QUASI MONTE CARLO PARTITIONED FILTERING FOR VISUAL HUMAN MOTION CAPTURE

FONTMARTY Mathias†‡, *DANÈS Patrick*†‡, *LERASLE Frédéric*†‡

† CNRS ; LAAS ; 7 avenue du colonel Roche, F-31077 Toulouse, France
‡Université de Toulouse ; UPS, INSA, INP, ISAE ; LAAS ; F-31077 Toulouse, France

ABSTRACT

Visual Human Motion Capture (HMC) is a motivating challenge in the Computer Vision community as it enables lots of applications. Many methods have been proposed among which Particle Filters (PF) meet a great success. In this paper, we propose a new algorithm, mixing advantages of the PARTITIONED scheme and quasi random methods. We use a trinocular visual system to propose a comparative study of this particle filter against four other classical ones with respect to a ground truth provided by a commercial HMC system.

***Index Terms*—** Motion capture, trinocular system, particle filters, quasi random sampling.

1. INTRODUCTION

The visual based HMC from multi-camera video data and advanced appearance-based approaches has improved in the last decade [2]. The global principle is to infer the configuration of a coarse 3D human kinematic model from its projection in monocular [3] or multi-ocular [4, 5, 6] image sequences. Such a principle enables to derive abundant appearance information from the image contents, especially with multi-view systems which are prone to estimate motion-in-depth more accurately and reliably. Regarding the estimation process, the particle filtering framework [7], first introduced for visual tracking in the form of the CONDENSATION algorithm [8], has proved well suited for HMC application. The key idea is to represent the posterior distribution by a set of samples—or particles—with associated importance weights. This particle set is recursively updated over time taking into account the visual data and the observation model. PFs make no restrictive assumption on the probability distributions entailed in the characterization of the problem, and permit a probabilistic principled fusion of diverse kinds of measurements. However, the main drawback for particle filters remains the high computational cost with regard to the state-space dimensionality. This still prevents real-time human motion capture from becoming a reality.

In this paper, we thus propose an original tracking framework especially designed to be more efficient than other ones for a low number of particles. Section 2 briefly presents the particle filter formalism and introduces our PARTITIONED QRS PF. Our system setup is then described in section 3, including human body model and likelihood function design. Section 4 exposes quantitative evaluations of our method with regard to four other strategies on various motion capture sequences. Finally, section 5 summarizes our contributions and discusses future works.

2. THE QRS PARTITIONED PARTICLE FILTER

2.1. Basics

In a stochastic Bayesian filtering approach to appearance-based motion capture, the 3D template situation and configuration parameters to be estimated are first incorporated in a state vector $\mathbf{x}_k$, whose (given) initial probability density function (pdf) and prior dynamics write as $p_0(\mathbf{x}_0)$ and $p(\mathbf{x}_k|\mathbf{x}_{k-1})$. At any time k, the available visual data, symbolized by $\mathbf{z}_k$, is related to $\mathbf{x}_k$ by the observation density $p(\mathbf{z}_k|\mathbf{x}_k)$. Due to the high number of degrees of freedom (DOF) of the underlying articulated 3D model and to the difficulty to assess its projection onto the current images, the posterior pdf $p(\mathbf{x}_k|\mathbf{z}_{1:k})$ to be estimated is multimodal, defined in an high-dimensional state space, and unavailable in closed-form. A point-mass (or particle) approximation $p(\mathbf{x}_k|\mathbf{z}_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)})$, $\sum_{i=1}^{N} w_k^{(i)} = 1$ is then recursively propagated along time through sequential Monte Carlo estimation methods [7, 8]. An approximation of the minimum mean-square error estimate (MMSEE) $\mathbb{E}(\mathbf{x}_k|\mathbf{z}_{1:k})$ follows.

The celebrated “Sampling Importance Resampling” (SIR) algorithm [7] operates in three major steps. First, the particles $\mathbf{x}_k^{(i)}$, $i = 1..N$ are propagated in the state space through an importance function $q(\mathbf{x}_k|\mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$, selected to adaptively explore relevant areas of the state space. Then, the weights $w_k^{(i)}$ are updated to ensure the consistency of the point-mass approximation taking into account the observation density. Last, when approximation tends to be degenerated, a resampling stage is inserted. Importantly, the SIR framework encompasses the CONDENSATION [8] as well as importance sampling from the current images.

