International Journal of Pattern Recognition
and Artificial Intelligence
Vol. 23, No. 7 (2009) 1–35
© World Scientific Publishing Company

World Scientific
www.worldscientific.com

EVALUATIONS OF PARTICLE FILTER BASED HUMAN MOTION VISUAL TRACKERS FOR HOME ENVIRONMENT SURVEILLANCE

MATHIAS FONTMARTY*, PATRICK DANÈS† and FRÉDÉRIC LERASLE‡

CNRS ; LAAS ; 7, Avenue du Colonel Roche
F-31077 Toulouse, France

and

Université de Toulouse ; UPS, INSA, INP
ISAE ; LAAS-CNRS ; F-31077 Toulouse, France
**mathias.fontmarty@laas.fr*
†*patrick.danes@laas.fr*
‡*frederic.lerasle@laas.fr*

This paper presents a thorough study of some particle filter (PF) strategies dedicated to human motion captured from a trinocular vision surveillance setup. An experimental procedure is used, based on a commercial motion capture ring to provide ground truth. Metrics are proposed to assess performances in terms of accuracy, robustness, but also estimator dispersion which is often neglected elsewhere. Relative performances are discussed through some quantitative and qualitative evaluations on a video database. PF strategies based on Quasi Monte Carlo sampling, a scheme which is surprisingly seldom exploited in the Vision community, provide an interesting way to explore. Future works are finally discussed.

Keywords: Tracking; trinocular vision; data fusion; particle filtering; performance evaluations.

1. Introduction

Achieving markerless Human Motion Capture (HMC) from wide-angle synchronized cameras is a motivating challenge and has been widely investigated during the last decade in the Computer Vision community. These investigations are especially motivated by supportive home environment surveillance applications. The broader technical aims of vision systems are to track the motions of humans in their daily life in order to monitor their behaviors. Such systems could enable elderly people to live longer independently and safely.

Commercial HMC systems are unsuitable as they are expensive, intrusive — due to marker use — and because their portability is limited to home-made environments. The surveillance application requires standard cameras to impose as few restrictions as possible on both the human performer and the viewing conditions. The principle is then to concentrate on advanced vision techniques as well as 3D human body modeling to make the problem tractable.

The development of this application clearly requires extensive performance evaluations and comparisons on surveillance video databases. Comparisons of trackers dedicated to human body parts do exist in the Vision literature but they are typically restricted to low-dimensional problems. This greatly limits their applicability to HMC where at least 20 degrees of freedom are necessary to parameterize the whole body joints. Even in the vision-based HMC literature, no systematic quantitative evaluations are reported, as most relevant papers classically present only qualitative evaluations on key-sequences.[9,35] Quantitative evaluation of human motion recovery requires both videos with ground truth and performance metrics. Except for hand-labeled joints positioning in videos,[16,26] commercial HMC systems have been marginally used[3,25,40] while they could clearly provide this ground truth for large video datasets in order to evaluate the performances and to assess the relative merits of the algorithms. Moreover, the selection of metrics is an open question in the HMC literature, contrarily to other vision problems, e.g. face classification. In the particle filtering context, Wang *et al.* in Ref. 40 proposed both image-based and tracked joint 3D distance metrics while Balan *et al.*[3] and Gupta *et al.*[15] considered multiple 3D distances which seem to be significantly more useful than image distance in many situations.

State-of-the-art HMC systems[13,23] developed in the Vision community consist in estimating the body model configuration in space which best fits the available visual data. They essentially differ in the associated data processing (3D reconstruction versus appearance based approaches), and the estimation framework (deterministic versus stochastic optimization). Reconstruction-based approaches aim at making the model fit with the 3D-point cloud issued from a 3D-sensor system, e.g. a stereo head,[44] a Swiss Ranger[19] or multiple calibrated cameras.[26,36] Besides, the appearance-based approaches infer the model configuration from its projection in monocular[2,22,35] or multi-ocular[2,9,33,43] image sequences. This strategy enables to derive abundant appearance information from the image contents but is prone to misestimate the motion-in-depth especially when using a single camera. Multicameras approaches are employed to increase reliability, accuracy, and reduce problems with self-occlusions.

Regarding the estimation process, deterministic approaches — usually based on local descent search[6,37,38] — are almost neglected, due to their difficulties in handling multimodality. A more general alternative is the particle filtering framework,[10] first introduced for visual tracking purpose in the form of the CONDENSATION algorithm.[17] The key idea is to represent the posterior distribution over the state space by a set of samples — or particles — with associated importance weights. This particle set is first sampled from the state vector initial probability distribution, then updated over time taking into account the measurements and a prior knowledge of the system dynamics/observation models. Particle filters have proved to be well suited to the above requirements. Indeed, they 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. The main drawback for particle filters remains the computational cost which increases exponentially (in terms of the required number of particles to ensure a fixed dimension free error) with the state-space dimensionality.

The vision literature has focused on the two following key strategies to improve particle filters: (i) modify the algorithms themselves to prevent particle wastage, (ii) design more suited dynamic and observation models for the tracked body parts. Though observation models have been intensively studied in order to deal with variable viewing conditions,[1,22,24,31,35] (ii) still constitutes an open problem. Some approaches[32,38,43] assume strongly constrained models to place further restrictions on possible poses, yet such assumptions are problematic for motion capture applications. Indeed, as the interest is to capture novel motions, we cannot rely on previous move models. Clearly, the unpredictability and unusual motions we need to capture limit the models which can be applied.

Alternatively, the methodology (i) aims to directly control the support of the samples. We can here mention search space decomposition techniques,[8] annealed particle filter (APF),[2,7] covariance sampling,[9,35] unscented particle filter (UPF),[39] and Quasi Monte Carlo sampling which has been surprisingly seldom exploited for visual tracking purpose.[27,29] Unfortunately, few studies[3,40] have so far been concerned with the balance against each other, as very few available results compare a restricted number of alternative algorithms (UPF,[40] APF[3]) to the original CONDENSATION approach on a small probe set of videos. Consequently, it is not clear which particle filtering scheme, possibly hybrid, performs best. A thorough study comparing the efficiency of most of the above filtering strategies in terms of the following metrics is carried out hereafter given the ground truth: (1) root mean square error error, (2) tracking failure rate, (3) dispersion of estimates (e.g. estimator variance), (4) bias of the estimates. The dispersion is surprisingly seldom taken into account despite the intrinsic stochastic nature of any PF tracker. The "stability" of the estimates, as opposed to their variability, is clearly an important property as multiple runs on any given surveillance video should lead to homogeneous results. Since the strategies can differ in their computational cost per sample, their relative comparisons are normalized with respect to the computation time. Recall that the ground truth is provided by a commercial HMC setup.[42] These evaluations are performed to exhibit the best filtering strategy to be considered in our near-real-time surveillance video system which is devoted to natural human-centered indoor environment.

This paper is organized as follows. Section 2 briefly sums up the well-known particle filtering formalism, and describes some relevant variants. Our trinocular vision based setup as well as the implementation of trackers are then described in Sec. 3. Evaluations and comparisons between distinct strategies on both synthetic and real sequences are respectively reported in Secs. 4 and 5. Finally, Sec. 6 summarizes our contributions and opens the discussion for future works.

2. Particle Filtering Algorithms

2.1. *Visual-based HMC as a Bayesian filtering problem*

Like many other tracking problems, visual-based human motion capture can be generically addressed in the context of stochastic (Bayesian) filtering. So, the information held in the camera images can be fused with a prior knowledge on the dynamics of the human limbs, enabling a seamless spatio-temporal motion analysis. A state vector $\mathbf{x}$ is first defined, which gathers the 3D model situation and configuration parameters to be estimated. The prior dynamics of the template between two consecutive instants $k-1$ and k must be expressed as a transition kernel $p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})$, which may imply to insert in $\mathbf{x}$ derivatives of the situation/configuration parameters. Besides, the measurement vector $\mathbf{z}_k$ and observation density $p(\mathbf{z}_k \,|\, \mathbf{x}_k)$ respectively symbolize the visual data available at time k and its relationship with $\mathbf{x}_k$. Then, upon the additional knowledge of the initial state probability density function (pdf) $p_0(\mathbf{x}_0)$, the aim is to estimate the posterior pdf $p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k})$, which expresses all the information on $\mathbf{x}_k$ captured by $\mathbf{z}_{1:k} = \mathbf{z}_1, \ldots, \mathbf{z}_k$.

When considered as a function of $\mathbf{x}_k$, the observation density $p(\mathbf{z}_k \,|\, \mathbf{x}_k)$ depicts the likelihood of the state vector regarding the measurement. In any appearance-based approach, as is the case for this study, the output equation cannot be expressed in closed-form. The likelihoods of distinct state vector values can just be evaluated separately, e.g. by assessing the projections of the corresponding figure model onto the current camera images. In addition, even if the environment is controlled, the likelihood function has multiple modes, some of which can be very sharp. So, the posterior $p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k})$ is multimodal and cannot be expressed analytically. One must then return to approximate filtering schemes. Among these, particle filtering constitutes a versatile recursive approach[4,10,11] in that it can handle any Markov stochastic system subject to any kind of noise, even if a constructive model of the state-output relationship is not available. Elements of the theory are hereafter recalled.

2.2. *Basics of particle filtering*

2.2.1. *Generalities*

The cornerstone of particle filtering is to represent the posterior probability density function $p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k})$ by the point-mass distribution

$$p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k}) \approx \sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)}), \quad \sum_{i=1}^{N} w_k^{(i)} = 1, \tag{1}$$

which represents the selection of a value — or particle — $\mathbf{x}_k^{(i)}$ with probability — or weight — $w_k^{(i)}$, $i = 1, \ldots, N$. An approximation of the conditional expectation

Table 1. Generic particle filtering algorithm (SIR).

