# Hard / soft switching particle filters for efficient real-time visual tracking

Takashi Bando[1,4] Tomohiro Shibata[1,2,4] Kenji Doya[2,3,4,1] Shin Ishii[1,4]

[1] Nara Institute of Science and Technology
[2] ATR Computational Neuroscience Laboratories
[3] Initial Research Project, Okinawa Institute of Science and Technology
[4] CREST, Japan Science and Technology Agency
takash-b@is.naist.jp

## Abstract

Particle filtering is an approach to Bayesian estimation of intractable posterior distributions from time series signals with non-Gaussian noises. Several particle filters employing different sampling methods have been proposed for approximating Bayesian computation with finite particles. Our previous work first studied the difference between two filters, Condensation and Auxiliary Particle Filter. They can be considered compensatory in terms of accuracy and robustness under severe circumstances having unknown occluders and/or distracters. We then proposed a new particle filtering scheme, called switching scheme, which allows robust and accurate visual tracking under severe circumstances. This scheme utilizes the two complementary sampling algorithms above by switching them based on the confidence of filtered state of the visual target. This article presents an alternative switching method called soft switching while we call the previously proposed method hard switching. The soft switching method softly changes the population ratio of each sampling method for assigning particles. We examine their properties in terms of accuracy and robustness via real visual tracking experiments as well as computer simulations.

*keywords* – particle filter, sampling, real time, visual tracking, switching

## 1 Introduction

Particle filtering is an approach to Bayesian estimation of intractable posterior distributions from time series signals with non-Gaussian noise. This approach has been attracting attention in various research areas, including real-time visual processing which deals with images contaminated by non-Gaussian noises due to not only signal noises but also the existence of obstacles and/or distracters. Several particle filters have been proposed for approximating Bayesian computation with finite particles. The performances of such algorithms have not, however, been fully evaluated under circumstances specific to real-time vision systems, i.e., there are unknown occluders and distracters. It is important for real-time visual tracking systems to realize high accuracy and robustness, coping with these difficulties.

Our previous work studied the difference between two filters, Condensation [1] and Auxiliary Particle Filter [2] (APF). Condensation employs the Sampling Importance Resampling (SIR) method, and can run on a cheap standard PC. Condensation, however, has a problem called an outlier problem, i.e., large difference between prior and true distribution causes crude approximation of the posterior distribution. Unfortunately the outlier problem often occurs in a real-time visual tracking task due to unknown occlusions, distracters and target dynamics. APF was proposed to solve the outlier problem using the information of current observation. These two filters can be considered compensatory in terms of accuracy and robustness under the circumstances specific to real-time vision systems. To exploit their advantages, we proposed a new filtering scheme that switches these filters according to a simple criterion [3].

This article presents an alternative switching method called soft switching while we call the previously proposed method hard switching. The soft switching method softly changes the population ratio of each sampling method for assigning particles. We, then, examine their properties of accuracy and robustness via computer simulations that model realistic circumstances include occlusion and distracters. Next, the effectiveness of our methods are demonstrated by real visual tracking experiments. In our tasks, the tracking target is a red ball, and the ball moves sinusoidally behind a board as an occluder. Because there are in addition many objects similar to the tracking target in the background as distracters, they are realistic visual tracking tasks.

## 2 Particle Filters

Particle filters are based upon point-mass representation of probability densities. A continuous state vector of a target object at time step $t$ is denoted by $\mathbf{x}_t \in \mathcal{R}^{N_x}$, and a measurement vector is $\mathbf{z}_t \in \mathcal{R}^{N_z}$. Using a sample set $\{(\mathbf{x}_t^{(n)}, \pi_t^{(n)}), n = 1, ..., N\}$ at time step $t$, the posterior density is approximated as

$$p(\mathbf{x}_t|\mathbf{z}_t) = k_t p(\mathbf{z}_t|\mathbf{x}_t) p(\mathbf{x}_t|\mathbf{z}_{t-1}), \tag{1}$$

where $k_t$ is the normalization term. Then, prediction density $p(\mathbf{x}_t|\mathbf{z}_{t-1})$ as prior is approximated as

$$p(\mathbf{x}_t|\mathbf{z}_{t-1}) \approx \sum_n \pi_{t-1}^{(n)} p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)}). \tag{2}$$

The weights $\pi_t^{(n)}$ are determined by $\pi_t^{(n)} = p(\mathbf{z}_t|\mathbf{x}_t^{(n)})$.

