

# Abrupt Motion Tracking Via Intensively Adaptive Markov-Chain Monte Carlo Sampling

Xiuzhuang Zhou, *Member, IEEE*, Yao Lu, Jiwen Lu, *Member, IEEE*, and Jie Zhou, *Senior Member, IEEE*

**Abstract**—The robust tracking of abrupt motion is a challenging task in computer vision due to its large motion uncertainty. While various particle filters and conventional Markov-chain Monte Carlo (MCMC) methods have been proposed for visual tracking, these methods often suffer from the well-known local-trap problem or from poor convergence rate. In this paper, we propose a novel sampling-based tracking scheme for the abrupt motion problem in the Bayesian filtering framework. To effectively handle the local-trap problem, we first introduce the stochastic approximation Monte Carlo (SAMC) sampling method into the Bayesian filter tracking framework, in which the filtering distribution is adaptively estimated as the sampling proceeds, and thus, a good approximation to the target distribution is achieved. In addition, we propose a new MCMC sampler with intensive adaptation to further improve the sampling efficiency, which combines a density-grid-based predictive model with the SAMC sampling, to give a proposal adaptation scheme. The proposed method is effective and computationally efficient in addressing the abrupt motion problem. We compare our approach with several alternative tracking algorithms, and extensive experimental results are presented to demonstrate the effectiveness and the efficiency of the proposed method in dealing with various types of abrupt motions.

**Index Terms**—Abrupt motion, intensive adaptation, Markov-chain Monte Carlo (MCMC), stochastic approximation, visual tracking.

## I. INTRODUCTION

VISUAL tracking in dynamic scenarios refers to establishing the correspondences of the object of interest between the successive frames. It is a fundamental research topic in video analysis and has a variety of potential applications, including teleconferencing, gesture recognition, visual surveillance, and motility analysis.

As an important topic in the computer vision community, visual tracking has been extensively studied in the past decades, and a number of approaches have been proposed in the literature. In general, these tracking approaches can be divided into

two categories [38], i.e., deterministic ones [11], [12] and sampling-based ones [1]–[3], [7]–[10]. The deterministic ones are a class of successful tracking methods that have been popular over the years due to their fast convergence speed and relatively lower computational cost. However, they are prone to getting trapped in local modes in case of background clutter, distractions, or rapid moving object, as discussed in [2].

By maintaining multiple hypotheses, the sampling-based tracking methods are able to deal with the multimodal distribution and recover from the tracking failure, among which the particle filter (PF), also known as the sequential Monte Carlo [4], [5], is the most popular one and is first utilized for visual tracking in [1]. The basic idea of the PF is to use a set of weighted particles to approximate the true filtering distribution. The PF has achieved considerable success in visual tracking and becomes a widely used framework that is highly extensible and offers the flexibility to handle nonlinearity and non-normality in the object models.

In real world, many tracking tasks suffer from the multimodal likelihood and posterior, high-dimensionality and inaccurate local evidence. To facilitate efficient tracking, in general, most existing approaches are based on a smooth motion assumption or an accurate motion model. However, abrupt motions are common in real-world scenarios, such as fast motion, camera switching, low-frame-rate videos, and unexpected object dynamic.

It is challenging for tracking methods, both deterministic ones and sampling-based ones, to deal with the large motion uncertainty induced by abrupt motions. Intuitively, a direct solution for the sampling-based tracking methods is to enlarge the sampling variance to cover the possible motion uncertainty. Nevertheless, there exists a problematic issue to be addressed, i.e., the sampling inefficiency. This is because the increase in the sampling volume may require a more expensive computational cost, particularly for the systems with the high-dimensional state space.

The success of the PF highly relies on its ability to maintain a good approximation to the posterior distribution. For complex filtering distribution with rugged energy landscape, a large number of particles are required to guarantee sufficient sampling in the broad state space. The high computational burden caused by a large number of particles often makes the PF infeasible for practical applications. In the recent years, many extensions [6] to PF have been made to reduce the computational cost and improve the sampling efficiency. Among them, Markov-chain Monte Carlo (MCMC) methods have received much attention in visual tracking. In [7], a simulated annealing process is incorporated into the conventional PF [1], which allows for the generation of the samples closer to the true modes of the pos-

Manuscript received November 11, 2010; revised April 10, 2011; accepted July 01, 2011. Date of publication September 19, 2011; date of current version January 18, 2012. This work was supported in part by the National Natural Science Foundation of China under Grant 11178017, by Beijing Natural Science Foundation under Grant 4112050, and by Beijing Nova Program under Grant 2008B57. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Jenq-Neng Hwang.

X. Zhou is with the College of Information Engineering, Capital Normal University, Beijing 100048, China (e-mail: zxz@xeehoo.com).

Y. Lu is with the School of Computer Science and Technology, Beijing Institute of Technology, Beijing 100081, China (e-mail: vis\_y\_l@bit.edu.cn).

J. Lu is with the Advanced Digital Sciences Center, Singapore 138632 (e-mail: jiwen.lu@adsc.com.sg).

J. Zhou is with the Department of Automation, Tsinghua University, Beijing, 100084, China (e-mail: jzhou@tsinghua.edu.cn).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TIP.2011.2168414

terior distribution and avoids the problem of getting trapped in local modes of the high-dimensional sample space for articulated body tracking. In [8]–[10], the widely used importance sampling (IS) is replaced with MCMC to sample from the posterior distribution. The benefit for this replacement lies in the fact that, by using a well-designed proposal distribution, MCMC is capable of generating posterior samples that converge to the posterior mode in a much efficient manner.

Although popular and commonly used, the plain Gibbs sampler or Metropolis–Hastings (MH) algorithm often gets trapped in the local optimum when the energy landscape of the target distribution is rugged. The inability of the simulated Markov chain to move around the energy landscape often leads to an inaccurate Bayesian inference. To overcome the local-trapped problem, many advanced MCMC algorithms have been recently developed in physics and statistics [28]. Among them, the adaptive MCMC algorithms [21]–[25] have shown more superiority in improving the mixing and acceptance rates, even if much research is still expected in this exciting area. In principle, an adaptive MCMC algorithm aims to simulate a good chain, the distribution of which is closer to the target distribution by using the historical samples, and thus reduces the variance of the estimate of interest.

In this paper, a novel sampling-based tracking scheme is proposed to effectively deal with the abrupt motion difficulty in the Bayesian filtering framework. Rather than simply adopt the sequential importance resampling (SIR) [1]–[3] or standard MCMC sampling algorithm, which is common for the state-of-the-art tracking methods [7]–[10], we propose a more effective dynamic IS scheme to sample from the filtering distribution by using the stochastic approximation Monte Carlo (SAMC) algorithm [21] and present a sequential SAMC sampling algorithm for the tracking of abrupt motion, which demonstrates superiority in dealing with the local-trap problem with less computational burden.

For the proposed sequential SAMC tracking method, however, to guarantee its robustness, a certain number of samples are still required to capture the abrupt motion due to the broadness of the whole state space. We propose a more effective sampling algorithm to further reduce the computational cost. We achieve this by introducing a density-grid-based predictive model, which carries the statistical information about the filter distribution, to predict the promising regions of the state space in sampling. Based on the predicted result, a more informative proposal is learned on the fly, which helps to bias the sampling toward the promising regions of the state space to improve the sampling efficiency.

The remainder of this paper is organized as follows: In Section II, we review some related tracking algorithms. The proposed sequential SAMC sampling algorithm is described in Section III. In Section IV, we elaborate the intensively adaptive MCMC (IA-MCMC) sampler for abrupt motion tracking. Implementation details and experimental results are presented in Section V, and Section VI concludes this paper. An earlier version of this paper appeared in [42].

## II. RELATED WORK

There is a rich literature in visual tracking [38]. Here, we review only the most relevant tracking approaches, focusing

on algorithms that directly aim to deal with the abrupt motion difficulty.

The simplest solution to the abrupt motion problem is searching the whole state space to fully cover the motion uncertainty. In practice, however, it is infeasible due to the large search space of the object state, which often results in an extremely expensive computational cost. Indeed, an accurate dynamic model can be used to estimate the search space based on the object state prediction. However, accurate dynamic models are often learned from the specific training data [13], [14], and this often makes the tracking approaches less flexible.

Multiscale and hierarchical sampling strategy is another solution to this problem. This tracking approach aims at alleviating the effect by the abrupt motion to improve the sampling efficiency in the solution space. In the Bayesian context, Sullivan *et al.* [15] combined observations from multiple scales to facilitate efficient searching in the fine scale. A potential problem of this method is that inaccurate inference in the large scale may cause the failure of searching in a fine scale. To overcome this problem of potential error propagation, Hua and Wu [16] proposed to design a collaborative searching scheme based on the dynamic Markov network and developed a sequential-belief-propagation scheme for more accurate Bayesian inference. Recently, Li *et al.* [17] proposed a cascade PF to deal with the large motion uncertainty of the target object. While this approach has shown high efficiency in several face tracking cases, it requires several reliable observation models and an additive offline learning process.

In contrast with the approaches based on the specific dynamic model [13], [14], we propose to cope with the abrupt motion based on an adaptive sampling strategy without using any prior knowledge about the object motion. Moreover, different from the multiscale layer sampling or cascade filtering approaches [15]–[17] that may still suffer from the local-trap problem in sampling from the complex target distribution with deep local modes, we explicitly address the abrupt motion tracking in the Bayesian filter framework and aim to propose an efficient sampling scheme to essentially avoid the local-trap problem in sampling. Approaches that are the most similar in sampling mechanism to ours may be the Wang–Landau (WL) based tracking algorithm [18]. In [18], Kwon and Lee proposed a novel tracking approach to effectively deal with the local-trap problem in tracking by using a well-designed annealing WL sampling scheme. Note that the WL algorithm [20] itself is a MCMC method with target distribution adaptation, which solves both the weight estimation and the sampling problem in a single run. Compared with [18], the algorithm proposed in this paper is more principled and adaptive. First, the proposed IA-MCMC algorithm is a dynamic IS method in the Bayesian filtering framework, in which the target distribution is adaptively estimated from iteration to iteration in a more efficient way by introducing a stochastic approximation process (SAP) [21] into the learning process of the working weight, and the filtering inference based on this sampling scheme is supported by the rigorous theory in statistics literatures [27], [35], [36]. Second, our sampling method is able to not only avoid the local-trap problem by target distribution adaptation but also improve the overall sampling efficiency by learning the proposal distribution on the fly.