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

1: IF $k = 0$, THEN Draw $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(N)}$ i.i.d. according to $p_0(\mathbf{x}_0)$, and set $w_0^{(i)} = \frac{1}{N}$. END IF
2: IF $k \geq 1$ THEN {— $[\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^{N}$ being a particle description of $p(\mathbf{x}_{k-1} \,|\, \mathbf{z}_{1:k-1})$ —}
3: FOR $i = 1, \ldots, N$, DO Independently sample $\mathbf{x}_k^{(i)} \sim q(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$. Then, update its weight
   by $w_k^{(i)} \propto w_{k-1}^{(i)} \frac{p(\mathbf{z}_k \,|\, \mathbf{x}_k^{(i)}) p(\mathbf{x}_k^{(i)} \,|\, \mathbf{x}_{k-1}^{(i)})}{q(\mathbf{x}_k^{(i)} \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)}$, prior to a normalization s.t. $\sum_i w_k^{(i)} = 1$. END FOR
4: Compute the MMSE estimate $\mathrm{E}_{p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k})}[\mathbf{x}_k]$ from the approximation $\sum_{i=1}^{N} w_k^{(i)} \delta(\mathbf{x}_k - \mathbf{x}_k^{(i)})$
   of the posterior $p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k})$.
5: At any time or occasionally, resample $[\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}]_{i=1}^{N}$ into the equivalent evenly weighted particle
   set $[\{\mathbf{x}_k^{(s^{(i)})}, \frac{1}{N}\}]_{i=1}^{N}$, by selecting in $\{1, \ldots, N\}$ the indexes $s^{(1)}, \ldots, s^{(N)}$ with probability
   $P(s^{(i)} = j) = w_k^{(j)}$. Then, set $\mathbf{x}_k^{(i)}$ and $w_k^{(i)}$ with $\mathbf{x}_k^{(s^{(i)})}$ and $\frac{1}{N}$.
6: END IF
```

of any function f of $\mathbf{x}_k$, e.g. the Minimum Mean Square Error (MMSE) estimate, immediately follows as $\mathrm{E}[\mathbf{x}_k \,|\, \mathbf{z}_{1:k}] f(\mathbf{x}_k \,|\, \mathbf{z}_{1:k}) = \sum_{i=1}^{N} w_k^{(i)} f(\mathbf{x}_k^{(i)})$.

For a system described by $p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})$ and $p(\mathbf{z}_k \,|\, \mathbf{x}_k)$ with initial prior $p_0(\mathbf{x}_0)$, the particle approximation (1) is propagated throughout time according to Table 1. The "Sampling Importance Resampling" (SIR) algorithm reported therein is fairly generic, in that most strategies can fit into this framework,[11] provided an adequate definition of the importance function $q(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ and of the resampling scheme respectively exist in steps 3 and 5.

The importance function $q(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k)$ governs the stochastic evolution of the particles $\mathbf{x}_k^{(i)}$, $i = 1, \ldots, N$, at each time k, and is selected so as to adaptively explore relevant areas of the state space. The mathematical expression of the weights $w_k^{(i)}$ then ensures the consistency of the approximation (1). Besides, inserting a resampling stage limits the degeneracy phenomenon experienced by all sequential Monte Carlo filters, i.e. the collapsing of the weights of all but one particle. Note that this redistribution should be fired only when the filter efficiency goes beneath a predefined threshold.[11]

2.2.2. *Elementary schemes*

The CONDENSATION[17] — for "Conditional Density Propagation" — is the instance of the SIR in which $q(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k) = p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)})$ and $w_k^{(i)} \propto w_{k-1}^{(i)} p(\mathbf{z}_k \,|\, \mathbf{x}_k^{(i)})$. Its performance may be weak, as drawing the particles according to the system dynamics regardless of the current measurement may well lead to assign many of them a low likelihood and thus a low weight. This drawback is highly plausible in the HMC context, because of the sharpness of the likelihood modes. Besides, the I-CONDENSATION[18] samples some particles according to an importance function $\pi(\mathbf{x}_k \,|\, \mathbf{z}_k)$ specified from the current image, e.g.

through $q(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}, \mathbf{z}_k) \;=\; \alpha\pi(\mathbf{x}_k \,|\, \mathbf{z}_k) + \beta p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}^{(i)}) + (1-\alpha-\beta)p_0(\mathbf{x}_k)$, with $(\alpha, \beta) \geq (0, 0)$. Generally, $\pi(\mathbf{x}_k \,|\, \mathbf{z}_k)$ relies on the detection of intermittent primitives which, despite their sporadicity, are very discriminant when present.[28]

2.3. *Enhanced particle filtering strategies*

Human motion capture, be it appearance-based or grounded on sparse reconstruction of 3D information, raises sharp problems which make classical strategies fail unless a very high number of particles is used, thus at the expense of efficiency and real-time performance.[a] First, the state space is of very high dimension, typically from 20 to 40 even for an elementary first-order dynamics, e.g. a random walk. Though some early references claimed that particle filters "beat the curse of dimensionality", i.e. converge at a rate independent of the state dimension, this property does not hold. Daum and Huang[5] raised theoretical issues which underlie this fact, and derived back-of-the-envelope formulae for complexity depending on whether the problem is "vaguely Gaussian" or not. Their illuminating paper shows that for general problems, the computation time of a particle filter is linear in the number of particles, yet this number is truly exponential in the system order so as to ensure a fixed dimension free error.

Among the other challenges are the aforementioned multimodality and sharp peaks of the likelihood, as well as the information loss through projection. So, more involved strategies have been envisaged, aiming at a better exploration of state space areas which are likely w.r.t. the measurements, while keeping the possibility for some less likely hypotheses to strengthen along time if they are consistent enough with the prior dynamics and the measurements. Some of them are described below, together with original proposals.

2.3.1. *Important requirements*

Two fundamental points have been acknowledged in the particle filter based HMC literature in order to catch the likelihood modes. On the one hand, as likely areas of the state space cannot be portrayed *a priori*, iterative local searches must be designed so as to drive the drawn particles towards these peaks. On the other hand, the diffusion of particles from a parent sample must itself be carefully designed. For instance, spreading particles isotropically from a parent sample is inoperating. Instead, the dynamics noise, though physically realistic, should be scaled by the posterior covariance or by a similar indication extracted from the likelihood evaluations, with the aim to lay more samples on the hard-to-estimate directions. This last feature is termed "Covariance based sampling".[34]

[a] Deutscher *et al.*[7] reported failures of CONDENSATION even with a huge number of particles.

2.3.2. *Partitioned sampling*

For filtering problems where the system dynamics comes as the sequence of partial evolutions and where intermediate likelihoods of the state vector can be assessed after applying each partial dynamics, partitioned[21] or hierarchical[28] schemes apply. 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 selected strategy assumes a partition of $\mathbf{x}_k$ into the M subvectors $\{\mathbf{x}_k^1, \ldots, \mathbf{x}_k^M\}$ such that the full dynamics and the likelihood respectively factorize into $p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}) = \prod_{m=1}^M p_m(\mathbf{x}_k^m \,|\, \mathbf{x}_{k-1}^m)$ and $p(\mathbf{z}_k \,|\, \mathbf{x}_k) = \prod_{m=1}^M l_m(\mathbf{z}_k \,|\, \mathbf{x}_k^1, \ldots, \mathbf{x}_k^m)$, where the intermediate likelihoods $l_m(.\,|\,.)$ concern a subset of the state vector all the more important as $m \to M$.[21] Its derivation from the SIR is schematized in Table 2. Its computational complexity is linear in the number of partitions instead of being exponential in the number of degrees of freedom. Note that the number of particles may sometimes be reduced as the partition index grows, thus lowering the cost. Applying this basic partitioned strategy to the HMC context implies that separate parts of the body can be independently localized. According to some authors,[9] geometrical information must be supplemented by color or other labeling cues to ensure a proper tracking.

2.3.3. *Annealed particle filter*

The Annealed Particle Filter — or in short APF — includes a local stochastic optimization stage inspired by simulated annealing.[7] The main idea is to replace step 3 of the SIR by the recursive run of L layers, as outlined in Table 3. Each particle $\mathbf{x}_{k,l-1}^{(i)}$ related to the lth layer, $l = 1, \ldots, L$, is sampled from a "layer dynamics function" $p_l(\mathbf{x}_{k,l} \,|\, \mathbf{x}_{k,l-1})$, then gradually trapped towards the likely areas of the state space thanks to the evaluation of "layer likelihood function" $p_l(\mathbf{z}_k \,|\, \mathbf{x}_{k,l})$. Step 32 proposes a way to gradually sharpen both the state space exploration and the likelihood modes. Notice that in the case when the prior dynamics is a Gaussian random walk $p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{x}_{k-1}, \Delta_k)$ with Δ_k diagonal, sampling from

Table 2. Partitioned particle filtering.

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

30: Replace step 3 of the SIR algorithm on Table 1 by the following, $\tau_0^{(i)}$ being set to $w_{k-1}^{(i)}$.
31: FOR $m = 1, \ldots, M$, DO
32: FOR $i = 1, \ldots, N$, DO Independently sample $\mathbf{x}_k^{m,(i)} \sim p_m(\mathbf{x}_k^m \,|\, \mathbf{x}_{k-1}^{m,(i)})$, and associate
    $(\mathbf{x}_k^{1,(i)}, \ldots, \mathbf{x}_k^{m,(i)})$ the weight $\tau_m^{(i)} \propto \tau_{m-1}^{(i)} l_m(\mathbf{z}_k \,|\, \mathbf{x}_k^{1,(i)}, \ldots, \mathbf{x}_k^{m,(i)})$. END FOR
33: IF $m < M$ THEN Resample $[\{(\mathbf{x}_k^{1,(i)}, \ldots, \mathbf{x}_k^{m,(i)}), \tau_m^{(i)}\}]_{i=1}^N$ s.t. $P(s^{(i)} = j) = \tau_m^{(j)}$; rename
    the obtained evenly weighted particle set as $[\{(\mathbf{x}_k^{1,(i)}, \ldots, \mathbf{x}_k^{m,(i)}), \tau_m^{(i)} = \frac{1}{N}\}]_{i=1}^N$. END IF
34: END FOR — Then, $[\{\mathbf{x}_k^{(i)}, w_k^{(i)} = \tau_M^{(i)}\}]_{i=1}^N$ is a consistent description of $p(\mathbf{x}_k \,|\, \mathbf{z}_{1:k})$.
```

Table 3. Annealed particle filtering.

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

30: Replace step 3 of the SIR algorithm on Table 1 by the following, $[\{\mathbf{x}_{k,0}^{(i)}, w_{k,0}^{(i)}\}]_{i=1}^N$ being set
    to $[\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^N$, with $1 < \alpha_1 < \cdots < \alpha_L$ and $\beta_1 < \cdots < \beta_L < 1$ being selected beforehand.
31: FOR $l = 1, \ldots, L$, DO
32: FOR $i = 1, \ldots, N$, DO Independently sample $\mathbf{x}_{k,l}^{(i)} \sim p_l(\mathbf{x}_{k,l} \,|\, \mathbf{x}_{k,l-1}^{(i)}) = [p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})^{\alpha_l}]_{\mathbf{x}_k=\mathbf{x}_{k,l};\mathbf{x}_{k-1}=\mathbf{x}_{k,l-1}^{(i)}}$.
    Associate it the weight $w_{k,l}^{(i)} \propto p_l(\mathbf{z}_k \,|\, \mathbf{x}_{k,l}^{(i)}) = [p(\mathbf{z}_k \,|\, \mathbf{x}_k)^{\beta_l}]_{\mathbf{x}_k=\mathbf{x}_{k,l}^{(i)}}$. END FOR
33: Normalize $\{w_{k,l}^{(i)}\}$ s.t. $\sum_i w_{k,l}^{(i)} = 1$. IF $l < L$, THEN resample $[\{\mathbf{x}_{k,l}^{(i)}, w_{k,l}^{(i)}\}]_{i=1}^N$. ENDIF
34: END FOR — The particles set $[\{\mathbf{x}_{k,L}^{(i)}, w_{k,L}^{(i)}\}]_{i=1}^N$ has just to be renamed $[\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}]_{i=1}^N$.
```

$p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})^{\alpha_l}$ is equivalent to sampling from $\mathcal{N}(\mathbf{x}_k; \mathbf{x}_{k-1}, \frac{1}{\alpha_l}\Delta_k)$. On a theoretical side, it must be pointed that in its above form, APF is not a mathematically sound Monte Carlo method.[7]

2.4. *Quasi Monte Carlo filtering methods*

In particle filtering, the stochastic nature of the particle locations enables to adaptively explore areas of the state space in which the posterior distribution is dense. Yet, especially in high-dimension spaces, pure random importance sampling leads to "gaps and clusters" in the particle support. An excessive Monte Carlo variation of the predictions can follow, making the filter unreliable or even leading to failures.

An alternative is to substitute the random particles by a low-discrepancy deterministic sequence, enjoying a guaranteed degree of uniformity. For a d-dimensional state space, the error in approximating integrals through such so-called Quasi Monte Carlo methods has been proved to converge at a rate of $\mathcal{O}(N^{-1} \log^d N))$ in the number N of particles, which is much better than $\mathcal{O}(N^{-\frac{1}{2}})$ for standard Monte Carlo.[12] Yet, analyzing the accuracy of deterministic QMC approximations is difficult and prevents the use of statistical procedures. So, randomized QMC methods have been developed in order to define low-discrepancy sequences whose elements taken individually have a given distribution. These enable the design of unbiased estimators with reduced variance, whose approximation error can surprisingly converge as $\mathcal{O}(N^{-\frac{3}{2}} \log^{\frac{d-1}{2}} N)$ for sufficiently smooth integrands.

In the context of visual tracking, the performances of a deterministic and of a randomized QMC particle filter against a particle filter based on standard random sampling have been assessed only in two references, respectively.[27,29] Therein, a high number of runs on low-order synthetic experiments, with filters entailing the same number N of particles, proves that for deterministic or randomized QMC filters, (1) the RMS estimation error w.r.t. ground truth is always smaller, (2) its standard deviation over the whole temporal sequence is also lower, (3) when N

increases, the RMS estimation error drops all the faster as N is low and/or there is much noise, (4) for a given tolerance to errors, QMC needs between half and third the number of particles, which speeds its execution. Visual tracking experiments on real sequences, where the state space dimension is of the order of 10, show that the deterministic QMC filter[29] performs well on the whole time history and can recover from failures due to sudden changes or partial occlusions. As for the randomized QMC filter,[27] it is proved unbiased w.r.t. the true posterior mean and enjoys a variability on each entry of the estimated state vector from 5% to 20% below that of Monte Carlo particle filter. The authors claim a savings in the number of particles between 20% and 60%, which is more pronounced when few particles are used, together with a significant reduction of the ten-dimensional volume to be explored at each step.

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. The quasi-random counterpart of CONDENSATION,[14] hereafter named QRS-CONDENSATION, is presented in Table 4, where M terms the dimension of $\mathbf{x}$. It is considered in this paper, together with its straight partitioned version, henceforth termed PARTITIONED QRS.

3. Problem Formulation

3.1. *Generalities*

Recall that our appearance-based approach infers the human body model from its projection in trinocular image sequences. The whole human body model used for

Table 4. Quasi-random sampling CONDENSATION (QRS-CONDENSATION).

```
$[\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}]_{i=1}^{N} = QRS - CONDENSATION([\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)},\}]_{i=1}^{N}, \mathbf{z}_k)$

