IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received October 18, 2019, accepted November 3, 2019, date of publication November 15, 2019,
date of current version November 27, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2953742

Uncertain Motion Tracking Combined Markov Chain Monte Carlo and Correlation Filters

HUANLONG ZHANG[1], GUOHAO NIE[1], JIAN CHEN[1], JIE ZHANG[1], AND GUOSHENG YANG[2]

[1]College of Electric and Information Engineering, Zhengzhou University of Light Industry, Zhengzhou 450002, China
[2]School of Information Engineering, Minzu University of China, Beijing 100081, China

Corresponding author: Guosheng Yang (gsyangmuc@163.com)

This work was supported in part by the National Natural Science Foundation of China under Grant 61873246, Grant 61501407, Grant 61672471, and Grant 61702462, and in part by the Scientific and Technological Project of Henan Province under Grant 182102210607 and Grant 192102210108.

ABSTRACT To address the uncertain motion tracking problem, a tracking method based on the Markov Chain Monte Carlo and correlation filters is proposed. Firstly, multi-scope marginal likelihood (MSML) strategy is introduced to Wang-Landau Monte Carlo (WLMC) tracking method for increasing the acceptance ratio of samples in the promising regions and obtaining a more reliable distribution of density-of-states (DOS). Secondly, in order to raise the efficiency of the tracker, DOS is used to mark the region of interest. Then correlation filters are used to simplify the iterative optimizing operation of the subregions, and eventually target positioning is achieved by maximum response in the promising regions. Finally, a unified tracking framework is designed to enable correlation filters and WLMC with MSML strategy to exploit and complement each other to cope with uncertain motion tracking. Extensive experimental results on uncertain Motion sequences and benchmark datasets demonstrate that the proposed method performs favorably against the state-of-the-art methods.

INDEX TERMS Markov chain Monte Carlo, correlation filters, uncertain motion, visual tracking.

I. INTRODUCTION

Visual tracking is an active research topic in computer vision community that finds numerous applications such as intelligent surveillance, medical research, motion analysis, and autonomous driving [1]–[3]. In the past decades, dramatic progress has been achieved [4], but there are still some very challenging and unresolved issues, including illumination changes, fast motions, pose variations, partial occlusions and background clutters and so on. In order to obtain better tracking results, many trackers based on deep learning or correlation filtering have been proposed in recent years. Examples include target-aware deep tracking [5], multi-task correlation particle filter [6], hierarchical convolutional features [7], adaptive hedging [8], spatial regularization [9], continuous convolutional operations [10], etc.

Despite the great success of the above methods in visual tracking, most existing approaches are usually based on a smooth motion assumption. In real-word scenarios, it is possible for these trackers not to work well when dealing with abrupt motion problems such as fast motion, camera switching, low-frame-rate videos, etc. Examples of uncertain Motion are shown in Fig. 1. Many methods are

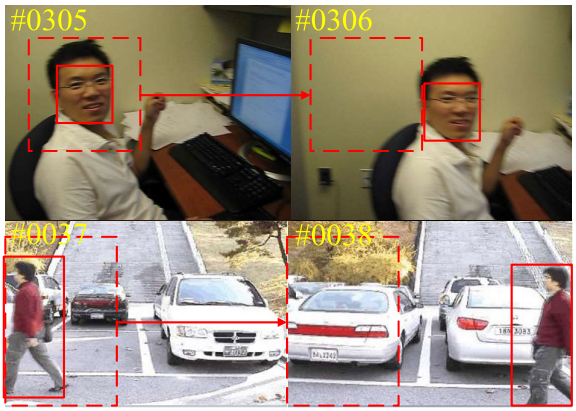


FIGURE 1. Examples of uncertain motion. Dotted and solid boxes mean search window and target, respectively. Tracking is no longer reliable when the target escapes from the search area due to uncertain motion.

The associate editor coordinating the review of this manuscript and approving it for publication was Shiping Wen.

proposed to solve these problems, like detection-based methods [11], dynamic model-based prediction method [12], multi-scale method [13], stochastic sampling-based tracking method [14], [15] and other methods.

In the above methods, the stochastic sampling based tracking method has been widely concerned. These methods are mostly under the framework of Markov Chain Monte Carlo (MCMC) or particle filtering (PF) and Bayesian filtering. Li *et al.* [16] proposed an abrupt motion tracking method based on stochastic search and level set methods for contour tracking. Li *et al.* [17] addressed the low frame rate problem from a view which integrates conventional tracking and detection. Su *et al.* [18] proposed a saliency embedded particle filter to detect the target region from salient regions when the object is lost for recovering tracking. But the above trackers still cannot obtain a robust tracking results under some kinds of abrupt motion scenarios. Zhou and Lu [14] and Zhou *et al.* [15] and other methods proposed an adaptive stochastic approximation Monte Carlo sampling (ASAMC) based tracking method for abrupt motion tracking, which can avoid the local-trap problem. Kwon and Lee [19], [20] introduced Wang-Landau algorithm into visual tracking. By estimating the DOS, it can track both smooth and abrupt motions accurately and robustly without loss of time. WLMC and ASAMC trackers are based on DOS which is divided into cells using a discretized grid. However, when each new state is generated, these samplers only consider the marginal likelihood (ML) score between regions or between samples. As a result, samples on the Markova chain will have a low quality so that, the sufficient accurate of the update DOS distribution can not be guaranteed. In addition, these methods need to visit the state space using the random walk method which often needs long iterations to reach the promising object state. It is difficult to have a good search efficiency, because of a large number of samples.

In this paper, we propose a tracking method (EWLCF-DP) based on the extend WLMC and correlation filters in order to solve the problems encountered by the traditional trackers. The contributions of the proposed tracking method are as follows:

1) We use MSML to estimate the sample similarity in two scopes, and then, introduce MSML to the acceptance formula of WLMC method. The acceptance probability of samples is increased in the promising regions, which ensures the reliability of the region proposal strategy based on DOS.
2) By replacing the traditional Bayesian iterative optimization process with KCF in promising subregions, the location can be realized in the frequency domain. This method can reduce the overall sample size and improve the tracking efficiency.
3) The uncertain motion tracking framework is designed, which unifies the methods of DOS and MSML guided Monte Carlo and correlation filters. The performance of the method is evaluated in the uncertain motion sequence and OTB benchmark datasets, which is proved to be superior to the state-of-the-art tracker.

II. RELATED WORK

In this section, we mainly discuss the methods closely related to this work, including stochastic sampling-based tracking methods and deep correlation filters tracking method.

A. STOCHASTIC SAMPLING BASED TRACKING METHODS

Stochastic sampling-based tracking methods are able to handle the multimodal distribution and recovering from the tracking failure. Among these frameworks, PF receives considerable attention because it can deal with nonlinearity and non-normality in the object models. Isard and Blake [21] first applied PF to cope with visual tracking difficulties encountered by Kalman filters. Del Bimbo and Dini [22] used PF to track long-term video sequences with first-order dynamic model and uncertainty adaptation. This method can continuously adapt the system model noise to balance uncertainty between the static and dynamic components of the state vector. Oron *et al.* [23] presented a locally orderless tracker using PF framework which showed better performance in tracking both demormable and rigid targets. But the high computational burden caused by a large number of particles often makes the PF infeasible for practical applications.

To reduce the computational cost and improve the sampling efficiency, MCMC methods gained widely attention in visual tracking. Khan *et al.* [24] proposed a MCMC-based PF method for tracking a variable number of objects, aiming at suppressing the particle degeneracy and particle impoverishment problems. Cong *et al.* [25] proposed a MCMC- based PF tracking method using the histogram of oriented gradients (HOG) based appearance model. It shows improved robustness in handling slight object occlusions. Kwon and Lee [26] proposed a robust tracker based on interactive MCMC sampling framework (Visual Tracking Decomposition, VTD tracker). VTD is proven to be robust for appearance variation induced by occlusion and illumination change. Although popular and commonly used, the plain Gibbs sampler or Metropolis-Hastings algorithm often gets trapped in the local optimum when the energy landscape of the target distribution is rugged.

To overcome these problems, many advanced MCMC algorithms have been developed. Roberts and Rosenthal [27] proposed the Adaptive MCMC algorithm to automatically adjust the proposal variance of MCMC as the Markov chain goes on. The proposal variance is tuned to produce an acceptance ratio as close as possible to the optimal value of 0.44 [28]. This adaptive scheme is very helpful in tracking the abrupt motion. Wang *et al.* [29], [30] proposed Langevin Monte Carlo sampling to overcome the problem of low sampling efficiency caused by random walk. Zhou *et al.* [31] utilized nearest neighbor field estimation to compute the importance proposal probabilities, which guide the Markov chain search towards promising regions. Mbelwa *et al.* [32]

integrated prior knowledge and objectness proposal into the smoothing stochastic approximate Monte Carlo to predict abrupt motion. In this work, we utilize the DOS distribution to propose candidate regions that may contain targets, and introduce MSML to further improve the reliability of region proposal strategy.