### III. SEQUENTIAL SAMC SAMPLING

Here, we present details of the proposed sequential SAMC sampling algorithm for the tracking of abrupt motion. We first make a brief review on the Bayesian filtering framework. Then, the basic idea behind dealing with the local-trap problem suffered by conventional sampling algorithms is discussed.

#### A. Sequential Bayesian Filtering

Visual tracking can be formulated as an inference task in a Markov model with hidden state variables, and it is often addressed in the Bayesian context. Let  $\mathbf{x}_t$  denote the object state at time  $t$ . Given a series of observations  $\mathbf{z}_{1:t} = \{\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_t\}$  up to time  $t$ , our aim is to estimate the hidden state variable  $\mathbf{x}_t$ . According to Bayes' rule, the filtering distribution  $p(\mathbf{x}_t|\mathbf{z}_{1:t})$  can be recursively estimated by

$$p(\mathbf{x}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1})p(\mathbf{x}_{t-1}|\mathbf{z}_{1:t-1})d\mathbf{x}_{t-1} \quad (1)$$

where  $p(\mathbf{x}_t|\mathbf{x}_{t-1})$  is the prior model describing the temporal evolution of the state variable and  $p(\mathbf{z}_t|\mathbf{x}_t)$  is the observation likelihood. For nonlinear and non-normal systems, integral (1) is often analytically intractable.

The PF is one of the most popular methods to address this filtering problem. It approximates the filtering distribution (1) with a set of weighted particles at each time step. While the PF and its variants have achieved considerable success in visual tracking, there exists the well-known sample impoverishment problem, due to their suboptimal sampling nature in the IS process, particularly for systems with high-dimensional state space. To improve the sampling efficiency, as discussed in Section I, many MCMC methods have been introduced to address this problem. For complex target distribution with rugged energy landscape, however, a more effective and efficient sampling algorithm is still desired for practical tracking systems.

#### B. Basic Idea of Our Approach

To achieve a good approximation to the filtering distribution, traditional MCMC algorithms can be used to simulate a Markov chain that converges to a stationary distribution, i.e., filtering distribution. MH is a commonly used MCMC algorithm, and many practical MCMC samplers can be regarded as the extensions of this sampler. Given the target distribution  $p(\mathbf{x}_t|\mathbf{z}_{1:t})$ , an MH sampling step includes drawing a candidate sample  $\mathbf{x}'$  based on the current sample  $\mathbf{x}_t$ , by using a proposal  $Q(\cdot; \mathbf{x}_t)$ . With these notations, a Markov chain grows and moves to  $\mathbf{x}'$  with the acceptance probability  $\alpha(\mathbf{x}'; \mathbf{x}_t)$ , i.e.,

$$\alpha(\mathbf{x}'; \mathbf{x}_t) = \min \left\{ 1, \frac{p(\mathbf{x}'|\mathbf{z}_{1:t})Q(\mathbf{x}_t; \mathbf{x}')}{p(\mathbf{x}_t|\mathbf{z}_{1:t})Q(\mathbf{x}'; \mathbf{x}_t)} \right\} \quad (2)$$

otherwise remains at  $\mathbf{x}_t$ . From (2), we can observe that the local-trap problem may occur when the sample drawn from the proposal distribution  $Q(\cdot; \mathbf{x}_t)$  lies in the tail of the filtering distribution (1) or the filtering distribution is peaked.

For abrupt motion tracking, the sample space of the filtering distribution may be much broader than that of the smooth tracking, and thus, its energy landscape may be rugged. From the perspective of energy space, if one sampler can make a random walk in each energy subregion of the filtering distribution, then the local-trap problem will be essentially overcome.

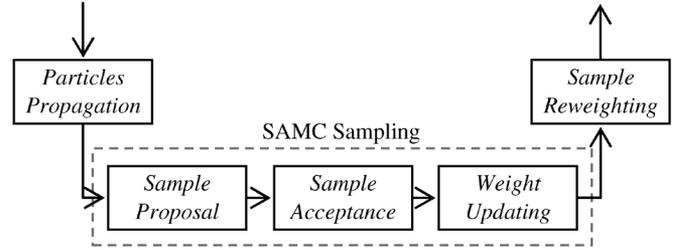


Fig. 1. Overview of the sequential SAMC sampling algorithm.

Let  $\mathcal{X}$  denote the state space of the filtering distribution (1). Suppose that  $\mathcal{X}$  has been partitioned into  $m$  disjoint subregions, i.e.,  $\mathcal{X} = \bigcup \mathcal{X}_i$ ,  $i = 1, 2, \dots, m$ , according to the real-valued energy function  $E(\mathbf{x}_t)$ . Here,  $E(\mathbf{x}_t) = -\log p(\mathbf{x}_t|\mathbf{z}_{1:t})$ . We now consider the following trial distribution:

$$p_{\mathbf{g}}(\mathbf{x}_t) \propto \sum_{i=1}^m \frac{p(\mathbf{x}_t|\mathbf{z}_{1:t})}{g_i} I(\mathbf{x}_t \in \mathcal{X}_i) \quad (3)$$

where  $g_i = \int_{\mathcal{X}_i} p(\mathbf{x}_t|\mathbf{z}_{1:t})d\mathbf{x}_t$  and  $I(\cdot)$  is the indicator function.  $\mathbf{g} = \{g_i, i = 1, 2, \dots, m\}$  can be regarded as the density of states (DoS) or the spectral density in physics. Intuitively, if  $\mathbf{g}$  can be learned during sampling, sampling from the trial distribution  $p_{\mathbf{g}}(\mathbf{x}_t)$  will result in a “free” random walk in the space of energy (by regarding each energy subregion as a single sample “point”). Hence, the local-trap problem can be essentially overcome.

The basic idea behind sampling from the trial distribution (3) instead of the original filtering distribution (1) is mainly motivated by the multicanonical algorithms [20], [21], [29], which have achieved great success in physics. We remark that most of them are devoted to the optimization problems in physics or statistics. In the context of Bayesian tracking, however, some relevant issues, such as particles propagation, IS, and choice of the proposal distribution, should be collectively addressed. In what follows, we elaborate how to incorporate the SAMC sampling [21] into the Bayesian filtering framework to give a novel tracking scheme.

#### C. Sequential SAMC Sampling

Motivated by the aforementioned discussion, we propose a sequential SAMC sampling framework. To give a clear view, the flowchart of the proposed sequential SAMC sampling framework is schematically depicted in Fig. 1. Similar to SIR [4], there are three major stages in the sequential SAMC, i.e., particles propagation, SAMC sampling, and sample reweighting.

1) *Particles Propagation*: In the particle propagation stage, a resampling process [28] is first run on the input particles set generated at the previous time step to give a new particles set with equal weight  $\{(\mathbf{x}_{t-1}^{(i)}, (1/N))\}_{i=1}^N$ . Based on this particles set, the filtering distribution (1) can be approximated by  $p(\mathbf{x}_t|\mathbf{z}_{1:t}) \propto p(\mathbf{z}_t|\mathbf{x}_t) \sum_{i=1}^N (1/N)p(\mathbf{x}_t|\mathbf{x}_{t-1}^{(i)})$ .

After that, we need to generate an initial sample  $\mathbf{x}_t^{(1)}$  for subsequent sampling operations. The initial sample is proposed based on a Gaussian transition  $\mathbf{x}_t^{(1)} \sim \mathcal{N}(\mu, \Sigma)$  with mean  $\mu = A_1\mathbf{x}_{t-1}^* + A_2\mathbf{x}_{t-2}^*$  and covariance  $\Sigma$ , where  $\mathbf{x}_{t-1}^*$  and  $\mathbf{x}_{t-2}^*$

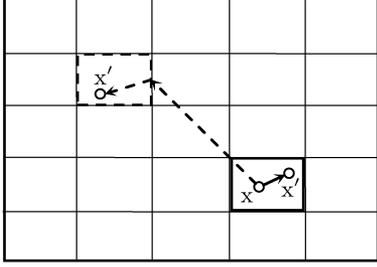


Fig. 2. Illustration of proposal  $Q(\mathbf{x}'; \mathbf{x})$  in a 2-D sample space, which is divided into cells using a discretized grid. This proposal accounts for (dashed arrow) the global hopping, as well as (solid arrow) the local random walk.

represent the maximum a posteriori (MAP) estimates of the previous states.  $A_1$  and  $A_2$  define the second-order autoregression (AR2) model with constant acceleration. It is sufficient for our particles propagation to use such a weak transition model since our objective is only to produce an initial sample to characterize the smooth motion, and the abrupt motion can be covered by subsequent SAMC sampling operations.

2) *SAMC Sampling*: In the sampling stage, we aim at simulating a chain converges to the filtering distribution (1) using the SAMC previously described. The sampling stage consists of three major consecutive steps, i.e., proposal, acceptance, and working-weight updating.