IF $k = 0$, THEN Generate a random QMC Sobol sequence $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(N)}$ in $[0,1)^M$ and convert it into $\mathbf{x}_0^{(1)}, \ldots, \mathbf{x}_0^{(N)} \sim p_0(\mathbf{x}_0)$. Set $w_0^{(1)} = \cdots = w_0^{(N)} = \frac{1}{N}$. END IF
IF $k \geq 1$ THEN {— $[\{\mathbf{x}_{k-1}^{(i)}, w_{k-1}^{(i)}\}]_{i=1}^{N}$ being a QMC weighted description of $p(\mathbf{x}_{k-1} \mid \mathbf{z}_{1:k-1})$ —}
  FOR $i = 1, \ldots, N$, DO
    Select $s^{(1)}, \ldots, s^{(N)}$ in $\{1, \ldots, N\}$ s.t. $P(s^{(i)} = j) = w_k^{(j)}$. Set $C_j$ = Cardinality$\{i : s^{(i)} = j\}$.
    FOR $j = 1, \ldots, N$, DO Generate a random QMC Sobol sequence $\mathbf{u}^{(1)}, \ldots, \mathbf{u}^{(C_j)}$ in $[0,1)^M$ and convert it into $\mathbf{x}_k^{(C_{j-1}+1)}, \ldots, \mathbf{x}_k^{(C_{j-1}+C_j)} \sim p(\mathbf{x}_k \mid \mathbf{x}_{k-1}^{j})$. END FOR
    Set $w_k^{(i)} \propto p(\mathbf{z}_k \mid \mathbf{x}_k^{(i)})$, prior to a normalization s.t. $\sum_i w_k^{(i)} = 1$.
  END FOR — $[\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}]_{i=1}^{N}$ is then a QMC weighted description of $p(\mathbf{x}_k \mid \mathbf{z}_{1:k})$ —
  Compute the MMSE estimate $\mathrm{E}_{p(\mathbf{x}_k \mid \mathbf{z}_{1:k})}[\mathbf{x}_k]$ from $[\{\mathbf{x}_k^{(i)}, w_k^{(i)}\}]_{i=1}^{N}$.
END IF
```

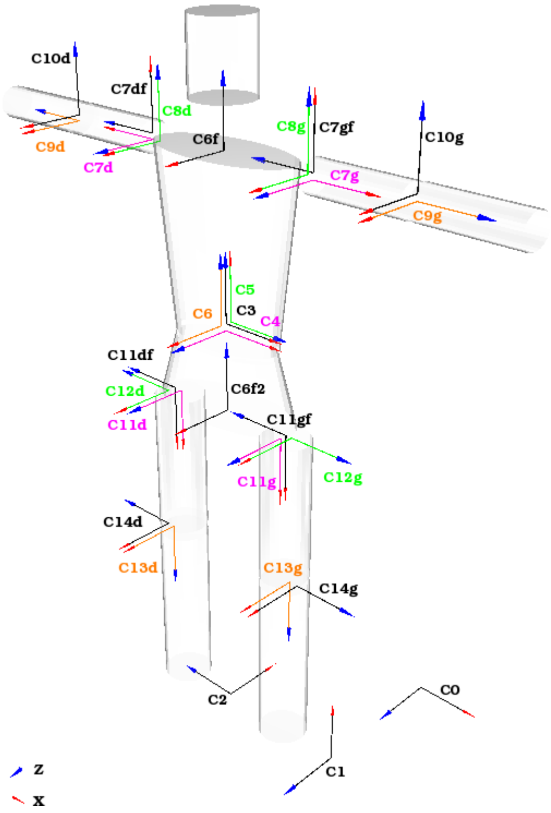


Fig. 1. 3D model and associated DOFs.

HMC is presented in Fig. 1. Each limb is fleshed out using truncated cones with fixed dimensions. These geometric primitives are easily handled, for their image projections and hidden parts removal can be obtained in closed form.[1] The model is also based on a kinematic tree consisting of nine body segments. Six degrees of freedom (DOF) are used for global position (t_x, t_y, t_z) and orientation (r_x, r_y, r_z). The shoulder and the thigh are treated as ball joints with three DOFs and the remaining joints are modeled as hinges requiring only one DOF.

All these 22 parameters are accounted for in the state vector $\mathbf{x}_k$ related to the three frames at instant k. With regard to the dynamics model $p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})$, the human limb motions are difficult to characterize over time. This weak knowledge is formalized by defining the state vector as $x_k = [t_x, t_y, t_z, r_x, r_y, r_z, dof_1, \ldots, dof_{16}]'$ and assuming that its entries evolve according to mutually independent Gaussian random walk models, viz. $p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1}) = \mathcal{N}(\mathbf{x}_k; \mathbf{x}_{k-1}, \Sigma)$, where $\mathcal{N}(.; \mu, \Sigma)$ is a Gaussian distribution with mean μ and diagonal covariance Σ.

Each tracker implementation relates to the modular architecture depicted in Fig. 2. The filtering module implements the PF schemes detailed in Sec. 2 while

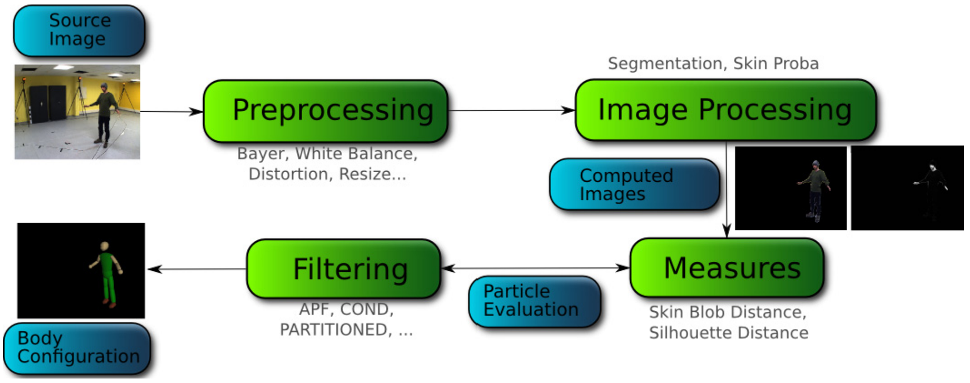


Fig. 2. Synopsis of our tracking modular architecture.

the measurement module is in charge of particle likelihood evaluations described here below.

3.2. *Observation likelihoods*

It can be argued[1,9,30,35] that appearance-based cues offer indeed a principled way to derive a plethora of measurements. We have constructed the weighting function on the basis of two visual features, namely foreground-background and skin blob segmentation. Mixing these cues offers a nice trade-off in terms of generality, simplicity, and complementarity.

3.2.1. *Silhouette-based likelihood*

In the vein of Ref. 7, the first feature extraction performed is foreground–background subtraction. A binary image $I_{l,s}$ is constructed for each camera l, with foreground pixels set to 1 and background set to 0 (Fig. 3(b)). Given the 3D model projection for the configuration $\mathbf{x}_k$, N_p pixels $\mathbf{p}_{l,i}, i \in \{1, \ldots, N_p\}$ are uniformly sampled in the interior of the projected truncated cones. The silhouette-based likelihood follows:

$$p(\mathbf{z}_k^{l,s_1} \,|\, \mathbf{x}_k) \propto \exp\left(-\frac{D_l^2}{2\sigma_{s_1}^2}\right), \quad D_l = \frac{1}{N_p}\sum_{i=1}^{N_p}(1 - I_{l,s}(\mathbf{p}_{l,i})), \tag{2}$$

where i indexes the N_p model points, $I_{l,s}(\mathbf{p}_{l,i})$ is the associated pixel value (0 or 1) in image $I_{l,s}$ for camera l, and σ_{s_1} is a standard deviation being determined *a priori*. Figure 4(b) shows the plot of the measurement function obtained by varying only one DOF (the right shoulder angle). The function is very sharp around the correct angle. This likelihood reaches its highest value if the projected model is inside the silhouette without demanding that the silhouette area is fully explained. Consequently, some uninformative model configurations situated inside the silhouette may peak this likelihood. This situation is alleviated by the use of the dual cue.

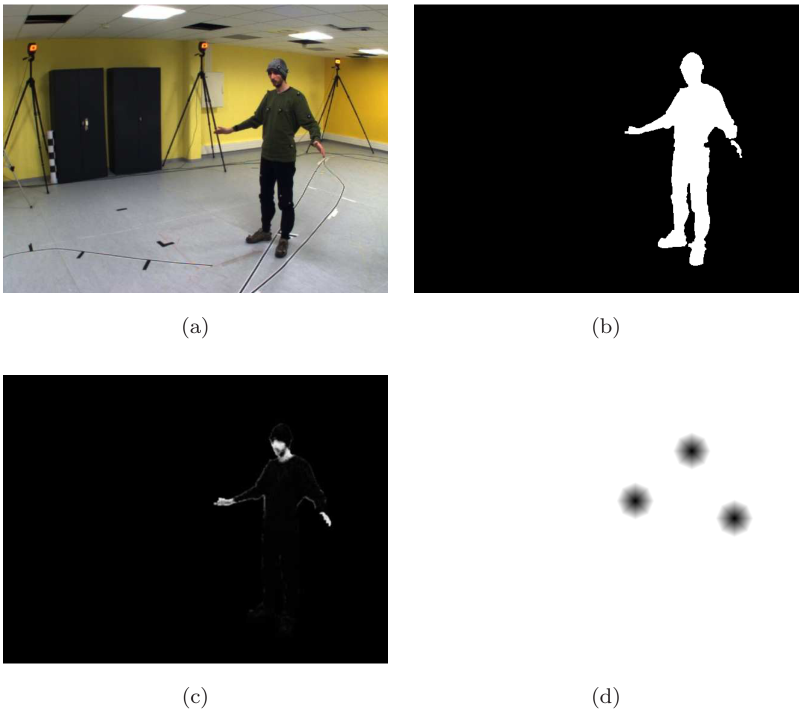

(a)
(b)
(c)
(d)

Fig. 3. (a) Feature extraction for the input image, (b) foreground segmentation, (c) skin blob segmentation, (d) skin blob distance image.

3.2.2. *Dual silhouette-based likelihood*

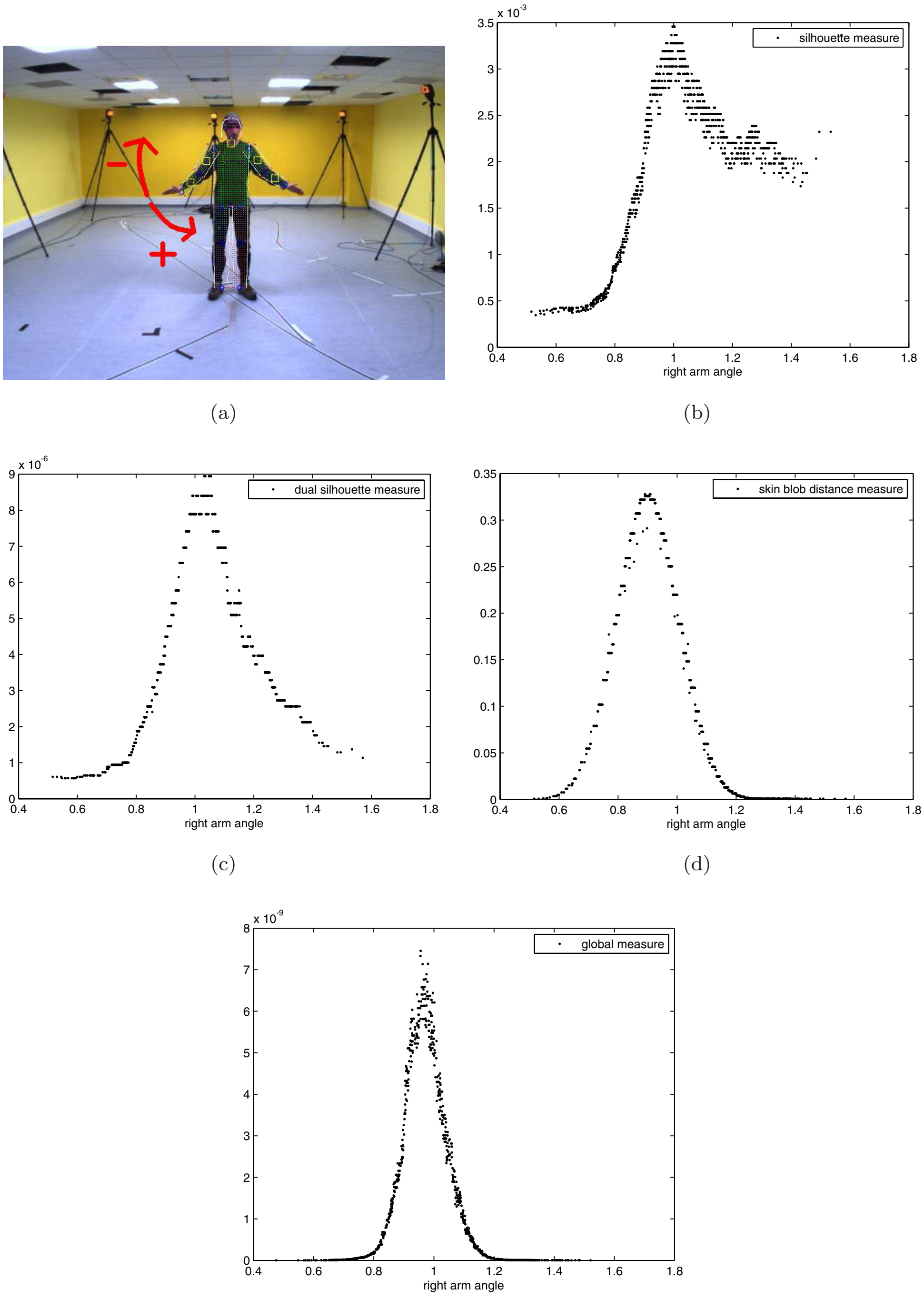
The principle is here to sample N_s points $\mathbf{p}_j$ from the segmented image $I_{l,s}$. The likelihood $p(\mathbf{z}^{l,s_2} \,|\, \mathbf{x}_k)$ has a form similar to (2) provided that σ_{s_2} corresponds to the standard deviation. The similarity distance D_l follows

$$D_l = \frac{1}{N_s} \sum_{j=1}^{N_s} (1 - f(\mathbf{p}_{l,j}, \mathbf{x}_k)),$$

where $f(\mathbf{p}_{l,j}, \mathbf{x}_k) = 1$ if the point $\mathbf{p}_{l,j}$ is in the silhouette corresponding to the model projection of $\mathbf{x}_k$ for camera l, 0 otherwise. Figure 4(c) shows the plot of this likelihood function for the same example.

3.2.3. *Skin blob-based likelihood*

To achieve more precision in the localization — especially for thin limbs such as arms — we set up an additional likelihood function involving skin blob detection. The image processing module computes a skin probability image thanks to an off-line learnt skin color histogram back-projection. Skin blobs are extracted (Fig. 3(c)) and a skin color blob distance image $I_{l,\text{skin\_dist}}$ (Fig. 3(d)) is computed.



(a)

(b)

(c)

(d)

(e)

Fig. 4. Example of a single DOF variation using the image and model configuration in (a) and likelihood functions (b)–(e) : $p(\mathbf{z}_k^{l,s_1} \,|\, \mathbf{x}_k)$, $p(\mathbf{z}_k^{l,s_2} \,|\, \mathbf{x}_k)$, $p(\mathbf{z}_k^{l,\text{skin}} \,|\, \mathbf{x}_k)$, $p(\mathbf{z}_k^{s_1,s_2,\text{skin}} \,|\, \mathbf{x}_k)$.

For each hypothesis $\mathbf{x}_k$, we compute the mean distance in pixels between the image projection of three virtual markers $\mathbf{p}_1, \mathbf{p}_2, \mathbf{p}_3$ — respectively corresponding to the head and the hands — and the nearest detected skin blobs, which results in the following similarity criterion:

$$D_l = \frac{1}{3}\sum_{i=1}^{3} I_{\text{skin\_dist}}(\mathbf{p}_i).$$

assuming σ_{skin} corresponds to *a priori* standard deviation in (2).

3.2.4. *The composite likelihood*

All these measurements are combined in a similar way whatever the time index k so as to account for the contents of all images. Assuming they are mutually independent, conditioned on the state $\mathbf{x}_k$, the composite likelihood factorizes as follows, given $C = 3$ is the number of cameras:

$$p(\mathbf{z}_k^{s_1,s_2,\text{skin}} \,|\, \mathbf{x}_k) = \prod_{l=1}^{C} p(\mathbf{z}_k^{l,s_1} \,|\, \mathbf{x}_k) \cdot p(\mathbf{z}_k^{l,s_2} \,|\, \mathbf{x}_k) \cdot p(\mathbf{z}_k^{l,\text{skin}} \,|\, \mathbf{x}_k).$$

This is illustrated in Fig. 4(e).

3.3. *Particle filter tuning*

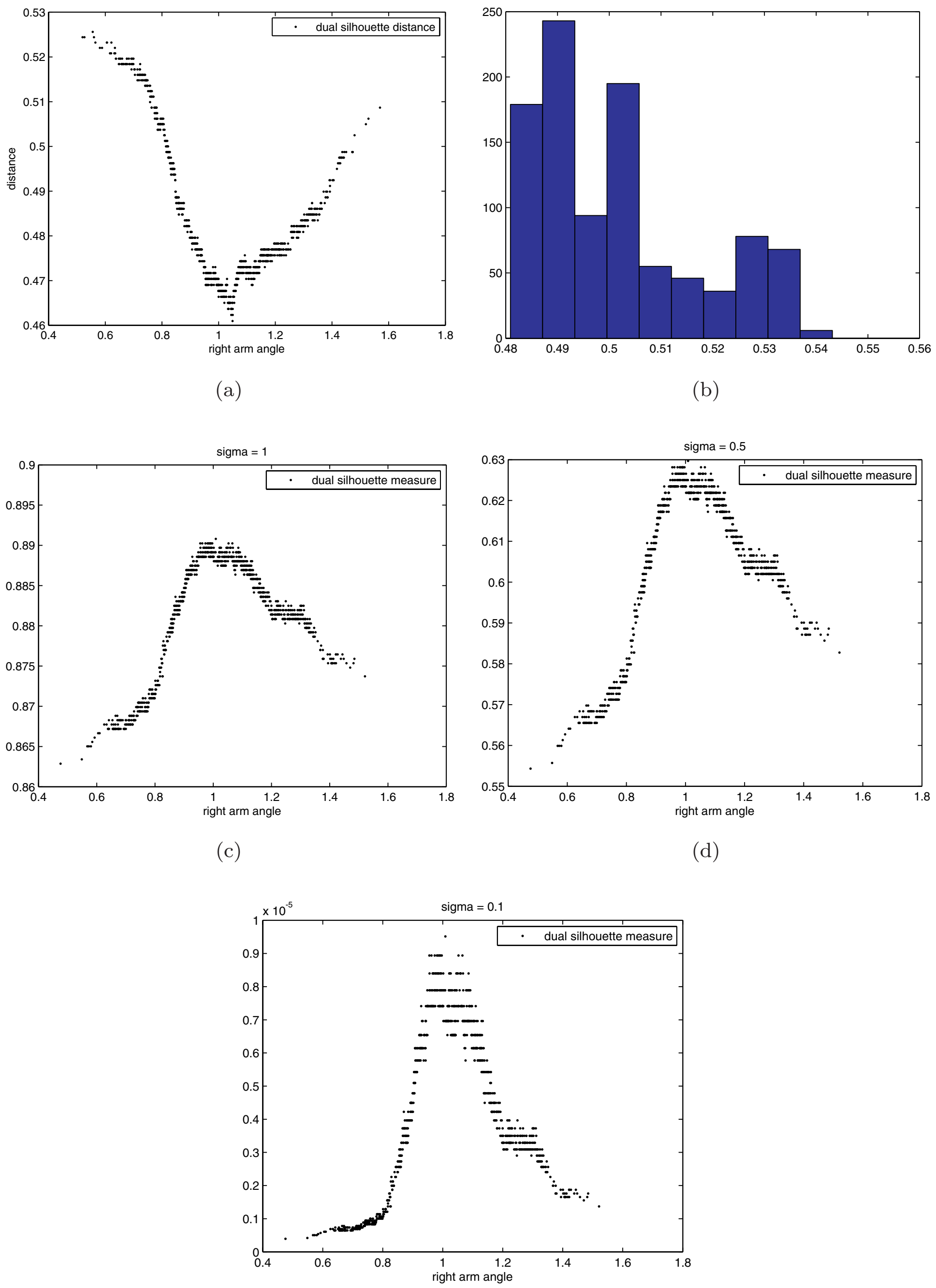
We hereafter discuss the tuning of the free parameters involved in the PF strategies, i.e. in the initial prior, the dynamics and the likelihood models, as well as the number of particles and of layers for layered strategies. Those are used in both synthetic and real sequences, what demonstrate the tracker's ability to deal with distinct data sources and to capture a wide range of human motions.