B. DEEP CORRELATION FILTERS TRACKING METHODS

Correlation filters have recently attracted considerable attention in visual tracking due to computational efficiency and robustness. Bolme *et al.* [33] modeled target appearance by learning an adaptive correlation filter which is optimized by minimizing the output sum of squared error (MOSSE). The tracker based on MOSSE filters is robust to variations in lighting, scale, pose, and non-rigid deformations. Henriques *et al.* [34] exploited the circulant structure of shifted image patches in a kernel space and propose the CSK method based on intensity features, and extended it to the KCF approach [35] with the HOG features. Its speed and precision achieve remarkable results. Danelljan et al. proposed the DSST method [36] with adaptive multi-scale correlation filters using HOG features to handle the scale change of the target object. Zhang *et al.* [37] utilized circulant property of target template to improve sparse based trackers (CST). High dimensional features can be embedded into CST to significantly improve tracking performance without sacrificing much computation time. Ma *et al.* [38] introduced an online random fern classifier as a re-detection component for long-term tracking, and used the correlation between temporal context to learn discriminative correlation filters from the most confident frames to estimate the scale change.

With the rise of neural networks [39]–[42], convolutional neural networks (CNNs) based trackers [43], [44] have obtained the excellent performance. Danelljan *et al.* [10] proposed a continuous convolution filters for tracking with multi-scale deep features to account for appearance variation caused by large scale change. Ma *et al.* [7] used hierarchical convolutional features of CNNs as target representations, and hierarchically inferred the maximum response of each layer to locate targets for handling large appearance variations and avoid drifting. Song *et al.* [45] presented the VITAL algorithm to address the problems of samples spatially overlapped and samples imbalance via adversarial learning. VIALT tracker can capture rich appearance variations in the tracking process. Lu *et al.* [46] applied residual connection to fuse multiple convolutional layers as well as their output response maps. This method further improves the performance of the regression tracker. Wang *et al.* [47] proposed Multi-Cue Correlation Filters which combining different types of features. It constructs multiple experts through Discriminative Correlation Filter (DCF) for robust tracking. In this work, we introduce KCF based on CNN features into WLMC methods as a local search method to improve the efficiency of the tracker and robustness about appearance variation.

III. WLMC AND KCF TRACKING ALGORITHMS

A. KCF TRACKING ALGORITHM

KCF is a discriminating tracking method, which generally trains a target detector during the tracking process and uses the target detector to detect whether the predicted position of the next frame is a target. The training set and the target detector (essentially regression) are updated with the new detection results. The target region is generally selected as a positive sample, and the surrounding region of the target as a negative sample. Of course, the closer it is to the target region, the more likely it is to be a positive sample. KCF uses the number in the range of [0,1] as the regression value of samples, so as to give different weights of samples under different offsets.

KCF constructs training samples by using cyclic shift of target window to form cyclic shift matrix of sample data. The purpose of this is, on the one hand, to maintain intensive sampling around the target, rather than random sampling; On the other hand, because of the particularity of cyclic shift matrix, the solution of the problem is transformed into the discrete Fourier domain. With the help of FFT transform, the convolution is transformed into the dot product of frequency domain, which avoids inverse matrix operation, reduces the complexity of operation and improves the processing speed.

At the same time, KCF maps ridge regression of linear space to nonlinear space through kernel function, and simplifies calculation by solving a dual problem in nonlinear space. The process of the KCF mentioned above can be described [35] as follows with the mathematical language.

In the linear ridge regression, the object of training in the linear ridge regression is to find a w that makes the function $f(z) = w^T z$ enable minimizing the squared error over samples x_i and their regression targets y_i

$$\min_w (f(x_i) - y_i)^2 + \lambda \|w\| \tag{1}$$

Let $\varphi(x)$ be a non-linear mapping from low dimension space to the high dimension space, and write the form of the dot product as $\varphi^T(x)\varphi(x') = \kappa(x, x')$, which is computed using the kernel function K (e.g., Gaussian or Polynomial), where x and x' are all different samples. Express the solution w as a linear combination of samples x_i. Consequently $f(z) = w^T z$ is converted into:

$$f(z) = w^T z = \sum_{i=1}^{n} \alpha_i \kappa(z, x_i) \tag{2}$$

where α_i is the element of the variable α which is becoming the alternative of w. Substituting Eq. (2) into Eq. (1), we can get the optimized solution which is given by [48]

$$\alpha = (K + \lambda I)^{-1} y \tag{3}$$

where K is the kernel matrix with the elements $K_{i,j} = \kappa\left(x_i, x_j\right)$ which represents the dot product between all pairs of samples, y means the regression target and λ is a regularization parameter that controls overfitting. It is natural for us to diagonalize Eq. (3) when we selected the appropriate

kernels to make K circulant. Thus we can easily calculate $\hat{\alpha}$ corresponding to the counterpart of α shown in Eq. (3).

$$\hat{\alpha} = \frac{\hat{y}}{\hat{K}^{xx} + \lambda} \tag{4}$$

where $\hat{K}^{xx}$ is the first row of the kernel matrix K, and "$\hat{}$" represents discrete Fourier transform (DFT) of a vector. If Gaussian kernel is adopted in dealing with multiple channels, and a vector x concatenating the individual vectors for C channels is given as $= [x_1, x_2, \cdots, x_c]$, then

$$\hat{K}^{xx} = \exp(-\frac{1}{\sigma^2}(\|x\|^2 + \|x'\|^2 - 2F^{-1}(\sum_c \hat{x}_c^* \odot \hat{x}_c'))) \tag{5}$$

where F^{-1} represents the inverse DFT, and σ is the variance of the Gaussian function. Suppose the new sample is z, a confidence map y can be obtained by:

$$y = F^{-1}(\hat{K}^{xz}) \odot z \tag{6}$$

where the multiplication $\odot$ is performed element-wise. The position with a maximum value y can be predicted as new position of the detecting object.

B. WLMC TRACKING ALGORITHM

The WLMC sampling method is one of the methods used to accurately estimate DOS in the statistical physics. Kwon Junseok et al. combined WLMC sampling method with MCMC method in abrupt motion tracking [19]. This method consists of three steps: the proposal step, the acceptance step, and the estimation step. First of all, the state space S is defined by a set of all possible states under Bayesian framework: $S = \{(X_t^x, X_t^y, X_t^s) \mid (X_t^x, X_t^y, X_t^s) \in D\}$, where D denotes the domain of the states. Let the state of target $X = X_t^p \times X_t^s$ at time t, where X_t^p and X_t^s represent the position and scale states, respectively. Fig. 2(a) illustrates an example of the state. Then in order to cope with the abrupt motion of the target, S^P is divided into d disjoint subregions: $S_i^p, i = \{1, \cdots, d\}$, as the shown in Fig. 2(b).

1) PROPOSAL STEP

There are many cases, in which an abrupt motion occurs. In this paper, we only consider abrupt changes in the state space of position and assume that the scale of the target is smooth over time. The algorithm uses two different types of proposal density to propose position and scales, respectively. Gaussian perturbation is used to set a large variance in proposal density an showen in Eq. (7):

$$\begin{cases} Q\left(X_t^{x'}; X_t^x\right) = G(X_t^x; {\sigma_x}^2) \\ Q\left(X_t^{y'}; X_t^y\right) = G(X_t^y; {\sigma_y}^2) \end{cases} \tag{7}$$

where $X_t^{x'}$ and $X_t^{y'}$ denote the new x and y positions, respectively, σ_x and σ_y denote the proposal variance of x and y coordinates, respectively. For dealing with smooth changes in scale, the second-order autoregressive model is adopted in proposal step [20].

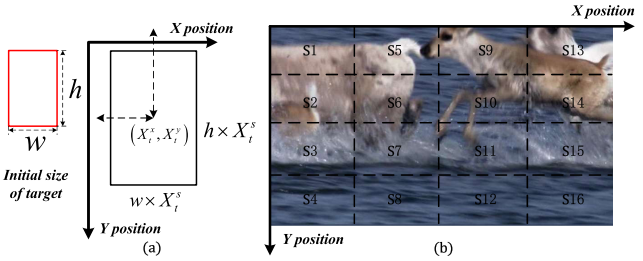


FIGURE 2. Example of state and subregion. (a) Target state representation $X = X_t^P \times X_t^S$. (b) State spaceS_i^P.