2.2. PARTITIONED particle filter

Contrarily to a common belief, the computation time of a particle filter for general problems, though linear in the number of particles, is exponential in the system order for a fixed dimension-free error [9]. To lower this complexity, many algorithms have been proposed. When the system dynamics comes as the sequence of M partial evolutions $p_m(\mathbf{x}_k^m|\mathbf{x}_k^{m-1})$ of the state vector $\mathbf{x}_k^m$ at step m and when intermediate likelihoods $l_m(\mathbf{x}_k^m|\mathbf{z}_k), m = 1..M$, can be assessed after applying each partial dynamics, PARTITIONED schemes apply [1].

From a succession of sampling operations followed by resampling based on the intermediate likelihoods, the particle cloud can be successively refined towards areas of the state space in which the posterior is dense. The computational complexity then becomes linear in the number of partitions.

2.3. Quasi Monte Carlo filtering methods

Pure random importance sampling leads to “gaps and clusters” in the particle support, especially in high-dimension spaces. An excessive Monte Carlo variation of the predictions can follow, making the filter unreliable or even leading to failures. Substituting the random particles by a deterministic or randomized low-discrepancy—or “Quasi Monte Carlo” (QMC)—sequence can lead to a better convergence rate w.r.t. the number of particles N [10], while lowering the root mean square (RMS) estimation error and leading to a variability reduction from 5% to 20% [11, 12].

Among the main issues on QMC filters are the difficulty to design low-discrepancy sequences in spite of the resampling steps, the exploitation of the current measurement in the definition of these sequences, and the possible trade-off between the reduction of the (quadratic) complexity and the mathematical soundness of the algorithms. A QMC counterpart of CONDENSATION, henceforth termed QRS (for Quasi Random Sampling), is proposed in [13]. We adapted this idea to the PARTITIONED filter proposed in [1]. The final algorithm is described Table 1. The key idea here is to gather propagating and resampling steps. This enables to generate low discrepancy samples from a particle to be resampled, thus resulting in a more regular state space exploration.

$$\{(\mathbf{x}_k^{(i)}, w_k^{(i)})\}_{i=1}^N = PARTITIONEDQRS(\{(\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)})\}_{i=1}^N, \mathbf{z}_k)$$

```
IF $k = 0$, THEN Sample a uniform randomized Sobol QMC sequence $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$ then turn it into $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(N)} \sim p_0(\mathbf{x}_0)$ - Set $w_0^{(i)} = \frac{1}{N}$. END IF
IF $k \geq 1$ THEN
  Set $\tau_0^{(i)} = w_{k-1}^{(i)}$ and $\mathbf{x}_k^{0,(i)} = \mathbf{x}_{k-1}^{(i)}$, $i = 1..N$
  FOR $m = 1..M$, DO
    Independently draw $s^{(1)}, \ldots, s^{(N)}$ into $1..N$ such that $P(s^{(i)} = j) = \tau_{m-1}^{(j)}$ - Set $C_j = card(\{i|s^{(i)} = j\})$
    FOR $j = 1..N$, DO
      Sample a uniform randomized Sobol QMC sequence $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(C_j)}$ then turn it into $\mathbf{x}_k^{m,(\sum_{l=1}^{j-1} C_l+1)}, \ldots, \mathbf{x}_k^{m,(\sum_{l=1}^{j-1} C_l+C_j)}$ i.i.d. according to $p_m(\mathbf{x}_k^m|\mathbf{x}_k^{m-1,(j)})$
    END FOR
    Update the weights $\tau_m^{(i)} \propto l_m(\mathbf{z}_k|\mathbf{x}_k^{m,(i)})$, then normalize them so that $\sum_i \tau_m^{(i)} = 1$
  END FOR
  Set $w_k^{(i)} = \tau_M^{(i)}$ and $\mathbf{x}_k^{(i)} = \mathbf{x}_k^{M,(i)}$ for $i = 1..N$
  Approximate the MMSEE $\mathbb{E}(\mathbf{x}_k|\mathbf{z}_{1:k})$ by $\sum_{i=1}^N w_k^{(i)} \mathbf{x}_k^{(i)}$
END IF
```

Table 1. PARTITIONED QRS: partitioned particle filter exploiting QMC techniques.