3.3.1. *Likelihood parametrization*

All parameters σ involved in the likelihoods were first tuned through simple heuristics, then sharpened from experimental data to achieve a stable configuration. Selecting the σ parameters is a very complex task, never thoroughly tackled in the literature so far.[20] If our measurement models were perfect, the "ideal" similarity distance when an hypothesis is located exactly on the ground truth should be zero. In practice, this is never the case. As a consequence, we have to choose well-balanced values for the σ parameters. We use the following heuristics to guide us in this choice.

3.3.1.1. Similarity distance

The evolution of this distance w.r.t. the state vector entries can constitute a good guideline to the tuning. Figures 5(a) and 5(b) present the similarity distance values for the experiment described in Fig. 4(c). As can be seen, values spread between 0.48 and 0.54. The early tuning of σ must be situated within this range.



(a)

(b)

(c)

(d)

(e)

Fig. 5. (a) The similarity distance of the "dual silhouette" cue described in Sec. 3.2 and (b) the associated histogram. (c)–(e) present the likelihood functions respectively for $\sigma = 1.0$, $\sigma = 0.5$ and $\sigma = 0.1$

3.3.1.2. Lowering the values

We can sharpen the initial choice by lowering it. Figures 5(c)–5(e) show the likelihood function with different values of sigma ($1.0, 0.5$ and 0.1). As sigma decreases, the likelihood function shows a higher peak, which results in a more accurate estimate and a more drastic selection of the fittest particles. However, values of the computed weights globally decrease too, which can lead to the following problem.

3.3.1.3. Computational limits

The value of σ cannot be decreased at will: indeed, too low a value results in a null likelihood function due to computer encoding limits (e.g. $\exp(\frac{1}{2}\frac{0.5^2}{0.01^2}) = 0$, so that selecting $\sigma = 0.01$ results in null weights and leads to the filter failure).

3.3.1.4. Well balanced multicues

This last mentioned phenomenon is amplified when we use multiple cues and/or multiple cameras as likelihood functions are multiplied. Consequently, one has to be very cautious at decreasing σ values.

From the above considerations, manual fine tuning on key-sequences has led to the empirical values

$$\sigma_{s_1} = 0.1, \quad \sigma_{s_2} = 0.07, \quad \sigma_{\text{skin}} = 5$$

though with no guarantee that they are optimum. The automatic tuning of these likelihood free parameters is undoubtedly an open question in the vision literature, all the more because it strongly affects the tracker performances. σ_{s_1} and σ_{s_2} values are quite the same as they represent a superimposition ratio between the segmented silhouette and the projected one. σ_{skin} seems higher but is expressed in pixels as D_{skin} usually lies between 0 and ~ 50 pixels. Finally, σ parameters are well-balanced.

3.3.2. *Filter strategies*

In this study, the SIR, PARTITIONED, APF, QRS and PARTITIONED QRS strategies are evaluated. All filters are initialized "by hand". Concerning the Gaussian random walk dynamics $p(\mathbf{x}_k \,|\, \mathbf{x}_{k-1})$ mentioned in Sec. 3.1, Table 5 gathers the selected standard deviations of the dynamics noise when the acquisition rate is 5 Hz.

The APF filter is run with three layers as more layers induce too few particles in each layer to allow an efficient tracking. α_l and $\beta_l, l \in \{1, \ldots, 3\}$, parameters are chosen according to the guide rules presented in Ref. 7 and sharpened by experiments resulting in $\alpha_1 = 1, \alpha_2 = 4, \alpha_3 = 4$ and $\beta_1 = 0.1, \beta_2 = 0.4, \beta_3 = 1$.

The PARTITIONED filter is used with two partitions. The first one is focused as usual on the six parameters needed to localize the torso, while the others consist of

Table 5. Standard deviation of the random walk dynamics applied on each DOF of the human body configuration.

Base translations	0.07 m
Base rotations	0.05 rad
Shoulder ball joints	0.1 rad
Elbow hinge joints	0.3 rad
Hip ball joints	0.1 rad
Knee hinge joints	0.1 rad

the limbs extremities.[9] This is an intuitive partitioning which seems fairly efficient according to the different tests we ran. For a more complete study, such parameters should be tested quantitatively to achieve the best configuration of each filtering scheme.

In addition, based on Ref. 40, each experiment in this study has been conducted by keeping constant the number of likelihood evaluations, which is the most time consuming part of the algorithms: when SIR trackers are run with N particles, PARTITIONED trackers are run with $N/2$ particles per partition, and APF use $N/3$ particles per layer. Actually our trackers perform more or less at 1 Hz for 500 likelihood evaluations on each step. This is not yet real-time, but performances can still be improved by optimizing the C++ code or by reducing the size of the processed images. Due to this strong temporal constraint, we limit our evaluations to a number $100 < N < 1500$.

3.3.3. *Metrics design for filter evaluation*

To evaluate the MMSE estimates $\widehat{\mathbf{x}_k}$ delivered by the filter with respect to ground truth $\mathbf{t}_k$, we set up the following metrics:

- Error w.r.t. true joint positions: The accuracy of a pose estimate is measured thanks to the average RMS error between the true positions of the J joints $\mathbf{m}_j^{\mathbf{t}_k}, j \in 1, \ldots, J$ and their estimates $\mathbf{m}_j^{\widehat{\mathbf{x}_{k,r}}}$:

$$\frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{R}\sum_{r=1}^{R}\|\mathbf{m}_j^{\mathbf{t}_k}-\mathbf{m}_j^{\widehat{\mathbf{x}_{k,r}}}\|_2^2}, \tag{3}$$

  where $r = 1, \ldots, R$ indexes each trial on a given single sequence. The RMS error is computed on all frames and all runs of the filter. This criterion seems to become a standard in Computer Vision literature.[3,40,41] One can note that error involves joint positions but no computation is done on the state vector itself. Indeed, comparing joint angles is a complex task and distinct configuration vectors can lead to the same body pose. This criterion will henceforth be denoted as RMSE (Root Mean Square Error).

- Bias: We have to check if multiple runs of the filters provide an estimate which is centered on the ground truth. Thus, we design the following criterion:

$$\frac{1}{J}\sum_{j=1}^{J}\left\|\frac{1}{K}\sum_{k=1}^{K}\left(\mathbf{m}_j^{\mathbf{t}_k}-\frac{1}{R}\sum_{r=1}^{R}\mathbf{m}_j^{\widehat{\mathbf{x}_{k,r}}}\right)\right\|_2, \tag{4}$$

  which is the average bias of the estimate with respect to the ground truth.
- Dispersion of the estimator: The variance of the estimated configuration must be analyzed over the filter runs. This is a key point as an average estimate close to ground truth does not necessarily imply a good quality of the estimate on a single run due to the stochastic nature of the tracker. This is problematic as any markerless HMC system must deliver stable outputs for a given video stream. The variance of the estimate is computed as follows:

$$\frac{1}{J}\sum_{j=1}^{J}\sqrt{\frac{1}{K}\sum_{k=1}^{K}\frac{1}{R}\sum_{r=1}^{R}\left\|\mathbf{m}_j^{\widehat{\mathbf{x}_{k,r}}}-\frac{1}{R}\sum_{r=1}^{R}\mathbf{m}_j^{\widehat{\mathbf{x}_{k,r}}}\right\|_2^2}. \tag{5}$$

- Tracking failure rate: To complete our evaluation, we set up a last criterion which focuses on the rate of failures. We consider the tracker fails each time one joint has a distance to the ground truth greater than T_{Failure} (in practice, we chose $0.2 < T_{\text{Failure}} < 0.3$ depending on the mean error). The number of tracking failures is computed at the following rate:

$$\frac{1}{R}\frac{1}{J}\frac{1}{K}\sum_{r=1}^{R}\sum_{j=1}^{J}\sum_{k=1}^{K}\text{fails}(\mathbf{m}_j^{\widehat{\mathbf{x}_{k,r}}}), \tag{6}$$

  where $\text{fails}(\mathbf{m}_j^{\widehat{\mathbf{x}_{k,(r)}}}) = 1$ if $\|\mathbf{m}_j^{\mathbf{t}_k} - \mathbf{m}_j^{\widehat{\mathbf{x}_{k,(r)}}}\|^2 > T^2_{\text{Failure}}$, 0 otherwise.

