TRACKING POORLY MODELLED MOTION USING PARTICLE FILTERS WITH ITERATED LIKELIHOOD WEIGHTING

Hammadi Nait-Charif and Stephen J. McKenna

Division of Applied Computing, University of Dundee, Dundee DD1 4HN, Scotland
{hammadi,stephen}@computing.dundee.ac.uk

ABSTRACT

Human motion in cluttered scenes is often tracked using particle filtering. However, poorly modelled inter-frame motion is not uncommon, resulting in poor priors for the filtering step. Alternatives to the Condensation algorithm in the form of an Auxiliary Particle Filter (APF) and Iterated Likelihood Weighting (ILW) are described. Experimental results comparing these filters' accuracy and consistency are presented for a scenario in which a person is tracked in an overhead view using an ellipse model with a likelihood based on colour and gradient cues. ILW is not intended to give unbiased estimates of a posterior but rather to reduce approximation error. It is shown to outperform both Condensation and the APF on sequences from this scenario.

1. INTRODUCTION

Consider tracking human motion in a scenario such as that shown in Figure 1. Here a person's head is tracked in an extremely cluttered overhead view using a method which will be described in Sections 5 and 6. Motion models will not always be reliable; in this example the person falls over, leading to relatively large and unpredictable inter-frame motion. In many applications it is important to track through these poorly modelled motions in order to allow rare but salient events such as falls to be recognised.

Visual tracking is often formulated from a Bayesian perspective as a problem of estimating some degree of belief in the state $\mathbf{x}_t$ of an object at time step t given a sequence of observations $\mathbf{z}_{1:t}$. Bayesian filtering recursively computes a posterior density using a Markov assumption:

$$p(\mathbf{x}_{t+1}|\mathbf{z}_{1:t+1}) \propto p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1})p(\mathbf{x}_{t+1}|\mathbf{z}_{1:t}) \quad (1)$$

where the prior is the previous posterior propagated using a dynamic model:

$$p(\mathbf{x}_{t+1}|\mathbf{z}_{1:t}) = \int p(\mathbf{x}_{t+1}|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t})d\mathbf{x}_t \quad (2)$$

Specific dynamic models for expected activities could be constructed using learning (e.g. [1, 2, 3]). However, when

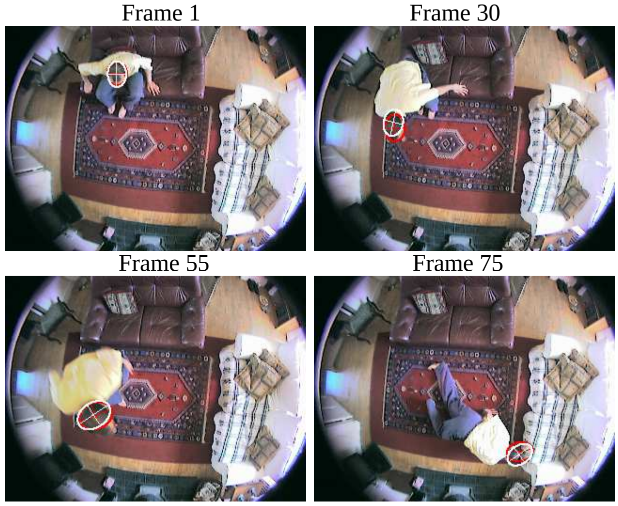


Fig. 1. Frames from a sequence in which the head is tracked using ILW. The person is tracked through a sudden fall. The white ellipse represents the most heavily weighted particle.

true motion deviates from that predicted by such specific models, tracking failure becomes likely. If the range of activities is varied and includes salient but rare movements, a more generic model results in more reliable tracking. Therefore, the approach taken in this paper is to adopt a very general motion model (a Gaussian random walk). Measurements from the current image are used to encourage exploration of high probability regions of the state space.

The posterior density in (1) cannot be computed analytically unless linear-Gaussian models are adopted. As is well-known, linear-Gaussian models are unsuitable for many visual tracking problems. Instead, simulation-based particle filters are often used to propagate what are often non-Gaussian, multimodal densities over time. This paper explores the use of particle filtering for tracking human motion. It compares, in the context of an overhead tracking application, alternatives to the widely used Condensation algorithm that can improve both the accuracy and consistency of tracking: an auxiliary particle filter (APF) [4] and iterated likelihood weighting (ILW) [5].

