Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2013, Article ID 323072, 10 pages
http://dx.doi.org/10.1155/2013/323072

Research Article

Vision Target Tracker Based on Incremental Dictionary Learning and Global and Local Classification

Yang Yang, Ming Li, Fuzhong Nian, Huiya Zhao, and Yongfeng He

School of Computer and Communication, Lanzhou University of Technology, Lan Zhou 730050, China

Correspondence should be addressed to Ming Li; lim3076@163.com

Received 28 February 2013; Accepted 2 April 2013

Academic Editor: Yong Zhang

Based on sparse representation, a robust global and local classification algorithm for visual target tracking in uncertain environment was proposed in this paper. The global region of target and the position of target would be found, respectively by the proposed algorithm. Besides, overcompleted dictionary was obtained and updated by biased discriminate analysis with the divergence of positive and negative samples at current frame. And this over-completed dictionary not only discriminates the positive samples accurately but also rejects the negative samples effectively. Experiments on challenging sequences with evaluation of the state-of-the-art methods show that the proposed algorithm has better robustness to illumination changes, perspective changes, and targets rotation itself.

1. Introduction

Visual target tracking in uncertain environment is an important component in the field of computer vision [1]. In uncertain scene, the negative impact on quality of target is mainly caused by occlusion, pose changes, significant illumination variations, and so on. Therefore, discrimination method with a strong robustness against the target and environment changes is required for accurate tracking. Visual target tracking can be treated as a binary classification problem between targets and backgrounds, target candidate set of which is established by affine transformation, and classifier is then used to discriminate the target from candidate set [2]. Therefore, classifier should be not only well discriminated to targets but also capable of rejecting the discrimination of background feature and even has better robustness to occlusions, pose changes, and illumination variations.

In this paper, an incremental tracking algorithm was proposed for resolving the target appearance variations and occlusions problems. The system chart as Figure 1. Object is represented with global and local sparse representation, and tracking task is formulated as sparse representation binary classification problem with dictionary incremental learning. For object representation, targets are treated as positive samples, whereas backgrounds are treated as negative samples. Then positive and negative samples are used to establish the discriminatory dictionary, where target appearance model is treated as linear combination with discriminatory dictionary and sparse coding. In the first frame, targets are affine-transformed to affine transformation subspace, where the target is found with the minimum reconstruction error. Global classifier with sparse representation is established to determine the global region of target from center-point collection, while sparse representation local classifier is used to set up discrimination to find the target location from global region. As we know, the appearance of the target itself and the external scenes vary in real time, so dictionary needs to be updated with features of the next frame by incremental learning to ensure tracking result accurately.

The rest of the paper is organized as follows. Section 2 reviews the related works. Section 3 proposes target motion model and sparse representation globe classifier and local classifier algorithm. Furthermore, dictionary learning and

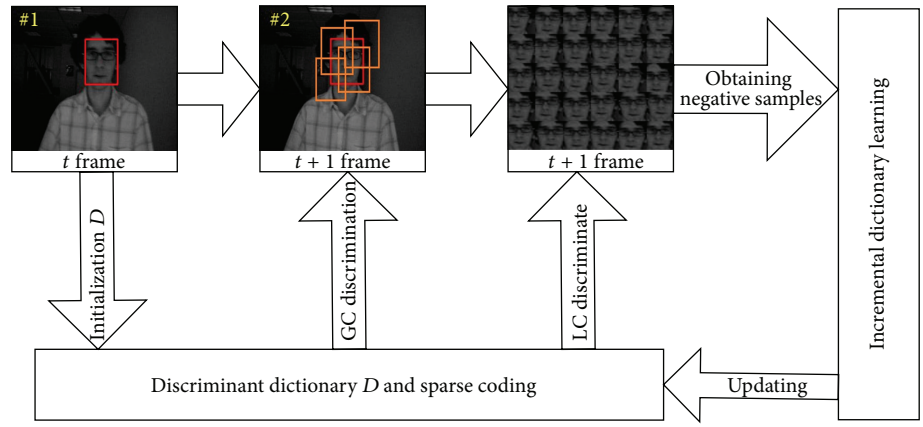


Figure 1: Diagram of targets tracking via sparse representation global and local classifier.

incremental updating are introduced in Section 4. Section 5 reports the experimental results. Finally, the conclusions are summarized in Section 6.

2. Related Works

Currently, according to the target appearance model, target tracking methods can be divided into two categories: generated and discriminative model methods. The generated model method use appearance model to replace the target observed template, tracking result get from the highest similarity search with the appearance model of the area. For example, the mean-shift [3] and the incremental tracking [4]. In [4], in order to make the algorithm adapt to the real-time changes of target appearance effectively, the target appearance models are incremental learned by a group of low-dimensional subspaces. Discriminative model method: cast the tracking as a binary classification problem. Tracking is formulated as finding the target location that can accurately separate the target from the background. In [5], online multiple instance learning methods improve the robustness of the target tracking system to the influence of occlusion. In [6], visual object tracking is construed as a numerical optimization problem and applies cluster analysis to the sampled parameter space to redetect the object and renew the local tracker. In [7], an ensemble of weak classifiers is trained online to distinguish between the object and the background, and the weak classifiers are combined into a strong classifier using AdaBoost; then, the strong classifier is used to label pixels in the next frame as either belonging to the object or the background.