2) ACCEPTANCE STEP

Due to the large variance, the proposal density proposes many unnecessary states. WLMC method inserts the novel DOS term into the acceptance ratio of MCMC to form the acceptance function, it can solve this inefficiency in acceptance step. The acceptance ratio a is calculated by the ML and DOS terms as follows:

$$a = \min\left[1, \frac{p\left(Y_t|S_i^{p'}\right)\frac{1}{g(S_i^{p'})}Q\left(X_t; X_t'\right)}{p\left(Y_t|S_i^{p}\right)\frac{1}{g(S_i^{p})}Q\left(X_t'; X_t\right)}\right] \tag{8}$$

where $S_i^{p'}$ and S_i^p denote the subregions that contains the proposed state $X_t{}'$ and previous state X_t, respectively, $p\left(Y_t|S_i^p\right)$ indicates the ML score of the subregion S_i^p, and $g(S_i^p)$ is the DOS score of the subregion S_i^p. By using $\frac{1}{g(S_i^{p'})}$ in Eq. (8), when the DOS score of the subregion which include new state is larger than the old one, the acceptance ratio is low, and vice versa. For the ML term $p\left(Y_t|S_i^p\right)$, if it has a high ML score in new subregion, then the state will be accepted with a greater probability. By trading off the ML term and the DOS term, WLMC method can accept or reject a new state, and exploitation and exploration can be efficiently achieved at the same time.

3) ESTIMATION STEP

The estimation step is performed after the acceptance step. This step is used to update the DOS score $g(S_i^p)$ and the corresponding statistical histogram $h(S_i^p)$. WLMC tracking method approximately estimates DOS by using the Monte Carlo simulation, and states-of-density indicates the number of states in a subspace. Initially, the DOS of each subregion is initially set to1, and the histogram $h(S_i^p)$ is initialized to 0. Sampling process constructs Markova chains by visiting each subregions through random walk. For each sampling, if the proposed state is accepted, the histogram and DOS of the subregion are updated according to Eq. (9) and Eq. (10). Otherwise, WLMC method does the same works for the subregion to which includes the original state:

$$h(S_i^p) \leftarrow h(S_i^p) + 1 \tag{9}$$

$$g(S_i^p) \leftarrow g(S_i^p) \times f \tag{10}$$

where f is the modification factor, which is larger than one. As the method progresses, DOS is constantly updated in

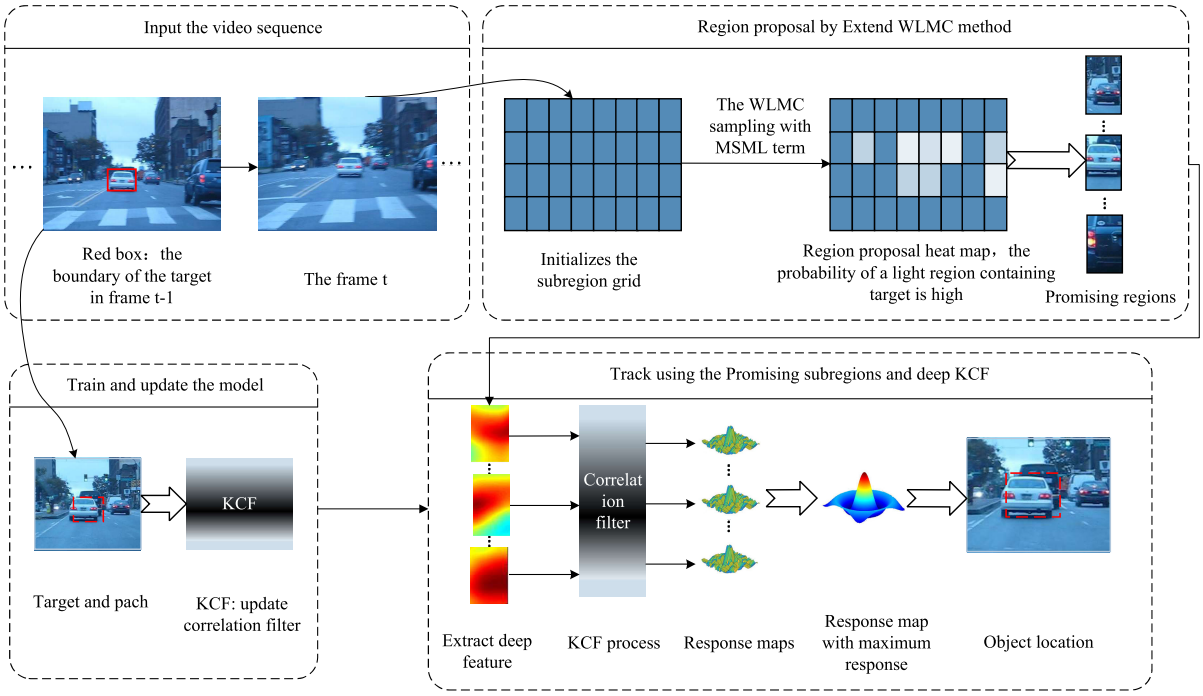


FIGURE 3. A brief illustration for our method.

accordance with the given setting, and the histogram $h(S_i^p)$ will be gradually increased in accordance with the algorithm, eventually generating a semiflat histogram. The semiflat histogram is defined as: if the value of the lowest bin is 80 percent larger than the average value of all bins in the histogram, then the histogram is considered to semiflat [49]. When the histogram has reached its semiflat one, the algorithm will adjust the modification factor to obtain more accurate DOS estimation, shown as Eq. (11):

$$f \leftarrow \sqrt{f} \tag{11}$$

At the same time, the histogram statistics are reset to 0 and the above estimation is redone until the histogram flattens again and starts over with a more accurate modification factor. When f approaches 1 or the number of iterations reaches the predefined value, the algorithm terminates.

IV. PROPOSED TRACKING ALGORITHM

In this section, we propose our tracking method. To obtain higher quality region proposal information in the process of WLMC sampling, we establish a reliable DOS distribution to propose the subregions that may contain the target. In this regard, we extend the WLMC method through the MSML, which can improve sample quality by introducing a multi-scope likelihood evaluation, so that the proposed subregions have a greater probability of containing the target. In order to improve efficiency, KCF is introduced into tracking to replace the complex sampling process, and the object location is obtained through the maximum response. At the same time, deep feature can improve the robustness and tracking precision of the tracker's appearance model. The proposed EWLCF-DP tracking framework is shown in Fig. 3. We describe each part of the algorithm in detail below.

In Fig. 3, when the tracker gets a new frame from the video, the 2-D density grid is initialized and the correlation filter is updated. In the preliminary sampling phase, the WLMC method with MSML is used to obtain the region proposal heat map (DOS distribution). According to the descending order of DOS score, T promising regions are found after the sampling phase. In the correlation filtering phase, extract the deep features of the promising regions. Then, the KCF process is carried out to obtain the T response maps of these subregions, respectively. Among all response maps, the location of the maximum response is considered the location of the target.

A. THE MULTI- SCOPE MARGINAL LIKELIHOOD

By using DOS and ML term, WLMC tracking method simulates the motions in the region far from the current local maximum or near the current local maximum, respectively. In the process of the simulating the motion near the current local maximum, the subregion ML term $p\left(Y_t|S_i^p\right)$ is used as the basis to decide whether to accept the new sample in the Eq. (8), or not. $p\left(Y_t|S_i^p\right)$ is approximated by:

$$p\left(Y_t|S_i^p\right) \leftarrow \frac{mp\left(Y_t|S_i^p\right)+p\left(Y_t|X_t^{m+1}\right)}{m+1}, \quad X_t^{m+1} \in S_i^p \tag{12}$$

where X_t^{m+1} is the $m+1-th$ sampled state of the subregion S_i^p, and m denotes the total number of the sampled states. Eq. (12) estimates the subregion ML score through the average likelihood degree of existing samples in the

subregion. The subregion ML is updated and used for the next sampling as the Monte Carlo simulation goes on. In the process of accepting new samples, Eq. (8) only compares the ML score between new subregion and old subregion. That is to say, the acceptance function seldom considers the similarity between samples and targets. Because of a large variance, Eq. (8) just considers the similarity between subregions may lead to the existence of many unnecessary samples. These low quality samples will reduce the ML score of the subregion which include the local maximum. And the probability of the sampler remaining in the region containing the local optimum is reduced. Under the updating strategy in Eq. (10), the DOS score for these subregions is still not prominent compared to other subregions. The reliability of region proposal strategy based on adaptive updating of the DOS distribution cannot be guaranteed. This problem causes the WLMC method to use a lot of low-quality regional recommendation information [19], [20].