In practice, the proposal distribution used for our sampling scheme should be carefully designed to account for the large motion uncertainty. Suppose the sample space of our filtering problem is compact and bounded. Different from the partition of the sample space for the DoS learning according to the real-valued energy function, we now give another partition method for the sample space to facilitate effective proposal operation. For the  $d$ -dimensional sample space  $\mathcal{X} \triangleq S_1 \times S_2 \times \dots \times S_d$ ,  $\mathcal{X} \subset \mathbb{R}^d$ , we use a discretized grid to divide it into multidimensional *cells*. For each dimension  $i$ , its space  $S_i$  is divided into  $n_i$  intervals, i.e.,  $S_i = \{S_{i,j}\}_{j=1}^{n_i}$ . Thus, the sample space is composed of  $n = \prod_i n_i$  cells, and each cell  $C$  can be represented by

$$C \triangleq S_{1,j_1} \times S_{2,j_2} \times \dots \times S_{i,j_i} \times \dots \times S_{d,j_d} \quad (4)$$

where  $j_i \in \{1, 2, \dots, n_i\}$ . Let  $C(\mathbf{x}) \triangleq \{j_1, j_2, \dots, j_i, \dots, j_d\}$  denote the cell into which sample  $\mathbf{x} = \{x_1, x_2, \dots, x_i, \dots, x_d\}$  falls. We have  $x_i \in S_{i,j_i}$ . With the foregoing notations, we define the following mixture proposal:

$$Q(\mathbf{x}'; \mathbf{x}) = \lambda \mathcal{N}(\mathbf{x}'; \mathbf{x}, \Sigma) + (1 - \lambda) Q_u(\mathbf{x}'; \mathbf{x}) \quad (5)$$

where  $\mathcal{N}(\mathbf{x}, \Sigma)$  is a normal distribution with mean  $\mu = \mathbf{x}$  and covariance  $\Sigma$ ,  $Q_u(\mathbf{x}'; \mathbf{x})$  denotes a uniform distribution on the cell set  $S' = \mathcal{X} \setminus C(\mathbf{x})$ , and parameter  $\lambda \in [0, 1]$  steers the balance between the local random walk and the global stochastic hopping. The proposal previously defined is illustrated in Fig. 2. It is easy to verify that  $Q(\mathbf{x}'; \mathbf{x})$  is a global proposal, i.e.,  $Q(\mathbf{x}'; \mathbf{x}) > 0$  for all  $\mathbf{x}, \mathbf{x}' \in \mathcal{X}$ .

Given the current sample  $\mathbf{x}_t^{(k)}$  at iteration  $k$ , suppose candidate  $\mathbf{x}'$  has been generated by using proposal  $Q(\mathbf{x}'; \mathbf{x}_t^{(k)})$ . We

then use a single MH update to determine whether the candidate sample  $\mathbf{x}'$  is accepted or not. The acceptance probability  $\alpha(\mathbf{x}'; \mathbf{x}_t^{(k)})$  is defined by

$$\begin{aligned} \alpha(\mathbf{x}'; \mathbf{x}_t^{(k)}) &= \min \left\{ 1, \frac{p_{\mathbf{g}}(\mathbf{x}')}{p_{\mathbf{g}}(\mathbf{x}_t^{(k)})} \frac{Q(\mathbf{x}_t^{(k)}; \mathbf{x}')}{Q(\mathbf{x}'; \mathbf{x}_t^{(k)})} \right\} \\ &= \min \left\{ 1, \frac{p(\mathbf{x}' | \mathbf{z}_{1:t})}{p(\mathbf{x}_t^{(k)} | \mathbf{z}_{1:t})} \frac{g_{J(\mathbf{x}_t^{(k)})}(\mathbf{x}_t^{(k)})}{g_{J(\mathbf{x}')}(\mathbf{x}_t^{(k)})} \frac{Q(\mathbf{x}_t^{(k)}; \mathbf{x}')}{Q(\mathbf{x}'; \mathbf{x}_t^{(k)})} \right\} \end{aligned} \quad (6)$$

where  $J(\mathbf{x})$  denotes the index of the subregion which sample  $\mathbf{x}$  belongs to.

Once a new sample, e.g.,  $\mathbf{x}_t^{(k+1)}$ , is simulated by the sampling process previously described, a updating step for the working weight  $\theta$  will be run to update the DoS of each energy subregion, i.e.,

$$\theta_i^{(k+1)} = \theta_i^{(k)} + \gamma_k \left[ I(\mathbf{x}_t^{(k+1)} \in \mathcal{X}_i) - \phi_i \right] \quad (7)$$

for  $i = 1, 2, \dots, m$ , where  $\theta_i = \log g_i$ .  $\phi = \{\phi_i, i = 1, 2, \dots, m\}$  denotes the desired sampling distribution on the energy subregions and  $\gamma_k$  is the gain factor that controls the speed of weight learning. In general, a positive and nonincreasing sequence  $\{\gamma_k\}$  is required to satisfy the following conditions [21]:

$$\text{i) } \sum_{k=1}^{\infty} \gamma_k = \infty \quad \text{ii) } \sum_{k=1}^{\infty} \gamma_k^\rho < \infty \quad (8)$$

for some  $\rho \in (1, 2)$ . The gain factor is empirically set by  $\gamma_k = k_0 / \max(k_0, k)$  for some  $k_0 > 0$ . The weight learning is based on a SAP that belongs to a general class of the stochastic approximation algorithms of the Robbins–Monro type [30]. Compared with the WL algorithm [20], SAMC shows more superiority in sampling efficiency due to its self-adjusting mechanism, which makes the sampling less trapped by local modes. In fact, from (6) and (7), we can see that, if a proposal is rejected, the weight (i.e., DoS) of the energy subregion that the current sample belongs to will be adjusted to a larger value, and thus, the probability of jumping out from the current subregion will increase in the next iteration. Meanwhile, the weights of other energy subregions will be adjusted to a smaller value, and thus, the probability of jumping to one of these energy subregions will increase in the next iteration.

#### D. Sample Reweighting

Let  $\{\mathbf{x}_t^{(i)}\}_{i=1}^N$  represent the sample set drawn by SAMC whose invariant distribution is  $p_{\mathbf{g}}(\mathbf{x}_t)$ . Notice that SAMC itself falls into the dynamic important sampling algorithm [27], [35], [36]. When used to sample from the trial distribution (3), it generates a set of weighted samples to approximate the filtering distribution (1). As a result, each sample  $\mathbf{x}_t^{(i)}$  is augmented by weight  $w_t^{(i)} = g_{J(\mathbf{x}_t^{(i)})}$ , where  $J(\mathbf{x}_t^{(i)})$  denotes the index of the subregion into which sample  $\mathbf{x}_t^{(i)}$  falls. Therefore, the filtering

distribution at each time  $t$  can be represented by the particles set  $\{(\mathbf{x}_t^{(i)}, w_t^{(i)})\}_{i=1}^N$ . It should be noticed that the sample's weight is dynamically estimated during the sampling process, since DoS  $\mathbf{g}$  is learned on the fly in sampling.

Based on the resulting particles set, the object state is typically estimated according to the minimum mean square error (MMSE) or MAP criterion. PFs often use the moments of the distribution, such as mean, to estimate the state vector. Although the mean is optimal in the MMSE sense, it may not be appropriate for the filtering distribution with complex multimodal feature. In this paper, we use the MAP criterion to estimate the object state at each time step.

#### IV. IA-MCMC

Indeed, the proposed sequential SAMC sampling algorithm can deal with the large motion uncertainty. However, to guarantee its robustness, a certain number of samples are still needed to capture the abrupt motion. This is because the “uninformative” proposal distribution for the sequential SAMC sampling often proposes samples with a low acceptance rate due to the broadness of the whole sample space. To further improve the overall sampling efficiency, we need to design a more efficient sampling algorithm.

By introducing a SAP into the MCMC sampling framework, SAMC can effectively overcome the local-trap problem during sampling even when the energy landscape is rugged. Even so, we believe that SAMC can be further adapted to improve its sampling efficiency. Our approach is partially motivated by the methodology of the adaptive MCMC algorithms [25], [34], the proposal distribution of which is adaptively updated using the past samples in simulations, and thus improves the convergence rate of the MCMC sampling. In this paper, we expect to estimate the promising regions (volumes) of the sample space and to thus learn more informative proposals to speed up the overall sampling process. We achieve this by introducing a density-grid-based predictive model into the SAMC sampling to give a sampling–prediction–sampling scheme. The density grid model carries the online statistical information about the filtering distribution. As applied to abrupt motion tracking, the promising regions correspond to those density regions with high posterior, i.e., regions that are most likely to compactly contain the target object.

The reason for the use of a density-grid-based predictive model in the SAMC sampling process lies in the fact that the density grid model aims to estimate the promising regions of the sample space so that the new samples generated from these promising regions will have more chances to reach the global optimum than one simply drawn by uninformative proposal operation over the broad sample space, whereas SAMC allows the sampling to explore the whole sample space and produce more representative samples. In principle, the proposed algorithm is an MCMC sampler, the target distribution adaptation of which is to overcome the local-trap problem in sampling, whereas the proposal adaptation aims to improve the overall sampling efficiency. The proposed IA-MCMC sampling scheme is depicted in Fig. 3.

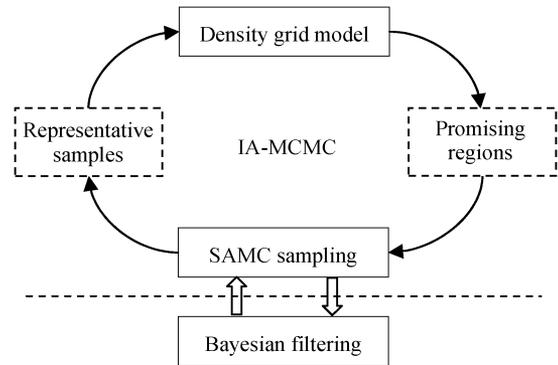


Fig. 3. Illustration of the IA-MCMC sampler in the sequential Bayesian filtering framework.