If we can prepare sufficient large number of particles, Eq.2 becomes accurate. However, we cannot prepare such a large number of particles for real-time processing.

In Condensation, $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ used as the proposal density is independent of observation $\mathbf{z}_t$ at each time step $t$, and the state space is explored regardless of $\mathbf{z}_t$. This property suffers from the outlier problem [2], i.e., model-implausible observations may occur when there are unexpected occluders, distracters, and changes in the target motion.

Auxiliary Particle Filter [2] proposed by Pitt and Shepherd employs elegant resampling method that solves the outlier problem. In their approach, likelihoods are calculated as weights at any likely point that characterizes $p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(n)})$, e.g., mean or mode. Then, by resampling with the weights that include information of current observation $p(\mathbf{z}_t|\mathbf{x}_t)$, the estimation around the likely points tends to be more accurate in APF than in Condensation. The diversity of particles, however, is lower in APF than in Condensation. For particle filters, we need to prepare an observation model, or likelihood function, of the target to be tracked. In this article, we employ the same observation model as that in [3].

## 3 Hard / Soft Switching Particle Filters

As described in the last section, Condensation emphasizes prediction of target transition (prior) more than APF, while APF emphasizes observation more than Condensation for estimating current target state. Emphasis on observation makes tracking accurate, while emphasis on prior would make tracking robust especially under the severe circumstances having unknown occluders and distracters. To exploit their advantages, we previously presented switching sampling scheme based on the confidence of filtered state of the visual target [3]. Our proposed method fully switched the two sampling algorithms according to a simple criterion, the variance of current estimated target states as confidence in the estimation. We call it hard switching method. The variance of the current estimated target states is defined as

$$\begin{aligned} V_{\text{est}}(t) =& (\sigma_{\text{est},1}^2, ..., \sigma_{\text{est},N_x}^2)^T & (3) \\ =& \frac{\sum_n \pi_t^{(n)} (\hat{\mathbf{x}}_t - \mathbf{x}_t^{(n)})(\hat{\mathbf{x}}_t - \mathbf{x}_t^{(n)})^T}{N-1}, & (4) \end{aligned}$$

where, $\hat{\mathbf{x}}_t = \sum_n \pi_t^{(n)} \mathbf{x}_t^{(n)}$. Then, using threshold vector $\gamma$, the hard switching method is described as

if $\sigma_{\text{est},i} > \gamma_i \; (\exists i)$, then use APF,
otherwise, use Condensation.

The vector $\gamma$ is a thresholding parameter that discriminates successful tracking from unsuccessful tracking based on the confidence of the current estimated target states. We set $\gamma$ at the variance of estimated target states during successful tracking. For example, in our experiments described later, $\gamma$ was set as follows. We assume that the ball size on the image plane is invariant and $x_i (i = 1, 2)$ are independent of each other for simplicity. Under such assumption, the covariance components of the current estimated target state are expected to be 0, and all we have to do is to compare the threshold vector $\gamma_i (i = 1, 2)$ to the standard deviation of the current estimated target state $\sqrt{V_{\text{est}}(t)}$. If we cannot assume the component-wise independence of state vector, we need to extend the variance $V_{\text{est}}(t)$ to the variance-covariance matrix. In our experiments we employ the target ball radius as threshold $\gamma$ because we found the standard deviation of the current estimated target states during successful tracking were almost equal to the ball radius.

In this article, we propose an alternative switching method called soft switching. Although the soft switching method is the same as the hard switching method except that it controls the population ratio of each sampling method for assigning particles, based on confidence level of estimation. The number of particles sampled by each sampling method, $N_{\text{cond}}$ and $N_{\text{apf}}$, are determined as

$$N_{\text{cond}} = N - N_{\text{apf}}, \; N_{\text{apf}} = \max_i N_{\text{apf},i},$$

$$\begin{cases} N_{\text{apf},i} = 0 & (\sigma_{\text{est},i} < \gamma_{\min,i}) \\ N_{\text{apf},i} = \frac{N(\sigma_{\text{est},i} - \gamma_{\min,i})}{\gamma_{\max,i} - \gamma_{\min,i}} & (\gamma_{\min,i} \le \sigma_{\text{est},i} \le \gamma_{\max,i}) \\ N_{\text{apf},i} = N & (\sigma_{\text{est},i} > \gamma_{\max,i}) \end{cases},$$

meaning that the more confident of tracking the tracker has, the larger number of particles are sampled by Condensation, while the less confident of tracking it has, the larger number of the particles are sampled by APF.