Several factors and assumptions affect the above metric estimates: 3D model shape/kinematic, dynamics model, measures, etc. We choose to assess all strategies gradually on real but also synthetic sequences. Synthetic data allow to control the image acquisition process. Thus, the filtering strategies can be characterized separately as the underlying system models and assumptions attempt to better fit the video contents.

4. Experiments on Synthetic Sequences

4.1. *Experimental design*

Synthetic sequences have been generated under an OpenGL environment. Three virtual cameras in a triangular configuration take images of configurations of the 3D human model in front of a uniform background (Fig. 6). Test videos including natural human motions like arm waving, jumping and walking are produced. Below, we present only the results on the "arm waving" sequence (Fig. 7) which supports the major results. Our synthesis sequences use a human model which perfectly fits our template. 80 runs have been performed on each sequence.

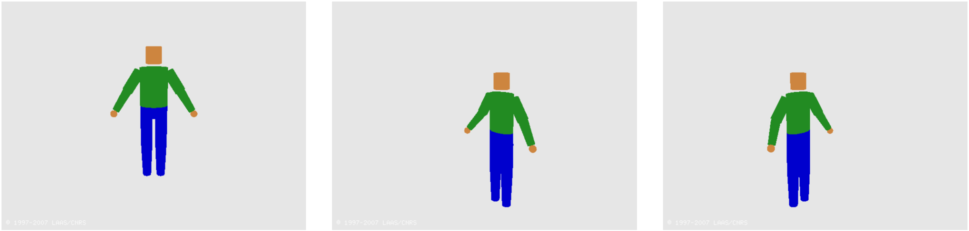


Fig. 6. Images from the three virtual cameras.

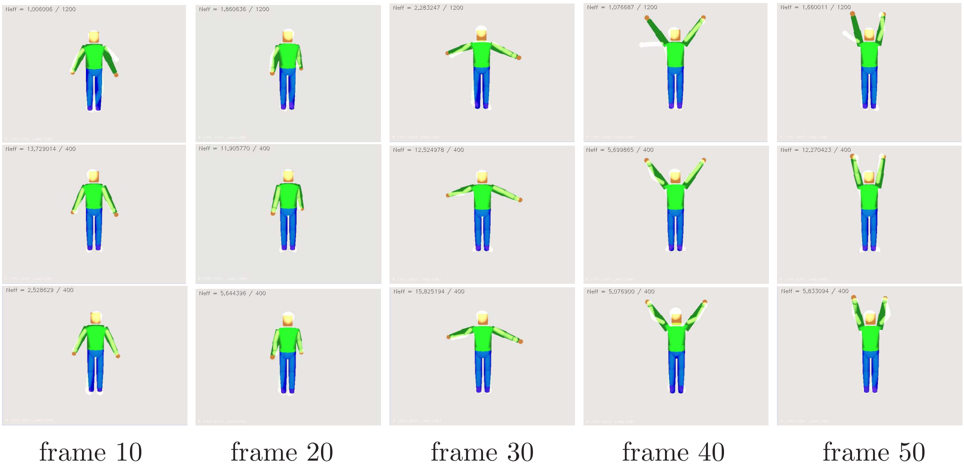


Fig. 7. Synthesis video and filter runs — Every 10th frame from one camera: CONDENSATION (top), PARTITIONED QRS (middle), APF (bottom). The 3D avatar represents the MMSE estimate.

4.2. *Experimental results*

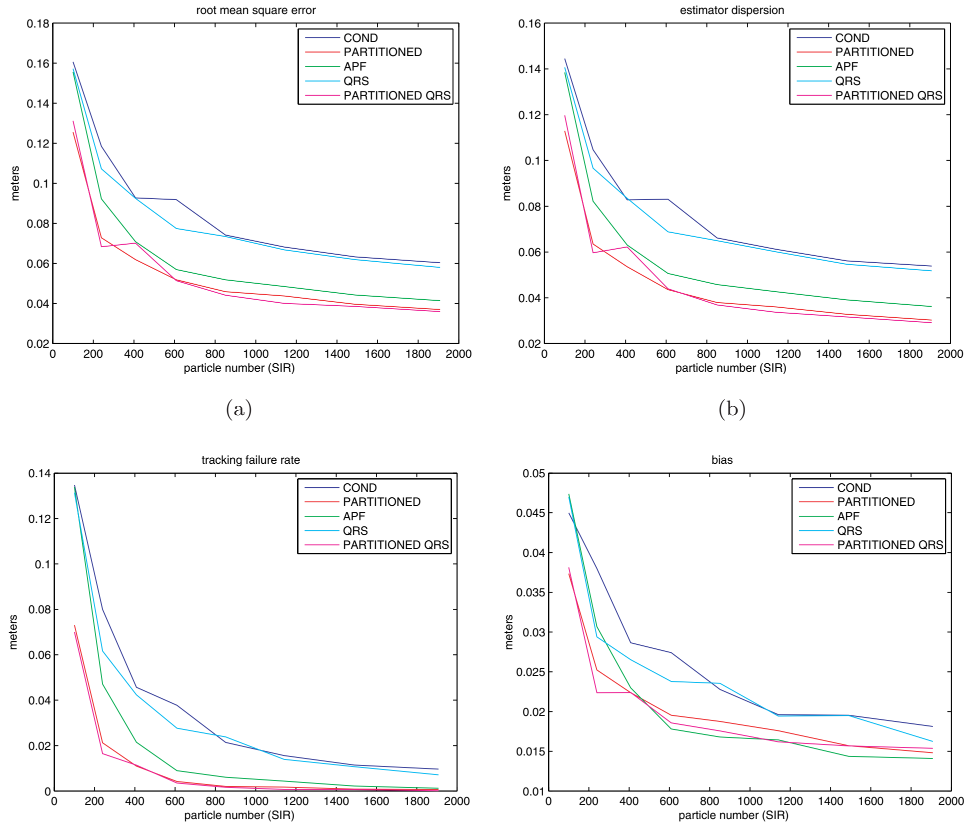
4.2.1. *Root mean square error*

This criterion measures the tracker accuracy relative to the ground truth. Figure 8(a) shows the average Euclidean error (Eq. (3)) between the MMSE estimate of the tracked joint positions and the ground truth values (RMSE).

According to Ref. 7, APF is statistically superior to any other strategy as soon as a minimum number of particles is used. Under this minimal value, the APF filter cannot perform better than the classical SIR.

PARTITIONED is efficient too in terms of error to the ground truth. This is quite logical as it divides the search space into smaller partitions for which the number of particles is enough to improve accuracy.

QMC approaches seem better anyway than their MC counterparts, even if accuracy improvement may be very slight. However, for a low number of particles, differences are more significant, confirming the observations made in Ref. 27. This is



(a)

(b)

(c)

(d)

Fig. 8. Quantitative evaluation of filter outcomes on synthesis sequences. (a) RMSE, (b) estimator dispersion, (c) tracking failure rate, (d) bias.

a consequence of the low-discrepancy of random QMC sequences. Search space is explored in a more uniform way, and thus, estimates are better. One should notice however that we use a QMC algorithm adapted to have an $\mathcal{O}(N)$ complexity though at the expense of mathematical correctness.

Figure 7 confirms these results. We can see that the CONDENSATION provides a poor tracking of arms while PARTITIONED QRS shows good results. APF estimates are fairly satisfactory. Nevertheless, APF results seem worse than those of PARTITIONED QRS, for example; but we have to recall here that these are images from one single run of the algorithms. This leads us to the next cue studied: the dispersion of the estimator.

4.2.2. *Variance of the estimator*

Given Eq. (5) and Fig. 8(b), one of the main conclusions of these preliminary experiments is that QMC methods provide more stable outputs than their MC counterparts. PARTITIONED strategies also show a lower estimate variability than other methods. Again, we posit that this is due to their splitting of the state space. Thus, their efficient number of particles, computed following Ref. 11 and visible in images from Fig. 7, is higher, resulting in a more "stable" estimate. The APF presents more stable results than the CONDENSATION, but still less than the PARTITIONED family for a reasonable number of particles.

Hence, while choosing a filter strategy, we must take into account the dispersion criterion. The APF and PARTITIONED QRS schemes constitute a good alternative with respect to both previous criteria.

4.2.3. *Failure rate*

To evaluate the robustness of each filter, Eq. (6) counts the number of times the error to ground truth exceeds a given threshold T_{failure} all over the frames processed by R runs of the filters. Of course, the failure number depends on the selected threshold, but the relative robustness of the strategies is independent of this value. The results presented in Fig. 8(c) are slightly linked to the two previous ones. This can be understood as the rate of "target loss" of the filters.

The PARTITIONED QRS is confirmed to present good performances. APF strategy also presents a low failure ratio; these two strategies definitively constitute a good choice with respect to our criteria. QRS strategies fare better in terms of failures.

4.2.4. *Bias*

Figure 8(d) shows the average bias of each filter, computed according to formula (4). It tends towards 0 as the number of particles grows, showing that estimates are globally centered on the ground truth. Thus, our measurement cues seem discriminant enough to enable a satisfying behavior of the filters.

We can see that APF has a lower bias than the others. This can be the consequence of its trend to model only the main mode of the posterior distribution.[3]

Table 6 sums up our evaluations on all sequences when considering unambiguous data. The commonly used APF, even if only few quantitative studies can be found in the literature, works well but surprisingly PARTITIONED QRS provides similar or better performances. However, those performances have been derived from a

Table 6. Sorting of the different strategies according to each criterion.

Name	Accuracy	Dispersion	Failure Rate	Bias
CONDENSATION	5	5	5	5
QRS	4	4	4	4
PARTITIONED	2	2	2	3
PARTITIONED QRS	1	1	1	2
APF	3	3	3	1

very friendly context provided by synthesis images. Real sequences provide a more complex problem to tackle as our models are not always fully adapted to the tracked target. The above results must be shaded by real data evaluation.

5. Experiments on Real Sequences

In this section, we develop two main evaluations. First, we present a quantitative assessment on real sequences with a ground truth provided by a commercial HMC system in order to enlighten the best filtering strategy with regard to our criteria. In the second part, we qualitatively study its behavior on various subjects in a cluttered real environment.

5.1. *Experiment design*

Some web databases of human movement video sequence already exist, such as Ref. 41, proposing four different subjects performing different standard movements. However, as we exploit skin color detection as a discriminant natural feature, subjects must wear long sleeve clothes in order to accurately localize hands, which is seldom the case in such databases. Beyond this practical consideration, we are interested in near real-time application, i.e. with video acquisition frequency far below 100 Hz. In addition, we with to have all control over data to free ourselves from generally heavy convention and model adaptation between third-party database and our softwares. This is the reason why we have set up our own sequence database.

The commercial HMC system[42] which provides us the ground truth is constituted of ten infra-red cameras acquiring data at 100 Hz. The subject wears special markers reflecting infra-red light, and the system localizes their 2D position on each image. 3D coordinates of each marker are then triangulated. Finally, a skeleton matching algorithm processes the data to retrieve the bone configurations and joint positions.

The systems are pre-calibrated using a 3D reference object where a few markers are fixed in order to estimate the extrinsic parameters between our video-based HMC system and the commercial HMC system. Thanks to this offline calibration and online synchronized acquisition between our system and the commercial one, ground truth spatio-temporal data is available for each analyzed video issued from the surveillance system.