IA-MCMC is a two-step sampling scheme that involves preliminary sampling and adaptive sampling. The preliminary sampling aims to discover the rough modes of the energy landscape, whereas the adaptive sampling is to refine the promising regions of the sample space and to thus guide the sampling around the posterior modes.

##### A. Density Grid

Here, we introduce the density grid structure and its associated operations, which found the basis of the predictive model for the proposal adaptation in IA-MCMC. Although our model here shares some common features as the cluster model described in [37], which operates on the sequential incoming unweighted samples, our aim is to perform prediction based on the online samples with energy (filtering) information.

We extend the discretized grid (discussed in Section III) to the density grid by introducing the concept of the density of the cell. Let  $G$  denote the density grid, the density of each cell of which is initialized with zero values. For each online sample  $\mathbf{x}_t$  drawn by SAMC, we assign it a density coefficient defined by  $d(\mathbf{x}_t) = e^{-E(\mathbf{x}_t)}$ , where  $E(\mathbf{x}_t) = -\log p(\mathbf{x}_t | \mathbf{z}_{1:t})$  is the energy function of the filtering distribution. Introducing the concept of the density coefficient allows us to effectively discover and refine the promising regions of the sample space by placing more density weights on those of samples with lower energy. Based on the density coefficient, the density value of a cell unit can be easily calculated. For the cell unit  $C$  at iteration  $k$ , its density is defined as the sum of the density coefficient of all samples falling into  $C$ , i.e.,

$$D(C, k) = \sum_{i=1}^k d(\mathbf{x}^{(i)}) I(\mathbf{x}^{(i)} \in C). \quad (9)$$

We now show how to construct the predictive model based on the density grid previously defined. We define *promising regions* to be those of cell units with relatively high density. Let  $D_0^{(k)}$  denote a density threshold at iteration  $k$ . The promising regions can be defined by

$$\Psi^{(k)} = \{C \in G | D(C, k) \geq D_0^{(k)}\}. \quad (10)$$

Typically, threshold  $D_0^{(k)}$  is set to average density of all cells at iteration  $k$ . As a result, the promising regions of the sample space are continuously refined as the sampling proceeds, since

the density of each cell of the density grid is dynamically updated in sampling. From the data-mining point of view, the density grid can be regarded as an online predictive model that partitions the whole sample space into two parts, i.e., promising regions and nonpromising regions. Although simple, we found that this model efficiently and effectively performs for the proposal adaptation.

### B. Proposal Adaptation

As illustrated in Fig. 3, there are two learning procedures in IA-MCMC, the working weight of which (i.e., DoS) is learned by SAMC, and it makes the sampling less trapped by local modes. Here, we consider another learning procedure, i.e., the proposal adaptation, to improve the sampling efficiency. Our purpose here is to learn a more informative proposal that biases the sampling toward promising regions of the sample space based on the density grid model previously described.

Notice that the promising regions  $\Psi^{(k)}$  found by the predictive model may contain multiple cell units. Let  $\hat{Q}^{(k)}$  denote the proposal distribution learned by our algorithm at iteration  $k$ . Given  $\mathbf{x}_t^{(k)}$ , we hope to draw a sample from the promising regions with a relatively large probability value. A natural definition for this proposal is

$$\hat{Q}^{(k)}(\cdot; \mathbf{x}_t^{(k)}) = \beta Q_p^{(k)}(\cdot; \mathbf{x}_t^{(k)}) + (1 - \beta) Q_u^{(k)}(\cdot; \mathbf{x}_t^{(k)}) \quad (11)$$

where  $\beta$  is the probability of sampling from promising regions,  $Q_p^{(k)}$  denotes the proposal used to draw samples in the promising regions  $\Psi^{(k)}$ , and  $Q_u^{(k)}$  is the uniform distribution on the nonpromising regions  $\mathcal{X} \setminus \Psi^{(k)}$ . Here, proposal  $Q_p^{(k)}$  takes the form of the proposal defined in (5) but limits the sampling within the promising regions  $\Psi^{(k)}$  instead of the whole sample space  $\mathcal{X}$ . This adaptive proposal accounts for the local random walk within each promising region, as well as stochastic hopping among multiple promising regions. In addition, the proposal biases the sampling toward the promising regions to improve the sampling efficiency. Note that the proposal can propose samples in nonpromising regions, but it proposes only limited samples there to help the sampling less trapped by the local optimum.

At the beginning of the adaptive sampling step, since the promising regions have not been well discovered, we hence update the proposal with a high probability value. However, as the sampling proceeds, the proposal learned from the predictive model becomes more and more informative, and it is unnecessary to update the proposal frequently. We therefore consider the following updating strategy:

$$Q^{(k)} = \xi_k \hat{Q}^{(k)} + (1 - \xi_k) Q^{(k-1)} \quad (12)$$

where  $\hat{Q}^{(k)}$  is the proposal learned at iteration  $k$ ,  $Q^{(k-1)}$  is the proposal used at iteration  $k - 1$ , and  $\xi_k$  is the learning rate. Specifically, the learning rate  $\xi_k$  is empirically defined by  $\xi_k = (k - M_0)^{-\eta}$ , where  $\eta$  controls the decreasing speed of the learning rate, and  $M_0$  is the number of samples used for the preliminary sampling.

### C. IA-MCMC Sampling

As previously discussed, the proposed IA-MCMC sampling scheme is comprised of two steps. In preliminary sampling, we perform the SAMC sampling without proposal adaptation, using a certain number of iterations to quickly explore the sample space to obtain the rough modes of the energy landscape. Based on the initial samples set generated by SAMC in preliminary sampling, the initial density grid is established, and the sample space is then initially grouped into promising regions and nonpromising regions. In the adaptive sampling step, to effectively find and refine the promising regions, it may consist of several stages. The number of iterations, i.e.,  $M$ , for each sampling stage depends on the complexity of the filtering distribution. As the sampling stage proceeds, the promising regions are gradually shrunk and refined based on the density grid. Specially, once the sampling operation steps into a new sampling stage, the density grid will be reconstructed with a higher resolution. For example, we can divide each cell unit into four equal-sized cells and then recalculate the density of all cells. The whole sampling process of IA-MCMC is illustrated in Fig. 4, and the sequential IA-MCMC sampling algorithm for abrupt motion tracking is outlined in Algorithm 1.

The proposed sampling–prediction–sampling scheme can be viewed as a data-mining-mode embedded sampling algorithm, which substantially speeds up the overall sampling process of abrupt motion tracking. Here, the density grid is selected to be the predictive model for searching the promising regions of the sample space because of its computational efficiency. It should be noticed that this density grid is repeatedly used in the adaptive sampling step.

We have also considered some other predictive models that might have better performance on prediction, such as the CART model, as used in [34]. However, they are generally computationally expensive for our tracking cases that are often required to be solved in a practical time scale. Finally, the proposed IA-MCMC algorithm, in theory, falls into the adaptive MCMC algorithms [25]. The satisfaction of the *diminishing adaptation* condition is necessary to ensure its ergodicity. The ergodicity of the proposed IA-MCMC algorithm can be verified based on the recent theoretical advances in the ergodicity of the adaptive MCMC algorithms [22]–[25].

---

#### Algorithm 1: Sequential IA-MCMC Sampling

---

**Input:** Particles set  $\{(\mathbf{x}_{t-1}^{(i)}, \pi_{t-1}^{(i)})\}_{i=1}^N$

**Output:** Particles set  $\{(\mathbf{x}_t^{(i)}, \pi_t^{(i)})\}_{i=1}^N$  and MAP estimate  $\mathbf{x}_t^*$

**Initialization:**

1. Sample space partition:  $\mathcal{X} = \bigcup \mathcal{X}_i, i = 1, 2, \dots, m$
2. Density grid initialization:  $s = 0, k = 0, D(C, k) = 0$
3. DoS initialization:  $g_i = 1, \text{ for } i = 1, 2, \dots, m$
4. Resample to give  $\{(\mathbf{x}_{t-1}^{(i)}, (1/N))\}_{i=1}^N$

**Preliminary Sampling:**

**for**  $k = 1$  to  $M_0$  **do**

1. Propose a candidate sample  $\mathbf{x}'$  using proposal (5).
2. Calculate the acceptance ratio  $\alpha$  according to (6).

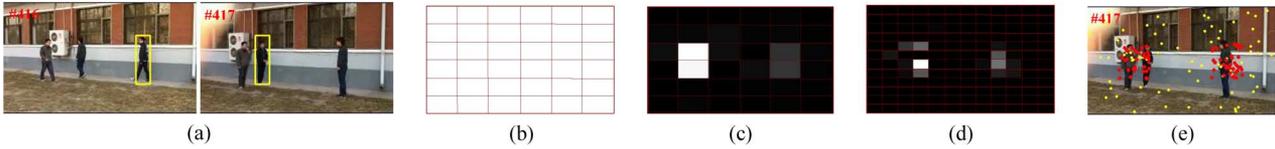


Fig. 4. Illustration of the overall sampling process of IA-MCMC. (a) Abrupt motion caused by camera switching. (b) Initial 2-D density grid ( $6 \times 6$  cells). (c) Rough promising regions found after the preliminary sampling step (at iteration  $k = 200$ ). (d) Promising regions are shrunk and refined as the sampling goes on (at iteration  $k = 300$ ). (e) Distribution of the samples drawn by IA-MCMC, where the samples in red are drawn in the adaptive sampling step.

3. Accept  $\mathbf{x}'$  with probability  $\alpha$ .
4. Update the working weights  $\mathbf{g}$  according to (7).

**end for**

#### Adaptive Sampling:

**for**  $k = M_0 + 1$  to  $N$  do

1. **if** iteration steps into a new sampling stage **then**

- 1.1. Reconstruct  $G$  with a higher resolution.
- 1.2. Update  $D(C, k)$  for all  $C \in G$ .
- 1.3.  $s \leftarrow s + 1$ .