This research was funded by UK EPSRC grant GR/R27419/01. Thanks to the reviewers and T. Roberts for helpful comments.

2. RELEVANT WORK

Particle filtering [6] has become popular for visual tracking since it was applied by Isard and Blake in the form of Condensation [7]. They later suggested using a secondary tracker to generate an importance function for sampling [8]. Condensation's behaviour with finite particle sets was investigated by King and Forsyth [9]. Several authors have suggested alternative sampling schemes (e.g. [4, 10, 11, 12, 13, 14, 15, 16, 17]). For example, Choo and Fleet [11] used a hybrid Monte Carlo filter to sample the posterior. Rui and Chen [16] used an unscented Kalman filter to generate importance densities. Deutscher *et al.* [12, 13] proposed annealed and partitioned particle filtering.

3. SAMPLING IMPORTANCE RESAMPLING

The widely used Sampling Importance Resampling (SIR) algorithm [6] (otherwise known as Condensation [7]) approximates $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ at each time step t by a set of N particles $\{\mathbf{x}_t^n, w_t^n\}_{n=1}^N$ where each particle is a weighted random sample and $\sum_{n=1}^N w_t^n = 1$. The filtered posterior is then

$$p(\mathbf{x}_{t+1}|\mathbf{z}_{1:t+1}) \propto p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}) \sum_{n=1}^{N} w_t^n p(\mathbf{x}_{t+1}|\mathbf{x}_t^n) \quad (3)$$

where the prior is now a mixture with N components. The SIR filter involves (i) selecting the n^{th} mixture component with probability w_t^n, (ii) drawing a sample from it, and (iii) assigning the sample a weight proportional to its likelihood. Resampling is used to obtain samples with equal weights in order to facilitate sampling from the mixture in (3). The dynamic (motion) model is encapsulated by the transition density $p(\mathbf{x}_{t+1}|\mathbf{x}_t^n)$. Typically, a sample can be drawn from it by adding random process noise and then applying deterministic dynamics (drift).

In general, sequential importance sampling filters operate by drawing samples from an importance density, $q(\mathbf{x})$, and weighting them using (4) to give a particle representation of the posterior density.

$$w_{t+1}^n \propto w_t^n \frac{p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}^n)p(\mathbf{x}_{t+1}^n|\mathbf{x}_t^n)}{q(\mathbf{x}_{t+1}^n|\mathbf{x}_t^n, \mathbf{z}_{t+1})} \quad (4)$$

SIR is a sequential importance sampling filter in which the prior is used as the importance density. This is a convenient choice because an unbiased, asymptotically correct estimate of the posterior can be obtained by simply weighting the samples with their likelihood. The resulting algorithm is therefore intuitive and easily implemented. However, the prior is certainly not the optimal choice of importance function since it does not take into account the most recent observation, $\mathbf{z}_{t+1}$. Sampling using SIR is particularly inefficient when the likelihood is in the tails of the prior or if the likelihood is narrow and peaked compared to the prior. Although SIR gives an asymptotically correct estimate of the posterior, its behavior with finite sample sets is often not good [9]. In human tracking, the dynamic models used can often result in poor priors due to unexpected motion. In such cases, SIR will place many samples in the wrong regions of the state space. As a result, very large particle sets can be required in order to achieve acceptable performance.

For $n = 1 \ldots N$
Compute μ_{t+1}^n
Compute $w_{t+1}^n = w_t^n p(\mathbf{z}_{t+1}|\mu_{t+1}^n)$
For each particle
Choose an index n with probability proportional to w_{t+1}^n
Draw a sample $\mathbf{x}_{t+1}^m$ from $p(\mathbf{x}_{t+1}|\mathbf{x}_t^n)$
Assign weight $w_{t+1}^m = \frac{p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}^m)}{p(\mathbf{z}_{t+1}|\mu_{t+1}^{n_m})}$
Normalise weights so that $\sum_{n=1}^N w_{t+1}^n = 1$