Some scholars have access to a stable target tracking system, which takes advantage of sparse representation classification model for target tracking. In [8], the set of trivial template are constructed with occlusion and corruption, target candidate is sparsely represented by target and trivial templates at new frames, then, the smallest projection error is taken to find target during tracking. In [9], sample-based adaptive sparse representation is proposed to address partial occlusion or deterioration; object representations are described as sample basis with exploiting L1-norm minimization. In [10], the paper proposed a dynamic group sparsity and two-stage sparse optimization to jointly minimize the target reconstruction error and maximize the discriminative power. In [11], tracking is achieved by minimum error bound and occlusion detection, the minimum error bound is calculated for guiding particle resampling, and occlusion detection is performed by investigating the trivial coefficients in the L1-norm minimization.

In addition, to resolve the problems that the overcompleted dictionary cannot discriminate sufficiently, Xuemei utilized target template to structure dictionary $[T, I, -I]$ [9, 11]; I and $-I$ are not related unit matrix, and T is a small piece of target templates. In [12], the use of the learning obtained in the sparse representation dictionary is more effective than a preset dictionary. In [13] proposed a dictionary learning function, the dictionary was obtained by K-SVD algorithm and linear classifier. In [14], dictionary learning problem is deemed as optimization of a smooth nonconvex over convex sets and proposed an iterative online algorithm that solves this problem by efficiently minimizing at each step a quadratic surrogate function of the empirical cost over the set of constraints. On this basis, paper [15] proposes a new discriminative DL framework by employing the Fisher discrimination and criterion to learn a structured dictionary.

3. Sparse Representation Global and Local Classifier

3.1. Motion Model of Targets. We denote affine transformation parameters $X_t = (x, y, s, r, \theta, \lambda)$ as target state in frame t, where X and Y are coordinates of center point, s is change of scale, r is bearing rate, θ is rotation angle, and λ is angle of inclination. The motion model of the object through

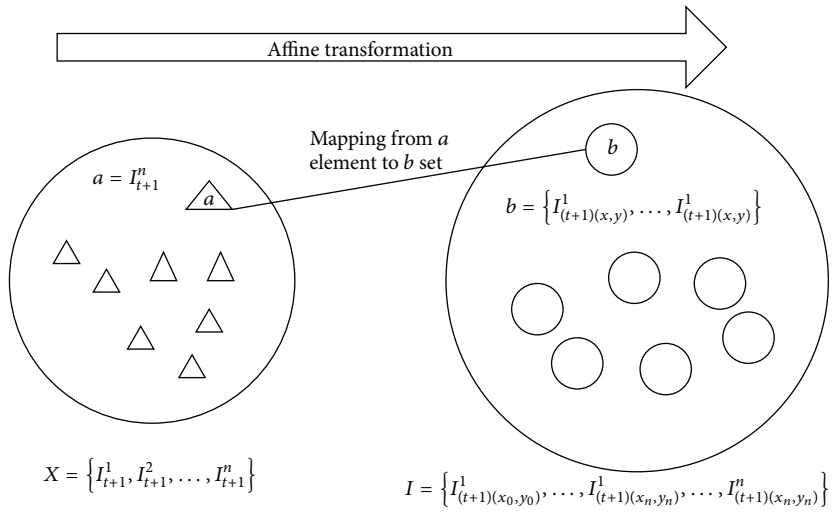


Figure 2: Affine transformation from X to I.

the transfer of the probability state of the affine transformation parameters is obtained, the function of motion model is as follows:

$$P(X_{t+1} \mid X_t) = N(X_{t+1} \mid X_t, \sigma), \quad (1)$$

where $N(X_{t+1} \mid X_t)$ is modeled independently by a Gaussian distribution, σ is a covariance diagonal matrix, and the elements of the diagonal matrix are the variance of each of the affine parameters. $\{X^1_{t+1}, X^2_{t+1}, \ldots, X^n_{t+1}\}$ is a group of affine parameter sets which are randomly generated by function (1), in current frame, and $\{I^1_{t+1}, I^2_{t+1}, \ldots, I^n_{t+1}\}$ is area of the target that may occur (candidate image area) which can be constructed by affine transformation from $\{X^1_{t+1}, X^2_{t+1}, \ldots, X^n_{t+1}\}$. Then, find the area of target from candidate image by sparse representation classifier; the classifier is trained by using previous tracking result.

3.2. Sparse Representation Classifier. Wright et al. [16] proposed the sparse representation-based classification (SRC) method for robust face recognition (FR). We denote $A = [A_1, A_2, \ldots, A_C]$ as the set of original training samples, where A_i is the subset of the training samples from class i. c is class numbers of subjects, and y is a testing sample. The procedures of sparse representation classifier are as follows:

$$\widehat{a} = \arg\min_a \left\{ \|y - A\alpha\|_2^2 + \gamma\|\alpha\|_1 \right\}, \quad (2)$$

where γ is a scalar constant,

classification via

$$\text{identity}(y) = \arg\min_i \{e_i\}, \quad (3)$$

where $e_i = \|y - A_i\widehat{\alpha}_i\|$, $\widehat{\alpha}_i = [\widehat{\alpha}_1; \widehat{\alpha}_2; \ldots; \widehat{\alpha}_c]$ and $\widehat{\alpha}_i$ is the coefficient vector associated with class i.