**end if**

2. Propose a candidate sample  $\mathbf{x}'$  using the proposal (12).
3. Calculate the acceptance ratio  $\alpha$  according to (6).
4. Accept  $\mathbf{x}'$  with probability  $\alpha$ .
5. Update the working weights  $\mathbf{g}$  according to (7).
6. Update  $D(C, k)$  for all  $C \in G$ .

**end for**

#### Reweighting:

1. Reweight the samples set  $w_t^{(i)} = g_{J(\mathbf{x}_t^{(i)})}$ .
2. Normalize the weights  $\pi_t^{(i)} = w_t^{(i)} / \sum_{i=1}^N w_t^{(i)}$ .

#### Inference:

1. MAP estimate  $\mathbf{x}_t^*$

**End**

## V. EXPERIMENT

### A. Implementation Details

To simplify notations, the tracker based on the sequential SAMC sampling algorithm is denoted by the SAMC tracker, and the IA-MCMC tracker refers to the tracking algorithm based on the sequential IA-MCMC algorithm. In this paper, we represent the object as a rectangular region defined by its position in the scenario and the scale of the region, i.e.,  $\mathbf{x} = \{x, y, s\}$ . Since our goal is to cope with the large motion uncertainty, it is unnecessary to learn the accurate parameters of the dynamic model. Therefore, we use a weak motion prior [18] defined by

$$p(\mathbf{x}_t | \mathbf{x}_{t-1}) = \begin{cases} U(\mathcal{X}_p) & \text{for object position} \\ \mathcal{N}(\mathbf{x}_t^s; \mu_s, \sigma_s^2) & \text{for object scale} \end{cases} \quad (13)$$

where  $U(\mathcal{X}_p)$  is a uniform distribution on the 2-D spatial space  $\mathcal{X}_p$  and  $\mathcal{N}(\mu_s, \sigma_s^2)$  is a normal distribution with mean  $\mu_s = A_1 \mathbf{x}_{t-1}^s + A_2 \mathbf{x}_{t-2}^s$  and variance  $\sigma_s^2$ , where  $A_1$  and  $A_2$  define the AR2 model with constant acceleration. In practice, it is appropriate to account for this weak motion prior for the object with abrupt motion changes since the object in real scenarios can move to any position, whereas its scale often approximately smoothly evolves. In this paper, the variance is set to  $\sigma_s^2 = 0.013^2$ .

Under the assumption of the smooth changes of the object scale, we modify the proposal distribution used in the SAMC and IA-MCMC trackers. The proposal for the object position takes the form as defined in (5) and (11), whereas the proposal for the object scale is modeled as an AR2 process that takes the form of the motion prior for the scale but uses the MAP estimate of the previous states instead of the state values  $\mathbf{x}_{t-1}$  and  $\mathbf{x}_{t-2}$ .

In this paper, we adopt the color-based appearance model [3], and the likelihood function for filtering distribution is based on the HSV color histogram (110 bins,  $N_h = N_s = N_v = 10$ ) similarity, which is defined by

$$p(\mathbf{z}_t | \mathbf{x}_t) = e^{-\lambda_c d(H_r, H(\mathbf{x}_t))} \quad (14)$$

where  $H_r$  is the reference appearance model,  $H(\mathbf{x}_t)$  is the candidate appearance at  $\mathbf{x}_t$ ,  $d$  is the Bhattacharyya distance on the HSV histogram, and  $\lambda_c$  is a predefined parameter that is empirically set to 20 in this paper. Since the proposed tracking approach is more about dealing with the abrupt motion difficulties, we only adopt a simple appearance model. We found that our tracking approach can effectively deal with the large motion uncertainty in our tracking cases even using such a simple observation model.

In our experiments, we empirically partition the sample space into 50 subregions with equal energy bandwidth, according to the real-valued energy function  $E(\mathbf{x}_t) = -\log p(\mathbf{x}_t | \mathbf{z}_{1:t})$ . As discussed in [21], the maximum energy difference in each subregion should not be larger than 2 to ensure that the local random walk within the same subregion has a reasonable acceptance rate. As to the gain factor  $k_0$ , it is set to a relatively large value, e.g.,  $2N$ , at the preliminary sampling steps since a large  $k_0$  value will enable the sampler to explore all the energy subregions quickly, even when the energy landscape of the filtering distribution is rugged. Accordingly, parameter  $\lambda$  is set to 0.2 to facilitate the global stochastic hopping of the proposal in the preliminary sampling step, while in the adaptive sampling step,  $k_0$  is set to a small value, e.g.,  $(N/4)$ , to facilitate the adaptive sampling by refining the rough modes of the filtering distribution, and the bias parameter  $\beta$  is set to 0.9 to bias the sampling toward the promising regions of the sample space. The desired sampling distribution is set to be uniform, i.e.,  $\phi_i = 1/m, i = 1, 2, \dots, m$ .

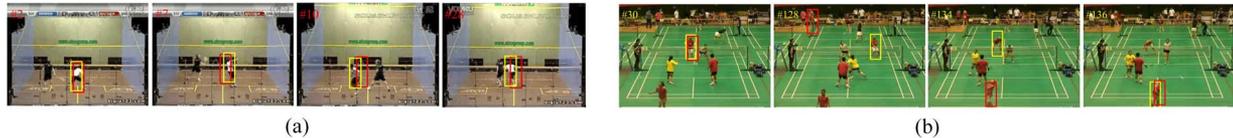


Fig. 5. Comparison of the proposed two trackers on the low-frame-rate video (a) *Squash* and (b) *Badminton*. (Red) SAMC. (Yellow) IA-MCMC.

For the density grid, it is initially set to a resolution with  $6 \times 6$  cells, as depicted in Fig. 4(b).

### B. Experiment Setup

To test the empirical performance of our tracking approach, we collected several sequences that involve abrupt motions in various scenarios. Implemented in MS VC++ based on OpenCV, our algorithm runs at about 20 frames per second with 300 particles on a standard 2.8-GHz personal computer. Here, we present only the sample tracking results. Note that the results can be better viewed on high-resolution displays or color printouts.

In our experiments, we first compare the tracking performance of our two trackers, i.e., SAMC and IA-MCMC, by qualitatively evaluating the impact of the proposal adaptation in sampling. We then compare our method to three state-of-the-art sampling-based tracking methods, i.e., MCMC-PF [8], A-WLMC [18], and the adaptive Metropolis (AM) based on [26]. For the sake of fair comparison, all the tracking algorithms adopt the same dynamic (prior) model and observation likelihood model previously defined. The proposal covariance for the four tracking algorithms are all set to  $\Sigma = \text{diag}(\sigma_x^2, \sigma_y^2, \sigma_s^2) = \text{diag}(4^2, 3^2, 0.013^2)$ . In IA-MCMC, the number of samples for the preliminary sampling is twice as much as that for each stage of the adaptive sampling step.

### C. Impact of the Proposal Adaptation

We begin by giving the results of the proposed two trackers on a low-frame-rate video *Squash*, which is obtained by keeping one frame in every 15 frames from a video with more than 400 original frames. As illustrated in Fig. 5(a), both using 300 samples, IA-MCMC more accurately performed than SAMC throughout the sequence. We also compare the proposed two trackers on another low-frame-rate video *Badminton*, which is downsampled from an original video with more than 1700 frames by keeping one frame in every ten frames. In this sequence, some objects (players) have similar color appearance to the object of interest. As shown in Fig. 5(b), IA-MCMC performed well except for a very few frames (e.g., #136), in which the object undergoes large appearance changes; whereas SAMC failed to track the player in some frames using the same number of samples.

We then perform a quantitative comparison of tracking accuracy between SAMC and IA-MCMC to further verify that the use of proposal adaptation in SAMC does help. The ground truths of the two test videos are manually annotated, and the comparison is based on the position error in pixels. As shown in Fig. 6, the position error of IA-MCMC is apparently lower than that of SAMC, and the former is less trapped by local optimal modes. We believe that the improved tracking performance of IA-MCMC is mainly due to the proposal adaptation during the

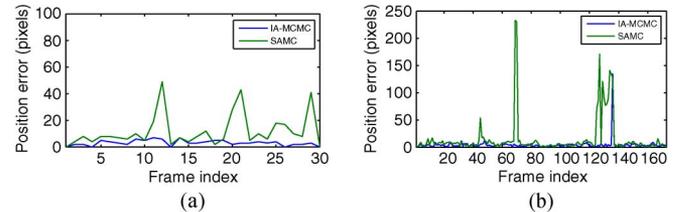


Fig. 6. Comparison of tracking accuracy between IA-MCMC and SAMC on the low-frame-rate video (a) *Squash* and (b) *Badminton*.

adaptive sampling step, which biases the SAMC sampling toward the promising regions of the sample space to speedup the overall sampling process.

### D. Qualitative Comparison

To qualitatively evaluate the tracking performance of IA-MCMC, we compare it with the other three trackers on several test videos that involve abrupt motions in various scenarios, including fast motion, camera switching, sudden dynamic changes, and the low-frame-rate videos.

**Fast Motion:** We first qualitatively evaluate the tracking performance of the four trackers on the sequence *Face* [43], in which a face quickly moves left and right. In this experiment, 50 samples are used for MCMC-PF and AM, and 25 samples for A-WLMC and ours. As illustrated in Fig. 7, even using only a few samples, A-WLMC and ours can effectively track the face with fast motion. On the other hand, MCMC-PF and AM fail to track the face in some frames, even using more samples. It is observed that at least 100 samples are required for MCMC-PF and AM to effectively track the face in this experiment.

