Vis Comput (2016) 32:307–320
DOI 10.1007/s00371-015-1067-1

ORIGINAL ARTICLE

Visual tracking with semi-supervised online weighted multiple instance learning

Zhihui Wang · Sook Yoon · Shan Juan Xie · Yu Lu · Dong Sun Park

Published online: 24 February 2015

Abstract Adaptive discriminative tracking is a new research topic that has attracted broad attention due to its extensive application value. To take full advantage of the information about targets and their surrounding background, we propose a novel single object tracking-by-detection tracker in this paper, combining semi-supervised learning, multiple instance learning and the Bayesian theorem. The tracker uses a block-based inconsistency function of the labeled and unlabeled training samples in the selection of optimal weak classifiers during the parameter updating phase of each frame. Experimental results showed that the proposed tracker has excellent performance over other eight state-of-the-art trackers for thirteen open-access video sequences.

Keywords Multiple instance learning · Semi-supervised learning · Weak classifier · Unlabeled sample · Inconsistency function

Z. Wang · Y. Lu
Department of Electronics Engineering, Chonbuk National University, Jeonju, South Korea
e-mail: zhihuiwangjl@gmail.com

Y. Lu
e-mail: luyu0311@gmail.com

S. Yoon
Department of Multimedia, Mokpo National University, Jeonnam, South Korea
e-mail: syoon@mokpo.ac.kr

S. J. Xie
Institute of Remote Sensing and Earth Science, Hangzhou Normal University, Hangzhou, China
e-mail: shanj_x@hotmail.com

D. S. Park (✉)
Division of Electronics Engineering, Chonbuk National University, Jeonju, South Korea
e-mail: dspark@jbnu.ac.kr

1 Introduction

Object tracking is a fundamental yet immature research topic in computer vision. During the past decades, a large variety of tracking algorithms have been proposed [1–9]. These tracking algorithms show excellent performances for some special scenarios, but not for general scenarios. Object tracking technology still faces challenges, such as shape and scale variations of targets and external factors, background and illumination changes, heavy occlusions and similar objects interference.

These discriminative tracking methods can track various kinds of objects effectively under some assumptions of scenarios [10–17]. These trackers update their parameters frame by frame with online methods. For each frame, a certain amount of positive and negative samples are collected surrounding the tracked object. These samples are utilized to update the parameters of these trackers. However, no tracker can effectively deal with all these internal and external factors during the tracking process. Moreover, a great deal of useful information is ignored during the parameter updating phase due to these training samples selection strategies. These trackers can enhance their adaptive capacities if they can take full advantage of the useful information given in each frame.

The online multiple instance learning (MIL) tracker based on the training of positive and negative instance bags and the NOR model shows high performance [18]. The online weighted multiple instance learning (WMIL) tracker combines multiple instance learning and instance weighting, and simplifies the weak classifiers selection [19]. The Semi-Boost tracker [14] combines the AdaBoost algorithm [20] and semi-supervised learning [21–23], but has drifting problems. Based on the theory of semi-supervised learning theory and the three aforementioned trackers, a novel single object

tracker, called the semi-supervised online weighted multiple instance learning (Semi-WMIL) tracker, is proposed in this paper. While collecting positive and negative training samples, it also collects the unlabeled samples between the regions of the positive and negative samples. These labeled and unlabeled training samples are used to select the weak classifiers of the proposed Semi-WMIL tracker. Therefore, the proposed tracker fully uses the information of each frame. The efficiency and robustness of the proposed Semi-WMIL tracker, compared with those of other state-of-the-art trackers, has been verified in simulation experiments.

The remainder of the paper is organized as follows. We start by introducing the semi-supervised MIL and online MIL algorithm in Sect. 2. System overview, training samples collection strategy, and the principle of strong classifier construction of the proposed Semi-WMIL tracker are demonstrated in Sect. 3. The experimental comparison of the proposed system with other state-of-the-art trackers is demonstrated in Sect. 4. Finally, we summarize the characteristics of the proposed system and discuss its superiority over other trackers in Sect. 5.

2 Related works

To expound the proposed tracking method clearly, the principles of semi-supervised MIL and online WMIL tracker are reviewed in this section.

2.1 Semi-supervised multiple instance learning

For the semi-supervised MIL [23], these training samples are mainly divided into positive samples $\aleph^+ = \{(x_i^+, 1)\}_{i=1}^{n_1}$, unlabeled samples $\aleph^u = \{x_i^u\}_{i=1}^{n_2}$ and negative samples $\aleph^- = \{(x_i^-, 0)\}_{i=1}^{n_3}$, where n_1, n_2, n_3 are numbers of instances on each class. The most commonly used criterion for semi-supervised multiple instance learning with manifold regularization terms is a combined objective functional:

$$F = \arg\min_F \left\{ \frac{1}{n_1+n_3} V\left(\aleph^+, \aleph^-, F\right) + \frac{\lambda_1}{n_2(n_1+n_3)} \cdot I_{LU}\left(\aleph^+, \aleph^-, \aleph^u, F\right) + \frac{\lambda_2}{(n_1+n_3)^2} I_{UU}\left(\aleph^u, \aleph^u, F\right) \right\}, \tag{1}$$

where $V\left(\aleph^+, \aleph^-, F\right)$ is the loss function of labeled training samples, $I_{LU}\left(\aleph^+, \aleph^-, \aleph^u, F\right)$ is the inconsistency function between labeled and unlabeled samples, $I_{UU}\left(\aleph^u, \aleph^u, F\right)$ is the inconsistency function among unlabeled samples, and λ_1, λ_2 is given weights for the latter two terms. The latter two terms in Eq. (1) can be combined if the inconsistency functions of these two terms are same.

2.2 Online weighted multiple instance learning (WMIL) tracker

Similar to [18,24], a random Haar-like feature vector function, $\mathbf{f}(x) = [f_1(x), \ldots, f_K(x)]^\top$ is utilized to represent each image patch x for the WMIL tracker [19]. Each feature value is a weighted sum of the pixels from 2 to 4 rectangles which are generated within the image patch randomly. The procedure of weak classifiers selection of WMIL tracker is based on Bayesian theorem, the posterior probability of sample x is estimated as follows:

$$p(y = 1 \mid x) = \sigma\left(H_K(x)\right), \tag{2}$$

where $H_K(x) = \ln\left(\frac{p(\mathbf{f}(x)|y=1)p(y=1)}{p(\mathbf{f}(x)|y=0)p(y=0)}\right)$ and $\sigma(\cdot)$ is the sigmoid function. With the assumption that the elements in $\mathbf{f}(x)$ are independently distributed and uniform prior $p(y=0) = p(y=1)$ [18,19], the classifier $H_K(\cdot)$ is described with the feature $\mathbf{f}(\cdot)$ as:

$$H_K(x) = \ln\left(\frac{p(\mathbf{f}(x) \mid y=1)}{p(\mathbf{f}(x) \mid y=0)}\right) = \sum_{k=1}^{K} h_k(x), \tag{3}$$

where the weak classifier $h_k(x) = \ln\left(\frac{p(f_k(x)|y=1)}{p(f_k(x)|y=0)}\right)$ with the Gaussian distribution assumption $p(f_k(x) \mid y=1) \sim N(\mu_1, \sigma_1)$ and $p(f_k(x) \mid y=0) \sim N(\mu_0, \sigma_0)$. The parameters μ_1, σ_1 are updated by the following scheme:

$$\begin{cases} \mu_1 \leftarrow \eta\mu_1 + (1-\eta)\overline{\mu}; \\ \sigma_1 \leftarrow \sqrt{\eta(\sigma_1)^2 + \frac{1-\eta}{N}\sum_{i=1}^{N}\left(f_k(x_{i,t}^+) - \overline{\mu}\right)^2 + \eta(1-\eta)(\mu_1 - \overline{\mu})^2}; \end{cases} \tag{4}$$

where η is a learning rate parameter and $\overline{\mu}$ is the average value of the kth feature from the positive samples at the current frame. μ_0 and σ_0 can be updated similarly.