3.3. Sparse Representation Global and Local Classifier. We divided the set of target states $X_t = (x, y, s, r, \theta, \lambda)$ into two parts: $X_t = (x, y)$ and $X_{t,(x,y)} = (s, r, \theta, \lambda)$, $X_t = (x, y)$ is center coordinates and $X_{t,(x,y)} = (s, r, \theta, \lambda)$ is local state in time t, $\{I^1_{t+1}, I^2_{t+1}, \ldots, I^n_{t+1}\}$ is obtained by affine transform from $X_t = (x, y)$, $I^i_{(t+1)(x,y)}$ is obtained by affine transform from $X_{t,(x,y)} = (s, r, \theta, \lambda)$, and the relationship between two sets is as shown in Figure 2.

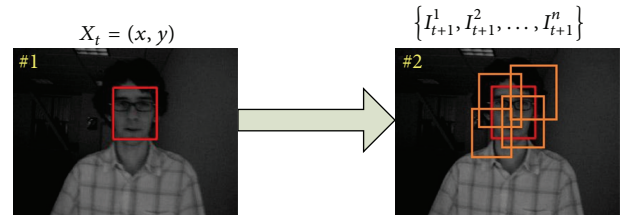


Figure 3: Image set via affine transformation.

Each element of the set I can be obtained from the affine transformation of $X_t = (x, y, s, r, \theta, \lambda)$; usually, the element numbers in set I are very large, and computational cost for discriminating the set immediately is the key issue. All of the subsets b in set I can be obtained by the affine transformation of element a in set X illustrated in Figure 3. In order to reduce the computational cost, search element a from set X first, and, then, search element b from the subset that is mapping of element a. However, that implies the need for training the two classifiers to role set X and a collection of I, and the computational cost for classifiers is raised once again. In sparse representation classifier models, the method updating completed dictionary can achieve the purpose of training multiple classification and then reduce the computation cost for the classifiers trained.

The X set constructed by center-coordinates affine transformation,in which, most of elements containing numerous negative samples features. Figure 3 shows that the classified algorithm for set X is equivalent to sparse representation global classifier, namely, SRGC. Considering target in two

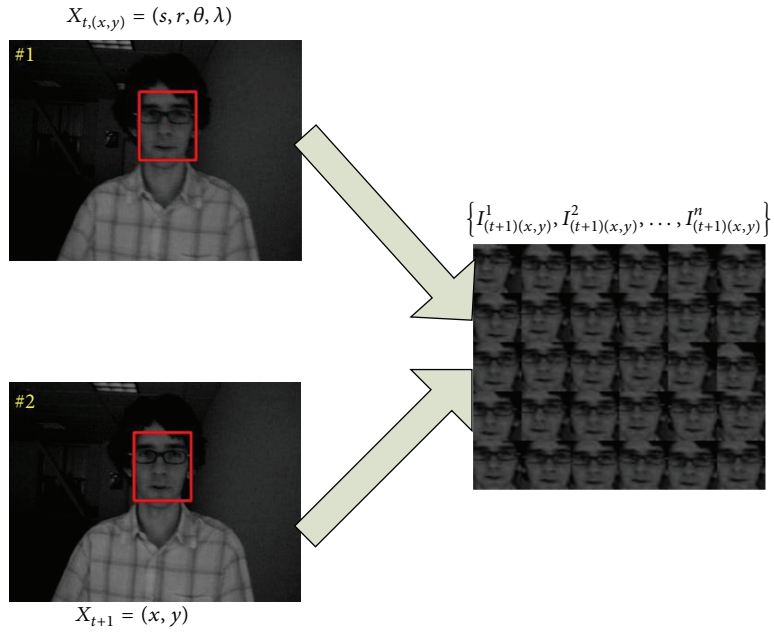


Figure 4: Obtaining target set of states.

frames with maximum likelihood, nonzero entries in the sparse coding are more concentrated in the same position; we add a constraint $\|\widehat{\alpha}_{\mathrm{GC}} - X\|_2^2$ in reconstruction error function to make sure of the maximum likelihood at sparse coding in current frames and prior to the tracking results. We define the metric for classification as follows:

$$\widehat{\alpha}_{\mathrm{GC}} = \arg\min_{a_{\mathrm{GC}}} \left\{ \|y - D\alpha_{\mathrm{GC}}\|_2^2 + \gamma \|\alpha_{\mathrm{GC}}\|_1 \right\},$$
$$e = \|y - D\widehat{\alpha}_{\mathrm{GC}}\|_2^2 + \omega \|\widehat{\alpha}_{\mathrm{GC}} - X\|_2^2, \tag{4}$$

where γ is a scalar constant, ω is a preset weight coefficient, α_{GC} is coding coefficient vector for determining the location of the center coordinates, X is coefficient vector and prior to the tracking results. The sparse representation global classifier is made by (3).

In current frame, fix the target center-point, then the set of target local status is constructed by affine transformation with the last tracking result, illustrated in Figure 4. Taking into account that most target features are imprisoned in elements of set I, we consider that discrimination in this part can be equivalent to sparse representation local classifier, as SRLC. The Objective function can be transformed into coding function over local dictionary by adding constraint $\|\alpha_{\mathrm{LC}} - X\|_2^2$ in. We add coding discriminant fidelity term $\|\alpha_{\mathrm{LC}}\|_1$ in reconstruction error function to ensure that the target is the most sparse coding in the local dictionary. We define the function for classification as follows:

$$\widehat{a}_{\mathrm{LC}} = \arg\min_{a_{\mathrm{LC}}} \left\{ \|y - D\alpha_{\mathrm{LC}}\|_2^2 + \gamma_1 \|\alpha_{\mathrm{LC}}\|_1 + \gamma_2 \|\alpha_{\mathrm{LC}} - X\|_2^2 \right\},$$
$$e = \|y - D\widehat{\alpha}_{\mathrm{LC}}\|_2^2 + \gamma_1 \|\alpha_{\mathrm{LC}}\|_1. \tag{5}$$

The sparse representation local classifier is made by (3). The proposed algorithm is summarized in Algorithm 1.