**Camera switching:** The second experiment is to track a walker in scenarios with camera switching, which causes abrupt motions of the target object. Our experiments indicated that IA-MCMC can successfully cope with this large motion uncertainty. As illustrated in Fig. 8, using 100 samples, A-WLMC and IA-MCMC successfully tracked the walker throughout the sequence. On the other hand, even using 1000 samples, MCMC-PF and AM cannot effectively estimate the abrupt motions when camera switching occurs in scenarios. The reason for this is mainly due to the fact that the two trackers have no effective mechanism to handle the local-trap problem in sampling from the multimodal filtering distribution. For this sequence, however, it is observed that IA-MCMC archived higher tracking accuracy than A-WLMC using the same number of samples, as shown in Fig. 8(c) and (d).

**Sudden dynamic changes:** The third scenario in our experiments is a pingpong that struck the racket and bounced back with sudden dynamic changes. The unexpected motion dynamic makes the tracking task rather hard by learning an accurate motion model. Our experiment showed that our approach can effectively deal with this difficulty only using a weak motion

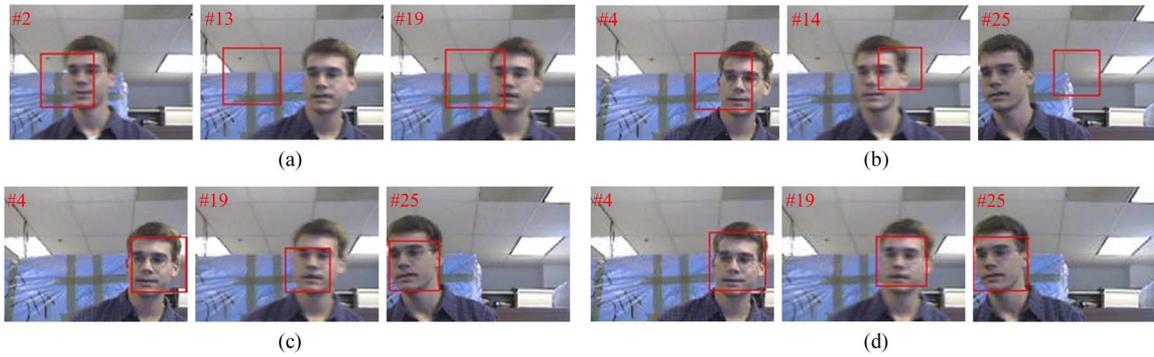


Fig. 7. Tracking results of the four trackers on the sequence *Face*. (a) MCMC-PF. (b) AM. (c) A-WLMC. (d) IA-MCMC.

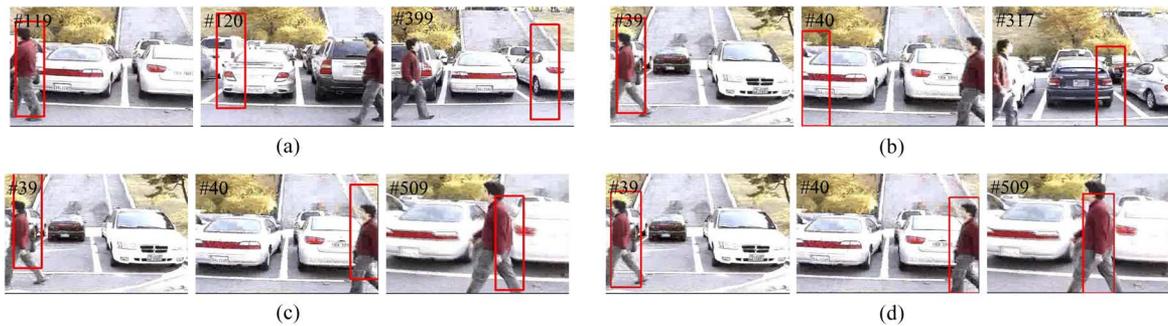


Fig. 8. Tracking results of the four trackers on the sequence *Youngki* [18]. (a) MCMC-PF. (b) AM. (c) A-WLMC. (d) IA-MCMC.

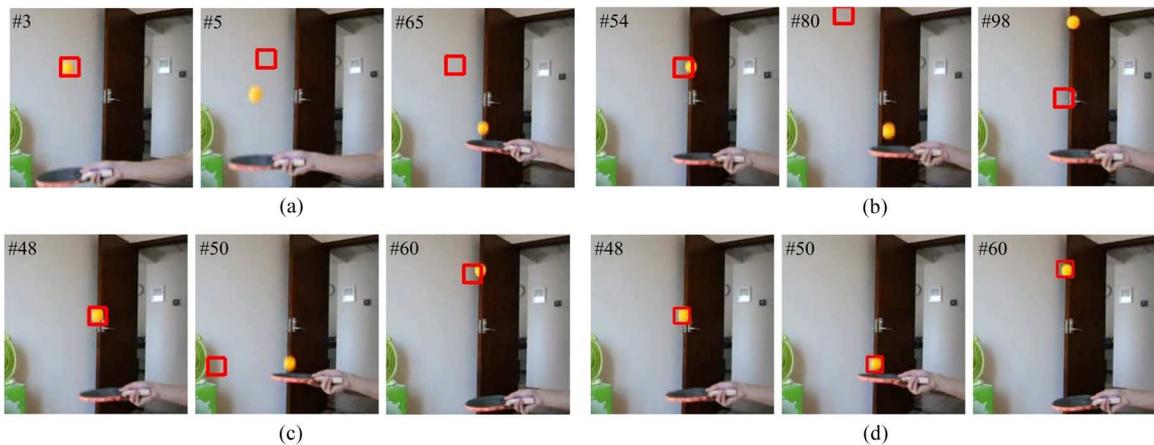


Fig. 9. Tracking results of the four trackers on the sequence *Pingpong*. (a) MCMC-PF. (b) AM. (c) A-WLMC. (d) IA-MCMC.

model. With 300 samples, our method successfully tracked the bouncing pingpong throughout the sequence. Sample frames are shown in Fig. 9. Note that, in this sequence, the object size is much smaller than that of the walker in previous experiments. Accordingly, the promising region of the sample space is much smaller, and thus, more samples are required for sufficient sampling in the larger sample space to discover the promising regions. Even using 1000 samples, MCMC-PF and AM poorly perform, experiencing a significant drift of the target object. Moreover, A-WLMC failed to track the pingpong in some frames using the same number of samples as ours.

*Low-frame-rate video:* The fourth experiment is to track a tennis player in a low-frame-rate video, which is downsampled from a video with more than 700 original frames, by keeping one frame in every 20 frames. We test four trackers on this se-

quence, and sample frames are shown in Fig. 10. Even using 1000 samples, MCMC-PF and AM frequently lost the track due to the abrupt motions caused by severe frame dropping; whereas IA-MCMC effectively dealt with this difficulty using only 300 samples. On the other hand, also using 300 samples, A-WLMC cannot effectively cope with the abrupt motions of the player in some frames, as shown in Fig. 10(c).

*Camera switching + Low-frame-rate video:* The final experiment is to qualitatively evaluate the tracking performance of the four trackers on a synthetic sequence (*Walk*) that involves the severe abrupt motion of the object caused by both camera switching and low-frame-rate video. In this sequence, three persons walk back and forth, and partial occlusions frequently occur. In this experiment, 600 samples are used for MCMC-PF and AM, and 300 samples are used for A-WLMC

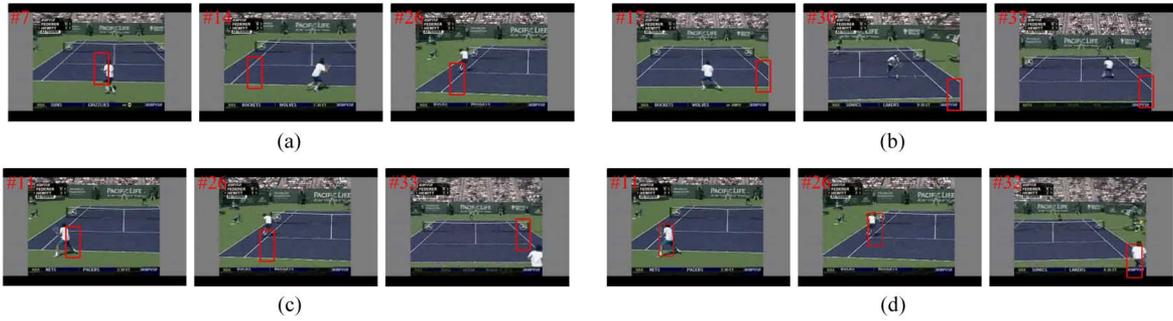


Fig. 10. Tracking results of the four trackers on the sequence *Tennis* [18]. (a) MCMC-PF. (b) AM. (c) A-WLMC. (d) IA-MCMC.

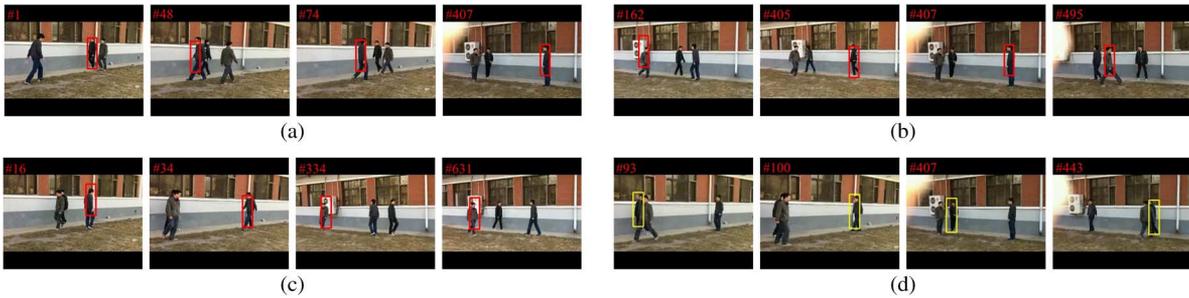


Fig. 11. Tracking results of the four trackers on the sequence *Walk*. (a) MCMC-PF. (b) AM. (c) A-WLMC. (d) IA-MCMC.

