Enhanced Importance Sampling: Unscented Auxiliary Particle Filtering for Visual Tracking

Chunhua Shen, Anton van den Hengel, Anthony Dick, and Michael J. Brooks

School of Computer Science, The University of Adelaide,
Adelaide, SA 5005, Australia
Cooperative Research Centre for Sensor Signal and Information Processing,
Mawson Lakes, SA 5095, Australia
{chhshen, anton, ard, mjb}@cs.adelaide.edu.au

Abstract. The particle filter has attracted considerable attention in visual tracking due to its relaxation of the linear and Gaussian restrictions in the state space model. It is thus more flexible than the Kalman filter. However, the conventional particle filter uses system transition as the proposal distribution, leading to poor sampling efficiency and poor performance in visual tracking. It is not a trivial task to design satisfactory proposal distributions for the particle filter. In this paper, we introduce an improved particle filtering framework into visual tracking, which combines the unscented Kalman filter and the auxiliary particle filter. The efficient unscented auxiliary particle filter (UAPF) uses the unscented transformation to predict one-step ahead likelihood and produces more reasonable proposal distributions, thus reducing the number of particles required and substantially improving the tracking performance. Experiments on real video sequences demonstrate that the UAPF is computationally efficient and outperforms the conventional particle filter and the auxiliary particle filter.

1 Introduction

The particle filter (PF), also known as sequential Monte Carlo or CONDENSATION [1], has been extensively studied for the sequential time series inference due to its relaxation of the linearity and Gaussianity constraints of the Kalman filter (KF). This method represents the posterior distribution of the states with a group of discrete samples/particles. During the filtering process, these particles are updated to maintain the posterior with the importance sampling technique. In theory the PF algorithm can deal with any nonlinearities or distributions.

In the context of computer vision, its applications includes visual tracking [2, 3, 4, 5, 6], robot localisation [7], structure from motion [8] *etc.* Despite its successful application in those cases, it usually requires a considerable amount of discrete particles to effectively approximate the continuous probabilistic distributions; this results in low efficiency and even prohibitive computational burden for high dimensional state space problems. One solution is to design "optimal" proposal sampling distributions for the importance sampling process, generating

G.I. Webb and Xinghuo Yu (Eds.): AI 2004, LNAI 3339, pp. 180–191, 2004.

less particles in the low posterior probability areas, which contribute little to the state estimation.

The conventional PF [2] for visual tracking adopts the dynamic transition prior (the probabilistic model of the states' evolution) as its proposal distribution. When the dynamic motion model can not capture the subject's motion accurately, the motion predicted by the prior deviates from the true trajectory. This occurs frequently when tracking with a conventional particle filter. In such cases, the likelihood distribution is situated in the prior's tail and sampling from such a prior fails to generate sufficient particles in the high posterior areas [9]. Several approaches have been advanced to relocate the particles to the dominant modes in the likelihood or posterior distribution space with a stochastic or deterministic optimisation method [3, 4, 5, 10, 11].

Alternatively, a learned dynamic model instead of a simple predefined autoregressive model yields better sampling [12, 13]. Nevertheless, such approaches are usually only available for specific motions such as cyclic human walking.

Another way is to design a better proposal. As an *ad hoc* approach, in [14] an auxiliary colour tracker is used to generate proposal distributions for the main tracker. The limitation of this method is that the auxiliary tracker might fail and produce false proposal distributions. A more general approach is "optimal filtering" [15]. It has been shown that the "optimal" importance function should take the most recent observations into consideration [15].

An elegant solution to this problem is the auxiliary particle filter (APF) [9] which generates particles from an importance distribution depending on the most recent observations and then sample from the estimated posterior using this importance distribution. APF has been shown to outperform traditional CONDENSATION algorithm in many applications [7, 9]. In [6] Nait-Charif *et al* have compared APF and CONDENSATION in the context of tracking a person in an overhead view, and better performances are observed when using the APF algorithm. But the improvement is very limited. The reason is that the APF algorithm cannot approximate accurately the one-step ahead likelihood which plays an important role in the algorithm. In this paper, we use an unscented transformation (UT) to alleviate this problem (more discussion is presented in Section 2).