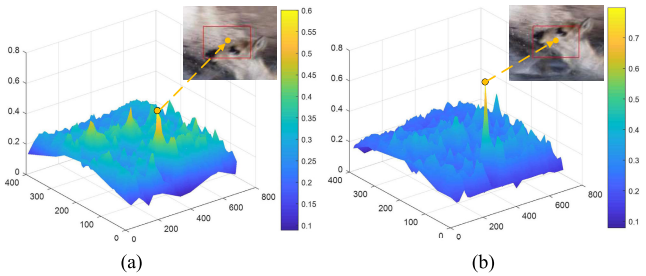
By improve the acceptance probability of samples in the subregion that containing the target state to improve samples quality, the reliability of DOS distribution can be guaranteed. To this end, we introduce MSML to guide the construction of Markov chain:

$$M\left(x_t{}';x_t\right)=\frac{p\left(Y_t;X_t{}'\right)p\left(Y_t|S_i^{p'}\right)}{p\left(Y_t;X_t\right)p\left(Y_t|S_i^{p}\right)} \tag{13}$$

where $x_t{}'$ donates new sample and x_t donates the old sample, $p\left(Y_t;X_t{}'\right)$ means the likelihood term over the sample $x_t{}'$, and $p\left(Y_t|S_i^{p'}\right)$ is calculated by Eq. (12). The MSML item contains the ML information in two scope of the sample and its subregion. Using MSML to improve Eq. (8), the MSML is inserted into the acceptance function:

$$a=\min\left[1,\frac{\frac{p(Y_t;X_t{}')p\left(Y_t|S_i^{p'}\right)}{g(S_i^{p'})}Q\left(X_t;X_t{}'\right)}{\frac{p(Y_t;X_t)p(Y_t|S_i^{p})}{g(S_i^{p})}Q\left(X_t{}';X_t\right)}\right] \tag{14}$$

In Eq. (14), we consider the similarity of samples and subregions to the target at the same time. When sampler is transferred between subregions, it tends to stay in subregions with higher regional likelihood, and samples within these regions will have a higher acceptance rate, and vice versa. However, even in the region with high regional likelihood, there are still a large number of low-quality samples, which will reduce the quality of the Markov chain. Therefore, we consider the transfer between samples to suppress the large acceptance of low-quality samples, the tracker will use the sample likelihood to control the samples that accept higher likelihood values, so as to ensure the quality of the samples accepted in the region. By defining the MSML in two scopes, the sample quality is introduced into the acceptance function, the subregion which include the target state will provide more valid samples. And the improved sampling quality makes the DOS distribution based on state transition update more reliable.



(a) (b)

FIGURE 4. Fitness distribution diagrams and matching results. X, Y and Z axes represent X, Y coordinates and fitness values of sample points, respectively. (a) Fitness distribution of WLMC (The maximum value is 0.592). (b) Fitness distribution of Extend WLMC containing MSML (The maximum value is 0.794).

Fig. 4 shows the impact of MSML on fitness distribution in tracking. Although the final result is still near the target, it only covers a part of targets. By analyzing the fitness distribution of the samples, the optimal state is not found, and the fitness distribution near the local optimal is relatively smooth. With the same sample size, the result of extended WLMC is closer to the target, and the crest is steeper. From the experimental results, the extended WLMC can accept more samples from the regions near the target, so that the DOS score in these subregions are higher. And ultimately the reliability of region proposal strategy is increased.

By using improved WLMC acceptance function, the reliability of promising regions that derived from DOS distribution is better guaranteed after the end of the sampling phase. Compared with AWLMC method, our tracker chooses $d/4$ subregions as promising regions. And these fewer and higher quality regions are used as search space for finding targets.

B. THE EWLCF-DP METHOD

Although region proposal strategy reduces the number of samples and improves the sampling efficiency of the algorithm to some degree, as shown in Fig. 5. However, this local search method based on random sampling has two disadvantages: 1) Due to the random walk, it is still inevitable that the tracker needs long iterations to reach the promising object state. And each iteration requires a relatively complex evaluation with low efficiency. 2) Traditional sampling trackers are mostly based on hand-crafted features (for example, HOG or color feature), and the robustness of the appearance model is insufficient.

To further improve the search efficiency of tracking algorithm, we introduce the KCF to the WLMC method. By replacing the traditional sampling method with the correlation filtering process, the target position is determined. In this phase, our tracker is to learn a correlation filters online so as to localize the object in consecutive frames by identifying the location of maximal correlation response from candidate regions of interest. The long iterations can be simplified. In addition, CF can transform the time-consuming convolution operation in time domain to an element-wise multiplication in Fourier domain, which enables the tracker to have a low computational cost. Recently, features based

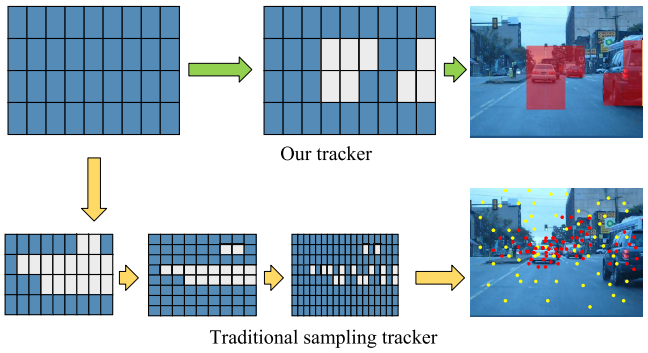


FIGURE 5. Examples of search space in traditional trackers and our tracker. The selected subregions color is white. Red mask means correlation filters. The yellow and red dots represent the samples at different sampling stages respectively. For traditional sampling trackers, proposal regions are gradually shrunk and refined as the sampling goes on. For our method, we are using preliminary region proposal to simplified iterative optimization process. The complex sampling process in the spatial domain is replaced by correlation filters in the frequency domain.

on convolutional neural networks (CNNs) have demonstrated state-of-the-art results on a wide range of visual recognition tasks. To further improve tracking precision and robustness, we use the convolutional feature maps, VGG-Net [44], to encode target appearance. Compared with hand-crafted features, KCF based CNNs tracker can significantly improve its robustness to geometric and appearance.

C. THE EWLCF-DP TRACKER

The proposed sampling-prediction-filter scheme substantially speeds up the overall sampling process of abrupt motion tracking. We utilize a certain number of iterations to quickly explore the whole state space (image). Here, the DOS is selected to be the predictive model for searching the promising regions of the sample space because of its computational efficiency. In local neighborhood-searching, deep KCF is introduced into tracking to locate the target.

For the appearance model, the HOG was selected as the feature in sampling step. It has the invariance to local geometric and photometric transformations. In the process of searching the whole state space, the efficiency of the algorithm can be guaranteed by using hand-crafted features which have less computation. In CF phase, tracker needs to obtain the accurate position of the target through the correlation filtering process with less evaluation times. The deep feature is selected in this step because it is more robust to appearance variation. The deep feature can further improve the precision of the tracker. The proposed tracking framework is summarized in Algorithm 1.

Algorithm 1 EWLCF-DP Tracking Framework

Input: Image sequence.
Output: Object state in each frame.
1: Sampling phase.
2: **for** 1 to N
3: Proposal step: Propose a new state using (7).
4: Acceptance step:
5: 1. Calculating acceptance ratio using (14).
6: 2. Accepting or rejecting sample.
7: Estimation step:
8: 1. Update the histogram using (9).
9: 2. Estimate the DOS using DOS using (10).
10: 3. Estimate the marginal-likelihoods using (12).
11: 4. Adjust the modification factor f by the process explained in Section 3.
12:**end**
13: Correlation Filters phase.
14: Using DOS score to sort subregions in descending order.
15: The first $d/4$ subregions was selected for promising regions.
16: Using KCF process to obtain d/4 response maps of promising regions.
17: Output the location of the maximum response as the target location.
18: Update correlation filters model.

V. EXPERIMENTS

A. EXPERIMENTAL SETUP AND METHODOLOGY

In this paper, our method is implemented in MATLAB R2017a on Windows 10. The experiments are conducted on a PC with Intel Core i3 3.60GHz, 8GB RAM and a 1060 GPU. The average tracking speed is 9.7 FPS at OTB-100. $N = 300$ particles were used for EWLMC step. σ_x and σ_x in Eq.(7) was set to 250. f in Eq.10 was set to 2.7. Correlation coefficient is used as the fitness function. And the state space was divided to 4×8 subregions. The values of parameters were fixed throughout all experiments. To prove performance of our EWLCFDP tracker, Experiment is divided into two groups. The first group tests the EWLCFDP tracker on 10 sequences which contains abrupt motion. And in the second group, the proposed method was compared to 10 state-of-the-art trackers in OTB dataset [50].

We follow the standard evaluation metrics from the benchmarks. For the OTB-50 and OTB-100, we use the one-pass evaluation (OPE) with precision and success plots metrics. The precision metric measures the rate of frame locations within a certain threshold distance from those of the ground truth. The threshold distance is set as 20 for all the trackers. The success plot metric measures the overlap ratio between predicted and ground truth bounding boxes.

B. ABRUPTION MOTION SEQUENCES

To prove the proposed method can track visual object successfully, especially for the object with the abrupt motion, we test our method on 10 challenging videos, whose motion displacement are more than 70 pixels. Targets displacement between image frames is listed in Table.1. At the same time, we compared our tracker EWLCFDP with 5 state-of-the-art trackers, including TADT [5], HCFTstar [51], CF2 [7], DCF_CA [52], KCFDP [53].