Table 1. The auxiliary particle filter

4. AUXILIARY PARTICLE FILTERS

The APF was proposed as a way of filtering with an importance density that depends on the most recent observation [4]. It is an extension of SIR that approximates the filtered posterior of (3) as

$$\hat{p}(\mathbf{x}_{t+1}|\mathbf{z}_{1:t+1}) \propto \sum_{n=1}^{N} w_t^n p(\mathbf{z}_{t+1}|\mu_{t+1}^n) p(\mathbf{x}_{t+1}|\mathbf{x}_t^n) \quad (5)$$

where μ_{t+1}^n is some value likely to be generated by the dynamic model $p(\mathbf{x}_{t+1}|\mathbf{x}_t^n)$. The algorithm consists of sampling $m = 1 \ldots N$ times from this mixture and then weighting the samples using

$$w_{t+1}^m \propto \frac{p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}^m)}{p(\mathbf{z}_{t+1}|\mu_{t+1}^{n_m})} \quad (6)$$

where $\mu_{t+1}^{n_m}$ is the value associated with the component $p(\mathbf{x}_{t+1}|\mathbf{x}_t^{n_m})$ from which the m^{th} sample was drawn. The algorithm is summarised in Table 1.

Specifically, if the dynamic model is zero-mean Gaussian noise and $\mu_{t+1}^{n_m}$ is taken to be the expected value of $p(\mathbf{x}_{t+1}|\mathbf{x}_t^n)$ then an APF is obtained by (i) choosing a component n with probability proportional to $w_t^n p(\mathbf{z}_{t+1}|\mathbf{x}_t^n)$, (ii) drawing a sample $\mathbf{x}_{t+1}^m$ from $p(\mathbf{x}_{t+1}|\mathbf{x}_t^n)$ and (iii) weighting the sample as:

$$w_{t+1}^m \propto \frac{p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}^m)}{p(\mathbf{z}_{t+1}|\mathbf{x}_t^n)} \quad (7)$$

APF generates particles from an importance density conditioned on the most recent observation and then samples the posterior using this importance density. When compared to SIR, this requires an extra likelihood evaluation per particle. However, this can be more than offset in terms of computational efficiency since fewer particles are likely to be needed due to the more efficient sampling of the posterior.

5. ITERATED LIKELIHOOD WEIGHTING

Great care is usually taken to ensure that an unbiased estimate of the posterior is obtained when applying particle filtering to tracking. The importance sampling steps of (4), (6) and (7) are bias-correcting schemes used to obtain such an unbiased estimate. However, it is well known in statistical inference that approximation error depends not only on the bias but on the variance. If the importance density is reasonably accurate, the correction step may in fact increase approximation error for all but very large particle sets [18].

Furthermore, the prior density is often poor and noisy and it therefore makes little sense to attempt to obtain a computationally expensive, high accuracy approximation to the posterior. This is particularly true in many human tracking applications where inter-frame motion is often poorly modeled by the dynamic model (transition density).

A scheme is proposed here in which only a subset of the particles at each time step is sampled from the 'posterior'. The remainder of the particles are used to increase sampling in regions of high likelihood via a simple iterative search using the most recent observation. This is useful when the prior (dynamic model) is poor. It can prevent tracking failure in the case of unexpected motion, for example. Rather than attempt to perform a (potentially expensive) bias-correction step for those particles used to search high-likelihood regions, they are weighted at each iteration based on their likelihood. The resulting algorithm (Table 2) is not an unbiased, Bayesian particle filter within the usual Markov framework. After an intial iteration of SIR, the sample set is split uniformly at random into two sets of equal size. One of these sets is propagated to the next time step unaltered while the samples in the other set are subjected to further iterations of diffusion, likelihood weighting and resampling. This has the effect of migrating half of the particles to regions of high likelihood while the other half are sampled using the prior as the importance function. The effectiveness of ILW is demonstrated empirically in Section 7 where it is compared to SIR and APF over multiple runs.

6. LIKELIHOOD MODEL

In order to apply the above filtering schemes to tracking, a state, $\mathbf{x}$, and a likelihood model, $p(\mathbf{z}|\mathbf{x})$, must be defined. Head shape is reasonably well approximated as an ellipse in the image irrespective of pose. Previous authors have used constrained ellipses to track frontal-profile views of the head [16, 19, 20]. As orientation and elongation vary with pose and position, particularly in an overhead view, all five ellipse parameters were estimated here.