The KF can be used to incorporate the current observation into the PF and consequently to improve the distribution proposal. The unscented particle filter (UPF) is a combination of the unscented Kalman filter (UKF) and the generic PF [16] which has proven to be a better solution for particle filtering based visual tracking [17, 18].

In the context of signal processing, Andrieu *et al* combines UT, UKF and APF for a *jump* Markov system and they obtain promising performances in the application of time-varying autoregressions [19]. Our work is motivated mostly by their strategy.

In this paper we introduce the enhanced importance sampling strategy — the UAPF technique, which uses the UKF to approximate the one-step ahead likelihood for the APF. At the same time, the UKF generates the proposal distributions as the UPF does. This step can utilise the results obtained by the previous UT approximation step, hence little extra computation is involved. This approach then yields efficient sampling and substantial improvement over the conventional PF and the APF algorithm. We apply the UAPF to the visual tracking problem. Experiments on real video sequences demonstrate that the UAPF is computationally efficient and outperforms both the conventional PF and APF.

The structure of the paper is as follows. After the introduction in Section 1, we discuss the optimal sampling technique and present the UAPF algorithm in detail in Section 2. In Section 3 we discuss the application to visual tracking. Experiments on real video sequences are also presented in Section 3, followed by concluding remarks in Section 4.

2 The Unscented Auxiliary Particle Filter

To make this paper self-contained, we first briefly review the technique of Monte Carlo Bayesian filtering, which is described in more detail in [1]. We then summarise the proposed auxiliary particle filter algorithm. Essentially it is an implementation of the APF depending on the UKF. This sampling strategy leads to an enhanced efficient importance sampling distribution and an accurate approximation over the conventional APF. Appealingly, this combination does not introduce much extra computation, which is a critically important factor in real-time visual tracking.

2.1 Bayesian Filtering and the Particle Filter

Visual tracking is usually formulated as Bayesian filtering. Given the Markovian dynamic model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$:

$$\mathbf{x}_t = \mathbf{g}(\mathbf{x}_{t-1}, \mathbf{u}_t) \tag{1}$$

and the observation model $p(\mathbf{z}_t|\mathbf{x}_t)$:

$$\mathbf{z}_t = \mathbf{h}(\mathbf{x}_t, \mathbf{v}_t) \tag{2}$$

at time t, the task is to infer the latent state vectors $\mathbf{x}_t$ based on the observation sequences $\mathbf{z}_{1:t}$. In Eqs. (1) (2), $\mathbf{g}(\cdot,\cdot)$ and $\mathbf{h}(\cdot,\cdot)$ are the system dynamics model and observation model, respectively. Usually they are highly nonlinear. The process and measurement noises at time t are given by $\mathbf{u}_t$ and $\mathbf{v}_t$.

The inference is achieved by

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t)p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) \tag{3}$$

where the prior is the previous posterior propagated across the temporal axis,

$$p(\mathbf{x}_t|\mathbf{z}_{1:t-1}) = \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{x}_{t-1}. \tag{4}$$

When the dynamic and observation models are nonlinear and/or non-Gaussian, the above posterior cannot be analytically computed and one has to resort to numerical approximations such as particle filters. In the visual tracking problem, the dynamic model can be approximated by a linear model while the observation model is usually highly nonlinear.

The essential idea of the particle filter is that the posterior is approximated by a series of discrete particles, each of which includes a state vector $\mathbf{x}_t$ and an associated weight w_t: $\{\mathbf{x}_t^{(n)}, w_t^{(n)}\}_{n=1}^N$, where $\sum_{n=1}^N w_t^{(n)} = 1$ holds. The posterior is formulated as $p(\mathbf{x}_t|\mathbf{z}_{1:t}) = \sum_{n=1}^N w_t^{(n)} \delta(\mathbf{x}_t - \mathbf{x}_t^{(n)})$, where $\delta(\cdot)$ is the Dirac function. Then the integral in Eq. (4) is tractable with this numerical approximation.