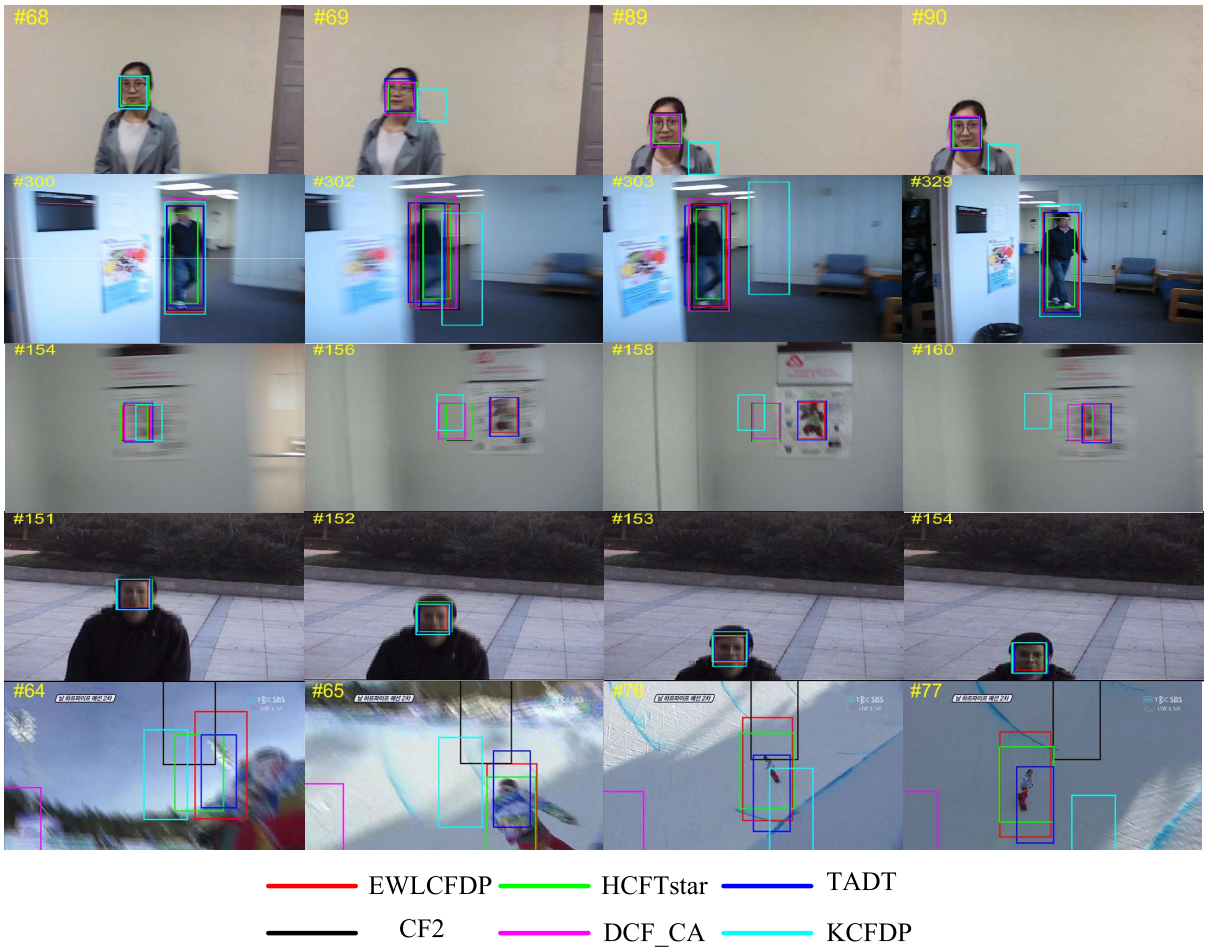


FIGURE 6. The qualitative results of trackers on 5 challenging sequences. (From top to down: Zxj, Blurbody, Blurowl, Face2 and Snowboard, respectively.)

TABLE 1. The image sequences and max displacement.

Sequences	Frame	X axis	Y axis	Max
Zxj	118	70	18	70
Blurbody	334	76	26	76
Blurowl	631	76	77	77
Face2	310	29	90	90
Snowboard	113	94	82	94
DragonBaby	113	94	96	96
Fhc	123	188	104	188
Zt	115	256	149	256
Boxing	813	287	18	287
Youngki	770	296	36	296

1) QUALITATIVE EVALUATION

Fig. 6 shows some results of the first five sequences. The displacement of these sequences is more than 70 pixels and less than 94 pixels. For Zxj sequence in the first line of figure 10, before the frame #68, all tracker work well. However, KCFDP loses the target at #69 due to the abrupt motion.

For Blurbody and Blurowl sequences in second and third line respectively, motion blur occurs simultaneously with the abrupt motion. KCFDP deviate the target since abrupt motion, but it can recover tracking. KCFDP, DCF_CA, CF2 and HCFTstar lose the target in BLUROWL sequence. Our method and TADT obtain the best performance.

For Snowboard sequence in the last line, a massive change has taken place in the appearance of the target with the camera switch. At #64 and #65, our method captures the target after camera switch.

Fig. 7 shows some results of the last five sequences. The displacement of these sequences is more than 96 pixels and less than 296 pixels. For DragonBaby sequence in first line, TADT, DCF_CA and KCFDP lose the target due to the abrupt motion at #42. And at #79 and #80, our tracker, HCFTstar and CF2 obtain the best performance after camera switch.

For Fhc and Zt sequences in second and third line respectively, all trackers have run at a high resolution video sequence and had better results. But due to the abrupt motion, KCFDP still drift in Zt.

For Boxing sequence in the fourth line, this video is a boxing match. The athlete moves very quickly and cam-

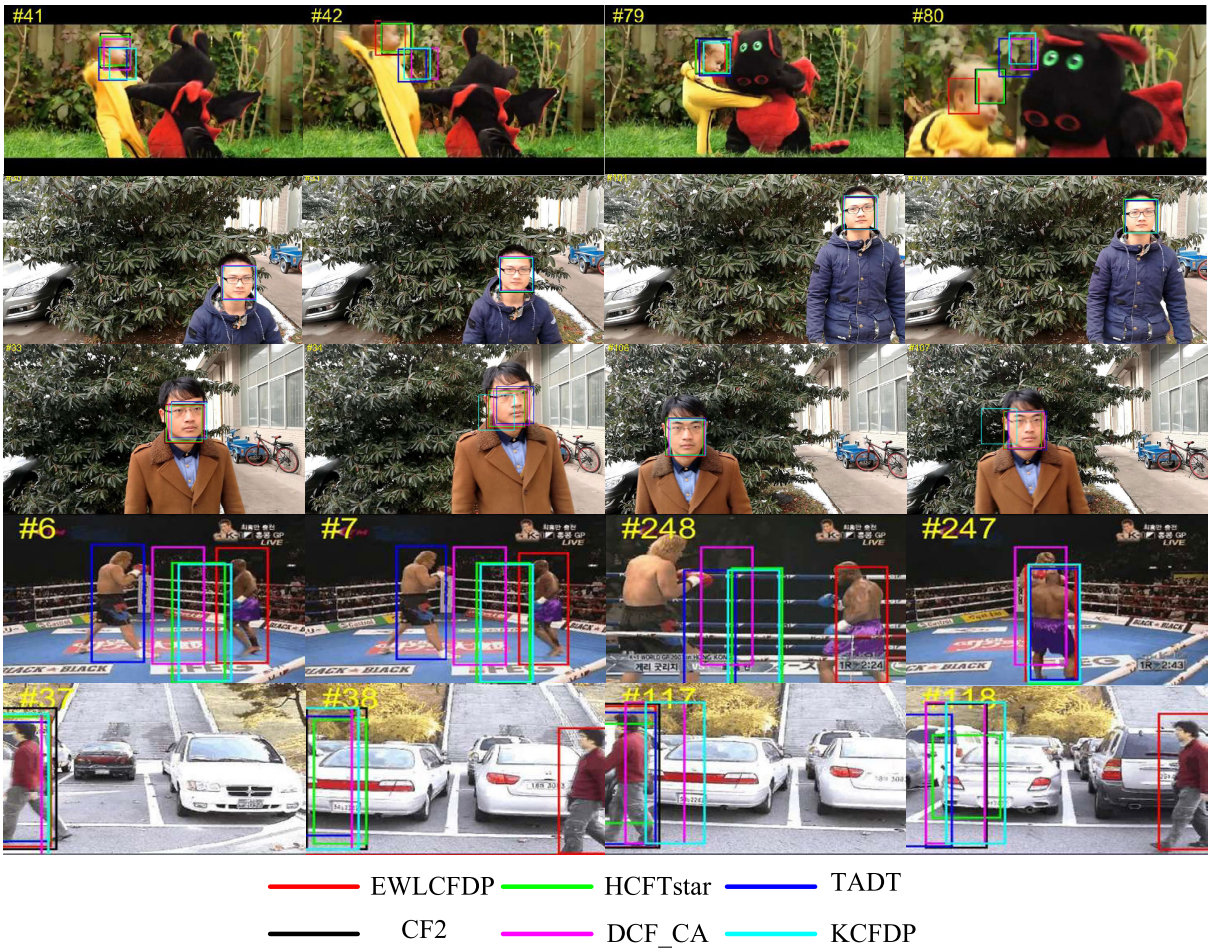


FIGURE 7. The qualitative results of trackers on 5 challenging sequences. (From top to down: DragonBaby, Fhc, Zt, Boxing and Youngki, respectively.)

eras change with the game. Because of abrupt motion, most trackers lose the target at #6 and #7, but our tracker captures the target. After camera switch at #247 and #248, other trackers lose the target, and our tracker obtain the best performance.