3. TRACKER IMPLEMENTATION

3.1. Human body model

Our appearance-based approach infers the human body model from its projection in trinocular image sequences. The whole human body model is fleshed out using truncated cones with fixed dimensions. These geometric primitives are easily handled and hidden part removal can be obtained in closed form. The model is based on a kinematic tree consisting of nine body segments and 22 DOF. We assume Gaussian random walk dynamics.

3.2. Observation likelihoods

It can be argued [4, 14] that appearance-based cues constitute a principled way to derive plethora of measurements, thus offering a nice trade-off in terms of generality, simplicity, and complementarity. Our tracker implementation relies on 3 visual cues involved in the likelihood definition $p(\mathbf{z}_k|\mathbf{x}_k)$:

Silhouette-based likelihood - In the vein of [4], we first perform a foreground-background silhouette segmentation in order to obtain the silhouette mask. We sample points inside the limbs of each projected particle to check whether or not the limbs are consistent with the segmented silhouette.

Dual silhouette-based likelihood - In order to complete this first cue, we counterbalance it by the one proposed in [3]. The principle is symmetric, consisting in sampling the segmented silhouette and check its consistency w.r.t. the current projected particle.

Skin blob-based likelihood - To improve localization accuracy—especially for thin limbs such as arms—, we set up an additional likelihood function involving skin blob detection. For each projected particle, we compute the distance between head and hands and the nearest detected skin blob. The best configurations w.r.t. this cue are the ones showing the lowest distances.

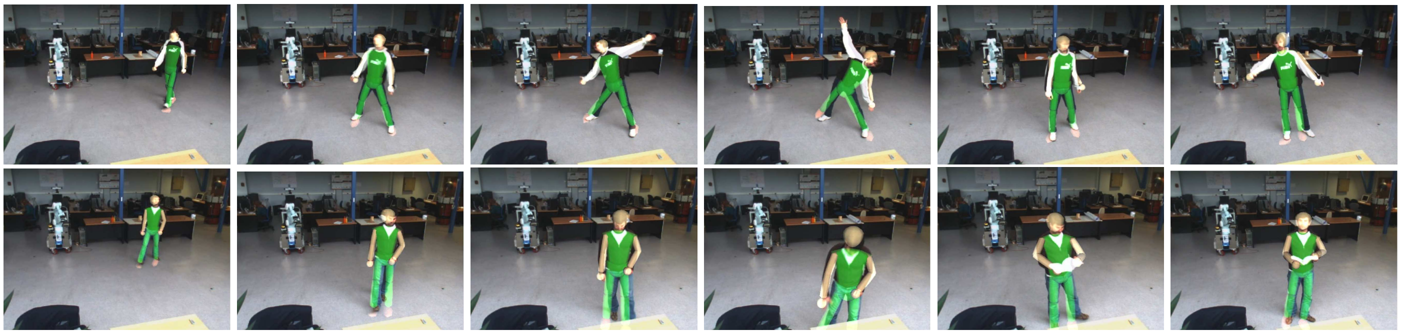

Fig. 1. Snapshots from 2 sequences: walking and gym movement (top), and walking and reading with a different subject (bottom). Only one of the three images is shown due to space limitation.

3.3. System setup

Our vision-based HMC system involves three IEEE 1394 “progressive scan” Flea 2 color cameras providing 640×480 images. We set up the system in a 4×3 m working area in an indoor environment surveillance context. Some representative results are shown in figure 1. We only show images from one of the 3 cameras due to space reasons, what can be somehow misleading about filter accuracy and movement complexity. For a more complete overview, the entire videos can be found at the URL: `www.laas/∼mfontmar`.

Generally speaking, we notice that tracking is correct as soon as a good segmentation of the silhouette is performed, which confirms the results of [15]. However, as we exploit skin segmentation, performances are damaged when the subject does not present hands or face on at least one camera. The results obtained with PARTITIONED QRS strategy are visually more stable than to the ones obtained with APF [4]. However, comparing performances in a qualitative way is not always an easy task, as the projection of the template in the image can be forked. This is why we propose in the next section a quantitative study of our algorithm.