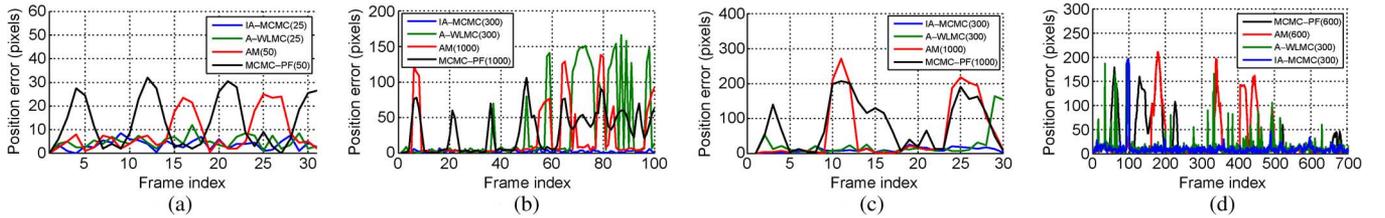


Fig. 12. Frame-by-frame comparison of the position error (in pixels) for the four trackers on the four sequences. (a) *Face*, (b) *Pingpong*, (c) *Tennis*, and (d) *Walk*.

and ours. Sample frames are illustrated in Fig. 11. It is observed that our approach can effectively track the object throughout the sequence (except for only a few frames, in which the target object is almost completely occluded by another walker, such as #100). As shown in Fig. 11, A-WLMC fails to track the object in quite a few frames when partial occlusions occur, although it has an effective mechanism to deal with the abrupt motion caused by camera switching and low-frame-rate video. On the other hand, AM and MCMC-PF frequently lose the track and poorly perform on this sequence due to the large motion uncertainty.

### E. Quantitative Comparison

To quantitatively evaluate the tracking performance of our tracker, we compare ours with the other three trackers on the test sequences. We first define tracking to be lost when the center position of the estimated rectangle is not in that of the manually labeled ground truth anymore. We quantitatively evaluate the performance of the four different trackers in terms of position error (in pixels) on the four test videos, which involve different types of abrupt motions, i.e., fast motion, sudden dynamic changes, low-frame-rate video, and camera switching. A frame-by-frame comparison of the position error in pixels for the four trackers is shown in Fig. 12, and the relative position

error is listed in Table I. Here, the relative position error is defined by  $d = \|(x, y) - (x_g, y_g)\|/s_g$ , where  $\mathbf{x}_g = (x_g, y_g, s_g)$  is the ground-truth state. The reason for the use of this error measurement is that it facilitates comparing the tracking accuracy for the target objects with different sizes [33]. As shown in Table I, both the mean relative error and the standard deviation of our approach are consistently smaller than those of other approaches, which indicated that the proposed approach is more accurate and stable even using a small number of samples.

To evaluate how the frame dropping rate affects the tracking performance of the different approaches, we test the four trackers on a series of low-frame-rate videos [18], i.e., *Tennis(10)*, *Tennis(15)*, *Tennis(20)*, *Tennis(25)*, and *Tennis(30)*, which are downsampled from a sequence with more than 700 original frames, by keeping one frame in every 10, 15, 20, 25, and 30 frames, respectively. In this experiment, 300 samples are used for IA-MCMC; other trackers use 600 samples. Fig. 13 shows the successful tracking rate versus the downsampling interval of the four trackers. It is observed that the tracking performance of our approach is better than others even using fewer samples. Moreover, the tracking performances of A-WLMC and our approach are less affected by the frame dropping rate. This experiment also verified the robustness of our algorithm in tracking the object with severe abrupt motions.

TABLE I  
RELATIVE POSITION ERROR OF THE FOUR TRACKERS ON THE TEST SEQUENCES

Method	Face		Pingpong		Tennis		Walk	
	$d_\mu$	$d_\sigma$	$d_\mu$	$d_\sigma$	$d_\mu$	$d_\sigma$	$d_\mu$	$d_\sigma$
MCMC-PF	0.5997	0.4887	1.5571	1.3633	2.0175	1.7516	1.2710	2.0469
AM	0.3896	0.3589	1.1502	1.8803	1.3132	2.0012	0.9732	1.9215
A-WLMC	0.2234	0.1197	1.8041	2.7661	0.4867	0.8858	0.4575	1.0337
IA-MCMC	0.1706	0.1024	0.1396	0.0821	0.1716	0.1256	0.3316	0.7847

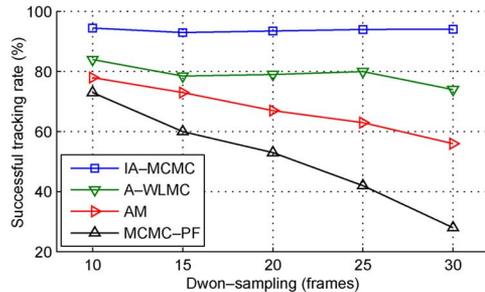


Fig. 13. Comparison of the four trackers affected by the frame dropping rate.

We also investigated how the parameter settings affect the tracking performance of the proposed tracker. In our experiments, we found that IA-MCMC can achieve good tracking performance, and it is insensitive to the number of samples for the preliminary sampling if  $M_0$  is set to  $(2/3)N \sim (4/5)N$ . This may be attributed to the fact that seeking the rough modes of the sample space often requires most of the samples to perform sufficient sampling in the preliminary sampling step. As to parameter  $\beta$  in (11), it is observed that IA-MCMC works well if it is set to  $\beta \geq 0.90$  in the experiments. Finally, parameter  $\lambda$  in proposal (5) is crucial to the tracking performance of IA-MCMC. A small  $\lambda$  is more effective for smooth motions, whereas a large value of  $\lambda$  is more suitable for abrupt motions. It is nontrivial to determine the optimal value of  $\lambda$  in tracking due to the lack of prior knowledge on the motion of the target object in real-world applications. Since we mainly focus on abrupt motion tracking in this paper, we perform the investigation on how parameter  $\lambda$  affects the performance of IA-MCMC on several test videos that involve various types of abrupt motions, and it is observed that IA-MCMC performs well if  $\lambda$  is chosen around 0.2 in the experiments. This also indicated that the tracking performance of IA-MCMC may benefit from a hybrid proposal that properly combines local random walk and global hopping.

#### F. Discussions

All the aforementioned experiments have validated the proposed tracking approach. When the target experiences abrupt motion changes, we can explain the reason why the two existing methods, i.e., AM and MCMC-PF, poorly perform in the experiments. This can be attributed to the fact that there is no mechanism in the two methods to effectively avoid the local-trap problem during sampling. Thus, they tend to lose the track even using a large number of samples when abrupt motion occurs. On the other hand, although A-WLMC is able to handle the local-trap problem, a certain number of samples are required to guarantee sufficient sampling in the broad sample space due to the

fact that it only adopts a less efficient proposal distribution. Although an annealing procedure is incorporated, its sampling efficiency is expected to be further improved for practical tracking applications. Unlike other methods, our algorithm utilizes the information of historical samples (particles) of the simulated chain for *intensive* adaptation, the target (distribution) adaptation of which helps to deal with the local-trap problem, and the proposal adaptation facilitates efficient sampling by learning more informative proposal on the fly. As a result, the proposed approach can be more effective and efficient to cope with the abrupt motion changes.

## VI. CONCLUSION

We have presented a novel approach for robust abrupt motion tracking in various scenarios. In the proposed tracking algorithm, we have introduced the SAMC sampling method into the Bayesian filtering framework to solve the local-trap problem in sampling suffered by many existing sampling-based tracking approaches. Furthermore, we have extended the SAMC algorithm to a MCMC sampler with intensive adaptation by learning the proposal on the fly in sampling. Extensive experiments have indicated that our method outperforms other alternatives and exhibit better efficiency and effectiveness in the tracking of abrupt motion. Since this paper has mainly focused on dealing with the abrupt motion problem, we have not considered the appearance adaptation. Nevertheless, we believe many state-of-the-art techniques on appearance adaptation [31]–[33], [40], [41] and object representation [39] can be integrated in our tracking scheme to further enhance the robustness to the appearance changes and the background distractions. Our further study will include the investigation for the situation when the object undergoes significant motion and appearance changes simultaneously.

## ACKNOWLEDGMENT

The authors would like to thank the associate editor and all anonymous reviewers for their insightful comments. They also thank the authors of [18] for providing the test videos and code.

## REFERENCES

- [1] M. Isard and A. Blake, "Condensation—Conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998.
- [2] K. Nummiaro, E. Koller-Meier, and L. Van Gool, "An adaptive color-based particle filter," *Image Vis. Comput.*, vol. 21, no. 1, pp. 99–110, Jan. 2003.
- [3] P. Perez, C. Hue, J. Vermaak, and M. Gangnet, "Color-based probabilistic tracking," in *Proc. ECCV*, 2002, pp. 661–675.
- [4] A. Doucet, N. de Freitas, and N. Gordon, *Sequential Monte Carlo Methods in Practice*. New York: Springer-Verlag, 2001.