Suppose we can sample the particles from an importance density $q(\cdot)$, *i.e.* $\mathbf{x}_t^{(n)} \sim q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t}), (n = 1, \ldots, N)$, then each particle's weight is set to

$$w_t^{(n)} \propto \frac{p(\mathbf{z}_t|\mathbf{x}_t^{(n)})p(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t})}. \tag{5}$$

Before or after the importance sampling step, a selective re-sampling step is adopted to ensure the efficiency of the particles' evolution [1].

To summarise, we present the complete algorithm for a conventional PF in Fig. 1.

- *Initialisation*:
  Set $t = 1$. Sample N particles $\{\mathbf{x}_{t-1}^{(n)}, w_{t-1}^{(n)}\}_{n=1}^N$ from the prior $p(x_0)$.
- *Re-sampling*:
  Re-sample to obtain N replacement particles $\{\mathbf{x}_{t-1}^{(n)}, \frac{1}{N}\}_{n=1}^N$, according to the weights $w_{t-1}^{(n)}$.
- *Importance sampling*:
  For $n = 1, \ldots, N$, sample N particles $\mathbf{x}_t^{(n)}$ from the importance proposal $q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t})$, and evaluate the weights according to Eq. (5). Then normalise the weights.
- Set $t = t + 1$, go to the *Re-sampling* step to process the next frame.

Fig. 1. The particle filtering algorithm

The proposal distribution $q(\cdot)$ is critically important for a successful particle filter because it concerns putting the sampling particles in the useful areas where the posterior is significant. It is known that sampling using the dynamic transition model as the proposal distribution is usually inefficient when the likelihood is situated in the prior's tail or it is highly peaked. In visual tracking the dynamical models cannot accurately predict the true motion trajectory due to unexpected motion. In such cases the conventional PF which samples from the dynamic model is quite likely to put most of the particles in the wrong areas of the state space.

2.2 The Auxiliary Particle Filter

The APF is an improved sampling strategy which performs approximately "optimal" sampling and utilises the most recent observations [9]. It is a one-step look-ahead procedure, in which a particle is propagated to the next time frame in order to help the sampling from the posterior. However, the predictive likelihood is only available by an approximation, *i.e.*, the predictive likelihood $p(\mathbf{z}_t|\mathbf{x}_{t-1}^{(n)})$ is approximated by $\widetilde{p}(\mathbf{z}_t|\mathbf{x}_{t-1}^{(n)})$.

In order to put the particles in useful areas of the state space, the APF re-samples the particles $\{\mathbf{x}_{t-1}^{(n)}, \frac{1}{N}\}_{n=1}^{N}$ according to the values $\pi_t^{(n)} = w_{t-1}^{(n)} \cdot \widetilde{p}(\mathbf{z}_t|\mathbf{x}_{t-1}^{(n)})$. Similar to the standard importance sampling, we sample from a proposal distribution $\mathbf{x}_t^{(n)} \sim q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t})$, then the weight associated to each particle should be set to

$$w_t^{(n)} \propto \frac{p(\mathbf{z}_t|\mathbf{x}_t^{(n)})p(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})}{q(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t})\widetilde{p}(\mathbf{z}_t|\mathbf{x}_{t-1}^{(n)})}. \tag{6}$$

In [9] Pitt *et al* use the values likely to be generated by the dynamic model $p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)})$ as the approximation. Although the APF outperforms the conventional PF in many applications [7, 9], as pointed out in [19], this approximation of the predictive likelihood could be very poor and lead to performance even poorer than the standard importance sampling if the dynamic model $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is quite scattered and the likelihood $p(\mathbf{z}_t|\mathbf{x}_t)$ varies significantly over the prior $p(\mathbf{x}_t|\mathbf{x}_{t-1})$. In [6] the authors observe that the APF achieves a little improvement over the stand particle filters in the context of visual tracking.