The likelihood model combined intensity gradient information along the head boundary with a colour model of the ellipse's interior region as described in an earlier paper [5]. The region likelihood $p(\mathbf{r}_t|\mathbf{x}_t^n)$ was based on divergence of

1. Draw N samples $\mathbf{x}_{t+1}^n \sim p(\mathbf{x}_{t+1}|\mathbf{x}_t^n)$
2. Assign weights $w_{t+1}^n = p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1}^n)$
3. Normalise weights so that $\sum_{n=1}^N w_{t+1}^n = 1$
4. Resample with replacement to obtain samples $\mathbf{x}_{t+1}^n$ with equal weights
5. Split the sample set at random into two sets of size $M = N/2$: $\{\mathbf{x}_{t+1,1}^m\}_{m=1}^M$ and $\{\mathbf{x}_{t+1,*}^m\}_{m=1}^M$
6. For $k = 1 \ldots K$
   - Draw M samples $\mathbf{x}_{t+1,k+1}^m \sim p(\mathbf{x}_{t+1,k+1}|\mathbf{x}_{t+1,k}^m)$
   - Assign weights $w_{t+1,k+1}^m = p(\mathbf{z}_{t+1}|\mathbf{x}_{t+1,k+1}^m)$
   - Normalise weights so that $\sum_{m=1}^M w_{t+1,k+1}^m = 1$
   - Resample with replacement to obtain M samples $\mathbf{x}_{t+1,k+1}^m$ with equal weights
7. For $m = 1 \ldots M$
   - $\mathbf{x}_{t+1}^m = \mathbf{x}_{t+1,*}^m$
   - $\mathbf{x}_{t+1}^{M+m} = \mathbf{x}_{t+1,K+1}^m$

Table 2. The ILW filter. Here $p(\mathbf{x}_{t+1,k+1}|\mathbf{x}_{t+1,k}^m)$ is a transition density with expected value $\mathbf{x}_{t+1,k}^m$.

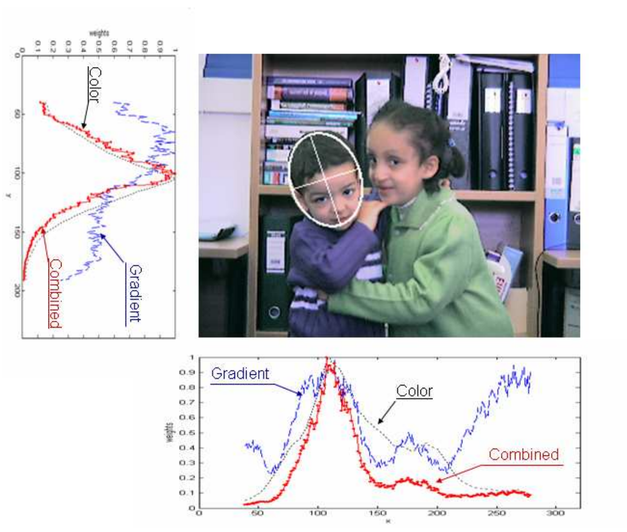
a colour histogram of the ellipse's interior from a model histogram. The boundary likelihood $p(\mathbf{b}_t|\mathbf{x}_t^n)$ was computed by searching for maximal gradient magnitude points near the ellipse boundary. Assuming conditional independence, the likelihood was obtained using Equation (8). Figure 2 illustrates the characteristics of the likelihood and compares it to the use of boundary cues and region cues alone. It varies in a well behaved manner under translation and scaling.

$$p(\mathbf{z}_t|\mathbf{x}_t^n) = p(\mathbf{b}_t|\mathbf{x}_t^n)p(\mathbf{r}_t|\mathbf{x}_t^n) \tag{8}$$

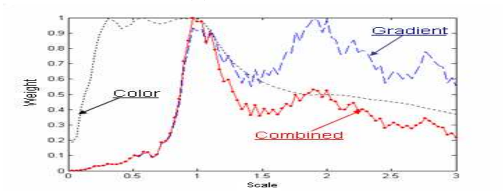
7. EXPERIMENTS

Results are reported here for scenarios in which a person is tracked moving around a home environment using a wide-angle, ceiling-mounted camera. Whilst the articulated structure of the body will not always be readily apparent, it can be assumed that the head will nearly always be visible. The target application is a monitoring system to help extend independent living for older people in their own homes. Here we give indicative results on typical sequences of interest.

Likelihood computation is the main computational expense during tracking and the different filters require different numbers of likelihood evaluations per frame. In order to obtain a fair empirical comparison, the number of particles used with each filter was chosen so that the number of likelihood evaluations per frame was equal. All filters were run with the same transition density and the same noise parameters. Particle set sizes for SIR, APF and ILW were 2000, 1000 and 400 respectively. ILW used an additional 8 iterations per frame giving a total of 2000 likelihood evaluations.