Figure 9(a) shows the projection of model joints from the commercial HMC system onto an image from our video system. Our three Firewire cameras are placed in a triangle configuration in front of the subject (Fig. 9(b)). They provide 640×480 color images at 5 Hz.

5.2. *Experimental results*

Tracking has been performed on four different sequences from 50 to 100 images (simple arm movement, walking, fitness and a mix of walking and arm movement).

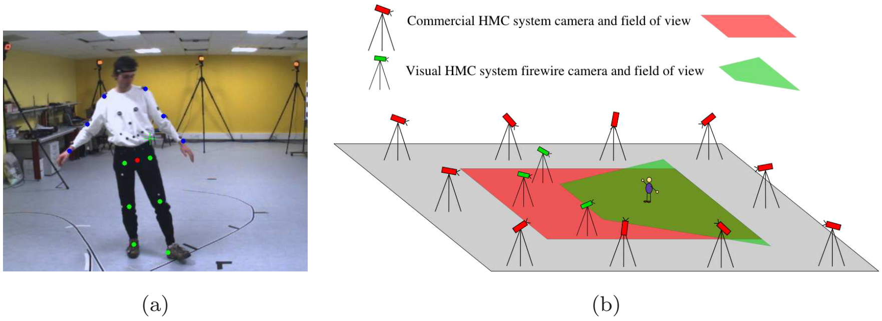

Commercial HMC system camera and field of view
Visual HMC system firewire camera and field of view
(a)
(b)

Fig. 9. (a) Projection of the HMC ground truth configuration blends into an image from our video camera system. The red point is the root of the model, green points are leg joints and blue points are arm joints. (b) Configuration of the two HMC systems.

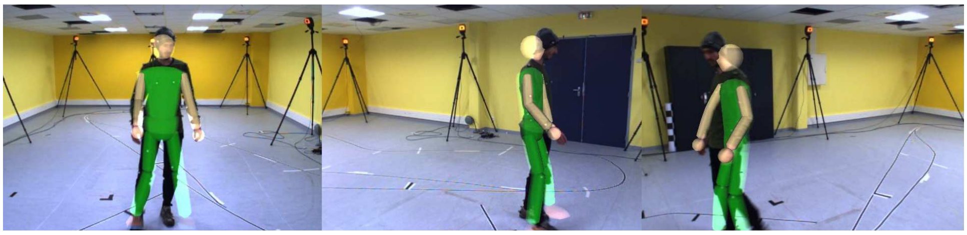


Fig. 10. Images from the three cameras with avatar reprojection taken during the walking sequence, which presents partial occlusions of body members.

Each tracker has been run between 50 and 80 times depending on the length of sequences. Figure 11 proposes some screenshots of the different sequences on which is run an APF strategy with 400 particles and three layers. Due to space reasons, we only show the frames from one camera, which can be somehow misleading about filter accuracy and movement complexity (cf. Fig. 10). For a more complete overview, the entire videos can be found at the URL: www.laas/∼mfontmar. The next subsections present the results with regard to the different metrics we set up, but also additional considerations derived from our evaluations. Due to lack of space, we only present the quantitative results obtained for two sequences which are representative of the global behavior.

5.2.1. *Accuracy*

Figures 12 and 13(a) present the average global RMSE of the estimated joint positions with respect to the ground truth. First of all, we must notice that all trackers are globally efficient. Errors generally lie below 12 cm, which is reasonable considering the modeling approximation and data acquisition. Computer vision literature classically presents errors below 10 cm.[16] This can be explained by our rough model of the human body and our simple measures. Moreover, to achieve better performance, the number of particles needs to grow exponentially. Our real-time

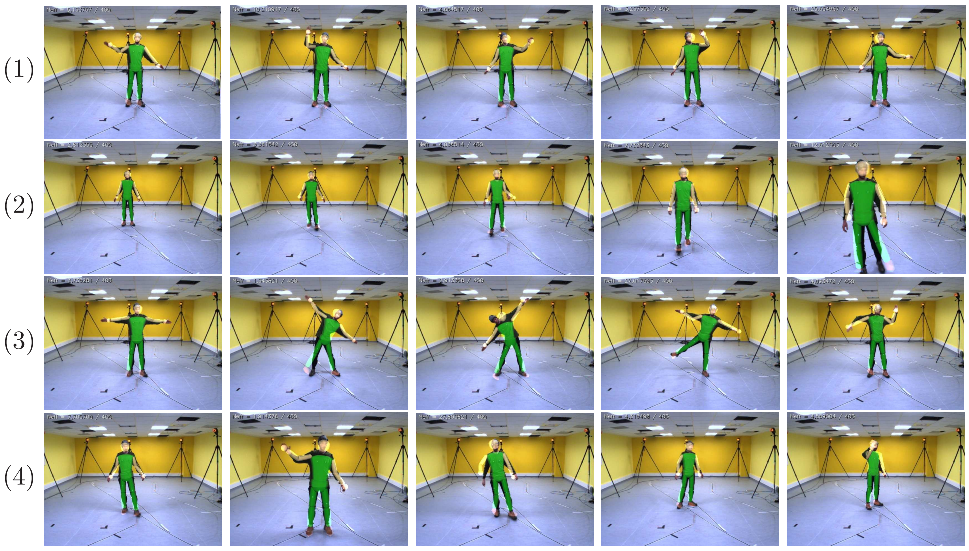


Fig. 11. Run of the APF strategy on four different sequences (one on each line); from top to bottom: simple arm movement (1), walking (2), fitness (3), mix of walking and arm movement (4). Screenshots are taken from the central camera.

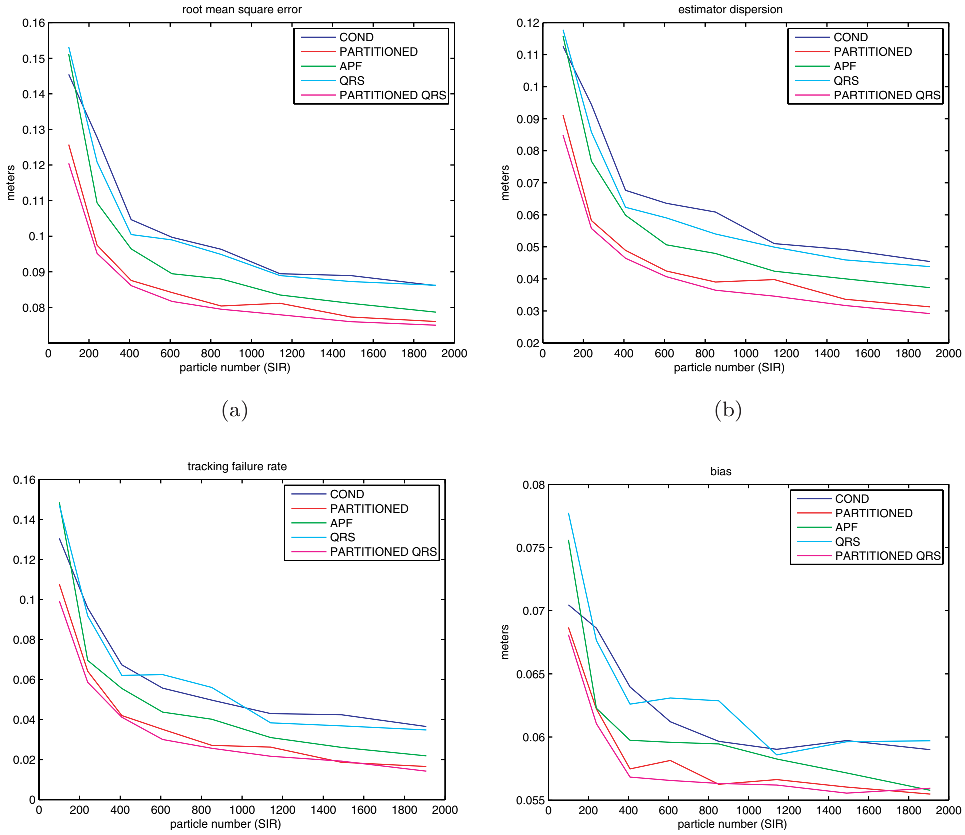
constraints force ourselves to limit the number of samples used, thus limiting accuracy. However, we are interested here in the relative behavior of the filters. We notice that advanced strategies perform better than the classical CONDENSATION, which seems consistent with the results obtained on synthetic data. APF and PARTITIONED relative performances may vary according to the sequence. In sequence (2) (walking), PARTITIONED strategies may be confused while tracking the torso as arms move along the upper body.

Generally speaking, QMC methods seem to provide better estimates than their MC counterparts, which is logical with regard to the synthesis sequence evaluations. For a same given error they can lead to a 25% reduction of the number of particles in the best case (sequence (2)), which can constitute a significant gain in computing time.

The avatar superimposition for the state vector estimate (Fig. 11) confirms that tracking provides acceptable results in most cases. Errors on sequence (2) are higher due to the long distance between the subject and the cameras on a long part of the sequence, but visually, tracking seems as satisfactory as on the other sequences.

5.2.2. *Joint error and measurements*

Figure 14 presents the average error of the APF on sequence (1) for each joint. First of all, we can see that errors are really different from one joint to the other. Nearly all joint errors are below 10 cm but feet present very high localization errors.



(a)

(b)

(c)

(d)

Fig. 12. Evaluation on sequence (1): (a) Mean error, (b) variance, (c) tracking failure rate (d) bias of the filters for a varying number of particles N.

This is not really surprising as members are always more difficult to track than torso. However hands present an acceptable localization error, due to our skin blob distance measure. Introducing a measure taking into account the position of the final points of the kinematic chain greatly improves their tracking. As we do not set up such a measurement for the feet, their tracking is less accurate. Furthermore, the segmented silhouette may present a bad quality especially on the ground because of the shadows. This damages feet localization too. Elbows are tracked fairly consistently, and this is due to their leanness with respect to the torso or even the legs. This is confirmed by the video sequences showing the projected estimated configuration (Fig. 11). Additionally, experiments made without using the skin blob distance measure confirmed the tracking of hands can reveal as poor as the foot one. This also reveals that our measurement functions need improvement to achieve a better estimate of the body pose.

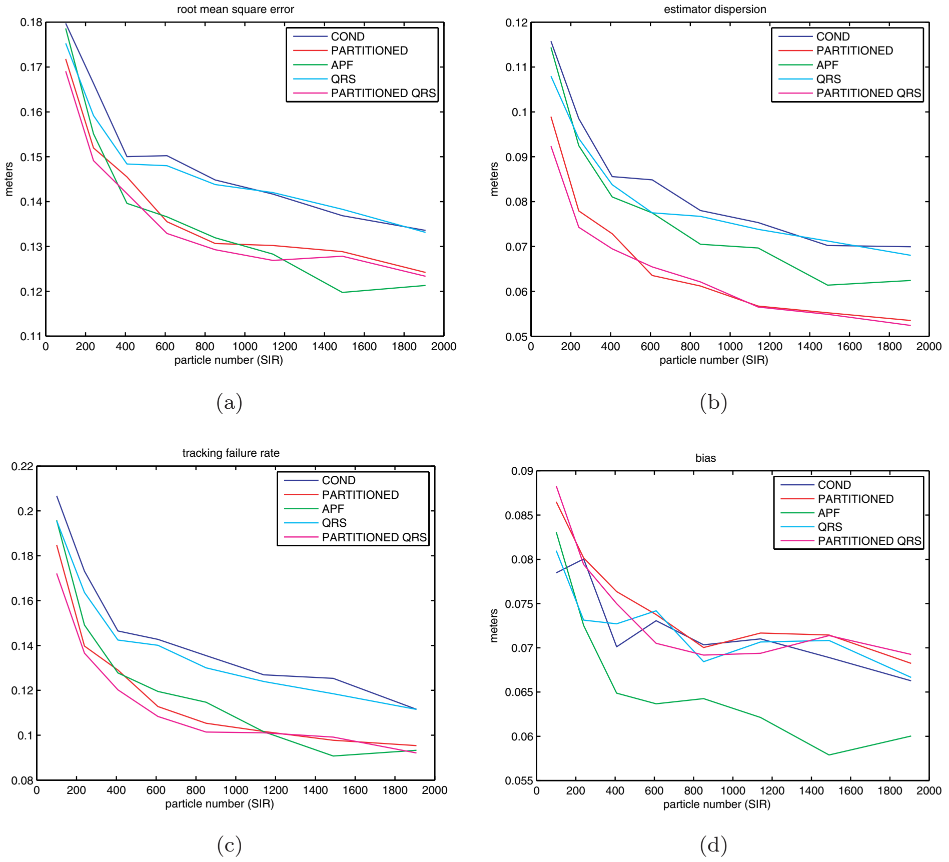


Fig. 13. Evaluation on sequence (2): (a) Mean error, (b) variance, (c) tracking failure rate (d) bias of the filters for a varying number of particles N.

5.2.3. *Variance*

Variances presented in Figs. 12 and 13(b) seem to be correlated with the plots obtained on synthetic data. This is quite logical as variance does not depend on the ground truth. However, it can still be influenced by the context and the model accuracy such as in sequence (2) for example.

As was the case for synthetic sequences, PARTITIONED and APF schemes generally present a lower variance. QMC versions of the filters tend to provide a more "stable" estimate than the classical MC counterparts. The evaluation of filter estimate variance globally seems consistent with the results obtained on synthesis sequences.

5.2.4. *Tracking failures*

The number of tracking failures is shown in Figs. 12 and 13(c). Results are consistent with those obtained on synthesis sequences: APF always presents a low number

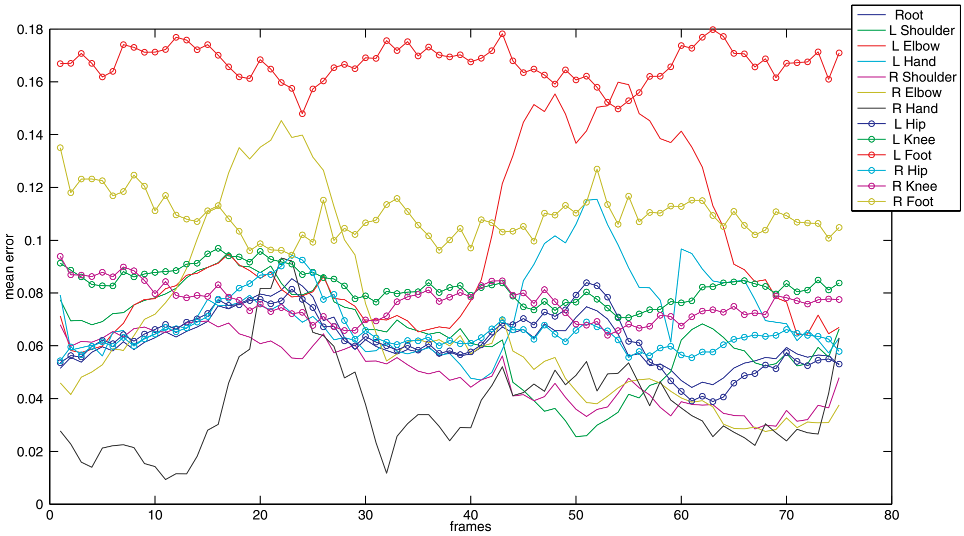


Fig. 14. Mean error of each joint for an APF strategy on a 75 image sequence.

1 of tracking failures, QMC strategies outperform MC ones and PARTITIONED strategies present a very acceptable failure rate with regard to classical ones 3 (SIR).

5.2.5. *Bias*

5 The bias norm is presented in Figs. 12 and 13(d). We notice that bias values are higher than in a synthetic context, showing that inadequacy of our models (human, 7 measurement cues) significantly degrades the filter behavior. APF globally seems to have a lower bias than the other strategies. The bias limit when $N \to \infty$ would 9 fix the physical limit below which any filter would never achieve.

Finally, Table 7 sums up this evaluation against all the proposed criteria, 11 keeping in mind that results may vary according to the movement type. We notice however that relative behaviors of the filter are independent of the num- 13 ber of particles. Thus, this table informs us about intrinsic properties of the filters.

Table 7. Sorting of the different strategies according to each criteria.

Name	Accuracy	Dispersion	Failure Rate	Bias
CONDENSATION	5	5	5	5
QRS	4	4	4	4
PARTITIONED	3	3	2	3
PARTITIONED QRS	2	1	2	1
APF	1	2	1	2

5.2.6. *Particles and efficiency*

Figures 15(a)–15(e) show the mean error (RMSE) for tracking sequence (1) with a varying number of particles for each filtering scheme. We see that — of course — the tracking is more accurate for a growing number of particles, but also that the trackers seem to converge towards an error which is not zero. This means that our likelihood function does not present a maximum for the real human pose whatever the chosen filtering strategy. This is due to our inaccurate models of the human body and measurement functions and it is consistent with our bias evaluations. Moreover, we see that there seems to exist a maximum number of particles above which the tracking is not improved.