Suppose that $\aleph_t^+ = \{(x_{i,t}^+, y^+)\}_{i=1}^{n_1}$ are positive sample bags and $\aleph_t^- = \{(x_{i,t}^-, y^-)\}_{i=1}^{n_3}$ are negative sample bags for frame t, where $y^+ = 1$ and $y^- = 0$. For the WMIL tracker, the positive and negative bag probabilities are defined as follows:

$$\begin{cases} p(y=1 \mid \aleph_t^+) = \sum_{i=1}^{n_1} \omega_i^+ p(y=1 \mid x_{i,t}^+); \\ p(y=0 \mid \aleph_t^-) = \sum_{i=1}^{n_3} \omega^- (1 - p(y=1 \mid x_{i,t}^-)); \end{cases} \tag{5}$$

where the weight function ω_i^+ is defined as:

$$\omega_i^+ = \frac{1}{c}\exp(- \mid l(x_{i,t}^+) - l(x_t) \mid), \tag{6}$$

where c is the normalization constant and $l(\cdot)$ is the location function. And x_t is a block including an object estimated at frame t and $l(x_t)$ is its center location. As all negative samples

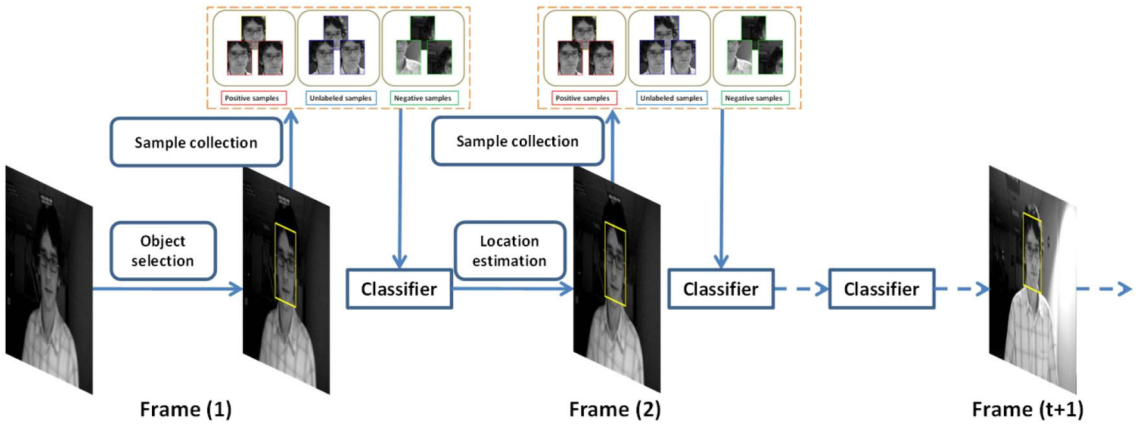

Fig. 1 Basic flow of the proposed Semi-WMIL tracker

are far from x_t, the weight ω^- for the negative samples is a positive constant.

During the weak classifiers selection process, K optimal weak classifiers are chosen from the weak classifier pool $\phi = \{h_1, \ldots, h_M\}$. For the WMIL tracker, a log-likelihood function is utilized to select these optimal weak classifiers by using the following criterion:

$$h_k = \arg\max_{h \in \phi} \langle h, \nabla \ell(H) \rangle \mid_{H=H_{k-1}}, \tag{7}$$

where $H_{k-1} = \sum_{m=1}^{k-1} h_m$ is a strong classifier with $k-1$ selected weak classifiers. And $\langle \cdot \rangle$ is the inner product and $\nabla \ell(H)$ is the derivative of the log-likelihood function and this equation has been deduced in [19]. $\forall x_{j,t} \in \{\aleph_t^+, \aleph_t^-\}$,

$$\nabla \ell(H)(x_{j,t}) = y_{j,t} \frac{\tilde{\omega}_{j,t} \sigma(H(x_{j,t}))(1 - \sigma(H(x_{j,t})))}{\sum_{m=1}^{n_1} \tilde{\omega}_{m,t} \sigma(H(x_{m,t}^+))} + (1 - y_{j,t}) \frac{\sigma(H(x_{j,t}))(1 - \sigma(H(x_{j,t})))}{\sum_{m=1}^{n_3} (1 - \sigma(H(x_{m,t}^-)))}, \tag{8}$$

where $\tilde{\omega}_{j,t} = exp(- \mid l(x_{j,t}) - l(x_t) \mid)$, $y_{j,t} \in \{0, 1\}$ is corresponding output of $x_{j,t}$.

3 Proposed semi-supervised online weighted multiple instance learning tracker (Semi-WMIL tracker)

In this section, the principle and the tracking procedure of the proposed Semi-WMIL tracker are presented in details. In Sect. 3.1, the system overview of the proposed Semi-WMIL tracker is presented. Then, the training samples collection strategy and the principle of strong classifier construction of the proposed Semi-WMIL tracker are introduced in Sects. 3.2 and 3.3. Estimation of object location is explained in 3.4.

3.1 System overview

The basic flow of the proposed Semi-WMIL tracker is demonstrated in Fig. 1. At each frame, the proposed method performs the following three main procedures: collection of training samples, construction of a strong classifier, and estimation of an object location. An object location at t th frame is estimated by a strong classifier constructed at $(t-1)$ th frame. And, based on the object location estimated at tth frame, training samples are collected to construct a strong classifier at t th frame. Collected training samples are classified into three sets (bags): (labeled) positive sample set, (labeled) negative sample set, and unlabeled sample set. The proposed Semi-WMIL constructs a strong classifier at tth frame, using these three training sets. A strong classifier constructed at t th frame is used to estimate an object location at $(t+1)$th frame. For the procedure, initial object information at $t = 0$ (the first frame) is given manually.

3.2 Collection of training samples

The existing tracking methods based on discriminative approaches classify the training samples for algorithm parameter updating mainly into positive and negative samples based on the distances of training samples and a given target center. Samples adjacent to the target center are regarded as positives and those far from the target center are regarded as negatives. On the other hand, as shown in Fig. 2, the proposed method collects three sets of training samples, such as positive sample set, negative sample set and unlabeled sample set from their corresponding regions surrounding the center location as follows:

$$\begin{cases} \text{Positive region:} & X_t^{\alpha} = \{x \mid \| l(x) - l(x_t) \| < \alpha\}; \\ \text{Unlabeled region:} & X_t^{\alpha,\zeta} = \{x \mid \alpha \leq \| l(x) - l(x_t) \| < \zeta\}; \\ \text{Negative region:} & X_t^{\zeta,\beta} = \{x \mid \zeta \leq \| l(x) - l(x_t) \| < \beta\}. \end{cases} \tag{9}$$

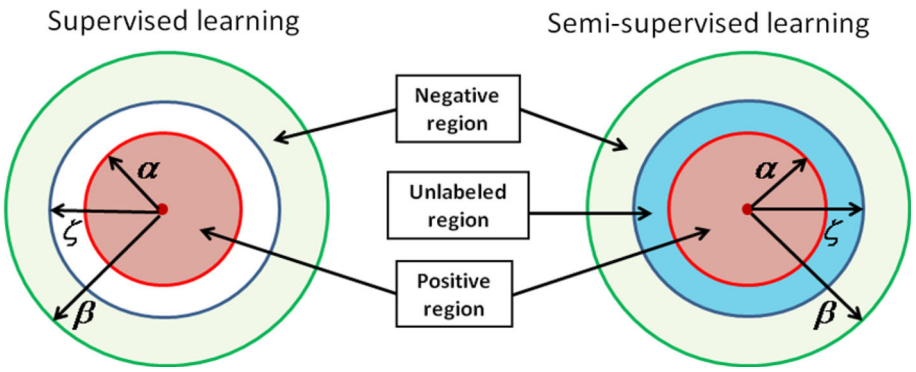

Fig. 2 Training sample regions comparison of supervised and semi-supervised learning

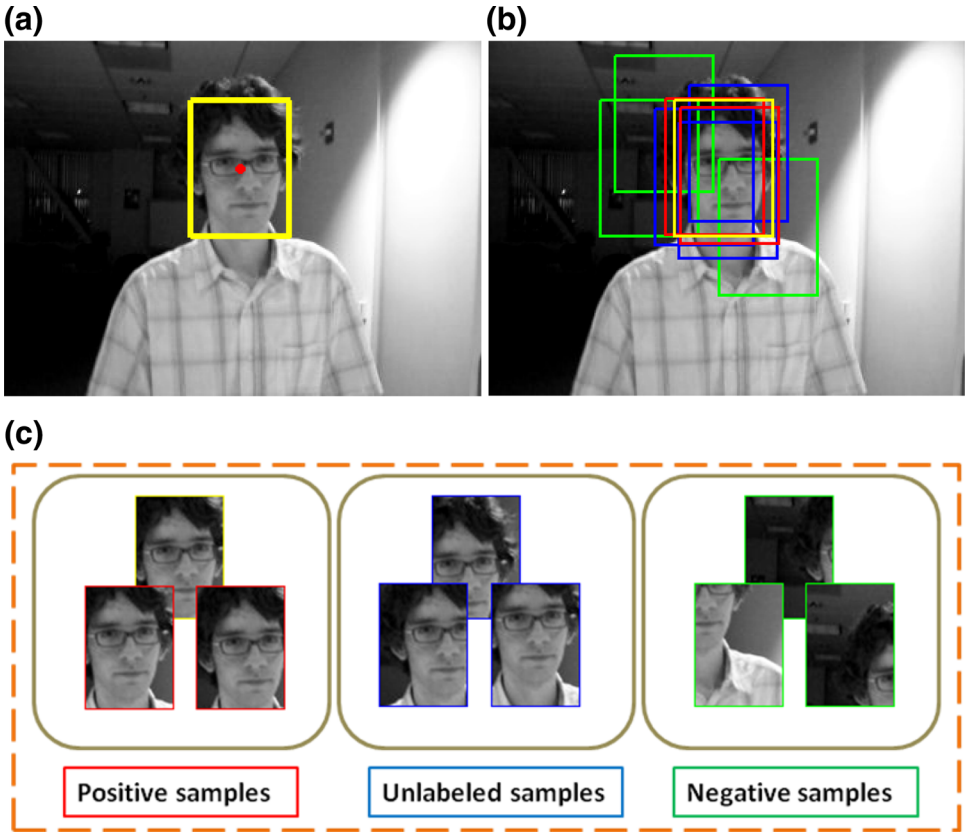

Fig. 3 Training samples collection of the proposed Semi-WMIL tracker. **a** Object location, **b** sample collection, **c** collected samples

where $\alpha < \zeta < \beta$. These positive, unlabeled, and negative samples are randomly collected in corresponding regions with given numbers n_1, n_2, n_3.