4. Learning and Incremental Updating for Dictionary

According to the aforementioned code and discriminant function mentioned in last section, the coefficient of α_{GC} and α_{LC} could not have prominent sparsity if overcompleted dictionary has bad discriminant results as well; all of the samples could probably be chosen as the sparsest code, which is bad for the classification performance of reconstruction error function.

Therefore, biased discriminant analysis [17] (BDA) method was introduced into the dictionary learning function in this paper, taking effect for objective and opposite for nonobjective. So the dispersity expressions of plus and minus samples of BDA (S_+ and S_-) are as follows:

$$S_+ = \sum_{i=1}^{N_+} (y_i^+ - u_+)(y_i^+ - u_+)^T,$$
$$S_- = \sum_{j=1}^{N_-} (y_j^- - u_+)(y_j^- - u_+)^T. \tag{6}$$

N_+ and N_- are the total number of plus and minus samples, respectively; y_i^+ and y_j^- are the ith and jth element of plus set $\{y_1^+, y_2^+, \ldots, y_{N_+}^+\}$ and minus set $\{y_1^-, y_2^-, \ldots, y_{N_+}^-\}$; u_+ is the mean value of plus sample set.

4.1. Dictionary Learning Using Biased Discriminant Analysis (BDA). Give a dictionary $D = [d_1, d_2, \ldots, d_n]$, where d_i is

Input: I_t is the tracking result of prior frame, $\{I_{t+1}^1, I_{t+1}^2, \ldots, I_{t+1}^n\}$ is set of candidate samples credible positions of the center-point coordinates in next frame. D is Over-complete dictionary, T is frame numbers.

(1) for $t = 1 : T$
(2) SRGC
calculate $\widehat{a}_{\mathrm{GC}} = \arg\min_{a_{\mathrm{GC}}} \left\{ \left\| y - D\alpha_{\mathrm{GC}} \right\|_2^2 + \gamma \left\| \alpha_{\mathrm{GC}} \right\|_1 \right\}$
(3) calculate identity $(y) = \arg\min_i \{e_i\}$; where $y = \{I_{t+1}^1, I_{t+1}^2, \ldots, I_{t+1}^n\}$,
(4) obtaining the center point (x, y) from $I_{(t+1)}^i$
(5) obtaining the $\left\{ I_{(t+1)(x,y)}^1, I_{(t+1)(x,y)}^2, \ldots, I_{(t+1)(x,y)}^n \right\}$ by (x, y) and Affine transformation of I_t
(6) SRLC to $\left\{ I_{(t+1)(x,y)}^1, I_{(t+1)(x,y)}^2, \ldots, I_{(t+1)(x,y)}^n \right\}$;
calculate $\widehat{a}_{\mathrm{LC}} = \arg\min_{a_{\mathrm{LC}}} \{ \left\| y - D\alpha_{\mathrm{LC}} \right\|_2^2 + \gamma_1 \left\| \alpha_{\mathrm{LC}} \right\|_1 + \gamma_2 \left\| \alpha_{\mathrm{GC}} - X \right\|_2^2 \}$;
(7) calculate identity $(y) = \arg\min_i \{e_i\}$, where $e = \left\| y - D\widehat{a}_{\mathrm{LC}} \right\|_2^2 + \gamma_1 \left\| \alpha_{\mathrm{LC}} \right\|_1$

Output: $I_{(t+1)(x,y)}^i$

Algorithm 1: Algorithm of Fisher discrimination dictionary learning.

an n-dimensional vector $d_i = [d_i^1, d_i^2, \ldots, d_i^n]^T$ and d_i^j is the jth element of ith vector which is called atom of dictionary. $A = [A_+, A_-]$ is the training sample set, where A_+ and A_- are the characterized and noncharacterized samples for objective, which are also called plus and minus samples.

However, for objective tracing, only the region of objective is interested, so the background characteristics, noises, occlusions, and so on are regarded as noncharacterized samples A_-. Let $X = [x_1, x_2, \ldots, x_n]$ be the code coefficient vector of sample set A of dictionary D, provided that the tested sample set can be denoted as $A \approx DX$. Furthermore, dictionary learning function is

$$J_{(D,X)} = \arg\min_{(D,X)} \left\{ \left\| A_+ - DX \right\|_F^2 + \lambda_1 \|X\|_1 + \lambda_2 f(X) \right\}, \quad (7)$$

where $\|A_+ - DX\|_F^2$ is the discriminant fidelity term which is only used for A_+, since the interesting thing in objective tracing is only the area of objective. $\|X\|_1$ is $l1$-norm sparse constraint term, and $f(X)$ is the discriminant constraint with respect to coefficient vector X.