Additionally, Fig. 15(f) presents the number of tracking failures for each strategy depending on the number of particles in log-scale. SIR family strategies present an exponential complexity to achieve a lowest error with regard to the number of particles, and not a linear one as this can be the case in simpler contexts.[5] PARTITIONED and APF have less than exponential complexity. Our system actually performs in 1 fps on a Pentium IV 3 GHz while any vision-based approach based on particle filtering is far from real time: 0.02 fps in Ref. 3, 0.03 fps in Ref. 16, 0.07 in Ref. 8. One limiting issue for such approaches is clearly the computational challenge of processing full video streams.

To sum up, it emerges that PARTITIONED QRS strategy is seldom exploited in the literature but provides results as good as the APF. Given our evaluations, these two algorithms outperform the other strategies. In the next section, we study the behavior of APF in a real human environment with various subjects.

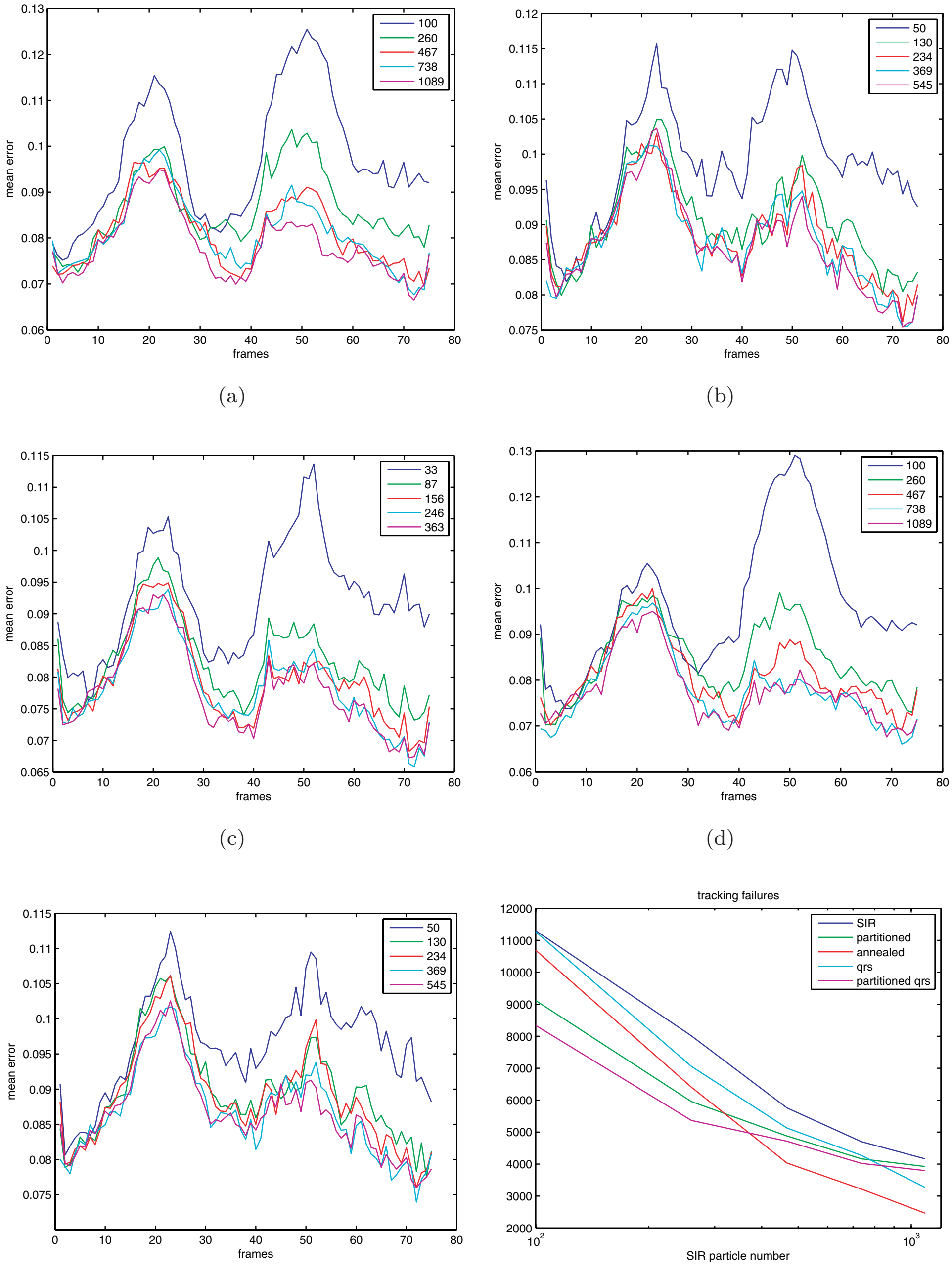
5.3. *Qualitative evaluations for real surveillance video application*

The previous evaluations have been made in using a commercial HMC context, i.e. in a clear dedicated room in order to obtain the ground truth in the best conditions. Moreover, we have assessed the algorithms performance on a single subject. To complete this quantitative study, we propose in this section a few analyses on three subjects evolving in a 4×3 m working area in a real indoor environment surveillance context. Some representative results are shown in Figs. 16 and 17. We only show images from two of the three cameras for lack of space.

Three Firewire Point Grey color cameras Flea 2 have been mounted in our robotic hall in order to assess tracking in a human indoor context. 640×480 pixel images are acquired at 6 Hz and algorithms are run offline.

Figure 16 presents images from a walk and gym movement performed in such an environment. The tracking is performed with an APF involving 500 particles and three layers. We notice that the tracking performs well, yet, the localization of the feet seems less accurate than the one of the hands. This confirms our previous results. Tracking seems quite repeatable on different runs.

Figure 17 shows another walking subject who crosses the scene to take a book and read it. Occlusion of hands occur during the sequence but the tracker



(a)

(b)

(c)

(d)

(e)

(f)

Fig. 15. Mean error of each filtering scheme for different number of particles: (a) SIR, (b) PARTITIONED, (c) APF, (d) QRS, (e) PARTITIONED QRS. (f) presents the number of tracking failures with respect to the number of particles in a log-scale.

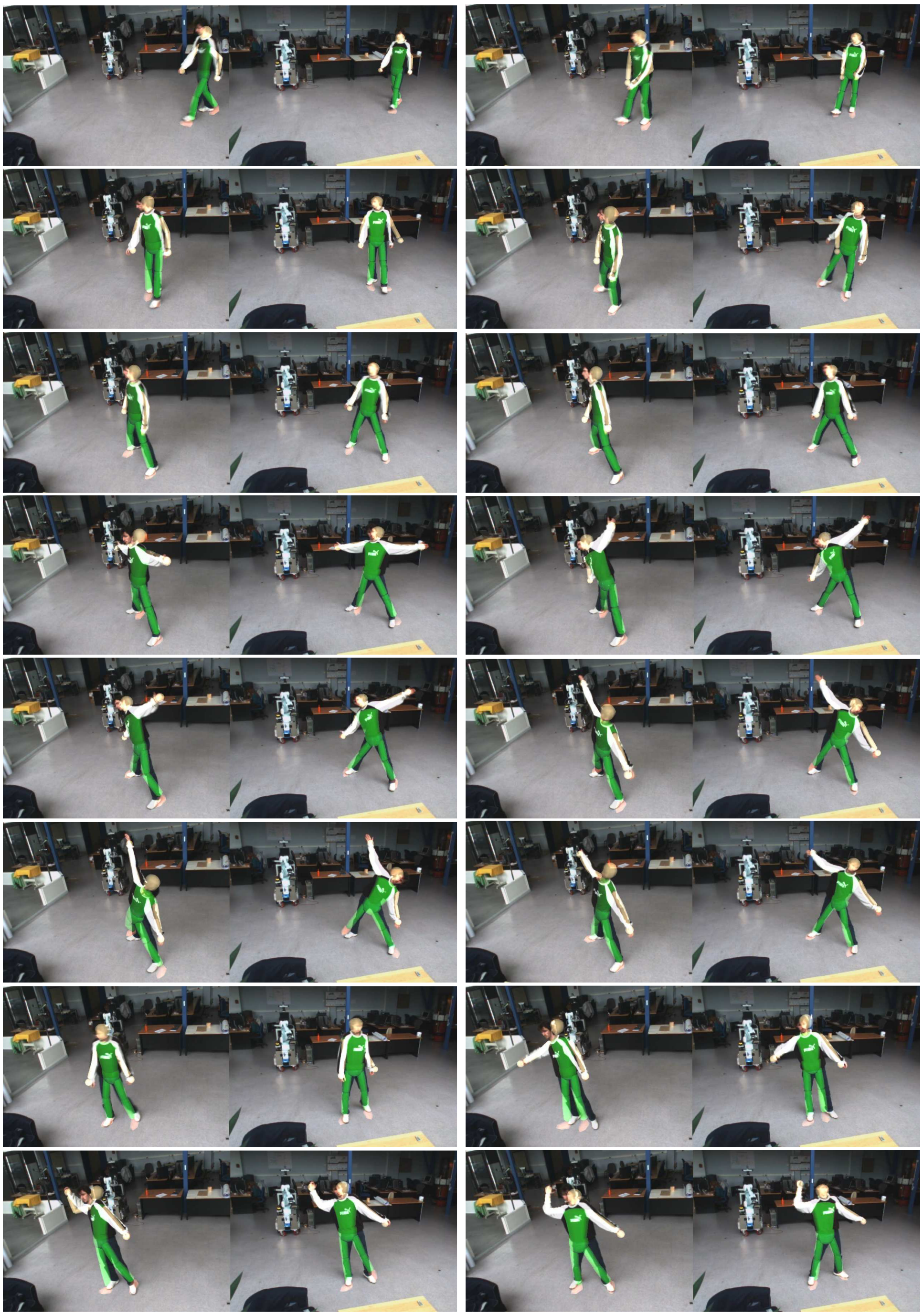


Fig. 16. Pairs of snapshots from a sequence of walking and gym movement performed in a natural human centered environment.

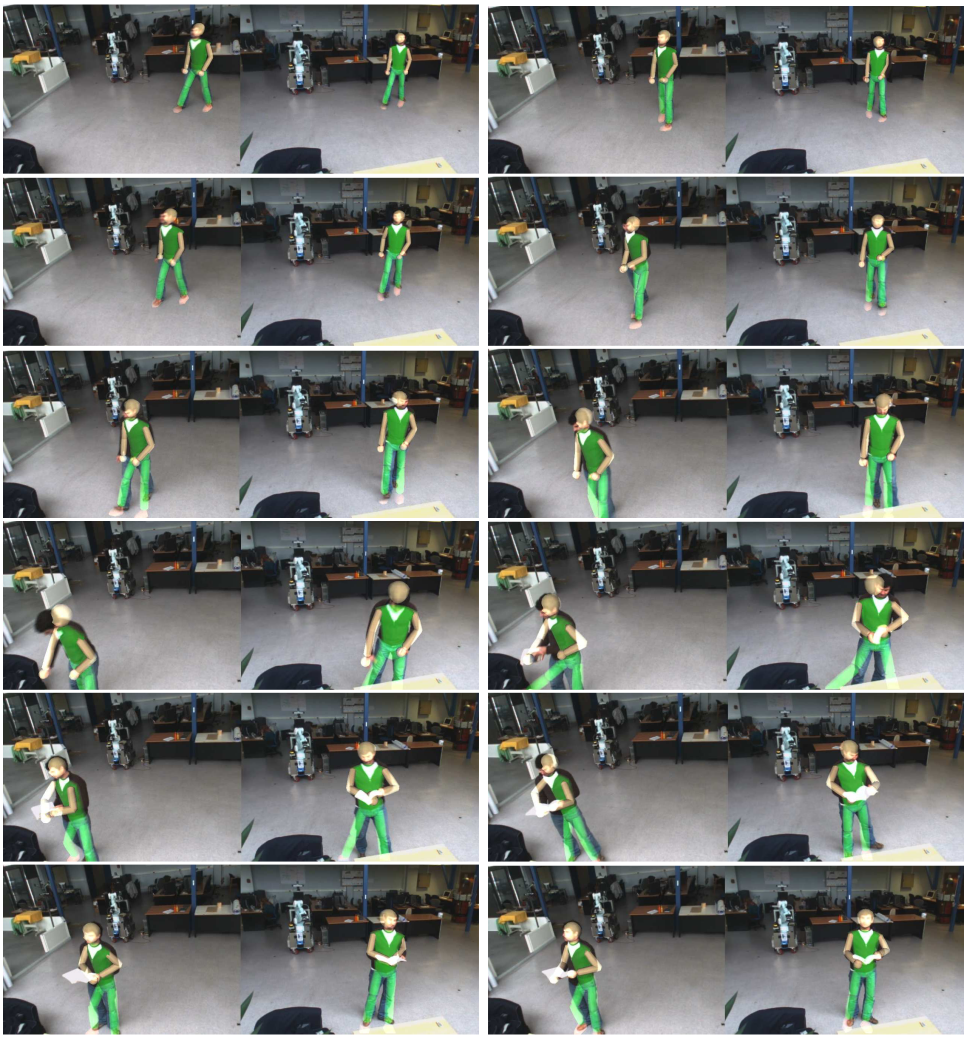


Fig. 17. Pairs of snapshots from a sequence of walking and reading with a different subject.

still performs well. However, the performance is a bit degraded when the subject approaches the border of the working area, but the target is locked again as he returns back. In addition, we notice that despite clothes being different on the above sequences, the tracker is not disturbed. The only condition our measurements impose is that the subject must wear long sleeved clothes. Our system is then robust to different subjects, even if, as is the case in this sequence, morphology is slightly different.

Generally speaking, we notice that tracking is correct as soon as a good segmentation of the silhouette is performed, which confirms the results in Ref. 3. However, as we exploit skin segmentation, performances are degraded when the subject does not present hands or face on at least one camera. The results obtained with PARTITIONED QRS strategy are visually similar to the ones obtained with APF, confirming that both consitute an interesting choice.

6. Conclusion and Future Works

This study aims to present an overview of the characteristics of some well-known particle filters. Four metrics have been set up to assess behaviors of the algorithms on synthesis sequences as well as real ones. Many conclusions have been deducted from those experiments.

Though behaviors of the filters are expected in friendly contexts (such as synthetic sequences), real sequences seem more complex to analyze and to understand, as some movements possibly spoil filter relative efficiency. Hence, one has to be very careful while comparing particle filters; the more "advanced" algorithms are not always better in practice. However, APF provides good results, and PARTITIONED strategies may be disturbed by relying on an inadequate likelihood function. Indeed, results should be enhanced with "labeling cues" enabling a more efficient mining of the image, but such measurements are difficult to set up in practice.

In addition, the average error is not the only criterion we must look at, as in our stochastic context, variance of the estimates must be taken into account. The number of tracking failures is influenced by both mean error and variance of the estimates in such a way that a scheme providing a worse error but a better variance than another can be of better interest in some contexts. With regard to this criterion, QMC strategies seem to provide an interesting way to explore, while they are not used much in tracking problems. In real context application, APF strategy performs well even for various subjects with a rough body model. However, foot localization may be poor due to segmentation errors, and specific measurements such as hand localization improve tracking results.

Obviously, trackers are sensitive to the chosen strategy, but a lot of work still has to be done on measurements, since if the filtering algorithm can enable a faster convergence according to the strategy or the number of particles, the limit of this convergence is defined by the chosen measurement. In other words, measurements control the point of convergence and filter strategies control the speed of convergence. If measurements are not well appropriate (due to a difficult context or rough models), no strategy will ever provide a good result. To endorse this idea, it appears that choosing likelihood functions focused also on end points of the kinematic chains have been proven to improve tracking significantly (hand positions are more precise than foot ones). Thus, designing likelihood functions must be done very carefully, under penalty of severely damaging filter results.

Regardless of these assumptions, further investigations will concern the tuning of the filter parameters. To our knowledge, no study tries to tackle the "σ-problem" which is one of the most difficult of the particle filter use. Influence of these variables should be tested to fully exploit particle filter possibilities. In a similar way, the number of layers and/or partitions of “advanced” strategies also constitutes a strategic choice and deserves a more complete study.

To expand this study, one should also take a look at importance sampling strategies, which takes into account the image while processing particle exploration of the state space. With a perspective to achieving real-time robust tracking, reducing the number of particles and setting up automatic re-initialization seem to constitute the key points especially in even more difficult contexts such as robotics. Indeed, in such a context, very informative measurements such as background segmentation may be compromised as embedded cameras are able to move. Additionally, the background can be very cluttered and lighting conditions may vary. All those characteristics undoubtedly influence tracking results, and also deserve a complete study.