(a) Translation



(b) Scaling

Fig. 2. Likelihoods as the ellipse (a) translates and (b) changes scale away from the correct ellipse. Dashed: gradient likelihood. Dotted: colour likelihood. Solid: combined likelihood (8).

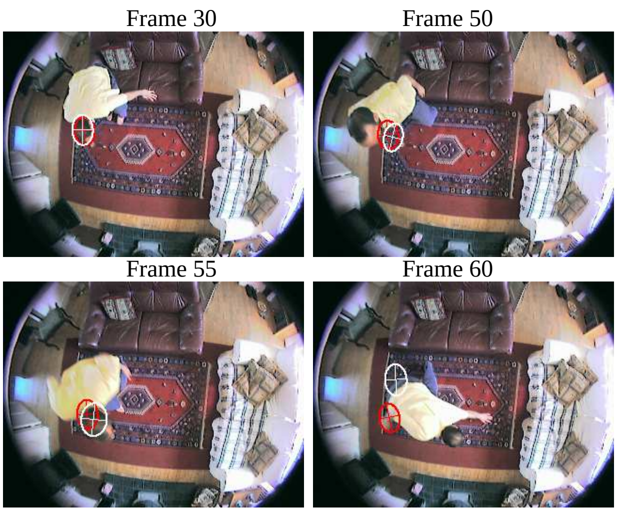


Fig. 3. Frames from the sequence in Figure 1 tracked using SIR. The tracker loses lock in frame 56 and is unable to recover.

Figures 1, 3, 4 and 5 show typical runs of SIR and ILW. A red ellipse indicates the mean estimated from the particle set and a white ellipse indicates the most heavily weighted particle for that frame. In Figure 3 the SIR filter loses track

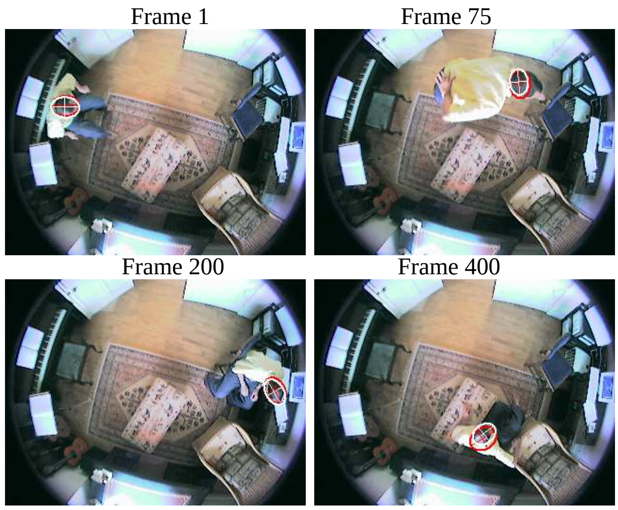


Fig. 4. Frames from a 400-frame sequence in which the occupant stands up, moves around the room, sits on a chair, leans over and finally sits on the floor. ILW tracked successfully throughout.

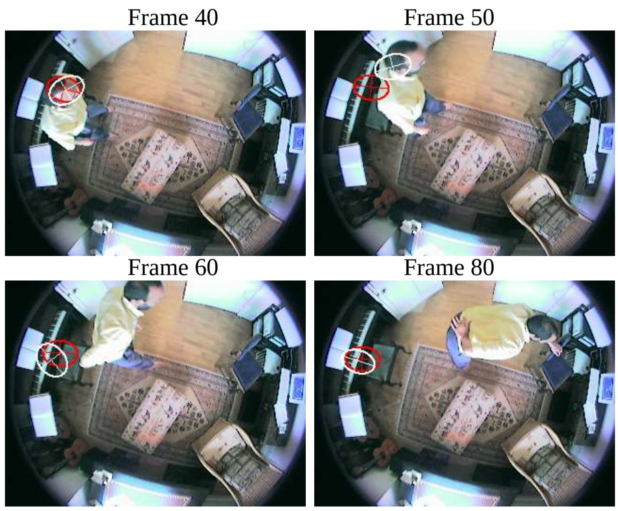


Fig. 5. Frames from the sequence of Figure 4 showing the SIR tracker losing lock after frame 50. It did not recover.

when the person falls due to the sudden, poorly modelled motion. However, Figure 1 shows this sequence being successfully tracked using ILW. Similarly, Figure 4 shows ILW successfully tracking a 400-frame sequence while the SIR filter was easily distracted by clutter (Figure 5).