According to BDA discriminant rule, let $f(X)$ be $\mathrm{tr}(S_+) - \mathrm{tr}(S_-)$. Let $\|X\|_F^2$ be added into $f(X)$ as a relaxed term because the function $f(X)$ is nonconvex and unstable, therefore

$$f(X) = \mathrm{tr}(S_+) - \mathrm{tr}(S_-) + \eta \|X\|_F^2, \quad (8)$$

where η is the control variable. Furthermore, the proposed BDDL method can be formed as

$$J_{(D,X)} = \arg\min_{(D,X)} \left\{ \left\| A_+ - DX \right\|_F^2 + \lambda_1 \|X\|_1 + \lambda_2 \left(\mathrm{tr}(S_+) - \mathrm{tr}(S_-) \right) + \eta \|X\|_F^2 \right\}. \quad (9)$$

Similar to [15], (D, X) is nonconvex for function J which is the convex function on set X when D is already known and also the convex function on set D when X is already known. So, J is in fact a biconvex function on sets D and X.

4.2. Dictionary Incremental Updating. A new plus and minus samples set $Y_{\mathrm{new}}^+ = \{y_1^+, y_2^+, \ldots, y_{M_+}^+\}_{\mathrm{new}}$ and $Y_{\mathrm{new}}^- = \{y_1^-, y_2^-, \ldots, y_{M_+}^-\}_{\mathrm{new}}$ can be obtained according to current objective tracing result, the mean value of which is $u_{\mathrm{new}}^+ = (1/m_+) \sum_{i=1}^{m_+}{}_{\mathrm{new}} y_i^+$ and $u_{\mathrm{old}}^+ = (1/n_+) \sum_{i=1}^{n_+}{}_{\mathrm{old}} y_i^+$, respectively; m_+ and n_+ are the number of new and old samples. Furthermore, the weighted mean of these two mean values of plus sample set is

$$u_+ = \frac{n_+ u_{\mathrm{old}}^+ + m_+ u_{\mathrm{new}}^+}{n_+ + m_+}. \quad (10)$$

Similarly, the new dispersity expression of plus sample set using weighted mean value u_+ is

$$S_{\mathrm{new}}^+ = \sum_{i=1}^{M_+} \left({}_{\mathrm{new}} y_i^+ - u_+ \right) \left({}_{\mathrm{new}} y_i^+ - u_+ \right)^T. \quad (11)$$

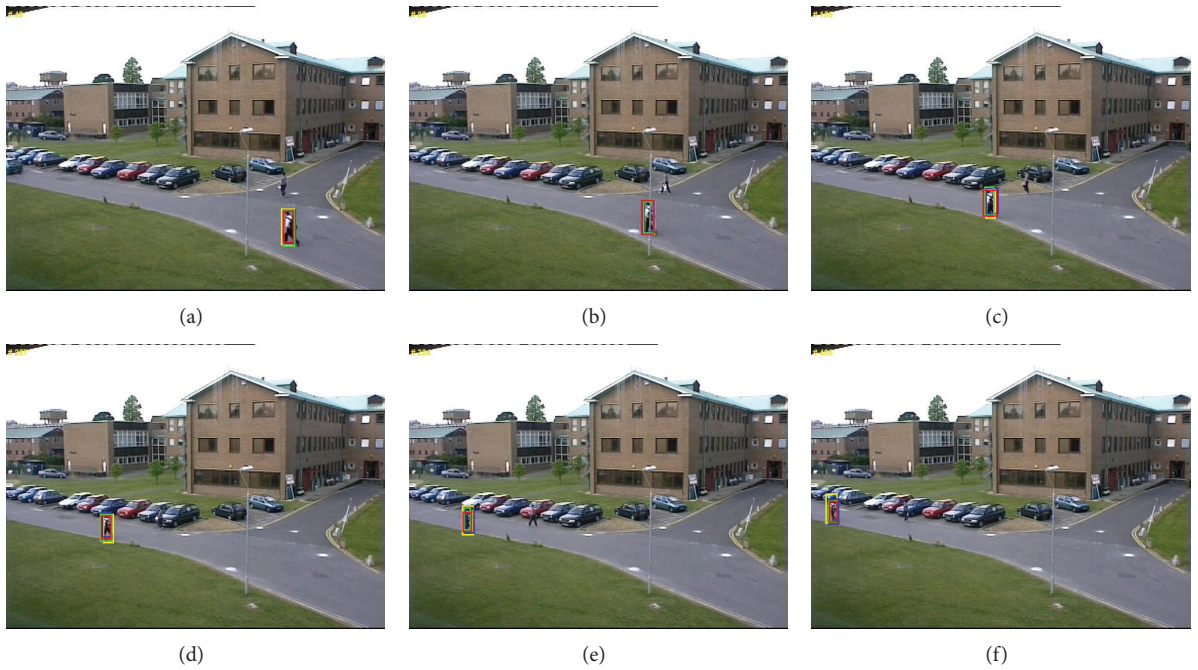
The dispersity expression of the updated plus sample S_+ is

$$S_+ = S_{\mathrm{old}}^+ + S_{\mathrm{new}}^+ + \frac{n_+ m_+}{n_+ + m_+} \left(u_{\mathrm{new}}^+ - u_{\mathrm{old}}^+ \right) \left(u_{\mathrm{new}}^+ - u_{\mathrm{old}}^+ \right)^T. \quad (12)$$

S_{old}^+ is the old dispersity expression of plus sample set.