For Youngki sequence in the last line, a person walks in video. He walks out from one side at #37 and appears from the other side at #38. All trackers lost their target except our tracker. This condition was repeated at #117 and #118, and our tracker still has best results.

2) QUANTITATIVE ANALYSIS

We compare our EWLCFDP tracker with 5 state-of-the-art trackers, including TADT [5], HCFTstar [51], CF2 [7], DCF_CA [52], KCFDP [53], on 10 challenge sequences which are listed in Table.1. Fig. 8 show the results of one-pass evaluation using the distance precision and overlap success rate. It indicates that our EWLCFDP tracker performs better than TADT which use the Siamese Network. The local search strategy makes it difficult for TADT to keep track of the target in the face of uncertain movement beyond the search scope, so the accuracy performance is lower than that of HCFTstar and our method. At the same time, because HCFTstar can detect target lost and recover tracking by the re-detected module, robust tracking results can still be obtained when facing a certain range of displacement. But is still cannot recover from excessive displacement. CF2, DCF_CA and KCFDP which based on correlation filter have poor adaptability to uncertain motion. Overall, our method achieves the best results in these challenge sequences.

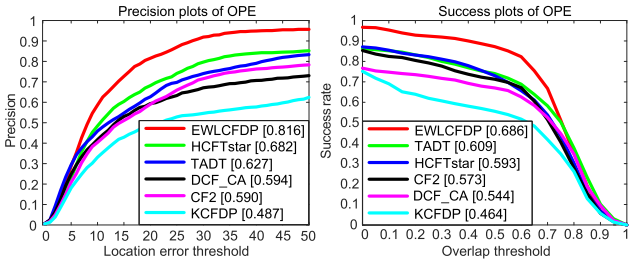


FIGURE 8. Precision and success plots over all 10 sequences using one-pass evaluation. The legend contains the area-under-the-curve score and the average distance precision score at 20 pixels for each tracker.

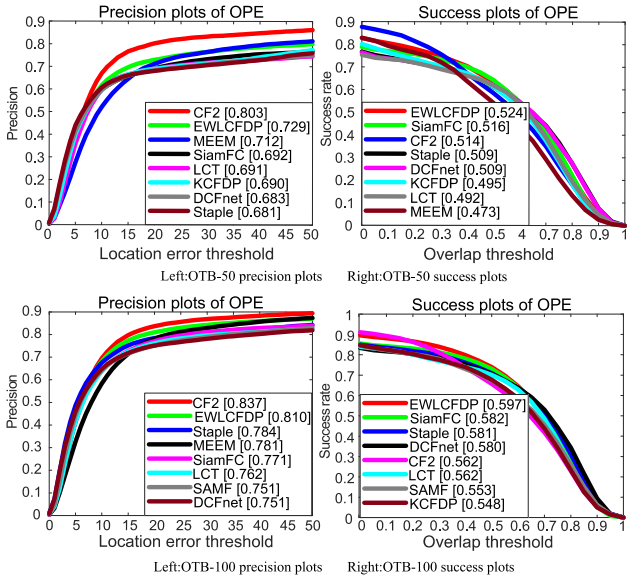


FIGURE 9. Average overall performance on OTB-50 and OTB-100.

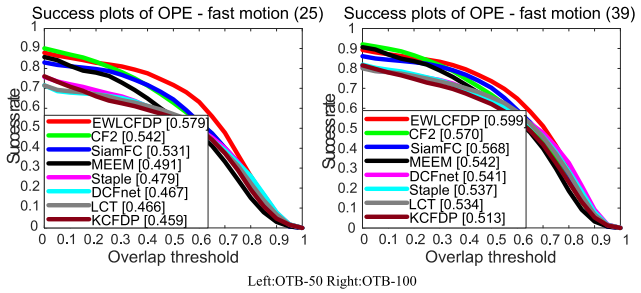


FIGURE 10. Fast motion sequences performance on OTB-50 and OTB-100.

C. OVERALL PERFORMANCE

We evaluate the proposed tracker on OTB-50 and OTB-100. The proposed algorithm is compared with 10 advanced trackers, such as CF2 [7], MEEM [54], SAMF [55], SiamFC [56], LCT [38], KCFDP [53], DCFnet [57], Staple [58], KCF [35], DSST [36]. In the following, we will present the results and analyses.

The OTB-50 dataset with 50 sequences and the extended OTB-100 dataset with additional 50 sequences are two widely used tracking benchmarks. The sequences in the OTB datasets include a wide variety of tracking challenging, such as illumination variation, scale variation, deformation, occlusion, fast motion, rotation, and background clutters. The OTB benchmark adopts Center Location Error (CLE) and Overlap Ratio (OR) as the base metrics. Based on CLE and OR, the precision and success plots are used to evaluate the overall tracking performance. The precision plot measures the percentage of frames whose CLE is within a given threshold, which is usually set to 20 pixels. The success plot computes the percentage of the successful frames whose OR is larger than a given threshold. The area under the curve (AUC) of the success plot is mainly used to rank tracking algorithms. This experiment is used one-pass evaluation (OPE) method to initialize the tracking algorithm. OPE is the most common evaluation method which runs trackers on each sequence once. It initializes the trackers with the ground truth object state in the first frame and reports the average precision or success rate of all the results. For more details about the adopted evaluation methodology, we refer the reader to [50].

1) QUANTITATIVE ANALYSIS

Fig. 9 shows the average precision scores at a threshold of 20 pixels and the average success scores at AUC model. The best result for each attribute is highlighted in red, and the sub-optimal result is in green. Among all the trackers, our EWLCFDP tracker performs favorably on both the distance precision and overlap success rate. Our tracker used similar feature layers as CF2, but our tracker has a good success score by Sampling-Predict- Filters model. When the target location undergoes obvious changes or escapes the search area, existing DCF based trackers can not accurately locate the target object.

Fig. 10 shows the success score of fast motion tracking challenges at OTB-50 and OTB-100 in AUC model. For the fast motion, our EWLCFDP tracker get the highest success score at OTB-100. DCFnet tracker treat discriminative correlation filters (DCF) as a special correlation filter layer added in a Siamese network, it develops a network to automatically learn the features. But DCFnet lacks target position prediction, so that it prone to failure after large movement of the target. And this automatic learning affects the subsequent tracking process to some extent. Our tracker obtains the approximate location of the target and provides it to the correlation filters. This strategy makes the update of the KCF model is as useful as possible. It is observed that the proposed approach almost achieves the best or sub-optimal result. This clearly demonstrates that the region proposal strategy by enhanced DOS distribution is very effective.

2) QUALITATIVE EVALUATION

Fig. 11 shows some results of the top performing trackers: CF2, DCFnet, Staple, MEEM, SiamFC and our EWLCFDP tracker on 6 challenging sequence. To prove the proposed method can track visual object successfully, especially for the object with the abrupt motion, these sequences all include motion displacement which more than 30 pixels. The Staple tracker does not perform well in all the presented sequences. Because it adopts empirical operations for feature extraction. Although it combines color changes and deformations, but usually abrupt motion is often accompanied by motion blur, handcrafted features with limited performance are not able to differentiate the target and background. In comparison, due to the deep learning, other trackers have a good performance when face to the motion blur by abruption. However, when targets move quickly between frames, these trackers cannot adapt well. For sequence 'DragonBaby', and 'Ironman', these sequences captured from the film are easily