- [5] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for on-line nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.
- [6] O. Cappe, S. J. Godsill, and E. Moulines, "An overview of existing methods and recent advances in sequential Monte Carlo," *Proc. IEEE*, vol. 95, no. 5, pp. 899–924, May 2007.
- [7] J. Deutcher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Proc. CVPR*, 2000, pp. 126–133.
- [8] Z. Khan, T. Balch, and F. Dellaert, "An MCMC-based particle filter for tracking multiple interacting targets," in *Proc. ECCV*, 2004, pp. 279–290.
- [9] T. Zhao and R. Nevatia, "Tracking multiple humans in crowded environment," in *Proc. CVPR*, 2004, pp. II-406–II-413.
- [10] K. Choo and D. Fleet, "People tracking using hybrid Monte Carlo filtering," in *Proc. ICCV*, 2001, pp. 321–328.
- [11] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. CVPR*, 2000, pp. 142–149.
- [12] C. Yang, R. Duraiswami, and L. Davis, "Efficient mean-shift tracking via a new similarity measure," in *Proc. CVPR*, 2005, pp. 176–183.
- [13] B. North and A. Blake, "Learning dynamical models by expectation maximization," in *Proc. ICCV*, 1998, pp. 384–389.
- [14] M. Isard and A. Blake, "A mixed-state condensation tracker with automatic model-switching," in *Proc. ICCV*, 1998, pp. 107–112.
- [15] J. Sullivan, A. Blake, M. Isard, and J. MacCormick, "Object localization by Bayesian correlation," in *Proc. ICCV*, 1999, pp. 1068–1075.
- [16] G. Hua and Y. Wu, "Multi-scale visual tracking by sequential belief propagation," in *Proc. CVPR*, 2004, pp. I-826–I-833.
- [17] Y. Li, H. Ai, T. Yamashita, S. Lao, and M. Kawade, "Tracking in low frame rate video: A cascade particle filter with discriminative observers of different life-spans," in *Proc. CVPR*, 2007, pp. 1–8.
- [18] J. Kwon and K. Lee, "Tracking of abrupt motion using Wang–Landau Monte Carlo estimation," in *Proc. ECCV*, 2008, pp. 387–400.
- [19] C. Andrieu, É. Moulines, and P. Priouret, "Stability of stochastic approximation under verifiable conditions," *SIAM J. Control Optim.*, vol. 44, no. 1, pp. 283–312, 2005.
- [20] F. Wang and D. P. Landau, "Efficient multiple-range random-walk algorithm to calculate the density of states," *Phys. Rev. Lett.*, vol. 86, no. 10, pp. 2050–2053, Mar. 2001.
- [21] F. Liang, C. Liu, and R. J. Carroll, "Stochastic approximation in Monte Carlo computation," *J. Amer. Statist. Assoc.*, vol. 102, no. 477, pp. 305–320, Mar. 2007.
- [22] W. Gilks, G. O. Roberts, and S. Sahu, "Adaptive Markov chain Monte Carlo through regeneration," *J. Amer. Statist. Assoc.*, vol. 93, no. 443, pp. 1045–1054, Sep. 1998.
- [23] Y. Atchad and J. S. Rosenthal, "On adaptive Markov chain Monte Carlo algorithms," *Bernoulli*, vol. 11, no. 5, pp. 815–828, Oct. 2005.
- [24] H. Haario, E. Saksman, and J. Tamminen, "An adaptive metropolis algorithm," *Bernoulli*, vol. 7, no. 2, pp. 223–242, Apr. 2001.
- [25] G. O. Roberts and J. S. Rosenthal, "Coupling and ergodicity of adaptive Markov chain Monte Carlo algorithms," *J. Appl. Probab.*, vol. 44, no. 2, pp. 458–475, Jun. 2007.
- [26] G. O. Roberts and J. S. Rosenthal, "Examples of adaptive MCMC," *J. Comput. Graph. Statist.*, vol. 18, no. 2, pp. 349–367, Jun. 2009.
- [27] J. S. Liu, F. Liang, and W. H. Wong, "A theory for dynamic weighting in Monte Carlo," *J. Amer. Statist. Assoc.*, vol. 96, pp. 561–573, 2001.
- [28] J. S. Liu, *Monte Carlo Strategies in Scientific Computing*. New York: Springer-Verlag, 2008.
- [29] B. A. Berg and T. Neuhaus, "Multicanonical algorithms for 1st-order phase-transitions," *Phys. Lett. B*, vol. 267, no. 2, pp. 249–253, Sep. 1991.
- [30] H. Robbins and S. Monro, "A stochastic approximation method," *Ann. Math. Statist.*, vol. 22, no. 3, pp. 400–407, Sep. 1951.
- [31] S. K. Zhou, R. Chellappa, and B. Moghaddam, "Visual tracking and recognition using appearance adaptive models in particle filters," *IEEE Trans. Image Process.*, vol. 13, no. 11, pp. 1491–1506, Nov. 2004.
- [32] J. Wang, X. Chen, and W. Gao, "Online selecting discriminative tracking features using particle filter," in *Proc. CVPR*, 2005, pp. 1037–1042.
- [33] M. Yang, Z. Fan, J. Fan, and Y. Wu, "Tracking nonstationary visual appearances by data-driven adaptation," *IEEE Trans. Image Process.*, vol. 18, no. 7, pp. 1633–1644, Jul. 2009.
- [34] Y. Ren, Y. Ding, and F. Liang, "Adaptive evolutionary Monte Carlo algorithm for optimization with applications to sensor placement problems," *Statist. Comput.*, vol. 18, no. 4, pp. 375–390, Dec. 2008.
- [35] F. Liang, "Dynamically weighted importance sampling in Monte Carlo computation," *J. Amer. Statist. Assoc.*, vol. 97, no. 459, pp. 807–821, Sep. 2002.
- [36] F. Liang, "On the use of stochastic approximation Monte Carlo for Monte Carlo integration," *Statist. Probab. Lett.*, vol. 79, no. 5, pp. 581–587, 2009.
- [37] Y. Chen and L. Tu, "Density-based clustering for real-time stream data," in *Proc. ACM SIGKDD*, 2007, pp. 133–142.
- [38] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, pp. 1–45, 2006.
- [39] E. Maggio and A. Cavallaro, "Accurate appearance-based Bayesian tracking for maneuvering targets," *Comput. Vis. Image Understand.*, vol. 113, no. 4, pp. 544–555, Apr. 2009.
- [40] H. Grabner, C. Leistner, and H. Bischof, "Semi-supervised on-line boosting for robust tracking," in *Proc. ECCV*, 2008, pp. 234–247.
- [41] B. Babenko, M.-H. Yang, and S. Belongie, "Visual tracking with online multiple instance learning," in *Proc. CVPR*, 2009, pp. 983–990.
- [42] X. Zhou and Y. Lu, "Abrupt motion tracking via adaptive stochastic approximation Monte Carlo sampling," in *Proc. CVPR*, 2010, pp. 1847–1854.
- [43] [Online]. Available: <http://vision.stanford.edu/~birch/headtracker/seq/>



**Xiuzhuang Zhou** (M'11) received the B.Sc. degree from Chengdu University of Information Technology, Chengdu, China, in 1996 and the M.Eng. and Ph.D. degrees from Beijing Institute of Technology, Beijing, China, in 2005 and 2011, respectively.

He is currently an Assistant Professor with the College of Information Engineering, Capital Normal University, Beijing. He has authored several scientific papers in peer-reviewed journals and conferences, including some top venues such as

the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE Conference on Computer Vision and Pattern Recognition, and the *Association for Computing Machinery Multimedia*. His research interests include computer vision, pattern recognition, and machine learning.



**Yao Lu** received the B.S. degree in electronics from Northeast University, Shenyang, China, in 1982 and the Ph.D. degree in computer science from Gunma University, Gunma, Japan, in 2003.

He was a Lecturer and an Associate Professor with Hebei University, China, from 1986 to 1998, and a foreign researcher with Gunma University in 1999. In 2003, he was an invited professor with the Engineering Faculty, Gunma University. He is currently a Professor with the Department of Computer Science, Beijing Institute of Technology, Beijing, China. He

has published more than 40 papers in international conferences and journals. His research interests include neural network, image and signal processing, and pattern recognition.



**Jiwen Lu** (S'10–M'11) received the B.Eng. degree in mechanical engineering and the M.Eng. degree in signal and information processing from Xi'an University of Technology, Xi'an, China, in 2003 and 2006, respectively, and the Ph.D. degree in electrical and electronic engineering from Nanyang Technological University, Singapore, in 2011.

He is currently a Post-Doctoral Research Fellow with the Advanced Digital Sciences Center, Singapore. He has authored more than 50 scientific papers in peer-reviewed journals and conferences, including

some top venues such as the IEEE TRANSACTIONS ON IMAGE PROCESSING, the IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS—PART B: CYBERNETICS, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, the International Conference on Computer Vision, the IEEE Conference on Computer Vision and Pattern Recognition, and the *Association for Computing Machinery Multimedia*. His research interests include computer vision, pattern recognition, machine learning, multimedia computing, and biometrics.

Dr. Lu was a recipient of the First Prize of the National Scholarship and the National Outstanding Student award given by the Ministry of Education of China in 2002 and 2003, respectively.



**Jie Zhou** (M'01–SM'04) was born in November 1968. He received the B.S. and M.S. degrees from Nankai University, Tianjin, China, in 1990 and 1992, respectively, and the Ph.D. degree from Huazhong University of Science and Technology, Wuhan, China, in 1995.

From 1995 to 1997, he was a Post-Doctoral Fellow with the Department of Automation, Tsinghua University, Beijing, China, where he has been a Full Professor since 2003. In the recent years, he has authored more than 100 papers in peer-reviewed journals and conferences. Among them, more than 30 papers have been published in top journals and conferences such as the IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, TRANSACTIONS ON IMAGE PROCESSING, and Conference on Computer Vision and Pattern Recognition. His research area includes computer vision, pattern recognition, and image processing.

Dr. Zhou is an associate editor for the International Journal of Robotics and Automation, the Acta Automatica, and two other journals.