A specific example of sample collection by the proposed Semi-WMIL tracker is demonstrated in Fig. 3. First, the center and scale of target object are determined in Fig. 3a. Then, several training samples are collected from three regions defined in Eq. (9), as shown in Fig. 3b. The training samples are divided into positive, unlabeled and negative samples based on their corresponding regions which are determined by the distance between each training sample and the target center (The samples with yellow and red rectangles are positives, samples with blue rectangles are unlabeled samples, and samples with green rectangles are negatives). The collected samples are demonstrated in Fig. 3c.

3.3 Construction of strong classifier

A strong classifier consists of a series of weak classifiers, and the method used to select weak classifiers in the proposed method is similar to that of the WMIL tracker in [19], but it uses the unlabeled samples for the semi-supervised learning as well as labeled samples (positive and negative samples) for the supervised learning. Suppose that $\aleph_t^+ = \{(x_{i,t}^+, y^+)\}_{i=1}^{n_1}$,

$\aleph_t^u = \{(x_{i,t}^u)\}_{i=1}^{n_2}$ and $\aleph_t^- = \{(x_{i,t}^-, y^-)\}_{i=1}^{n_3}$ are positive, unlabeled and negative sample bags for frame t, where $y^+ = 1$ and $y^- = 0$.

During the process of visual tracking, the efficiency of the tracker must be guaranteed to enhance its practicability in practical applications. Therefore, on the premise of high tracking accuracy, the reduction of the computational complexity is particularly important. In the main procedures of the proposed tracking system at each frame, the strategies of training sample collection and object location estimation are difficult to be optimized. The procedure of the strong classifier construction, by contrast, is more feasible to be simplified than the aforementioned two terms.

There are two different terms $I_{LU}(\cdot)$ and $I_{UU}(\cdot)$ of inconsistency functions in Eq. (1). Since the labeled training samples contain more definite information than the unlabeled training samples, we add only the term $I_{LU}(\cdot)$ to the inconsistency function of the proposed tracking system. Under the criterion of semi-supervised MIL, a strong classifier is constructed using optimal weak classifiers obtained by maximizing the log-likelihood function but minimizing the inconsistency function:

$$h_k = \arg\max_{h\in\Phi} \left\{ \frac{1}{n_1+n_3} \langle h, \nabla \ell(H) \rangle \mid_{H=H_{k-1}} - \frac{\lambda}{n_2(s_1+s_2)} I'_{LU}(\aleph_t^+, \aleph_t^-, \aleph_t^u, H_{k-1}+h) \right\}, \quad (10)$$

where the derivative $\nabla\ell(H)$ is defined in Eq. (8) and $I'_{LU}(\cdot)$ is the reduced form of the inconsistency function $I_{LU}(\cdot)$ in Eq. (1), and λ is the weight to balance the loss function and inconsistency function, s_1 and s_2 are the numbers of positive and negative samples blocks in the following Eq. (11). The loss function is depended on labeled samples and the inconsistency function is depended on the similarities between labeled and unlabeled samples. Therefore, the weight λ can be regarded as the tradeoff between the input/output relationship of labeled samples and these input feature similarities. The proposed Semi-WMIL tracker is degenerated to the WMIL tracker when $\lambda = 0$ [19]. Therefore, the original WMIL tracker can be regarded as a special case of the proposed Semi-WMIL tracker.

In the weak classifiers selecting phase of the proposed Semi-WMIL tracker, the log-likelihood function has already been simplified with the strategy of first-order Taylor expansion as the WMIL tracker [19]. At each frame, the positive training samples are collected from X_t^α, which is a circular region around the target location with radius α. These positive training samples have high similarity, especially when the radius α is rather small. The negative training samples are collected from the annular region $X_t^{\zeta,\beta}$, which is distributed more discretely than the positive training samples. The feature values of some negative training samples are varied largely with each other. Therefore, the strategy of sample blocks can be used to optimize the inconsistency function. To further enhance the efficiency of the tracking system and decrease the negative influence of outliers in extracted features of training samples, the simplified inconsistency function has been proposed and adopted in the proposed Semi-WMIL tracker.

In the weak classifiers selecting phase of the proposed Semi-WMIL tracker, only the inconsistency function between the labeled and unlabeled samples is considered and positive, and negative samples are divided into several blocks:

$$\begin{cases} \text{Positive blocks: } \left\{ \{(x_{i,t}^+, y^+)\}_{i=1}^{q}, \{(x_{i,t}^+, y^+)\}_{i=q+1}^{2q}, \ldots \right\}; \\ \text{Negative blocks: } \left\{ \{(x_{i,t}^-, y^-)\}_{i=1}^{q}, \{(x_{i,t}^-, y^-)\}_{i=q+1}^{2q}, \ldots \right\}. \end{cases} \quad (11)$$

with equal number q and the average values of these blocks $\{\bar{x}_{j,t}^+\}_{j=1}^{s_1}, \{\bar{x}_{j,t}^-\}_{j=1}^{s_2}$, $(s_1 = \lceil \frac{n_1}{q} \rceil, s_2 = \lceil \frac{n_3}{q} \rceil)$ are utilized to estimate the inconsistency function.

The reduced inconsistency function is defined as:

$$I'_{LU}(\aleph_t^+, \aleph_t^-, \aleph_t^u, H_{k-1}+h) = \min \left\{ \sum_{i=1}^{n_2} \sum_{j=1}^{s_1+s_2} S(x_{i,t}^u, \bar{x}_{j,t}^{\pm}) \exp\left(-2(H_{k-1}+h)(x_{i,t}^u) y_j^{\pm}\right), \Theta \right\}, \quad (12)$$

where $S(x_{i,t}^u, \bar{x}_{j,t}^{\pm}) = \exp\left(\frac{\|x_{i,t}^u - \bar{x}_{j,t}^{\pm}\|^2}{\sigma^2}\right)$ is the dissimilarity equation, σ is the scale parameter controlling the spread of the radial basis function, and $\Theta = \frac{n_2\theta(s_1+s_2)}{\lambda}$ is the given threshold for the inconsistency function, where θ is the threshold parameter for the inconsistency function. The unlabeled samples are more similar to the positive samples than to the majority of the negative samples, as the unlabeled region is closer to the positive region in our experiment. Therefore, we assume all these unlabeled samples are different from the negative samples; note that $\exp\left(-2(H_{k-1}+h)(x_{i,t}^u) y^-\right) \equiv 1$ in Eq. (12) as $y^- = 0$. This assumption can offset the negative effects of these features with similar properties between the unlabeled and negative samples during the weak classifiers selection.

The main steps of the refinement process of weak classifiers selection at frame t can be summarized in Algorithm 1.

3.4 Estimation of object location

As mentioned the above, in the proposed method, each training sample set is updated at each frame. Positive and negative samples are utilized to train a classifier with an online boosting algorithm and unlabeled samples are utilized to assist the selection of optimal weak classifiers. These samples at frame t are obtained using Eq. (9), based on an object

Algorithm 1 The proposed Semi-WMIL Boost

Require: All collected samples $\{\aleph_t^+, \aleph_t^-, \aleph_t^u\}$ at frame t;
1: update all M weak classifiers in the pool $\Phi = \{h_1, \ldots, h_M\}$ with data $\{\aleph_t^+, \aleph_t^-\}$;
2: initialize $I_0 = 0$, $H_0(x) = 0$ for $\forall x \in \{\aleph_t^+, \aleph_t^-\}$;
3: **for** $k = 1$ to K **do**
4: calculate $\nabla\ell(H)(x)\mid_{H=H_{k-1}}$ by Eq. (8);
5: **for** $m = 1$ to M **do**
6: $\ell_m = \sum_{x\in\{\aleph_t^+,\aleph_t^-\}} h(x)\nabla\ell(H)(x)\mid_{H=H_{k-1}}$;
7: $I_m = I'(\aleph_t^+, \aleph_t^-, \aleph_t^u, H_{k-1} + h_m)$;
8: **end for**
9: $m^* = \arg\max_m \left(\frac{\ell_m}{n_1+n_3} - \frac{\lambda I_m}{n_2(s_1+s_2)}\right)$;
10: $h_k(x) \leftarrow h_{m^*}(x)$;
11: $H_k(x) = H_{k-1}(x) + h_k(x)$;
12: **end for**
Ensure: Strong classifier $H_K(x) = \sum_{k=1}^K h_k(x)$ and $P(y = 1 \mid \cdot) = \sigma(H_K(\cdot))$.

location at frame t estimated using a strong classifier constructed at $(t-1)$th frame. To estimate the object location, some samples are collected from the predefined search range at frame t, $X_t^\gamma = \{x \mid \| l(x) - l(x_{t-1}) \| \leq \gamma\}$, surrounding an object location at $(t-1)$th frame and their confidences are estimated by the classifier trained at $(t-1)$th frame. Finally, the location $l(x_t)$ with maximum confidence $x_t = \arg\max_x P(y = 1 \mid x)$ is estimated as a new object location at frame t.