Although the above runs were typical for these sequences, isolated runs of particle filters are not sufficient to evaluate performance. The filters were compared over multiple runs on the sequence shown in Figure 1. This is a challenging sequence in several respects. The carpet, in particular, contains strongly structured edge clutter and many regions with similar colour distributions to the head being tracked. The sequence also contains large inter-frame motion when the

(a) Sampling Importance Resampling (SIR)

(b) Auxiliary Particle Filter

(c) Iterated Likelihood Weighting

Fig. 6. The mean ellipse (left) and the most highly weighted particle (right) after 55 frames for each of twenty runs.

person falls over. Figure 6 compares the mean and strongest ellipses obtained after 55 frames in 20 separate runs of the three filters on the sequence. SIR failed in the great majority of runs. In only 3 of the 20 runs did it provide a reasonable estimate in terms of the most heavily weighted particle. The mean did not provide good estimates of the state indicating that the distribution was clearly multimodal due to clutter. APF gave reasonable estimates in 9 of the 20 runs. ILW gave good estimates in terms of both the mean and the most heavily weighted particle in all but one run.

Figure 7 shows trajectories obtained by 20 separate runs of each of the three methods from identical initial conditions. Ground-truth data were acquired by manually fitting an ellipse to the head in each frame. The distances of the estimated ellipse centres from the ground-truth centres were computed for each frame. Figure 8 plots these errors for each filter averaged over 20 runs. At about frame 40 some of the trackers lost lock due to a strong mode in the distribution over the carpet but many subsequently recovered at around frame 50. After frame 55 the person begins to fall over causing many of the trackers to fail.

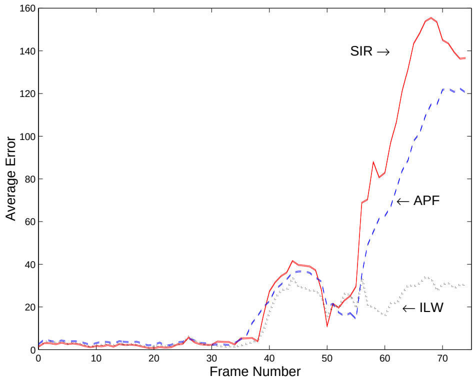


Fig. 8. Distances between estimated and ground-truth ellipse centres averaged over 20 runs of each filter.

8. DISCUSSION

King and Forsyth [9] point out that expectations computed using Condensation have high variance so that different runs of the tracker lead to very different answers. They also comment that "the tracker will appear to be following tight peaks in the posterior even in the absence of any meaningful measurement". The experiments conducted here show that the variance can indeed be high while the approximation accuracy is often poor. The use of an APF improved matters a little. However, ILW (a simply implemented modification to SIR) yielded better accuracy and lower variance. In particular, it was able to successfully track motion that was poorly accounted for by the dynamic model.

It should be stressed that the effect of the ILW algorithm is not the same as that obtained by simply increasing the variance of the motion model (or adopting a model with heavier tails) in the SIR filter. Asymptotically, the implied dynamics in ILW is a K times convolved version of the original dynamic kernel but because the particle set is finite and a sampling step is applied at each iteration of ILW, the outcome is very different.

This paper has compared ILW with SIR and an unbiased filter motivated by exploring regions of high likelihood (APF). Other particle filters have been suggested that share some of the motivations discussed here. It would be interesting to compare these in furture work. Partitioned sampling [14, 15] was proposed for tracking multiple or articulated objects and as such is not appropriate for the application presented here. Layered sampling [17] can reduce the complexity of factored sampling when the likelihood function is narrow. It was developed to address the problem of "overloading" when observations are made at a fine spatial scale. Annealed particle filtering [12] uses an heuristic annealing process to avoid Markov chains becoming trapped in a mode near the starting point. It is useful when the likelihood is very peaked. Rather than propagate

SIR APF ILW

Fig. 7. Trajectories obtained by 20 runs of each of SIR, APF and ILW.

the posterior, it finds the configuration that gives the maximum value using a weighting function. Both annealed and layered sampling were proposed to cope with peaked likelihoods. In contrast, the likelihood function used here was carefully designed to be broad and ILW search was effective. The likelihood function and sampling scheme are intimately related and should be considered in conjunction. ILW, in common with annealed filtering, is partly heuristic in nature. Future work should more fully explore the bias-variance trade-offs made by this and related schemes, shifting the emphasis from asymptotic properties to the more important approximation performance with finite, typically small, sample sets.

