.84	3.50	6.92
5	PS	0.71	1.26	2.33	3.34	4.43	5.43	11.62
	SPS	0.74	1.32	2.42	3.43	4.51	5.60	12.05
7	PS	0.95	1.72	3.21	4.7	6.22	7.86	17.10
	SPS	1.00	1.76	3.28	4.84	6.46	7.99	17.73
9	PS	1.21	2.18	4.12	6.16	8.12	10.35	22.06
	SPS	1.25	2.23	4.25	6.32	8.40	10.59	22.77
11	PS	1.41	2.61	4.99	7.38	10.02	12.69	26.37
	SPS	1.46	2.77	5.11	7.61	10.33	13.01	27.18

7. Hauberg, S., Sommer, S., Pedersen, K.: Gaussian-like spatial priors for articulated tracking. ECCV pp. 425–437 (2010)
8. Hauberg, S., Pedersen, K.S.: Stick it! articulated tracking using spatial rigid object priors. In: Kimmel, R., Klette, R., Sugimoto, A. (eds.) ACCV. Lecture Notes in Computer Science, vol. 6494, pp. 758–769 (2010)
9. Isard, M., Blake, A.: Condensation - conditional density propagation for visual tracking. International Journal of Computer Vision 29, 5–28 (1998)
10. MacCormick, J.: Probabilistic modelling and stochastic algorithms for visual localisation and tracking. Ph.D. thesis, Oxford University (2000)
11. MacCormick, J., Blake, A.: A probabilistic exclusion principle for tracking multiple objects. In: ICCV. pp. 572–587 (1999)
12. MacCormick, J., Isard, M.: Partitioned sampling, articulated objects, and interface-quality hand tracking. In: ECCV. pp. 3–19 (2000)
13. Murphy, K.: Dynamic Bayesian Networks: Representation, Inference and Learning. Ph.D. thesis, UC Berkeley, Computer Science Division (2002)
14. Pearl, J.: Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference. Morgan Kaufman Publishers, inc (1988)
15. Pérez, P., Hue, C., Vermaak, J., Gangnet, M.: Color-based probabilistic tracking. In: ECCV. pp. 661–675. London, UK (2002)
16. Qu, W., Schonfeld, D.: Real-time decentralized articulated motion analysis and object tracking from videos. IEEE Transactions on Image Processing 16(8), 2129–2138 (2007)
17. Rose, C., Saboune, J., Charpillet, F.: Reducing particle filtering complexity for 3D motion capture using dynamic bayesian networks. AAAI pp. 1396–1401 (2008)
18. Sánchez, A., Pantrigo, J., Gianikellis, K.: Combining Particle Filter and Population-based Metaheuristics for Visual Articulated Motion Tracking. Electronic Letters on Computer Vision and Image Analysis 5(3), 68–83 (2005)
19. Shen, C., van den Hengel, A., Dick, A., Brooks, M.: 2D articulated tracking with dynamic Bayesian networks. CIT pp. 130–136 (2004)
20. Smith, K., Gatica-perez, D.: Order matters: a distributed sampling method for multi-object tracking. In: BMVC. pp. 25–32 (2004)
21. Widynski, N., Dubuisson, S., Bloch, I.: Introducing fuzzy spatial constraints in a ranked partitioned sampling for multi-object tracking. In: ISVC. vol. LNCS 6453, pp. 393–404. Las Vegas, Nevada, USA (2010)