The threshold vectors in the soft switching method are also determined using the target ball radius. We set $\gamma_{\min}$ at smaller than the ball radius, and $\gamma_{\max}$ at larger than the ball radius. We will examine the performance of the soft switching method with various combinations of $\gamma_{\min}$ and $\gamma_{\max}$ in the next section.

## 4 Simulation Results

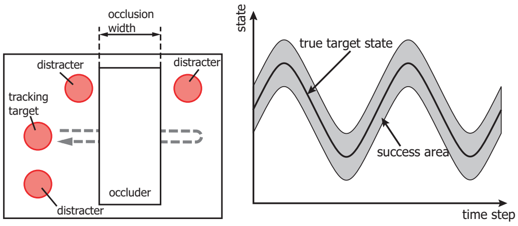


Figure 1: The setting of the computer simulations (left) and the success area defined encompassing the true target state (right).

To investigate the performance under realistic circumstances, we prepared simulated image sequences in which occlusion width, distracter size and positions varied (cf. Figure 1 (left)). The radius of the target was 5 pixels, and the target moved sinusoidally with amplitude 200 pixels and frequency 1 Hz. The occlusion width varied from 5 to 100 pixels with step size 5 pixels. For each image sequence, three distracters were randomly chosen from six candidate positions, and their radius were varied from 1 to 20 pixels. The length of each image sequence was set to 60 frames.

First, the hard switching method ($\gamma = 5$) and the soft switching method ($\gamma_{\min} = 4, \gamma_{\max} = 20$) were tested for the simulation image sequences. We examined robustness of the two methods in terms of the tracking success rate, where "success" means the number of failure frames in which estimated target state deviated from the success area (cf. Figure 1 (right)) was smaller than 15 frames. Their accuracies were also examined based on the mean tracking error in successful trials.

To examine difference in the success rate, we used the $\chi^2$ test. Figure 2 (left) shows results of the $\chi^2$ test ($p < 0.01$), indicating the hard switching method is more robust against unknown occluder and distracters except for too severe conditions, i.e., large occlusion and distracters, or very simple conditions, i.e., small occlusion and distracters.

To examine difference in the tracking error, we applied the Wilcoxon test because the tracking errors and the number of successful trials were too small for the t-test under the severe condition with low success rate. Figure 2 (right) shows results of the Wilcoxon test ($p < 0.05$), indicating the hard switching method

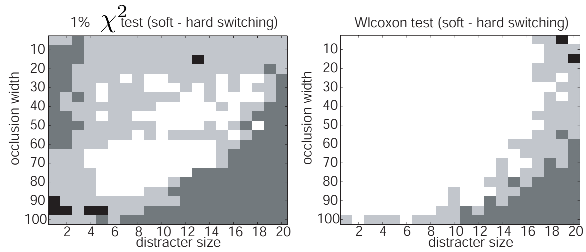


Figure 2: Result of the $\chi^2$ test (left) and the Wilcoxon test (right). White, black, gray, dark gray colors mean that the hard switching method was better significantly, the soft switching method was better significantly, there was no significant difference between their performances, and neither the $\chi^2$ test nor the Wilcoxon test was not applied because of the too small number of successful trials, respectively.

maintain its accuracy of tracking under the broad conditions better than the soft switching method.

In order to use switching methods, switching threshold $\gamma$ must be set beforehand. Although we found that the target ball radius was enough in our experiments, the switching methods depends on this parameter, which is more salient for the hard switching method. An inappropriate parameter causes failure tracking because the tracker misses timing of switching and then the balance of divergence and convergence of particles is broken. As for the soft switching method, setting $\gamma_{\min}$ at smaller than the ball radius, and $\gamma_{\max}$ at larger than the ball radius adding a margin to the timing of switching leads to robust tracking. The soft switching method becomes equivalent to the hard switching method in terms of the number of particles sampled by each sampling method if $\gamma_{\min} = \gamma_{\max}$. When sufficient resource cannot be allocated as in real-time visual tracking, the hard switching method is more efficient than the soft switching method if the precise switching threshold can be predetermined.

## 5 Real Experiments

In this section, we compare the performance of the hard and soft switching methods with Condensation, APF and Unscented Particle Filter (UPF) [4]. Similar to APF, UPF proposed by Merwe et al. emphasizes the information of current observation. Its proposal distribution is estimated by Unscented Kalman Filter for maintaining accurate estimation of target states. Because we cannot calculate $\mathbf{x}_t$ from $\mathbf{z}_t$, we use weighted mean by redness at each pixel as an observation in UPF. We evaluate the effectiveness by comparing the number of lost-target frames and mean error in successful trials.