2.3 The Unscented Transformation and the Unscented Kalman Filter

In this subsection we briefly introduce the UT and UKF. The UKF is the best Kalman filter for nonlinear estimation applications. By including the noise component in the state space, the UKF can be implemented naturally using the UT, which is the basis of UKF and UPF.

Using the UT, the mean and covariance of the Taylor expansion of a nonlinear transformation can be guaranteed to be accurate up to second order. Instead of linearizing using Jacobian matrices, the UT/UKF uses a deterministic sampling strategy to capture the mean and covariance with a small set of carefully selected points named "sigma points" [20]. Therefore, in accuracy, the UT/UKF is better than the extended Kalman filter which approximates the nonlinear transformation with a first-order linearisation. The UKF is also more computationally efficient due to its avoiding the calculation of Jacobian matrices. Note that, unlike the PF, both the EKF and UKF assume unimodal distributions.

Consider the nonlinear tracking problem modelled by the state-space equations (1) and (2), we select $2L+1$ scaled sigma points $\{\widehat{\mathbf{x}}_{t-1}^{(l)}, \widehat{w}_{\mathbf{m},t-1}^{(l)}, \widehat{w}_{\mathbf{P},t-1}^{(l)}\}_{l=0}^{2L}$ at time frame $t-1$, where $\widehat{w}_{\mathbf{m},t-1}^{(l)}$ and $\widehat{w}_{\mathbf{P},t-1}^{(l)}$ are the weights associated with each sigma point[1]. Then the sigma points are propagated through the nonlinear dynamic system eq. (1) ($l = 0, \ldots, 2L$),

$$\widehat{\mathbf{x}}^{(l)}_{t|t-1} = \mathbf{g}(\widehat{\mathbf{x}}^{(l)}_{t-1}, \mathbf{0}). \tag{7}$$

Compute the scaled mean and covariance of $\widehat{\mathbf{x}}^{(l)}_{t|t-1}$

$$\widetilde{\mathbf{x}}_{t|t-1} = \sum_{l=0}^{L} \widehat{w}^{(l)}_{\mathbf{m},t-1} \widehat{\mathbf{x}}^{(l)}_{t|t-1}, \tag{8}$$

$$\widetilde{\mathbf{P}}_{t|t-1} = \sum_{l=0}^{L} \widehat{w}^{(l)}_{\mathbf{P},t-1} (\widehat{\mathbf{x}}^{(l)}_{t|t-1} - \widetilde{\mathbf{x}}_{t|t-1})^{\mathrm{T}} (\widehat{\mathbf{x}}^{(l)}_{t|t-1} - \widetilde{\mathbf{x}}_{t|t-1}). \tag{9}$$

Note that we approximate $\mathbf{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ with the sample mean and covariance of the sigma points set. With the measurement model Eq. (2), we can similarly approximate $\widetilde{\mathbf{z}}_{t|t-1}$ and the covariance by calculating the sample mean and covariance of $\widehat{\mathbf{z}}^{(l)}_{t|t-1} = \mathbf{h}(\widehat{\mathbf{x}}^{(l)}_{t|t-1}, \mathbf{0})$, $l = 0, \ldots, 2L$. In order to implement the UKF, the cross-covariance needs to be calculated in the same way. Finally through *time update* and *measurement update* we can estimate $\mathbf{x}_{t|t}$ and $\mathbf{P}_{t|t}$ with the Kalman filter [20]. With Eq. (8) and the likelihood model the predicted likelihood $\widetilde{p}(\mathbf{z}_t|\mathbf{x}^{(n)}_{t-1})$ can be easily calculated. Moreover, this method also has the advantage of being able to compute the proposal $q(\mathbf{x}^{(n)}_t|\mathbf{x}^{(n)}_{t-1}, \mathbf{z}_{1:t})$ [16, 17].