However, we need just refused-discriminant to negative samples, instead of discriminant it in real time, then the dispersity of negative samples is as follows:

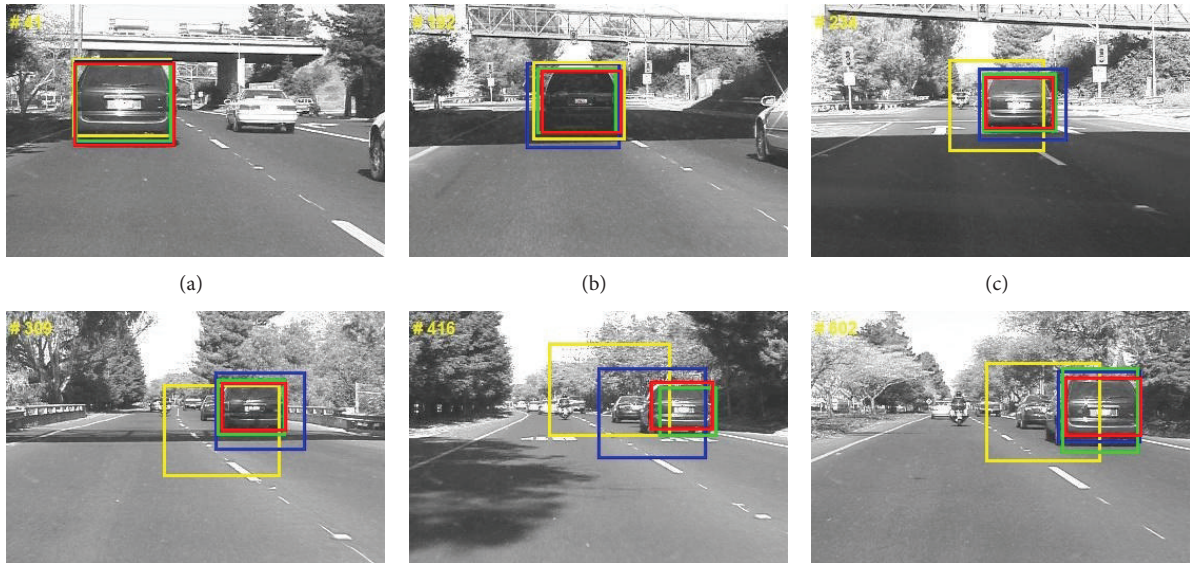
$$S_- = S_{\mathrm{old}}^- + S_{\mathrm{new}}^-,$$
$$S_{\mathrm{new}}^- = \sum_{j=1}^{M_-} \left(y_j^- - u_+ \right) \left(y_j^- - u_+ \right)^T. \quad (13)$$



(a) (b) (c)

(d) (e) (f)

Figure 5: Tracking results of the PETS01D1Human1 sequence (MIL is yellow, IVT is blue, L1 is green, and our tracker is red).



(a) (b) (c)

(d) (e) (f)

Figure 6: Tracking results of the Car4 sequence (MIL is yellow, IVT is blue, L1 is green, and our tracker is red).

If we take S_- and S_+ into consideration, then the updated function $f(X) = \mathrm{tr}(S_+) - \mathrm{tr}(S_-) + \eta\|X\|_F^2$ can be represented as

$$f(X) = \mathrm{tr}\left(S_{\mathrm{old}}^{+} + S_{\mathrm{new}}^{+} + \Psi\right) - \mathrm{tr}\left(S_{\mathrm{old}}^{-} + S_{\mathrm{new}}^{-}\right) + \eta\|X\|_F^2, \tag{14}$$

where $\Psi = (n_+ m_+/(n_+ + m_+))(u_{\mathrm{new}}^{+} - u_{\mathrm{old}}^{+})\ (u_{\mathrm{new}}^{+} - u_{\mathrm{old}}^{+})^T$.

According to (14), fix D_{old}, and then compute X; D is reconstructed by obtaining X that is updating D, where X is not used for discriminant immediately and is just reconfigurable coding coefficient matrix:

$$J_{(X)} = \arg\min_{X} \left\{\left\|A_+ - D_{\mathrm{old}}X\right\|_F^2 + \lambda_1\|X\|_1 + \lambda_2 f(X)\right\}, \tag{15}$$

where $f(X) = \mathrm{tr}(S_{\mathrm{old}}^{+} + S_{\mathrm{new}}^{+} + \Psi) - \mathrm{tr}(S_{\mathrm{old}}^{-} + S_{\mathrm{new}}^{-}) + \eta\|X\|_F^2$, D_{old} is the old dictionary; A_+ is the joint matrix of samples in current and previous frames, which is represented