### 5.1 Experimental Setup

The task was to track a red ball moved sinusoidally either by hand or by a robot manipulator. Real video images were taken by a digital video camera and processed by a Pentium 4 (2 GHz) Linux PC. These images were downloaded to the PC by Video for Linux at a resolution of 640 $\times$ 480 pixels, then reduced to 320 $\times$ 240. We also need to determine the number of particles. From simulation results using artificial images, we set the number of particles to 800 to maintain the accuracy and real-time quality. Our system processed one frame in about 8 msec.

### 5.2 Tracking with an Occluder and distracters

To statistically compare the performance of different particle filters under usual circumstances for vision systems, we prepared an environment in which we placed a board as an occluder between a red ball and a camera, cf. Figure 3. There were also some distracters, i.e., red objects around the trajectory and the board.

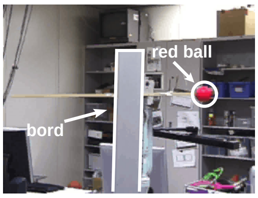


Figure 3: A sample image in the our task.

Figure 4 (left) shows histograms of "lost" frames (length of video sequence is 423 frames for 14 sec). "Lost" means that the estimated target state deviated from the success area defined around the true target state (cf. Figure 1 (right)). Figure 4 (right) shows the mean absolute position error of each tracker in successful trials. A successful trial means that the number of lost frames was less than 100.

These figures show that our switching scheme outperforms the other filters including UPF. Because APF takes particular note of the current observation, it sometimes failed in tracking the target due to distracters, especially as it crossed behind the board. Condensation was more robust for crossing the board, but subsequent convergence to the target was very slow and not promising. The UPF was more robust and accurate than those two filters, but its accuracy was much worse than that of the switching methods. Because the current target state cannot be observed directly from the image (only the redness at each pixel is observed), we choose the weighted mean of pixels using their redness as the observed state in UPF. In spite of inaccurate observation, UPF worked well. In a severe circumstance with occlusion and distracters, however, UPF was not able to keep accurate tracking.

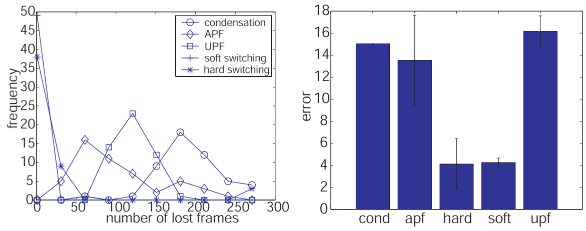


Figure 4: Lost frames (left) and mean absolute position errors (right).

## 6 Summary

We have examined new particle filtering methods, hard and soft switching particle filters, devised for duality with real-time vision systems. These filters employ two particle filters, Condensation and Auxiliary Particle Filter (APF), which incorporate different resampling algorithms. To exploit their advantages, our proposed filters switch them dynamically according to a simple criterion: the confidence level of the current estimated target state. The two methods, hard switching and soft switching, were examined in their properties in terms of accuracy and robustness via computer simulations that model realistic circumstances include occlusion and distracters. The simulation results showed that the soft switching method is more tolerant to the switching threshold, although the hard switching method is more efficient if precise switching threshold is pre-determined, especially when sufficient resource cannot be allocated as in real-time visual tracking. We also demonstrated that the switching methods outperform other well-known particle filters through real visual tracking experiments.

Automatic determination of the threshold parameter can be regarded as learning of a hyper-parameter, which will be our future work.

## References

[1] M. Isard and A. Blake, "CONDENSATION-conditional density propagation for visual tracking," *IJCV*,Vol. 29, pp. 5-28, 1998.

[2] M. k. Pitt and N. Shephard, "Filtering via Simulation: Auxiliary Particle Filter," *J. of American Statistical Association*, Vol. 99, pp. 590-599, 1999.

[3] T. Bando, T. Shibata, K. Doya, and S. Ishii, "Switching Particle Filters for Efficient Real-time Visual Tracking," *ICPR*, Vol. 2, pp.720-723, 2004.

[4] R. v d Merwe, A. Doucet, N. de Freitas, E, Wan, "The Unscented Particle Filter," *NIPS 13*, 2001.