2.4 The Unscented Auxiliary Particle Filter

Based on the APF and UKF techniques presented above, we use the UT and UKF to compute both the predicted likelihood $\widetilde{p}(\mathbf{z}_t|\mathbf{x}^{(n)}_{t-1})$ and the importance proposal $q(\mathbf{x}^{(n)}_t|\mathbf{x}^{(n)}_{t-1}, \mathbf{z}_{1:t})$. The second step is performed exactly the same as the UPF does. See [16] for more details. This combination yields an enhanced importance sampling strategy—the unscented auxiliary particle filter. For clarity we summarise the unscented auxiliary particle filter algorithm in Fig. 2. Note that a Markov transition kernel with the posterior distribution, such as Metroplis or Gibbs kernel, can be applied to each particle to rejuvenate the trajectory of the particles. This additional step can be used to deal with complex high-dimensional models [16, 22].

3 Visual Tracking with the Unscented Auxiliary Particle Filter

In this section, we describe a visual contour tracking system based on the UAPF technique.

[1] Refer to [21] for how to select the sigma points and calculate the weights.

- *Initialisation*:
  Set $t = 1$. Sample N particles $\{\mathbf{x}_{t-1}^{(n)}, w_{t-1}^{(n)}\}_{n=1}^{N}$ from the prior.
- *Auxiliary re-sampling*:
  Re-sample to obtain N replacement particles $\{\mathbf{x}_{t-1}^{(n)}, \frac{1}{N}\}_{n=1}^{N}$, according to the auxiliary weights $\pi_t^{(n)} = w_{t-1}^{(n)} \cdot \widetilde{p}(\mathbf{z}_t|\mathbf{x}_{t-1}^{(n)})$, with normalisation $\sum_{i=1}^{N} \pi_t^{(n)} = 1$. The predicted likelihood is obtained by unscented transformation.
- *Unscented importance sampling*:
  1. Update particles $\mathbf{x}_t^{(n)}$. Estimate the mean and covariance of the state vector, $\widetilde{\mathbf{x}}_t^{(n)}$ and $\widetilde{\mathbf{P}}_{t|t-1}^{(n)}$, with the unscented Kalman filter.
  2. Sample N particles $\{\mathbf{x}_t^{(n)}, w_t^{(n)}\}_{n=1}^{N}$ from the proposal distribution
  $$q(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t}) = \mathcal{N}(\mathbf{x}_t; \widetilde{\mathbf{x}}_t^{(n)}, \widetilde{\mathbf{P}}_{t|t-1}^{(n)}),$$
  where $\mathcal{N}(\cdot;\cdot,\cdot)$ is the Gaussian distribution.
  3. Compute the particle weights as
  $$w_t^{(n)} = \frac{p(\mathbf{z}_t|\mathbf{x}_t^{(n)})p(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)})}{\widetilde{p}(\mathbf{z}_t|\mathbf{x}_{t-1}^{(n)})q(\mathbf{x}_t^{(n)}|\mathbf{x}_{t-1}^{(n)}, \mathbf{z}_{1:t})}.$$
- Set $t = t + 1$, go to the *Auxiliary re-sampling* step to process the next frame.

Fig. 2. The unscented auxiliary particle filtering algorithm

3.1 The State Space and Dynamics Model

The visual contours we track are elliptic-shaped. Instead of using B-spline representations to model relatively complex shapes [18, 23], we model the shape representations with an ellipse which can be modelled by 5 parameters $\mathbf{x} = \{(O_x, O_y), (R_x, R_y), \theta\}$, which are the centre, axis lengths and orientation angle, respectively, as depicted in Fig. 3.

As most of the motions are full of uncertainties, it is quite difficult to model these motions with an auto-regression (AR) model (or more complex, a mixture of AR models), except for some cyclic motions such as human walking [24]. Even worse, in many scenarios, the camera itself is moving randomly, so we have to consider a combination of the camera's complex movement (pan/tilt/zoom) and the tracked object's motion, it is more unrealistic to learn the dynamic model. Thus it could be problematic to use a predefined first- or second-order AR model to capture the motion of a long video sequence. Consequently to use a motion prior as the sampling proposal could fail to generate particles in the area where the likelihood is significant. However, it is not this paper's intention to explore an elegant motion model. Rather, we assume a constant velocity model for the ellipse centre's motion while a Gaussian random walk for the scale change and the orientation angle's motion, because compared with the global motion of the object, these changes are relatively slower and more difficult to capture.