4. PARTICLE FILTER EVALUATIONS

4.1. Evaluation setup

Ground truth positions of the template joints are given by a commercial HMC system from Motion Analysis [16]. It is software calibrated and synchronized with our own trinocular visual tracking system. In order to analyze the average behavior of the filters, 30 runs are performed on each set of data. As we tend to set up a nearly real-time system, evaluations are done limiting N to $100..2000$. We assess the performance on various sequences of ~ 20 s including walking, arm waving, pointing, and fitness. Trackers are initialized by hand.

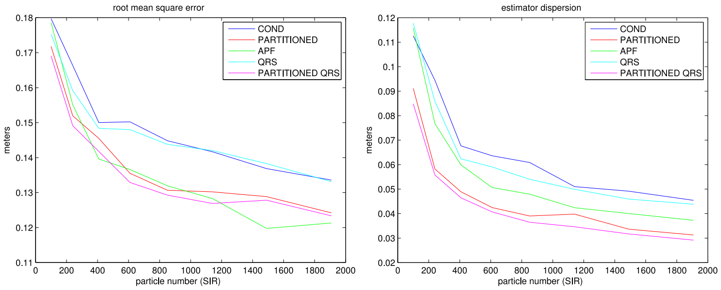

Fig. 2. RMS error (left) and estimator dispersion (right) of the filters for a varying number of particles.

Name	**RMSE**	**Dispersion**	**Failure**	**Bias**
CONDENSATION	5	5	5	5
QRS	4	4	4	4
PARTITIONED	3	2	3	3
PARTITIONED QRS	**1**	**1**	**1**	**2**
APF	**2**	**3**	**2**	**1**

Table 2. Sorting of the different PF according to each criteria.

4.2. Filter comparison

In this section, we present a comparison of our PARTITIONED QRS strategy with CONDENSATION [8], QRS [13], APF [4] and PARTITIONED [1] algorithms. All strategies are normalized with respect to the number of likelihood evaluations which is the most time consuming part. APF is used with 3 layers in order to achieve the minimum number of particles to be efficient [4], and free parameters are tuned following recommendations in [4]. PARTITIONED strategies involve 2 partitions: one for torso position and configuration, and one for member configuration. A summary of the results is proposed in table 2, relatively to the following criteria:

Accuracy - Figure 2 (a) presents the average global RMS error of the estimated joint positions with respect to ground truth, computed over all frames and filter runs. All trackers are globally efficient. Errors are reasonable considering our rough models of the human body, the simple measures and our nearly real-time context. Looking at the relative performances, we can notice that advanced strategies perform better than the classical CONDENSATION, however APF is less efficient than PARTITIONED QRS for a low number of particles. According to [4], APF needs a minimum number of particles per layer to be efficient. Generally speaking, QMC methods provide better estimates than their MC counterparts. For a same given error they can lead to a $20\ \%$ reduction of the number of particles in the best case, which can constitute an important gain in computing time.

Dispersion - Figure 2 (b) presents the variance of the estimated configuration over all frames and filter runs. PARTITIONED schemes present the lowest dispersion. QMC versions of the filters tend to provide a more "stable" estimate than the classical MC ones due to their low discrepancy sampling, even considering our simplified versions designed to match an $\mathcal{O}(n)$ complexity. This seldom exploited criterion reveals that our PARTITIONED QRS methods seems to provide the best results in terms of dispersion, even for a high number of particles. It can improve the dispersion of $3\ cm$ per joint in friendly cases, which constitutes a significant enhancement.

Failure and Bias - These two criteria are not shown in a figure due to lack of space. The failure rate is computed by counting each time one estimated joint position presents an error w.r.t. the ground truth higher than a given threshold. It is fairly consistent with RMSE and dispersion observations. The bias represents the distance between the average joint positions over filter runs and the groundtruth joint positions. It is slightly lower for APF as soon as $N > 1000$ in our context. Under this threshold, PARTITIONED schemes and APF present practically identical results.

Our system actually performs in 1 fps on a Pentium IV $3\ GHz$ while any vision-based approach using particle filtering is far from real time: 0.02 fps in [15], 0.03 fps in [5], 0.07 in [4],... To sum up, it comes out that PARTITIONED QRS provides a better accuracy than the APF for a low number of particles, while proposing the lowest dispersion of the estimates among the five tested PFs. Thus, it can constitute a good alternative for systems with strong time computation constraints.