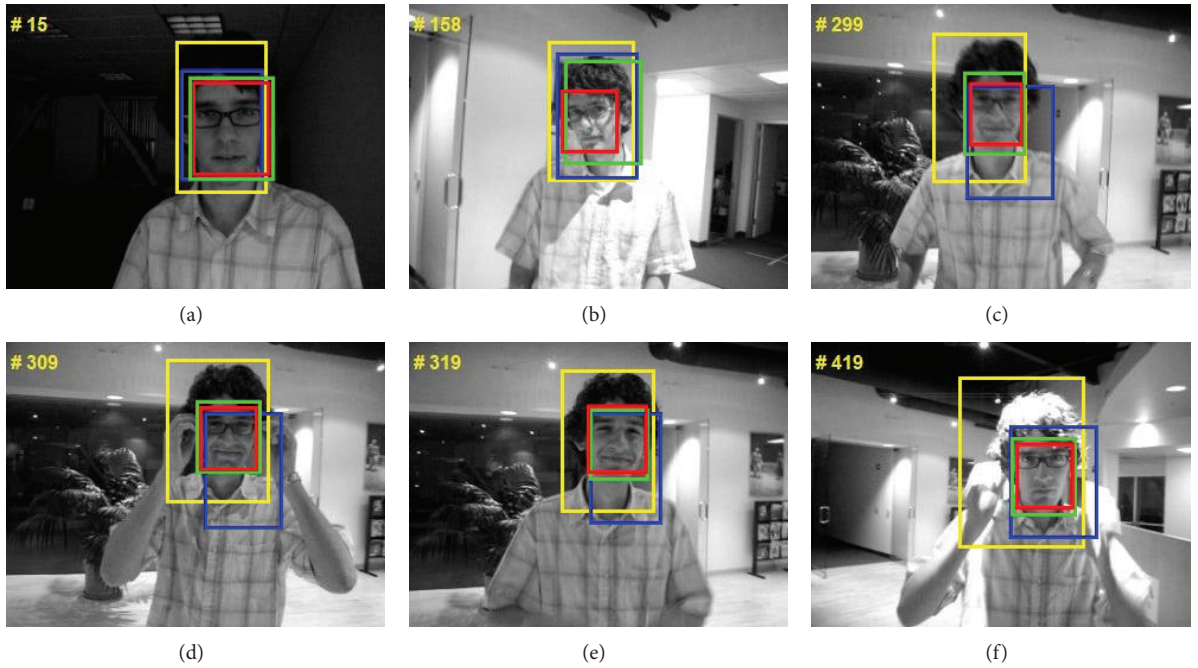


Figure 7: Tracking results of the David Indoor sequence (MIL is yellow, IVT is blue, L1 is green, and our tracker is red).

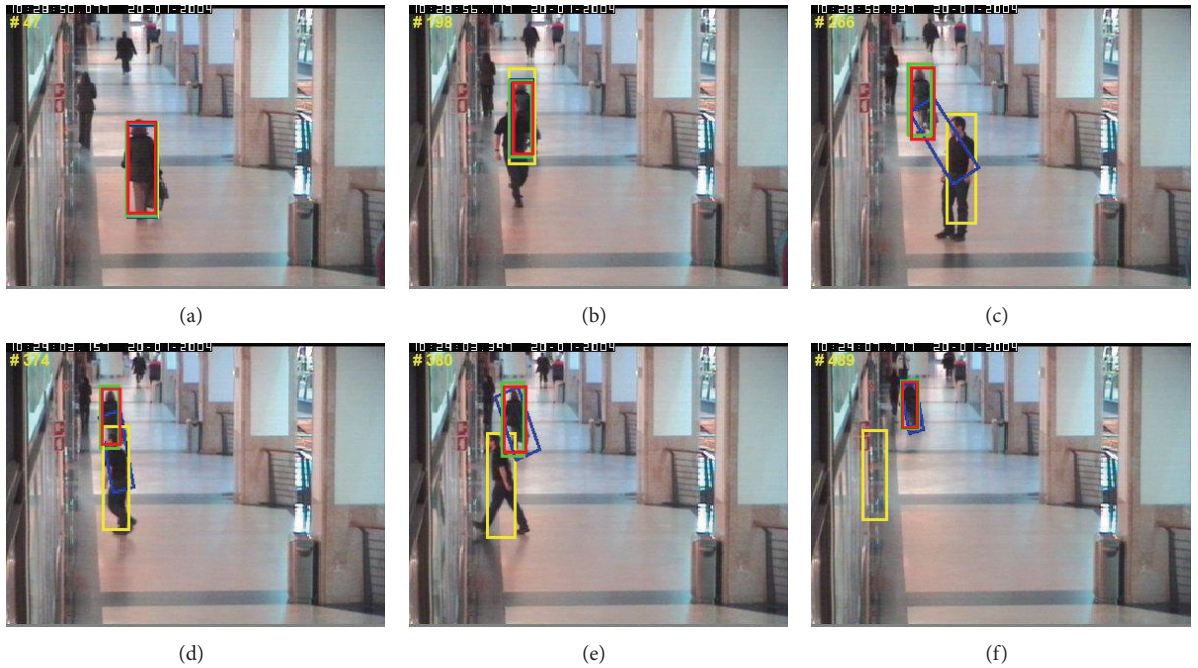


Figure 8: Tracking results of the OneLeaveShopReenter2cor sequence (MIL is yellow, IVT is blue, L1 is green, and our tracker is red).

as $A_+ = [Y^+_{\text{new}_{t-1}}; Y^+_{\text{new}_t}]$, $Y^+_{\text{new}_{t-1}} = \{y^+_1, y^+_2, \ldots, y^+_{M_+}\}_{\text{new}_{t-1}}$ and $Y^+_{\text{new}_t} = \{y^+_1, y^+_2, \ldots, y^+_{M_+}\}_{\text{new}_t}$. Then function $J_{(D,X)}$ could be rewritten as

$$J_{(D)} = \arg\min_D \|A_+ - DX\|_F^2. \tag{16}$$

In first frames we need initialization; the target is framed manually, the Y_0^+ is set of initial moment positive samples, Y_0^- is the set of initial moment negative sample, u_0^+ is the mean value of initial moment positive sample, compute S_+ and S_- by (6). We initialize all atoms p of dictionary D as random vector with $l2$-norm, solve X by solving (15), and then fix X and solve D by solving (16).