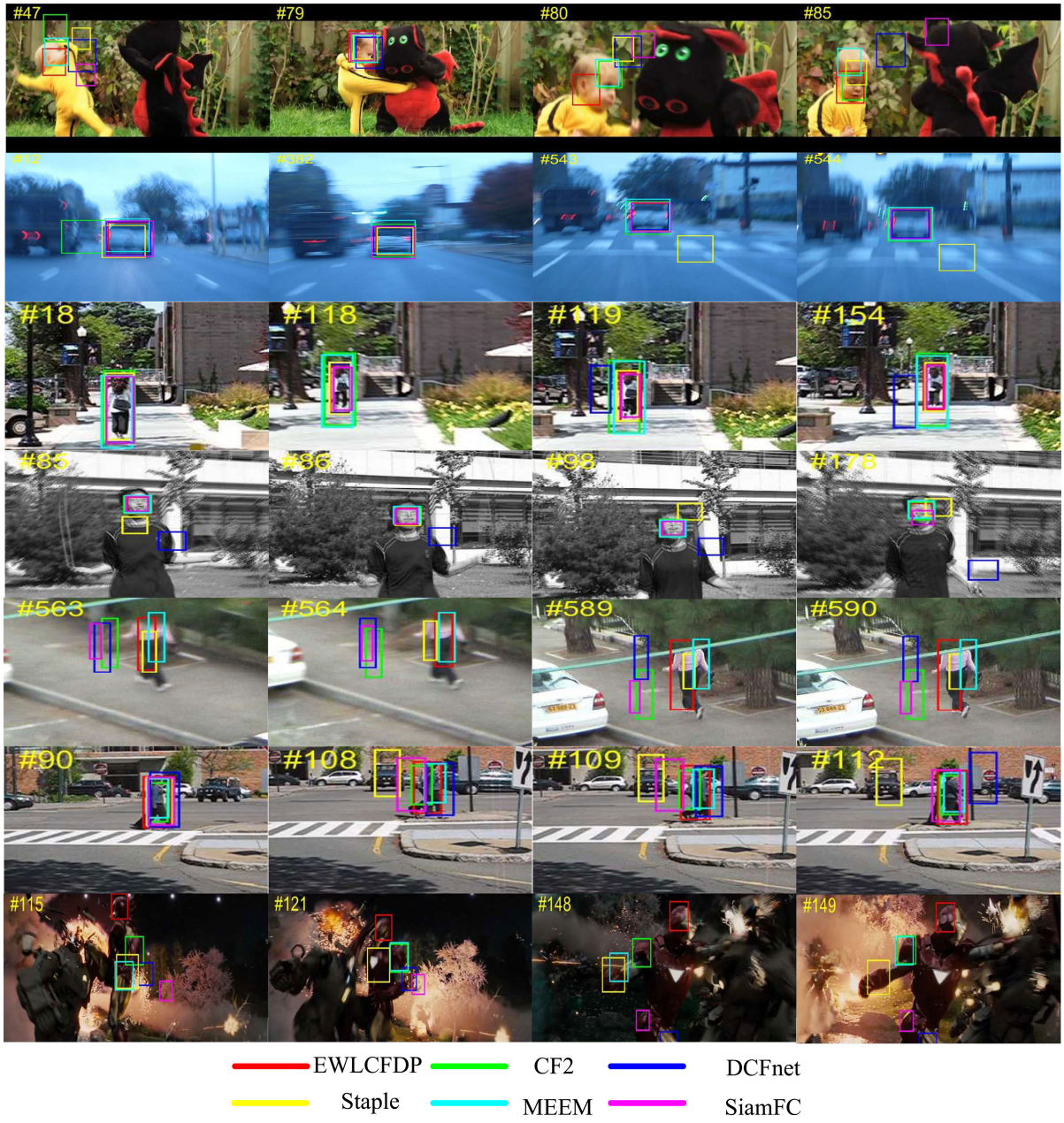


FIGURE 11. The qualitative results of trackers on 7 sequences (from top to down: DragonBaby, Blurcar1, Human7, Jumping, Woman, Couple and Ironman, respectively).

switching camera. Our tracker predicts the location of the targets through EWLMC model, allows the tracker to adapt to such large changes in position from camera switching. Our EWLCFDP algorithm further exploits the model potential through region proposal strategy and strengthen this strategy through multi-scope assessments. As a result, the response map predicted by our tracker is more accurate for target exploration, especially in the presented abrupt motion scenarios. Overall, the visual evaluation indicates that our EWLCFDP tracker performs favorably against advanced trackers.

VI. CONCLUSION

We have given an approach for uncertain motion tracking. It is inspired by fuzzy control [59], [60] and chaos control [61] to deal with uncertain problems, we have introduced the MSML into the acceptance function of WLMC to guarantee the reliability of proposal regions that derived from DOS distribution. Furthermore, we have introduced the KCF algorithm to WLMC sampler. Compere with WLMC tracker, this method can reduce the sample size and increase the tracking efficiency. Considering the appearance adaptation, we utilize deep feature to further strengthen the robustness

to the appearance changes and the background distractions. Extensive experiments have indicated that our method outperforms other alternatives and exhibit better efficiency and effectiveness in the tracking of abrupt motion.

REFERENCES

[1] T. Zhou, H. Bhaskar, F. Liu, J. Yang, and P. Cai, "Online learning and joint optimization of combined spatial-temporal models for robust visual tracking," *Neurocomputing*, vol. 226, pp. 221–237, Feb. 2016.

[2] T. Zhou, F. Liu, H. Bhaskar, J. Yang, H. Zhang, and P. Cai, "Online discriminative dictionary learning for robust object tracking," *Neurocomputing*, vol. 275, pp. 1801–1812, Jan. 2018.

[3] T. Zhou, H. Bhaskar, F. Liu, and J. Yang, "Graph regularized and locality-constrained coding for robust visual tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 10, pp. 2153–2164, Oct. 2017.

[4] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2411–2418.

[5] X. Li, C. Ma, B. Wu, Z. He, and M.-H. Yang, "Target-aware deep tracking," Apr. 2019, *arXiv:1904.01772*. [Online]. Available: https://arxiv.org/abs/1904.01772

[6] T. Zhang, C. Xu, and M.-H. Yang, "Multi-task correlation particle filter for robust object tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 4335–4343.

[7] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3074–3082.

[8] Y. Qi, S. Zhang, L. Qin, H. Yao, Q. Huang, J. Lim, and M.-H. Yang, "Hedged deep tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4303–4311.

[9] M. Danelljan, G. Hager, F. S. Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2015, pp. 58–66.

[10] M. Danelljan, A. Robinson, F. S. Khan, and M. Felsberg, "Beyond correlation filters: Learning continuous convolution operators for visual tracking," in *Proc. 14th Eur. Conf.*, Amsterdam, The Netherlands, Oct. 2016, pp. 472–488.

[11] Y. Hua, K. Alahari, and C. Schmid, "Online object tracking with proposal selection," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 3092–3100.

[12] H. Zhang, Y. Wang, L. Luo, X. Lu, and M. Zhang, "SIFT flow for abrupt motion tracking via adaptive samples selection with sparse representation," *Neurocomputing*, vol. 249, pp. 253–265, Aug. 2017.

[13] J. Deutscher, A. Blake, and I. Reid, "Articulated body motion capture by annealed particle filtering," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, vol. 2, Jun. 2000, pp. 126–133.

[14] X. Zhou and Y. Lu, "Abrupt motion tracking via adaptive stochastic approximation Monte Carlo sampling," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1847–1854.

[15] X. Zhou, Y. Lu, J. Lu, and J. Zhou, "Abrupt motion tracking via intensively adaptive Markov-chain Monte Carlo sampling," *IEEE Trans. Image Process.*, vol. 21, no. 2, pp. 789–801, Feb. 2012.

[16] W. Li, X. Zhang, and W. Hu, "Contour tracking with abrupt motion," in *Proc. 16th IEEE Int. Conf. Image Process. (ICIP)*, Nov. 2009, pp. 3593–3596.

[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," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 30, no. 10, pp. 1728–1740, Oct. 2008.

[18] Y. Su, Q. Zhao, L. Zhao, and D. Gu, "Abrupt motion tracking using a visual saliency embedded particle filter," *Pattern Recognit.*, vol. 47, no. 5, pp. 1826–1834, 2014.

[19] J. Kwon and K. M. Lee, "Tracking of abrupt motion using wang-landau Monte Carlo estimation," in *Proc. 10th Eur. Conf. Comput. Vis.*, Marseille, France, Oct. 2008, pp. 387–400.

[20] J. Kwon and K. M. Lee, "Wang-Landau Monte Carlo-based tracking methods for abrupt motions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 35, no. 4, pp. 1011–1024, Apr. 2013.

[21] M. Isard and A. Blake, "CONDENSATION—Conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, Aug. 1998.

[22] A. Del Bimbo and F. Dini, "Particle filter-based visual tracking with a first order dynamic model and uncertainty adaptation," *Comput. Vis. Image Understand.*, vol. 115, no. 6, pp. 771–786, Jun. 2011.

[23] S. Oron, A. Bar-Hillel, D. Levi, and S. Avidan, "Locally orderless tracking," *Int. J. Comput. Vis.*, vol. 111, no. 2, pp. 213–228, Jan. 2015.

[24] Z. Khan, T. Balch, and F. Dellaert, "MCMC-based particle filtering for tracking a variable number of interacting targets," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 11, pp. 1805–1819, Nov. 2005.

[25] D.-N. Cong, F. Septier, C. Garnier, L. Khoudour, and Y. Delignon, "Robust visual tracking via MCMC-based particle filtering," in *Proc. IEEE Int. Conf. Acoust., Speech Signal Process. (ICASSP)*, Mar. 2012, pp. 1493–1496.