5. CONCLUSION AND FUTURE WORK

We proposed a new hybrid PARTITIONED QRS PF markerless HMC which we assessed with regard to four other strategies. Advanced filtering techniques provide better results for the same visual cues, nevertheless PARTITIONED QRS outperforms all other algorithms for a low number of particles. Additionally, it leads to the lowest dispersion of the estimates. This definitely makes it well-suited for real-time applications where computational power is limited. A nearly real-time applicative system is proposed in a human cluttered environment to complete the study on several subjects, showing that reactive visual HMC systems seem within reach.

Some interesting future lines of investigation from this work could involve more advanced visual cues, as the ones chosen here are fairly classical. In addition, and to complete this study, one should also take into account importance sampling methods which enable automatic initialization, in order to propose a fully automatic system.

6. REFERENCES

[1] J. MacCormick and M. Isard, "Partitioned sampling, articulated objects, and interface-quality hand tracking," in *European Conference on Computer Vision (ECCV'00)*, Dublin, Ireland, 2000, pp. 3–19. 2, 3

[2] T. Moeslund, A. Hilton, and V. Krüger, "A survey of advanced vision-based human motion capture and analysis," *Computer Vision and Image Understanding (CVIU'06)*, vol. 104, pp. 174–192, Dec. 2006. 1

[3] C. Sminchisescu and B. Triggs, "Estimating articulated human motion with covariance scaled sampling," *International Journal on Robotic Research (IJRR'03)*, vol. 6, no. 22, pp. 371–393, May 2003. 1, 2

[4] J. Deutscher and I. Reid, "Articulated body motion capture by stochastic search," *International Journal of Computer Vision (IJCV'05)*, vol. 21, no. 3, pp. 185–205, 2005. 1, 2, 3, 4

[5] A. Gupta, A Mittal, and L.S. Davis, "Constraint integration for efficient multiview pose estimation of humans with self-occlusions," *Transactions on Pattern Analysis Machine Intelligence (PAMI'08)*, vol. 30, no. 3, pp. 493–506, Mar. 2008. 1, 4

[6] P. Wang and J. Rehg, "A modular approach to the analysis and evaluation of particle filters for figure tracking," in *IEEE Conference on Computer Vision and Pattern Recognition (CVPR'06)*, New York, USA, 2006, pp. 790–797. 1

[7] A. Doucet, N. De Freitas, and N. J. Gordon, *Sequential Monte Carlo Methods in Practice*, Series Statistics For Engineering and Information Science. Springer-Verlag, New York, 2001. 1

[8] M. Isard and A. Blake, "CONDENSATION – Conditional density propagation for visual tracking," *International Journal on Computer Vision (IJCV'98)*, vol. 29, no. 1, pp. 5–28, 1998. 1, 3

[9] F. Daum and J. Huang, "Mysterious computational complexity of particle filters," in *Signal and Data Processing of Small Targets*, Bellingham, MA, USA, Aug. 2003, vol. 4728 of *Proceedings of SPIE*. 2

[10] K.-T. Fang, Y. Wang, and P. M. Bentler, "Some applications of number-theoretic methods in statistics," *Statistical Science*, vol. 9, no. 3, pp. 416–428, 1994. 2

[11] V. Philomin, R. Duraiswami, and L. S. Davis, "Quasi-random sampling for CONDENSATION," in *European Conference on Computer Vision (ECCV'00)*, Dublin, Ireland, 2000, pp. 134–149. 2

[12] D. Ormoneit, C. Lemieux, and D.J. Fleet, "Lattice particle filters," in *Proceedings of the 17th Conference in Uncertainty in Artificial Intelligence (UAI'01)*, San Francisco, CA, USA, 2001, pp. 395–402. 2

[13] D. Guo and X. Wang, "Quasi-Monte Carlo filtering in nonlinear dynamic systems," *IEEE transactions on signal processing*, vol. 54, no. 6, pp. 2087–2098, June 2006. 2, 3

[14] D. Ramanan, D. Forsyth, and A. Zisserman, "Strike a pose: Tracking people by finding stylized poses," in *IEEE Conference on Vision and Pattern Recognition (CVPR'05)*, San Diego, USA, June 2005, pp. 271–278. 2

[15] A. Balan, L. Sigal, and M. Black, "A quantitative evaluation of video-based 3D person tracking," in *International Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS'05)*, Washington, USA, October 2005, pp. 349–356. 3, 4

[16] "http://www.motionanalysis.com — Motion Analysis Corporation," . 3