4 Experiments

The proposed Semi-WMIL tracker was tested on thirteen publicly available video sequences [18,25,26]. These video sequences include challenging factors: complex backgrounds, illumination and object pose variations, object rotation and non-rigid deformation, etc. The specific properties of the tested sequences are listed in Table 1. Eight state-of-the-art trackers, which included MIL [18], WMIL [19], ODFS [27], RTCT [28], IVT [29], DFT [30], LIAPG [31], MTT [32] were compared with the proposed Semi-WMIL tracker for these video sequences. The parameters of the proposed Semi-WMIL tracker are demonstrated in Sect. 4.1 and analyzed in Sect. 4.2, and the parameters of the other state-of-the-art trackers follow the original settings of the papers describing them or are chosen by tuning for their best performance. Quantitative and qualitative analysis are shown in Sects. 4.3 and 4.4.

4.1 Experimental setup

The outer radii of the positive, unlabeled and negative regions of all thirteen video clips are set to $\alpha = 4$, $\zeta = 6$ and $\beta = 50$, respectively. The radius of the unlabeled region is almost the same as that of the positive region, which guarantees a large overlap between the unlabeled samples and the positive samples and thus, their similarity. The corresponding numbers of positive, unlabeled and negative samples are $n_1 = 45, n_2 = 10$, and $n_3 = 50$. The radius of searching region in the next frame is $\gamma = 25$, half of the outer radius of the negative region. The number of candidate features in a feature pool is $M = 100$, and $K = 12$ features are selected from the feature pool to construct the weak classifiers. Compared with the MIL tracker ($M = 250$ and $K = 50$) and WMIL tracker ($M = 150$ and $K = 15$), the proposed Semi-WMIL tracker carried out less computationally complex feature extraction to balance the computing time of the inconsistency function between the labeled and unlabeled samples. Since the log-likelihood function works well on these sequences in [18], we ignored the coefficient on the log-likelihood function for these sequences. During the procedure of inconsistency function estimation, the number of positive or negative samples in each block is $q = 6$. The learning parameter η in Eq. (4) was set to $\eta = 0.85$. The squared scale parameter σ^2 in Eq. (12) was set to $\sigma^2 = 10^{10}$ for all the video clips, expect sequence *Occlusion1* with $\sigma^2 = 10^{11}$. The weight of the similarity term in Eq. (10) is $\lambda = 0.1$, and the threshold parameter θ for the similarity matrix in Eq. (12) is defined as a step function, which depends on the scale of target object P (amount of occupied pixels):

$$\theta = \begin{cases} 0.6, & P < 2500; \\ 0.65, & 2500 \leq P < 10{,}000; \\ 0.7, & P \geq 10{,}000; \end{cases} \tag{13}$$

Although the initialized scale of target in sequence *Twinnings* is larger than 2500, the threshold parameter θ was set to 0.6, which depended on its average scale in all these frames. The threshold parameter θ of sequence *Shaking* was also set to 0.6 with optimal performance.

4.2 Parameter analysis

All of the parameters adopted in the proposed tracking system has been demonstrated in Sect. 4.1. Part of these parameters have been inherited from [18,19], and fixed in our experiments, such as learning rate η and the collecting regions of positive and negative training samples, etc. Under the assumption of smooth motion, these sample collecting radii $\{\alpha, \zeta, \beta\}$, sample amount $\{n_1, n_3\}$, and search radius γ are robust enough to various kinds of testing sequences, which has been verified in [18,19] and our experiments. The unlabeled training samples are collected from the region between the positive and negative sample collecting regions, which have certain similarities with these training samples. The threshold parameter θ has been defined as a step function based on the scale size of the target object. The number of training samples in each block q is based on the num-

Table 1 The properties of the tested sequences

No.	Sequence	Frames	Mov. camera	Partial occ.	Pose change	Illum. change	Scale change	Similar objects	Fast mov.
1	*Basketball*	725	Yes	Yes	Yes	No	Yes	Yes	Yes
2	*Cliffbar*	329	No	Yes	Yes	No	Yes	Yes	Yes
3	*Couple*	140	Yes	No	Yes	No	Yes	Yes	Yes
4	*Crossing*	120	No	No	Yes	Yes	Yes	No	Yes
5	*DavidIndoorOld*	462	Yes	Yes	Yes	Yes	Yes	No	No
6	*Dog1*	1350	No	No	Yes	No	Yes	No	No
7	*Occlusion1*	886	Yes	Yes	No	No	No	No	No
8	*Occlusion2*	816	No	Yes	Yes	No	No	No	No
9	*Shaking*	365	Yes	Yes	Yes	Yes	Yes	Yes	Yes
10	*Sylvester*	1345	No	No	Yes	Yes	Yes	No	Yes
11	*Tiger1*	354	No	Yes	Yes	No	Yes	No	Yes
12	*Tiger2*	365	No	Yes	Yes	No	Yes	No	Yes
13	*Twinnings*	472	No	No	Yes	No	Yes	No	No

bers of positive samples n_1 and negative samples n_3, which is slightly smaller than the square root of n_1 and n_3. The assigned value of scale parameter σ is proportional to feature values of training samples basically. Since the weak classifiers are selected based on the log-likelihood function and the inconsistency function, the selected weak classifiers are more robust than MIL and WMIL trackers. Therefore, we selected $K = 12$ features to construct the weak classifiers from the feature pool with dimensions $M = 100$. The feature pool is nine times of the selected features, and large enough to construct the proper weak classifiers. As presented in Table 1, there are various kinds of challenges in all these tested sequences. And, the assigned parameters are rather robust to these challenges of all these tested sequences.

The mean failure rates of all these testing sequences with varied learning rate η and weight parameter λ have been demonstrated in Fig. 4a, b, correspondingly. The learning rate η is tested from 0.75 to 0.95, with a step size of 0.05. The weight parameter λ is tested from 0.05 to 0.15, with a step size of 0.025. The mean failure rates achieve the minimum values when learning rate $\eta = 0.85$, and weight parameter $\lambda = 0.1$. The mean failure rates are increased when the values of these two parameters varied. These parameters can be adjusted slightly for particular scenarios appropriately, based on the conditions of shape and scale variations of the targets itself and scene changes.

4.3 Quantitative analysis

The performances of these nine trackers were evaluated based on the criteria of failure rate and center location error. The tracking result is considered a failure if the score $\frac{\text{area}(R_t \cap R_g)}{\text{area}(R_t \cup R_g)} < 0.5$, where R_t is the tracking bounding box and R_g is the ground truth bounding box. For all these video sequences, the tracking results were evaluated once every five frames.

The tracking speeds, in term of average frame per second (FPS), of Semi-WMIL, Semi-WMIL*, WMIL, ODFS, RTCT, IVT, MIL, DFT, L1APG, and MTT, are 21, 13, 22, 21, 27, 13, 2, 11, 1, 1, correspondingly.

Figure 5 shows the performances of all trackers with respect to failure rate. To demonstrate the efficiency of the proposed Semi-WMIL tracker, the results of Semi-WMIL tracker with unreduced inconsistency function (Semi-WMIL*) are also demonstrated in Fig. 5. The Semi-WMIL* is the same as Semi-WMIL but, in this method, the positive and negative samples are utilized directly to evaluate the inconsistency of classifiers. In other words, $s_1 = n_1$ and $s_2 = n_3$ ($q = 1$) are used in Eq. (10). And their own optimized parameters are used in the Semi-WMIL* tracker and the Semi-WMIL tracker, respectively.

In Fig. 5, compared to the Semi-WMIL* tracker, the proposed Semi-WMIL tracker show similar or better performances for all these 13 tested video sequences with different situations as shown in Table 1. Furthermore, the reduced Semi-WMIL tracker is more efficient in speed than the unreduced Semi-WMIL* tracker while keeping similar or better performances in accuracy, since the FPS of the reduced one is 21 and the FPS of the unreduced one is 13. Therefore, we can say that the efficiency of the reduced Semi-WMIL tracker is improved due to the contribution of the reduced inconsistency function.