9. REFERENCES

[1] H. Ning, L. Wang, W. Hu, and T. Tan, “Articulated model based people tracking using motion models,” in *IEEE International Conference on Multimodal Interfaces*, 2002.

[2] D. Ormoneit, H. Sidenbladh, M. J. Black, and T. Hastie, “Learning and tracking cyclic human motion,” in *Advances in Neural Info. Proc. Sys. (NIPS)*, 2001, vol. 13.

[3] V. Pavlovic, J. M. Rehg, and J. MacCormick, “Learning switching linear models of human motion,” in *Advances in Neural Info. Proc. Sys. (NIPS)*, 2001, vol. 13.

[4] M. Pitt and N. Shephard, “Filtering via simulation: Auxiliary particle filters,” *J. American Statistical Association*, vol. 94, no. 446, pp. 590–599, 1999.

[5] H. Nait-Charif and S. J. McKenna, “Head tracking and action recognition in a smart meeting room,” in *4th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS-ICVS)*, 2003.

[6] N. Gordon, D. Salmond, and A. F. M. Smith, “Novel approach to nonlinear and non-Gaussian Bayesian state estimation,” *IEE Proceedings-F*, vol. 140, pp. 107–113, 1993.

[7] M. Isard and A. Blake, “Contour tracking by stochastic propagation of conditional density,” in *European Conference on Computer Vision*, 1996, vol. 1, pp. 343–356.

[8] M. Isard and A. Blake, “Icondensation: Unifying low-level and high-level tracking in a stochastic framework,” in *European Conference on Computer Vision*, Freiburg, Germany, 1998, vol. I, pp. 893–908.

[9] O. King and D. A. Forsyth, “How does Condensation behave with a finite number of samples?,” in *ECCV*, 2000, pp. 695–709.

[10] J. Carpenter, P. Clifford, and P. Fearnhead, “An improved particle filter for non-linear problems,” *IEE Proceedings on Radar and Sonar Navigation*, vol. 146, pp. 2–7, 1999.

[11] K. Choo and D. J. Fleet, “People tracking using hybrid Monte Carlo filtering,” in *ICCV*, 2001, pp. 321–328.

[12] J. Deutscher, A. Blake, and I. Reid, “Articulated body motion capture by annealed particle filtering,” in *IEEE CVPR*, 2000, vol. 2, pp. 126–133.

[13] J. Deutscher, A. Davison, and I. Reid, “Automatic partitioning of high dimensional search spaces associated with articulated body motion capture,” in *IEEE CVPR*, Hawaii, 2001, vol. 2, pp. 669–676.

[14] J. P. MacCormick and A. Blake, “A probabilistic exclusion principle for tracking multiple objects,” in *ICCV*, Corfu, Greece, 1999, pp. 572–587.

[15] J. P. MacCormick and M. Isard, “Partitioned sampling, articulated objects, and interface-quality hand tracking,” in *ECCV*, 2000, vol. 2, pp. 3–19.

[16] Y. Rui and Y. Chen, “Better proposal distributions: object tracking using unscented particle filter,” in *IEEE CVPR*, 2001, pp. 786–793.

[17] J. Sullivan, A. Blake, M. Isard, and J. MacCormick, “Object localization by Bayesian correlation,” in *ICCV*, Corfu, Greece, 1999, vol. 2, p. 1068.

[18] M. Zlochin and Y. Baram, “The bias-variance dilemma of the Monte Carlo method,” in *Artificial Neural Networks: ICANN*. 2001, Springer Verlag.

[19] K. Nummiaro, E. Koller-Meier, and L. Van Gool, “A color-based particle filter,” in *First International Workshop on Generative-Model-Based Vision*, 2002, pp. 53–60.

[20] S. Birchfield, “Elliptical head tracking using intensity gradients and color histograms,” in *IEEE CVPR*, 1998.