[26] J. Kwon and K. M. Lee, "Visual tracking decomposition," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 1269–1276.

[27] G. O. Roberts and J. S. Rosenthal, "Examples of adaptive MCMC," *J. Comput. Graph. Statist.*, vol. 18, no. 2, pp. 349–367, 2009.

[28] G. O. Roberts, A. Gelman, and W. R. Gilks, "Weak convergence and optimal scaling of random walk metropolis algorithms," *Ann. Appl. Probab.*, vol. 7, no. 1, pp. 110–120, 1997.

[29] F. Wang, B. Lin, J. Zhang, and X. Li, "Object tracking using langevin Monte Carlo particle filter and locality sensitive histogram based likelihood model," *Comput. Graph.*, vol. 70, pp. 214–223, Feb. 2018.

[30] F. Wang, P. Li, X. Li, and M. Lu, "Ordered over-relaxation based langevin Monte Carlo sampling for visual tracking," *Neurocomputing*, vol. 220, pp. 111–120, Jan. 2017.

[31] T. Zhou, Y. Lu, F. Lv, H. Di, Q. Zhao, and J. Zhang, "Abrupt motion tracking via nearest neighbor field driven stochastic sampling," *Neurocomputing*, vol. 165, pp. 350–360, Oct. 2015.

[32] J. T. Mbelwa, Q. Zhao, Y. Lu, H. Liu, F. Wang, and M. Mbise, "Objectness-based smoothing stochastic sampling and coherence approximate nearest neighbor for visual tracking," *Vis. Comput.*, vol. 35, no. 3, pp. 371–384, Mar. 2019.

[33] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, Jun. 2010, pp. 2544–2550.

[34] J. A. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proc. 12th Eur. Conf. Comput. Vis.*, Florence, Italy, Oct. 2012, pp. 702–715.

[35] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[36] M. Danelljan and G. Häger, F. Khan, and M. Felsberg, "Accurate scale estimation for robust visual tracking," in *Proc. Brit. Mach. Vis. Conf.*, Sep. 2014, pp. 1–5.

[37] T. Zhang, A. Bibi, and B. Ghanem, "In defense of sparse tracking: Circulant sparse tracker," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 3880–3888.

[38] C. Ma, X. Yang, C. Zhang, and M.-H. Yang, "Long-term correlation tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 5388–5396.

[39] X. Xie, S. Wen, Z. Zeng, and T. Huang, "Memristor-based circuit implementation of pulse-coupled neural network with dynamical threshold generators," *Neurocomputing*, vol. 284, pp. 10–16, Apr. 2018.

[40] G. Ren, Y. Cao, S. Wen, Z. Zeng, and T. Huang, "A modified Elman neural network with a new learning rate scheme," *Neurocomputing*, vol. 286, pp. 11–18, Apr. 2018.

[41] S. Xiao, X. Xie, S. Wen, Z. Zeng, T. Huang, and J. Jiang, "GST-memristor-based online learning neural networks," *Neurocomputing*, vol. 272, pp. 677–682, Jan. 2017.

[42] S. Wen, X. Xie, Z. Yan, T. Huang, and Z. Zeng, "General memristor with applications in multilayer neural networks," *Neural Netw.*, vol. 103, pp. 142–149, Jul. 2018.

[43] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proc. Adv. Neural Inf. Process. Syst. (NIPS)*, 2012, pp. 1097–1105.

[44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," Sep. 2014, *arXiv:1409.1556*. [Online]. Available: https://arxiv.org/abs/1409.1556

[45] Y. Song, C. Ma, X. Wu, L. Gong, L. Bao, W. Zuo, C. Shen, R. W. Lau, and M.-H. Yang, "VITAL: Visual tracking via adversarial learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 8990–8999.

[46] X. Lu, C. Ma, B. Ni, X. Yang, I. Reid, and M.-H. Yang, "Deep regression tracking with shrinkage loss," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, Sep. 2018, pp. 353–369.
[47] N. Wang, W. Zhou, Q. Tian, R. Hong, M. Wang, and H. Li, "Multi-cue correlation filters for robust visual tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 4844–4853.
[48] R. Rifkin, G. Yeo, and T. Poggio, "Regularized least-squares classification," *Nato Sci. Ser. Sub Ser. Comput. Syst. Sci.*, vol. 190, pp. 131–154, May 2003.
[49] 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, p. 2050, 2001.
[50] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.
[51] C. Ma, J.-B. Huang, X. Yang, and M.-H. Yang, "Robust visual tracking via hierarchical convolutional features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 11, pp. 2709–2723, Nov. 2019.
[52] M. Mueller, N. Smith, and B. Ghanem, "Context-aware correlation filter tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1396–1404.
[53] D. Huang, L. Luo, M. Wen, Z. Chen, and C. Zhang, "Enable scale and aspect ratio adaptability in visual tracking with detection proposals," in *Proc. Brit. Mach. Vis. Conf. (BMVC)*, X. Xie, M. W. Jones, and G. K. L. Tam, Eds. BMVA Press, Sep. 2015, pp. 185.1–185.12, doi: 10.5244/C.29.185.
[54] J. Zhang, S. Ma, and S. Sclaroff, "MEEM: Robust tracking via multiple experts using entropy minimization," in *Proc. 13th Eur. Conf.*, Zurich, Switzerland, Sep. 2014, pp. 188–203.
[55] Y. Li and J. Zhu, "A scale adaptive kernel correlation filter tracker with feature integration," in *Proc. Eur. Conf. Comput. Vis.*, Zurich, Switzerland, Sep. 2014, pp. 254–265.
[56] L. Bertinetto, J. Valmadre, J. F. Henriques, A. Vedaldi, and P. H. Torr, "Fully-convolutional siamese networks for object tracking," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, Oct. 2016, pp. 850–865.
[57] Q. Wang, J. Gao, J. Xing, M. Zhang, and W. Hu, "DcFNet: Discriminant correlation filters network for visual tracking," Apr. 2017, *arXiv:1704.04057*. [Online]. Available: https://arxiv.org/abs/1704.04057
[58] L. Bertinetto, J. Valmadre, S. Golodetz, O. Miksik, and P. H. S. Torr, "Staple: Complementary learners for real-time tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 1401–1409.
[59] S. Wen, Z. Q. M. Chen, X. Yu, Z. Zeng, and T. Huang, "Fuzzy control for uncertain vehicle active suspension systems via dynamic sliding-mode approach," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 1, pp. 24–32, Jan. 2017.
[60] S. Wen, T. Huang, X. Yu, Z. Q. M. Chen, and Z. Zeng, "Aperiodic sampled-data sliding-mode control of fuzzy systems with communication delays via the event-triggered method," *IEEE Trans. Fuzzy Syst.*, vol. 24, no. 5, pp. 1048–1057, Oct. 2016.
[61] J. Sun, X. Zhao, J. Fang, and Y. Wang, "Autonomous memristor chaotic systems of infinite chaotic attractors and circuitry realization," *Nonlinear Dyn.*, vol. 94, no. 4, pp. 2879–2887, 2018.

HUANLONG ZHANG received the Ph.D. degree from the School of Aeronautics and Astronautics, Shanghai Jiao Tong University, China, in 2015. He is currently an Associate Professor with the College of Electric and Information Engineering, Zhengzhou University of Light Industry, Henan, Zhengzhou, China. His research has been funded by the National Natural Science Foundation of China (NSFC), the Key Science and Technology Program of Henan Province *et al.* He has published more than 40 technical articles in refereed journals and conference proceedings. His research interests include pattern recognition, machine learning, image processing, computer vision, and intelligent human–machine systems.

GUOHAO NIE was born in Zhengzhou, Henan, China, in 1996. He is currently pursuing the degree with the Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include object detection and visual tracking.

JIAN CHEN was born in Luohe, Henan, China, in 1995. He is currently pursuing the degree with the Zhengzhou University of Light Industry, Zhengzhou, China. His research interests include deep learning and visual tracking.

JIE ZHANG received the Ph.D. degree in control theory and control engineering from the Harbin Institute of Technology, in 2018. He is currently with the Zhengzhou University of Light Industry. His research interests include compressed sensing, image denoising, and signal processing.

GUOSHENG YANG received the M.S. degree from the Harbin Institute of Technology, China, in 1988, and the Ph.D. degree in control theory and control engineering from the Beijing Institute of Technology, China, in 2002. He is currently a Professor with the School of Information Engineering, Minzu University of China, Beijing, China, where he directs the Pattern Recognition and Intelligent Systems Research Laboratory. He has published several books and more than one hundred technical articles in refereed journals and conference proceedings. His research interests mainly include computer vision, brain inspired computing, multimodality information fusion, pattern recognition, and artificial intelligence. His research has been funded by the National Natural Science Foundation of China (NSFC), the National Science and Technology Support Program, China, and the National High-tech Research and Development Program, China. He has been a Reviewer for many professional journals and conferences.