Compared with the other eight state-of-the-art trackers, the proposed Semi-WMIL tracker has the lowest failure rates for most of these thirteen video sequences. Especially, for the sequences *DavidIndoorOld*, *Occlusion1* and *Occlusion2*, the proposed Semi-WMIL tracker is 100 % successful with them. The average FPS is estimated using MATLAB for the same

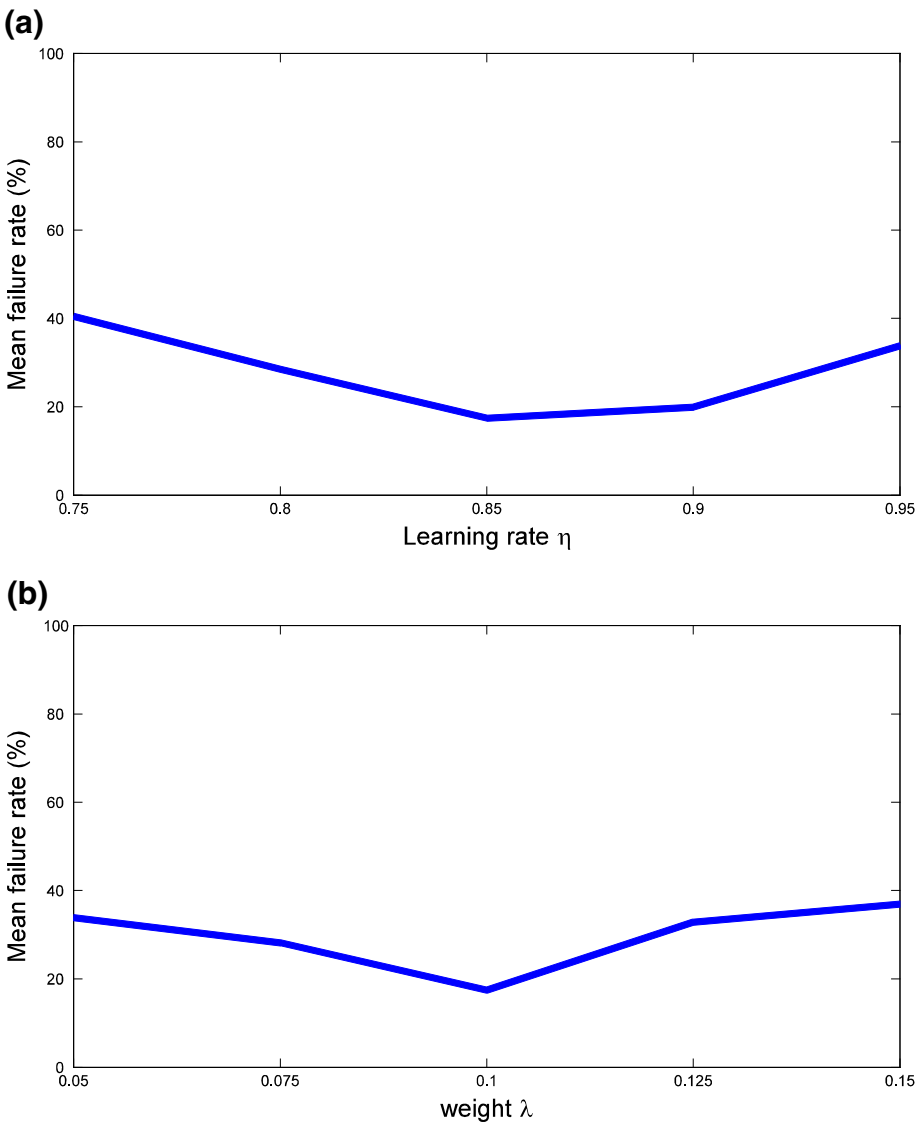

Fig. 4 Comparison of the mean failure rates with varied parameters. **a** Mean failure rates with varied learning rate, **b** mean failure rates with varied weight parameter

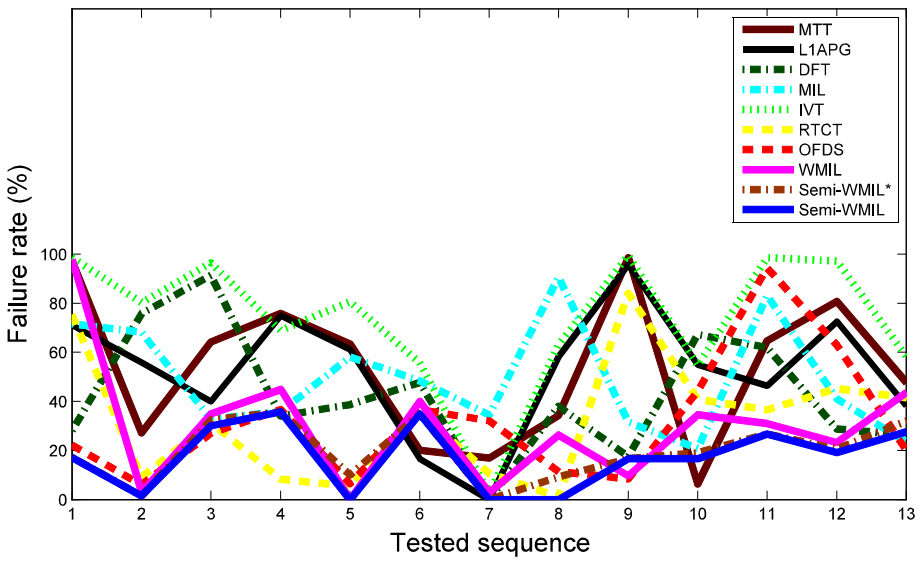

Fig. 5 Comparison of the mean failure rates for all tested sequences

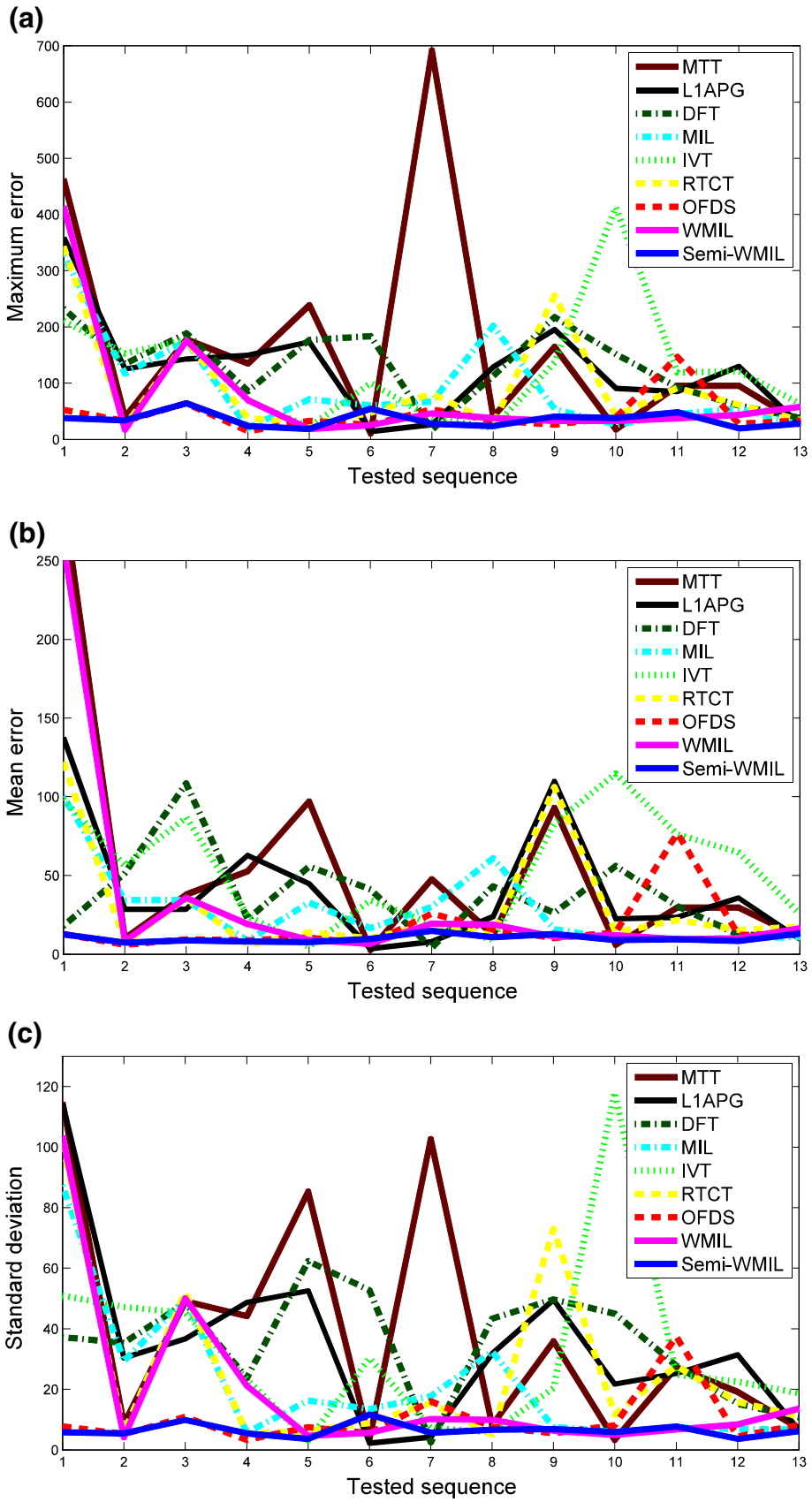

Fig. 6 Comparison of the center location errors of all nine trackers. **a** Maximum values, **b** mean values, **c** standard deviations