Acknowledgments

The authors thank Ms. Caroline Carvalho for generating all the synthesis images. This work was partially conducted within the french ANR project AMORCES.

References

1. P. Azad, A. Ude, T. Asfour and R. Dillmann, Stereo-based markerless human motion capture for humanoid robot systems, *Int. Conf. Robotics and Automation (ICRA’07)*, Roma, Italy, 2007, pp. 3951–3956.
2. A. Balan, M. J. Black, H. Haussecker and L. Sigal, Shining a light on human pose: on shadows, shading and the estimation of pose and shape, *Int. Conf. Computer Vision (ICCV’07)*, Rio de Janeiro, BRAZIL, October 2007.
3. A. Balan, L. Sigal and M. Black, A quantitative evaluation of video-based 3D person tracking, *Int. Workshop on Visual Surveillance and Performance Evaluation of Tracking and Surveillance (VS-PETS’05)*, Washington, USA, October 2005, pp. 349–356.
4. Z. Chen, Bayesian filtering: from Kalman filters to particles filters, and beyond, available on http://www.math.u-bordeaux.fr/∼delmoral/chen_bayesian.pdf, 2003.
5. F. Daum and J. Huang, Mysterious computational complexity of particle filters, *Signal and Data Processing of Small Targets*, *Proc. SPIE*, Vol. 4728, Bellingham, MA, USA, August 2003.
6. Q. Delamarre and O. Faugeras, 3D articulated models and multi-view tracking with physical forces, *Comput. Vis. Imag. Underst.* **81**(3) (2001) 328–357.
7. J. Deutscher, A. Blake and I. Reid, Articulated body motion capture by annealed particle filtering, *IEEE Conf. Computer Vision and Pattern Recognition (CVPR’00)*, Vol. 2, Hilton Head Island, South Carolina, USA, 2000, pp. 126–133.
8. J. Deutscher, A. Davison and I. Reid, Automatic partitioning of high dimensional search spaces associated with articulated body motion capture, *IEEE Conf. Computer Vision and Pattern Recognition (CVPR’01)*, Kauaii Marriott, Hawaii, USA, 2001, pp. 669–676.
9. J. Deutscher and I. Reid, Articulated body motion capture by stochastic search, *Int. J. Comput. Vis.* **21**(3) (2005) 185–205.
10. 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).

11. A. Doucet, S. Godsill and C. Andrieu, On sequential Monte-Carlo sampling methods for Bayesian filtering, *Stat. Comput.* **10**(3) (2000) 197–208.
12. K.-T. Fang, Y. Wang and P. M. Bentler, Some applications of number-theoretic methods in statistics, *Stat. Sci.* **9**(3) (1994) 416–428.
13. D. M. Gavrila, The visual analysis of human movement: a survey, *Comput. Vis. Imag. Underst.* **73**(1) (1999) 82–98.
14. D. Guo and X. Wang, Quasi-Monte Carlo filtering in nonlinear dynamic systems, *IEEE Trans. Sign. Process.* **54**(6) (2006) 2087–2098.
15. A. Gupta, T. Chen, F. Chen, D. Kimber and L. S. Davis, Context and observation driven latent variable model for human pose estimation, *IEEE Conf. Vision and Pattern Vision Recognition (CVPR'08)*, June 2008.
16. A. Gupta, A Mittal and L. S. Davis, Constraint integration for efficient multiview pose estimation of humans with self-occlusions, *Trans. Patt. Anal. Mach. Intell. (PAMI'08)* **30**(3) (2008) 493–506.
17. M. Isard and A. Blake, Contour tracking by stochastic propagation of conditional density, *European Conf. Computer Vision (ECCV'96)*, Cambridge, UK, April 1996, pp. 343–356.
18. M. Isard and A. Blake, I-CONDENSATION: Unifying low-level and high-level tracking in a stochastic framework, *European Conf. Computer Vision (ECCV'98)*, Freiburg, Germany, 1998, pp. 893–908.
19. S. Knoop, S. Vacek and R. Dillman, Sensor fusion for 3D human body tracking with an articulated 3D body model, *Int. Conf. Robotics and Automation (ICRA'06)*, Orlando (USA), May 2006, pp. 1686–1691.
20. J. Lichtenauer, M. J. T. Reinders and E. A. Hendriks, Influence of the observation likelihood function on particle filtering performance in tracking applications, *Automatic Face and Gesture Recognition (FGR'04)*, Seoul, KOREA, May 2004, pp. 767–772.
21. J. MacCormick and M. Isard, Partitioned sampling, articulated objects, and interface-quality hand tracking, *European Conf. Computer Vision (ECCV'00)*, Dublin, Ireland, 2000, pp. 3–19.
22. P. Menezes, F. Lerasle and J. Dias, Data fusion for 3D gesture tracking using a camera mounted on a robot, *Int. Conf. Pattern Recognition (ICPR'06)*, Vol. 1, Hong-Kong, August 2006, pp. 464–467.
23. T. Moeslund, A. Hilton and V. Krüger, A survey of advanced vision-based human motion capture and analysis, *Comput. Vis. Imag. Underst. (CVIU'06)*, **104** (2006) 174–192.
24. H. Moon and R. Chellappa, 3D shape-encoded particle filter for object tracking and its application to human body tracking, *EURASIP J. Imag. Vid. Process.* 2008.
25. L. Mündermann, S. Corazza and T. P. Andriacchi, Accurately measuring human movement using articulated ICP with soft-joint constraints and a repository of articulated models, *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'07)*, June 2007, pp. 1–6.
26. K. Ogawara, X. Li and K. Ikeuchi, Markerless human motion estimation using articulated deformable model, *Int. Conf. Robotics and Automation (ICRA'07)*, Roma, Italy, 2007, pp. 46–51.
27. D. Ormoneit, C. Lemieux and D. J. Fleet, Lattice particle filters, *Proc. 17th Conf. Uncertainty in Artificial Intelligence (UAI'01)*, San Francisco, CA, USA, 2001, pp. 395–402.
28. P. Pérez, J. Vermaak and A. Blake, Data fusion for visual tracking with particles, *Proc. IEEE* **92**(3) (2004) 495–513.

29. V. Philomin, R. Duraiswami and L. S. Davis, Quasi-random sampling for CONDENSATION, *European Conf. Computer Vision (ECCV'00)*, Dublin, Ireland, 2000, pp. 134–149.
30. D. Ramanan, D. Forsyth and A. Zisserman, Strike a pose: tracking people by finding stylized poses, *IEEE Conf. Vision and Pattern Recognition (CVPR'05)*, San Diego, USA, June 2005, pp. 271–278.
31. H. Sidenbladh, M. J. Black and D. J. Fleet, Stochastic tracking of 3D human figures using 2D image motion, *European Conf. Computer Vision (ECCV'00)*, Dublin, Ireland, 2000, pp. 702–718.
32. H. Sidenbladh, M. J. Black and D. J. Fleet, Implicit probabilistic models of human motion for synthesis and tracking, *European Conf. Computer Vision (ECCV'02)*, Copenhagen, Denmark, 2002, pp. 784–800.
33. L. Sigal, S. Bhatia, S. Roth, M. J. Black and M. Isard, Tracking loose-limbed people, *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'04)*, Washington, DC, USA, 2004, pp. 421–428.
34. C. Sminchisescu and B. Triggs, Covariance scaled sampling for monocular 3D body tracking, *IEEE Conf. Pattern Vision Recognition, (CVPR'01)*, Kauaii Marriott, Hawaii, USA, December 2001, pp. 447–454.
35. C. Sminchisescu and B. Triggs, Estimating articulated human motion with covariance scaled sampling, *Int. J. Robot. Res.* **6**(22) (2003) 371–393.
36. A. Sundaresan and R. Chellappa, Model driven segmentation of articulating humans in laplacian eigenspace, *IEEE Trans. Patt. Anal. Mach. Intell. (PAMI'08)* **30**(10) (2008) 1771–1785.
37. R. Urtasun, D. J. Fleet and P. Fua, Motion models for 3D people tracking, *Comput. Vis. Imag. Underst.* **104**(2–3) (2006) 157–177.
38. R. Urtasun and P. Fua, 3D human body tracking using deterministic temporal motion models, *European Conf. Computer Vision (ECCV'04)* (2004), pp. 92–106.
39. R. Van Der Merwe, N. De Freitas, A. Doucet and E. Wan, The unscented particle filter, *Adv. Neural Inform. Process. Syst.* **13** (2001).
40. P. Wang and J. Rehg, A modular approach to the analysis and evaluation of particle filters for figure tracking, *IEEE Conf. Computer Vision and Pattern Recognition (CVPR'06)*, New York, USA, 2006, pp. 790–797.
41. http://vision.cs.brown.edu/humaneva/ — humaneva.
42. http://www.motionanalysis.com — Motion Analysis Corporation.
43. X. Xu and B. Li, Learning motion correlation for tracking articulated human body with a rao-blackwellised particle filter, *Int. Conf. Computer Vision (ICCV'07)*, Rio de Janeiro, October 2007, pp. 1–8.
44. J. Ziegler, K. Nickel and R. Stiefehagen, Tracking of the articulated upper body on multi-view stereo image sequences, *Int. Conf. Computer Vision and Pattern Recognition (CVPR'06)*, New York, USA, 2006, pp. 774–781.