3.2 The Observation Model and Likelihood Model

The ellipse is centred at (O_x, O_y) and K measurement lines $\varphi^{(k)}$, $(k = 1, \ldots, K)$ are constructed passing through the intersection point $(C_x^{(k)}, C_y^{(k)})$, $(k = 1, \ldots, K)$ and the ellipse centre (See Fig. 3). Given the current state $\mathbf{x}$, the ellipse in the image coordinate is determined. The measurement line $\varphi^{(k)}$ is also determined. By jointly solving the ellipse equation and the measurement line equation, the intersection point $(C_x^{(k)}, C_y^{(k)})$ can be easily obtained. The observation Eq. (2) is written as,

$$\mathbf{z}_t = \mathbf{h}(\mathbf{x}_t, \mathbf{v}_t) = (C_x^{(k)}, C_y^{(k)}) + \mathbf{v}_t, \quad (k = 1, \ldots, K), \tag{10}$$

as stated in Eq. (2), $\mathbf{v}_t$ is the measurement noise. In this paper we assume it is a Gaussian distribution.

Calculating the likelihood is based on this ellipse. As in [18, 23] a Canny edge detector is applied along each measurement line $\varphi^{(k)}$ which has a fixed length L_M. Due to the cluttered background, multiple hypothesis points $\mathbf{z}^{(k)} = \{z_1^{(k)}, \ldots, z_{n_l}^{(k)}\}$[2], may be detected. Among them, at most one point is the true observation belonging to the tracked shape. Under the assumption that (1) the true edge point is normally distributed with zero mean and variance σ^2, (2) the clutter is a Poisson process with density λ, and (3) the density of the clutter features is uniform on the measurement line, the likelihood of the observation at measurement line k is [2, 18],

$$p(\mathbf{z}_t^{(k)}|\mathbf{x}_t) \propto 1 + \frac{1}{\sqrt{2\pi}\sigma h_p \lambda} \sum_{i=1}^{n_l} \exp\left[-\frac{|z_i^{(k)}|^2}{2\sigma^2}\right], \tag{11}$$

where h_p is the prior probability that no true contour edge is detected and $|z_i^{(k)}|$ is the distance of the detected feature point i from the contour. n_l is the number of detected feature points.

Assuming that the feature outputs on a distinct normal line are statistically independent, the overall likelihood for K lines which are roughly even around the ellipse is

$$p(\mathbf{z}_t|\mathbf{x}_t) = \prod_{k=1}^{K} p(\mathbf{z}_t^{(k)}|\mathbf{x}_t). \tag{12}$$

Note that the innovation calculation in [17], which is needed by the measurement update, involves estimation of the mixing weight associated with each detected point; we avoid this estimation by assuming the clutter is uniformly distributed along the measurement line and the detected edges have the same intensities.

---

[2] In the 2D image coordinate, $z_i^{(k)}$ is $(C_x^{(k)}, C_y^{(k)})$. For clarity we omit the time index t.

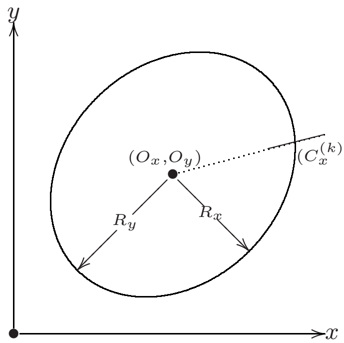

Fig. 3. The state space and the observation model

3.3 Evaluation

We evaluate the UAPF tracking system on two video sequences[3]. The resolution of both of the image sequences is 128×96 and they are sampled at 10 frames per second, both in typical office environments.