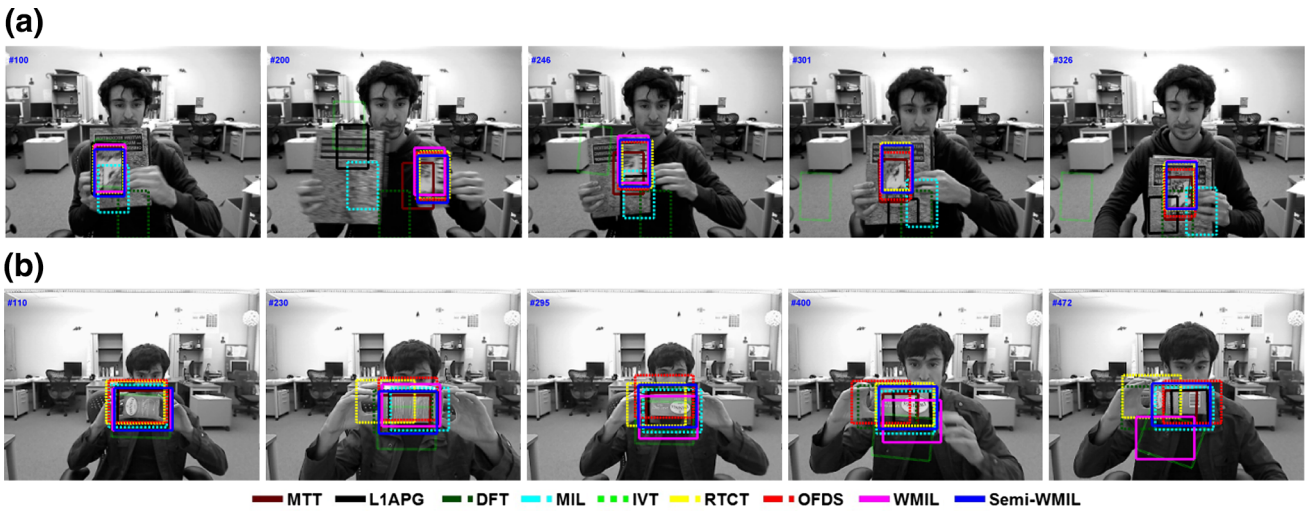

Fig. 7 Screenshots of some sampled tracking results for tested sequences: *Cliffbar* and *Twinnings* (*from top to bottom*). Figure best viewed in *color* and *line style*. **a** Cliffbar, **b** Twinnings

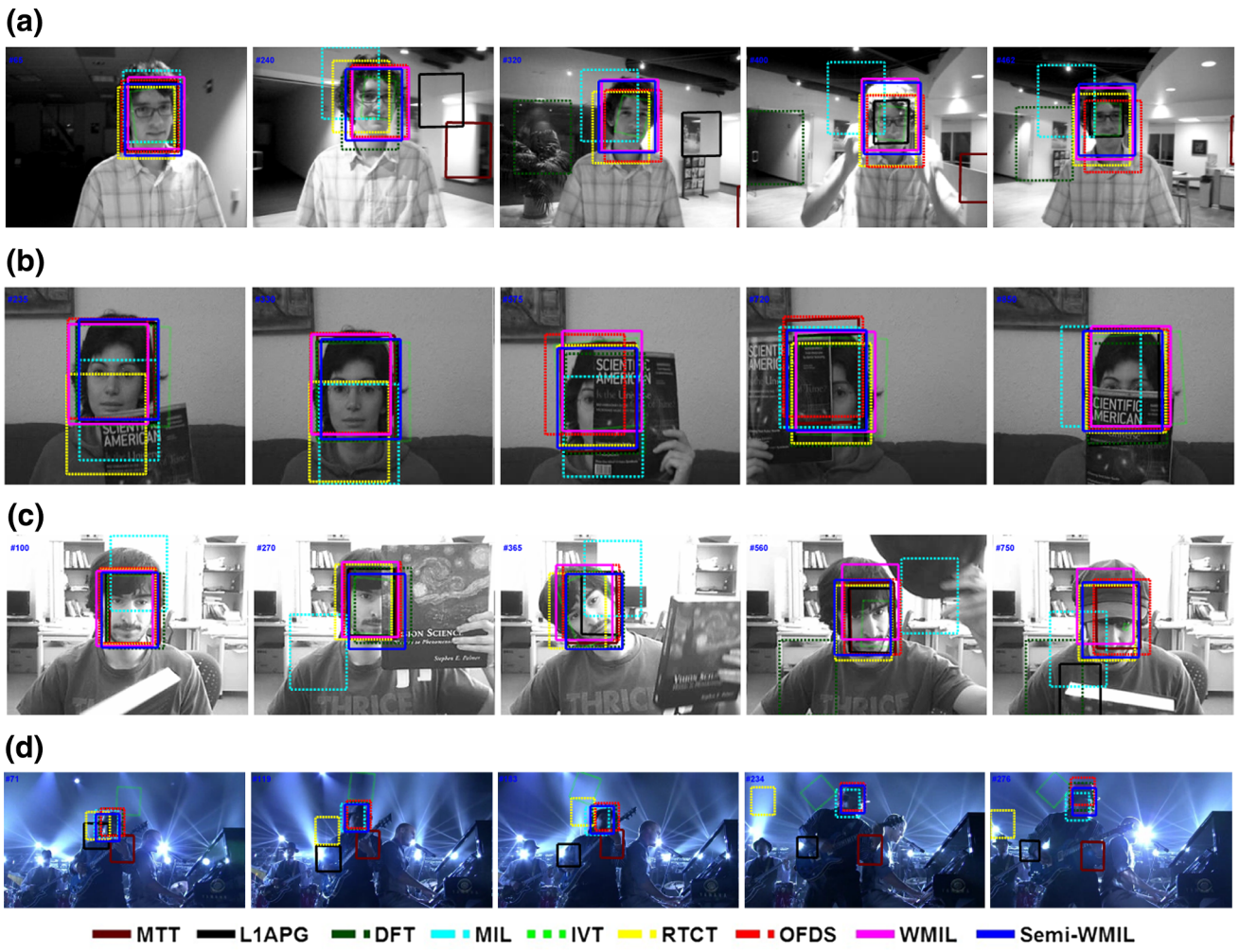

Fig. 8 Screenshots of some sampled tracking results for tested sequences: *DavidIndoorOld, Occlusion1, Occlusion2* and *Shaking* (from *top to bottom*). Figure best viewed in *color* and *line style*. **a** DavidIndoorOld, **b** Occlusion1, **c** Occlusion2, **d** Shaking

hardware conditions. Although the speed of the proposed tracker is not the fastest among these trackers, the average FPS of the proposed Semi-WMIL tracker reaches 21 and efficient enough to deal with online tracking issues, which is on the same order of magnitude with the WMIL, ODFS and RTCT trackers.

The maximum values, mean values and standard deviations of the center location errors of all nine trackers are compared in Fig. 6a–c, correspondingly. The proposed Semi-WMIL tracker outperforms the other eight trackers in most cases. The mean values of the Semi-WMIL tracker are all around 10.3 and the standard deviations are all around 6.5 for all thirteen video sequences. Moreover, the maximum values are all less than 65. Both the proposed Semi-WMIL tracker and the WMIL tracker show excellent overall performance, but the proposed Semi-WMIL tracker is even better. Other trackers work well on some video sequences, but terribly on other sequences. These results demonstrate that the proposed Semi-WMIL tracker has superior performance in accuracy and robustness, compared with the other eight trackers.

4.4 Qualitative analysis

4.4.1 Cliffbar and Twinnings

These two video sequences suffer from large scale variations, rotations and complicated backgrounds. Screenshots of some sampled tracking results for these two sequences are shown in Fig. 7. For the sequence *Cliffbar*, the texture of the background is similar to that of the target. The DFT, L1APG, MIL and IVT trackers undergo drift at frame #100 and lose the tracking target at frames #200, #246, #301, #326. The OFDS and MTT tracker undergo a slight drift, but the other three trackers work well. For the sequence *Twin-*

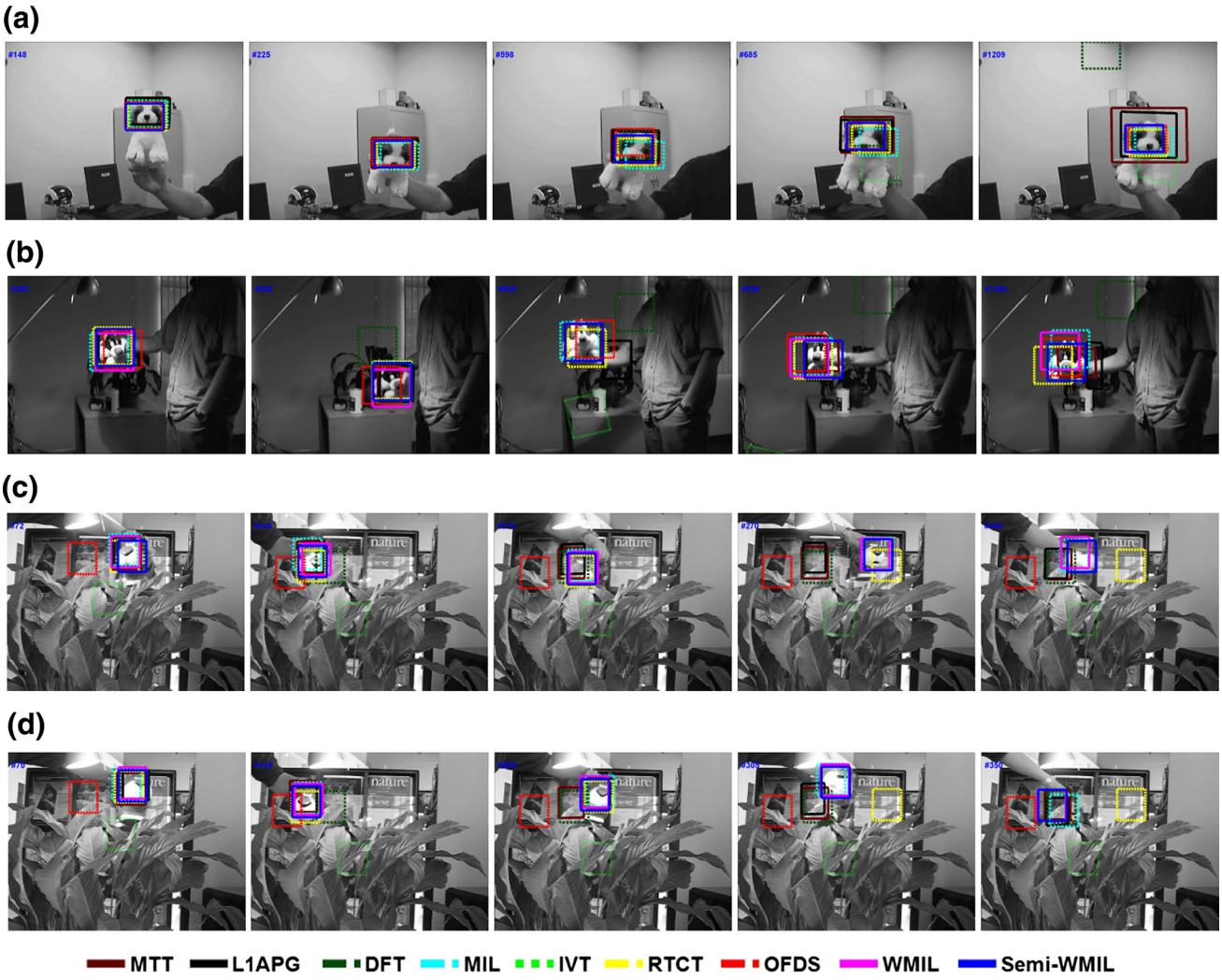

Fig. 9 Screenshots of some sampled tracking results for tested sequences: *Dog1, Sylvester, Tiger1* and *Tiger2* (from *top to bottom*). Figure best viewed in *color* and *line style*. **a** Dog1, **b** Sylvester, **c** Tiger1, **d** Tiger2

nings, all these trackers follow the target effectively. At the frames #230, #295, #400, #472, IVT, RTCT, OFDS, MTT and WMIL trackers undergo different degrees of drifts during the process of target movement. Compared with the other eight trackers, the proposed Semi-WMIL tracker performs well for both of these sequences.

4.4.2 *DavidIndoorOld, Occlusion1, Occlusion2 and Shaking*

The targets of these four sequences are all human faces. The sequence *DavidIndoorOld* and *Shaking* are taken by a mobile camera, while the sequence *Occlusion1* and *Occlusion2* are taken by fixed cameras. The main challenges presented by the sequence *DavidIndoorOld* are background changes, scale variation and head internal rotation. Moreover, the human takes off his/her glasses at frame #294 and puts it on again from frame #392. In the latter two sequences, the main issue is a heavy occlusion of either books or a hat; the sequence *Occlusion2* additionally suffers from head rotation. As shown in Fig. 8, the MIL tracker always lose the targets for these front three sequences, and suffers drifting problem in sequences *Shaking*. For the sequence *DavidIndoorOld*, the average center location error of the MIL tracker is less than those of the other trackers. However, the failure rate reaches 80.7 % as the tracking bounding box becomes smaller at frames #240, #320, #400, #462. The IVT tracker even loses the target in sequences *Occlusion1*, *Occlusion2* and *Shaking*. Both of the L1APG and MTT trackers cannot follow the target exactly in sequence *Shaking*. The RTCT and OFDS trackers undergo some drifting problems in most case. The WMIL tracker works well for the sequence *DavidIndoorOld*, and works poorly when faced with heavy occlusion in the latter two sequences. The proposed Semi-WMIL tracker has superior performance for all three sequences.

4.4.3 *Dog1, Sylvester, Tiger1 and Tiger2*

The targets of four sequences are all toys. These four sequences mainly suffer from multi-aspect rotation and fast motion, and the latter two sequences even suffer from a heavy occlusion. Sampled tracking results for these three sequences are shown in Fig. 9. The DFT and IVT trackers loses the target at frame #685 of the sequence *Dog1* and frames #650, #830, #1240 of the sequence *Sylvester*, and seldom follows the tracking target for the latter two sequences. For the sequence *Tiger1*, the OFDS tracker misses the target at all frames demonstrated in the third row of Fig. 9 and RTCT tracker loses the target at frames #270, and #340. The

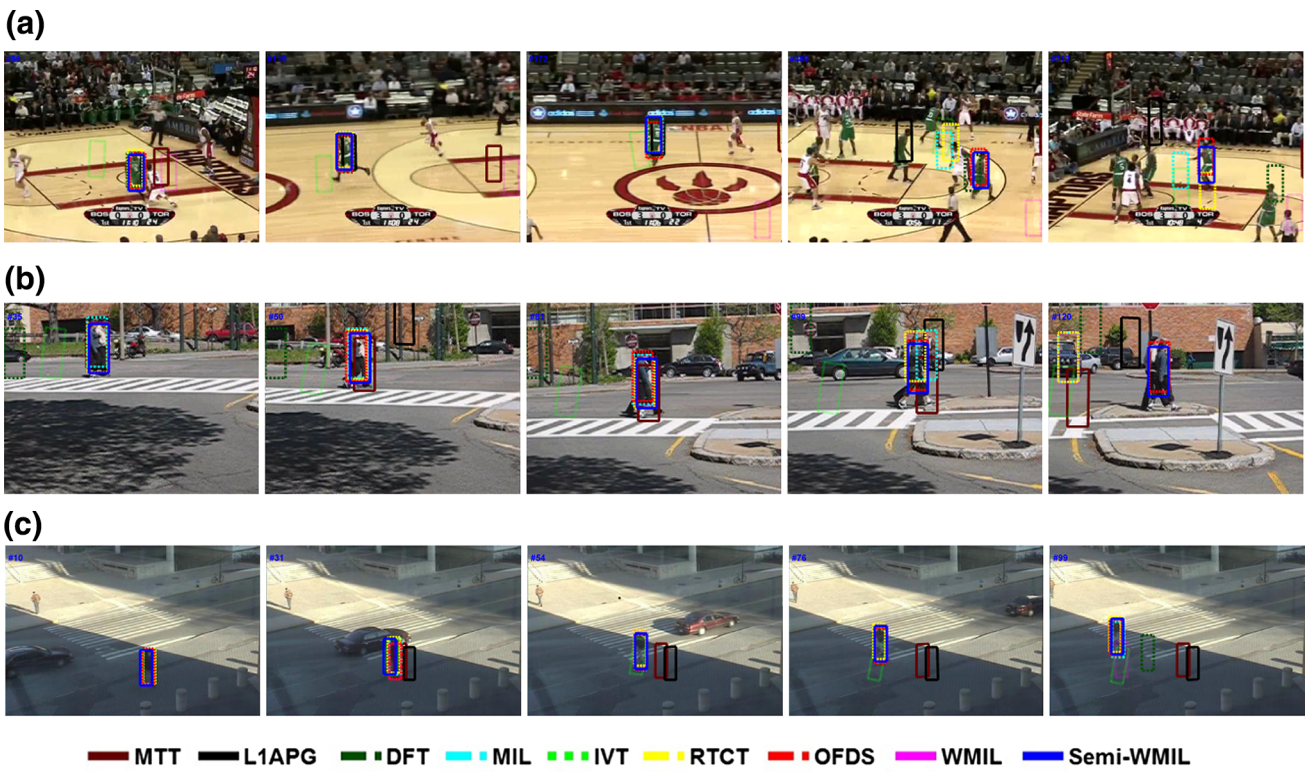

Fig. 10 Screenshots of some sampled tracking results for tested sequences: *Basketball, Couple* and *Crossing* (from *top to bottom*). Figure best viewed in *color* and *line style*. **a** Basketball, **b** Couple, **c** Crossing

MTT, L1APG and IVT trackers always lose the target in the Fig. 8d. The MIL and WMIL trackers suffer from a certain degree of drifting for all these four sequences.

4.4.4 Basketball, Couple and Crossing

The targets of four sequences are all humans, such as basketball players and pedestrians. In sequence *Basketball*, the proposed Semi-WMIL tracker has the best performance than other eight trackers. Sampled tracking results for these three sequences are shown in Fig. 10. Most of the trackers lose the target at frame #489 and #712, expect the Semi-WMIL and OFDS trackers. The IVT and DFT trackers lose the target at all these four frames of sequence *Couple* in Fig. 10b. The MTT, L1APG, RTCT and WMIL trackers lose the target at frame #120 of sequence *Couple*. For the sequence *Crossing*, most of these trackers work well excluding the MTT, L1APG, IVT and WMIL trackers.

Although other trackers work better on some sequences, the proposed Semi-WMIL tracker has excellent overall performance for all the thirteen video sequences. The proposed Semi-WMIL tracker shows outstanding adaptability and stability when faced with challenges, such as background changes, fast motion, and heavy occlusion.