In these experiments, all the parameters concerned with the particle filter and the measurement process are the same: $N = 500$ particles, the length of measurement line $L_M = 8$ pixels, $\sigma = 4$ pixels and $K = 25$ measurement rays. We start the particle filter by hand. That means the initial states of the particles are manually tuned and the parameters about the dynamical AR model are obtained by analysing the first several frames.

As pointed out in [6], similar results are observed in our experiments that the auxiliary particle filter could merely trivially improve the tracking performances over the conventional PF, when tracking poorly modelled motion. So in this section we only present the comparison between the results obtained by conventional PF and those by the UAPF.

Fig. 4 shows the tracking results on the face tracking with a cluttered background. We see that at frame 18 the conventional PF is easily distracted by the cluttered background when the the head moves in a direction different from what the dynamical model predicts. For the UAPF technique, the current observation is taken into account, so it can get a relatively accurate predicted likelihood which is utilised to generate better proposal distribution. Consequently the particles are placed effectively in those useful areas with significant posterior. Then the UAPF can track the image sequence successfully.

Fig 5 shows the results on a challenging image sequence. The human face is moving back and forth quite quickly. The tracker is easily confused by the clutters when the prior fails to predict the object's motion. The conventional PF tracker fails and never recovers from frame 7 on. However, with the better proposal

[3] The test image sequences (courtesy of Dr. Birchfield) are available at the URL address: `http://robotics.stanford.edu/~birch/headtracker/seq/`

Fig. 4. Tracking results with the conventional PF (top) and the UAPF (bottom). From left to right, the frame numbers are 3, 18, 31, 40 and 50. From frame 18 on, the motion of the tracked object moves towards the inverse direction of the pre-assumed motion direction which makes the conventional PF lose the target. In contrast, the UAPF can track the whole image sequence successfully

Fig. 5. Tracking results with the conventional PF (top) and the UAPF (bottom). From left to right, the frame numbers are 4, 7, 15, 30. At frame 7, the conventional PF tracker is trapped in a false region when the subject's face changes its motion direction suddenly. The conventional PF tracker fails for most the remaining frames and never be recovered

distributions, the UAPF tracker can track the whole sequence successfully. Please see the tracking video for details[4].

4 Discussion and Conclusion

We have introduced an enhanced importance sampling technique, termed UAPF, for particle filtering in the framework of visual tracking. The UAPF uses the unscented transformation to efficiently predict one-step ahead likelihood and produces more reasonable proposal distributions. We further apply this strategy

[4] The tracking results described in this paper can be accessed at the URL address: `http://www.cs.adelaide.edu.au/~vision/demo/`

to a variety of visual tracking sequences. Experiments on these real-world video sequences show that the UAPF is computationally efficient and outperforms the conventional particle filter and the auxiliary particle filter.

As pointed out in Section 2.4 this method can integrate Markov chain Monte Carlo (MCMC) steps to explore the posterior space, yielding an improved hybrid Monte Carlo filtering [5]. We plan to apply this combination to a high-dimensional problem such as articulated human tracking in which the conventional particle filter is usually deficient [4, 5, 25].

References

[1] Doucet, A., de Freitas, N., Gordon, N., eds.: Sequential Monte Carlo Methods in Practice, New York, Springer-Verlag (2001)

[2] Isard, M., Blake, A.: Condensation – Conditional density propagation for visual tracking. International Journal of Computer Vision **29** (1998) 5–28

[3] Cham, T.J., Rehg, J.M.: A multiple hypothesis approach to figure tracking. In: IEEE Conference on Computer Vision and Pattern Recognition. Volume 2., Fort Collins, Colorado (1999) 239–245

[4] Deutscher, J., Blake, A., Reid, I.: Articulated body motion capture by annealed particle filtering. In: IEEE Conference on Computer Vision and Pattern Recognition. (2000)