5 Conclusion and future work

This paper proposed a semi-supervised online weighted multiple instance learning tracker for single object tracking. The tracker uses a block-based inconsistency function between the labeled and unlabeled training samples to improve its tracking performance. Target objects of these thirteen open-access video sequences in simulation experiments suffer from various complex and challenging situations, which are authoritative to verify the tracking capacity of these single object tracking algorithms. The proposed semi-supervised WMIL tracker can deal with these issues effectively, and shows the best overall performance over the other eight state-of-the-art trackers. The high tracking speed of the proposed semi-supervised WMIL tracker guarantees its usefulness in online practical applications. Future work will include improving the selecting criterion of weak classifiers using modified semi-supervised learning algorithms. Moreover, the training sample collection strategy can be further improved to enhance the robustness and adaptability of the tracker.

Acknowledgments This work is supported by the Basic Science Research Program through the Brain Korea 21 PLUS Project and Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Education (2013R1A1A2013778).

References

1. Okuma, K., Taleghani, A., De Freitas, N., Little, J.J., Lowe, D.G.: A boosted particle filter: multitarget detection and tracking. Computer Vision-ECCV 2004, pp. 28–39, Springer (2004)
2. Comaniciu, D., Ramesh, V., Meer, P.: Kernel-based object tracking. IEEE Trans. Pattern Anal. Mach. Intell. **25**(5), 564–577 (2003)
3. Babu, R.V., Suresh, S., Makur, A.: Online adaptive radial basis function networks for robust object tracking. Comput. Vis. Image Underst. **114**(3), 297–310 (2010)
4. Tian, M., Zhang, W., Liu, F.: On-line ensemble svm for robust object tracking. Computer Vision-ACCV 2007, pp. 355–364, Springer (2007)
5. Yang, H., Shao, L., Zheng, F., Wang, L., Song, Z.: Recent advances and trends in visual tracking: a review. Neurocomputing **74**(18), 3823–3831 (2011)
6. Wu, H., Li, G., Luo, X.: Weighted attentional blocks for probabilistic object tracking. Vis. Comput. **30**(2), 229–243 (2014)
7. Zhang, S., Yao, H., Zhou, H., Sun, X., Liu, S.H.: Robust visual tracking based on online learning sparse representation. Neurocomputing **100**, 31–40 (2013)
8. Ma, Z., Wu, E.: Real-time and robust hand tracking with a single depth camera. Vis. Comput. **30**, 1–12 (2014)
9. Li, Z., He, S., Hashem, M.: Robust object tracking via multi-feature adaptive fusion based on stability: contrast analysis. Vis. Comput. 1–19 (2014). doi:10.1007/s00371-014-1014-6
10. Avidan, S.: Ensemble tracking. IEEE Trans. Pattern Anal. Mach. Intell. **29**(2), 261–271 (2007)
11. Wang, Q., Chen, F., Xu, W., Yang, M.-H.: Object tracking via partial least squares analysis. IEEE Trans. Image Process. **21**(10), 4454–4465 (2012)
12. Wang, D., Lu, H., Yang, M.-H.: Least soft-thresold squares tracking. In: Proceedings of the 2013 IEEE conference on computer vision and pattern recognition, Portland, Oregon, USA, pp. 2371–2378 (2013)
13. Xie, Y., Qu, Y., Li, C., Zhang, W.: Online multiple instance gradient feature selection for robust visual tracking. Pattern Recognit. Lett. **33**(9), 1075–1082 (2012)
14. Quan, W., Chen, J.X., Yu, N.: Robust object tracking using enhanced random ferns. Vis. Comput. **30**(4), 351–358 (2014)
15. Grabner, H., Leistner, C., Bischof, H.: Semi-supervised on-line boosting for robust tracking. Computer Vision-ECCV 2008, pp. 234–247, Springer (2008)
16. Kalal, Z., Mikolajczyk, K., Matas, J.: Tracking-learning-detection. IEEE Trans. Pattern Anal. Mach. Intell. **34**(7), 1409–1422 (2012)
17. Zhan, J., Su, Z., Wu, H., Luo, X.: Robust tracking via discriminative sparse feature selection. Vis. Comput. 1–14 (2014). doi:10.1007/s00371-014-0984-8
18. Babenko, B., Yang, M.-H., Belongie, S.: Robust object tracking with online multiple instance learning. IEEE Trans. Pattern Anal. Mach. Intell. **33**(8), 1619–1632 (2011)
19. Zhang, K., Song, H.: Real-time visual tracking via online weighted multiple instance learning. Pattern Recognit. **46**, 397–411 (2013)
20. Freund, Y., Schapire, R.E.: Experiments with a new boosting algorithm. In: Proceedings of the 13th international conference on machine learning, Bari, Italy, pp. 148–156, 1996
21. Mallapragada, P.K., Jin, R., Jain, A.K., Liu, Y.: Semiboost: boosting for semi-supervised learning. IEEE Trans. Pattern Anal. Mach. Intell. **31**(11), 2000–2014 (2009)
22. Zhu, X.: Semi-supervised learning literature survey. Technical report 1530. University of Wisconsin-Madison, Computer Science (2005)
23. Xu, X.-S., Jiang, Y., Xue, X., Zhou, Z.-H.: Semi-supervised multi-instance multi-label learning for video annotation task. In: Pro-

ceedings of the 20th ACM international conference on multimedia, Nara, Japan, pp. 737–740 (2012)
24. Collins, R., Liu, Y., Leordeanu, M.: Online selection of discriminative tracking features. IEEE Trans. Pattern Anal. Mach. Intell. **27**, 1631–1643 (2005)
25. Wu, Y., Lim, J., Yang, M.H.: Online object tracking: a benchmark. In: Proceedings of the 2013 IEEE conference on computer vision and pattern recognition, Portland, Oregon, USA, pp. 2411–2418 (2013)
26. Zhang, K., Zhang, L., Yang, M.-H.: Fast compressive tracking. IEEE Trans. Pattern Anal. Mach. Intell. **36**(10), 2002–2015 (2014)
27. Zhang, K., Zhang, L., Yang, M.: Real-time object tracking via online discriminative feature selection. IEEE Trans. Image Process. **22**, 4664–4677 (2013)
28. Zhang, K., Zhang, L., Yang, M.-H.: Real-time compressive tracking. Computer Vision-ECCV 2012, pp. 864–877: Springer (2012)
29. Ross, D.A., Lim, J., Lin, R.-S., Yang, M.-H.: Incremental learning for robust visual tracking. Int. J. Comput. Vis. **77**(1–3), 125–141 (2008)
30. Sevilla-Lara, L., Learned-Miller, E.: Distribution fields for tracking. In: Proceedings of the 2012 IEEE conference on computer vision and pattern recognition, Providence, RI, USA, pp. 1910–1917 (2012)
31. Bao, C., Wu, Y., Ling, H., Ji, H.: Real time robust L1 tracker using accelerated proximal gradient approach. In: Proceedings of the 2012 IEEE conference on computer vision and pattern recognition, Providence, RI, USA, pp. 1830–1837 (2012)
32. Zhang, T., Ghanem, B., Liu, S., Ahuja, N.: Robust visual tracking via multi-task sparse learning. In: Proceedings of the 2012 IEEE conference on computer vision and pattern recognition, Providence, RI, USA, pp. 2042–2049 (2012)

Zhihui Wang received the B.Sc. degree in applied mathematics from Qufu Normal University, China, in 2009, and the M.Sc. degree in applied mathematics from China Jiliang University, in 2012. He is currently working towards the Ph.D. degree in electronics and information engineering from Chonbuk National University, Korea. His research interests include artificial intelligence, pattern recognition, and visual tracking.

Sook Yoon received the B.S., M.S., and Ph.D. degrees in engineering from Chonbuk National University, Jeonbuk, Korea, in 1993, 1995, and 2003, respectively. Until June 2006, she conducted her postdoctoral research work in electrical engineering at the University of California, Berkeley. She is presently a associate professor at Department of Multimedia Engineering, Mokpo National University, Jeonnam, Korea. Her current research interests include image processing, pattern recognition, machine learning, and multimedia computing.

Shan Juan Xie is a lecturer in Hangzhou Normal University, who received her M.S. and Ph.D. degree from Chonbuk National University, Republic of Korea. Her areas of interest are biometrics, pattern recognition, image processing, and machine learning, in which she has authored almost 40 papers in related international journals and conferences.

Yu Lu received the B.S. degree from the Nanchang Institute of Technology, Nanchang, China, in 2008, and M.S. degree from Jiangxi University of Finance and Economics, Nanchang, China, in 2011. He is currently a PhD student in Chonbuk National University, Jeonju, South Korea. His research interests include image processing, biometrics, pattern recognition, computer vision, and intelligent transportation system.

Dong Sun Park received the B.S. degree from Korea University, Seoul, Korea, in 1979, and the M.S. and Ph.D. degrees from the University of Missouri, Columbia, in 1984 and 1990, respectively. He is currently a Professor with the School of Electronic Engineering, Chonbuk National University, Jeonbuk, Korea. His current research interests include image processing, pattern recognition, computer vision, and artificial intelligence.