[5] Choo, K., Fleet, D.: People tracking using hybrid Monte Carlo filtering. In: IEEE International Conference on Computer Vision. Volume 2., Vancouver, Canada (2001) 321–328

[6] Nait-Charif, H., McKenna, S.J.: Tracking poorly modelled motion using particle filters with iterated likelihood weighting. In: Asian Conference on Computer Vision, Jeju Island, Korea (2004)

[7] Vlassis, N., Terwijn, B., Krose, B.: Auxiliary particle filter robot localization from high-dimensional sensor observations. In: IEEE International Conference on Robotics and Automation. Volume 1. (2002) 7–12

[8] Qian, G., Chellappa, R.: Structure from motion using sequential Monte Carlo methods. International Journal of Computer Vision **59** (2004) 5–31

[9] Pitt, M.K., Shephard, N.: Filtering via simulation: Auxiliary particle filter. Journal of the American Statistical Association **94** (1999) 590–599

[10] Chang, C., Ansari, R.: Kernel particle filter: Iterative sampling for efficient visual tracking. In: IEEE International Conference on Image Processing, Barcelona, Spain (2003)

[11] Shan, C., Wei, Y., Tan, T., Ojardias, F.: Real time hand tracking by combining particle filtering and mean shift. In: 6th International Conference on Automatic Face and Gesture Recognition, Seoul, Korea (2004)

[12] Blake, A., North, B., Isard, M.: Learning multi-class dynamics. In: Advances in Neural Information Processing Systems, The MIT Press (1999) 389–395

[13] Pavlovic, V., Rehg, J., MacCormick, J.: Learning switching linear models of human motion. In: Advances in Neural Information Processing Systems, The MIT Press (2000) 981–987

[14] Isard, M., Blake, A.: ICONDENSATION: Unifying low-level and high-level tracking in a stochastic framework. In: 5th European Conference Computer Vision. Volume 1. (1998) 893–908
[15] Doucet, A., Godsill, S., Andrieu, C.: On sequential sampling Monte Carlo sampling methods for Bayesian filtering. Statistics and Computing **10** (2000) 197–208
[16] van der Merwe, R., Doucet, A., de Freitas, N., Wan, E.: The unscented particle filter. Technical Report CUED/F-INFENG/TR 380, Cambridge University Department of Engineering, UK (2000) available at http://cslu.cse.ogi.edu/nsel/research/ukf.html.
[17] Rui, Y., Chen, Y.: Better proposal distributions: Object tracking using unscented particle filter. In: IEEE Conference on Computer Vision and Pattern Recognition. Volume 2., Kauai, Hawaii (2001) 786–793
[18] Li, P., Zhang, T., Pece, A.E.C.: Visual contour tracking based on particle filters. Image and Vision Computing **21** (2003) 111–123
[19] Andrieu, C., Davy, M., Doucet, A.: Efficient particle filtering for jump Markov systems. application to time-varying autoregressions. IEEE Transactions on Signal Processing **51** (2003) 1762–1770
[20] Wan, E.A., van der Merwe, R.: The unscented Kalman filter. In Haykin, S., ed.: Kalman Filtering and Neural Networks. Wiley Publishing (2001)
[21] Julier, S.J., Uhlmann, J.K.: The scaled unscented transformation. In: IEEE American Control Conference, Anchorage AK, USA, IEEE (2002) 4555–4559
[22] Andrieu, C., de Freitas, N., Doucet, A., Jordan, M.I.: An introduction to MCMC for machine learning. Machine Learning **50** (2003) 5–43
[23] Blake, A., Isard, M.: Active Contours. Spinger, Berlin (1998)
[24] Agarwal, A., Triggs, B.: Tracking articulated motion using a mixture of autoregressive models. In: 8th European Conference on Computer Vision. Volume 3., Prague, Czech Republic (2004)
[25] Deutscher, J., Davison, A.J., Reid, I.: Automatic partitioning of high dimensional search spaces associated with articulated body motion capture. In: IEEE Conference on Computer Vision and Pattern Recognition. (2001)