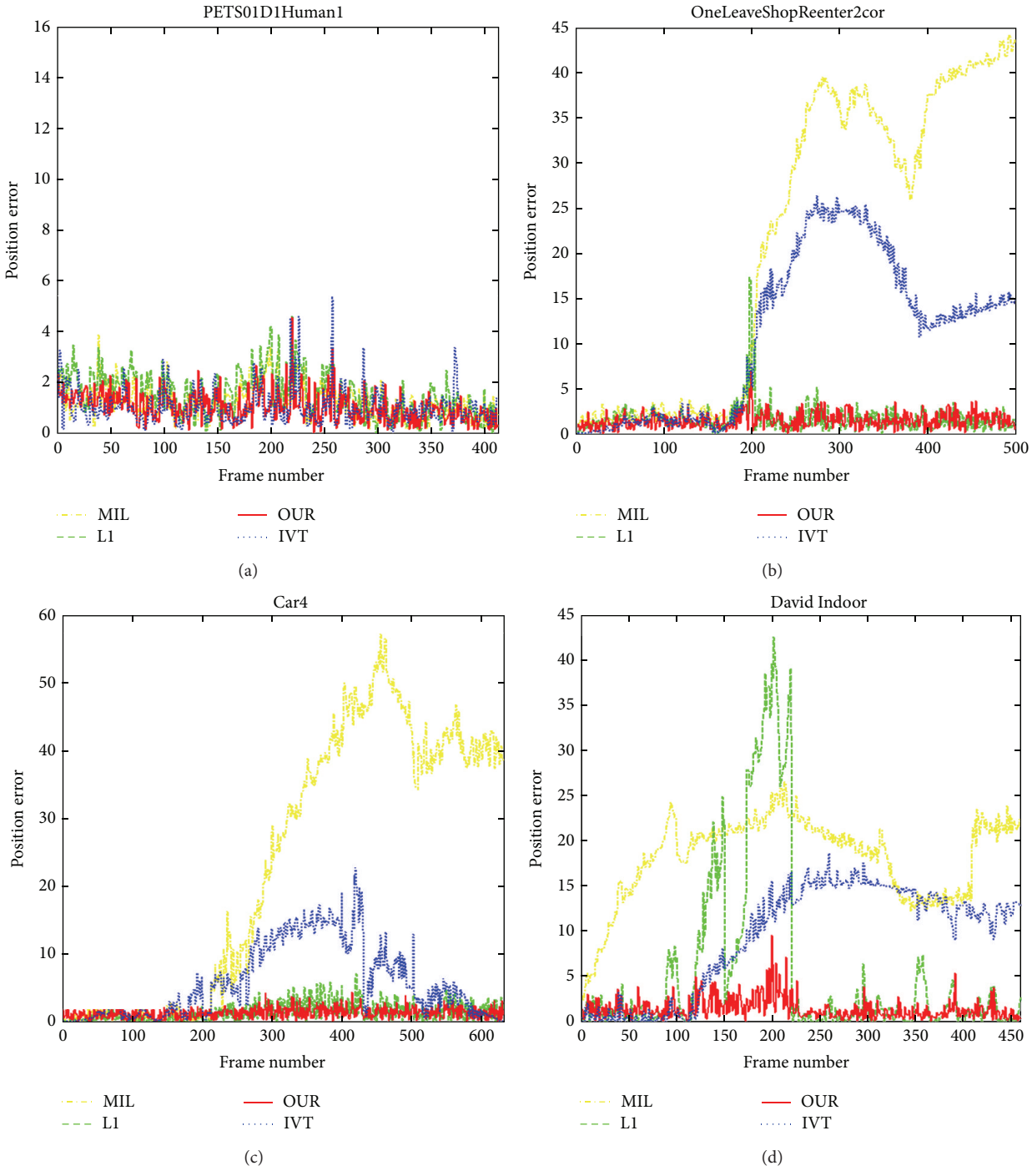


Figure 9: Position error plots of the tested sequences.

5. Experiments

Experiments were performed with four video sequences, which included occlusion, illumination variation, appearance variation and other corruption factors. In the experiment, target location of the first frame is framed manually, and initial dictionaries were randomly generated, and track results were to be released as rectangular boxes. Figures 5, 6, 7 and 8 show a representative sample of tracking results. finally, target tracking method of this paper contrasts with incremental visual tracking (IVT) [4], multiple instance Learning (MIL) [5] and L1 tracker (L1) [8]. To further evaluate the proposed method, each method applies to four video sequences and then compares the track results, that need to be evaluated qualitatively and quantitatively. The paper uses gray histogram as a way of presentation characteristics of

Table 1: Analysis of location errors.

	IVT			L1			MIT			Proposed		
	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std	Max	Mean	Std
OneLeaveShopReenter2cor	26.32	11.26	8.82	7.33	2.46	**1.72**	44.21	21.67	16.68	*6.65*	*2.02*	1.86
David Indoor	18.76	9.43	5.98	42.59	57.64	9.93	18.76	20.51	4.67	*6.71*	*1.45*	*1.21*
Car4	22.71	5.89	2.22	8.71	2.89	**1.44**	59.54	23.07	6.37	*7.41*	*1.36*	1.45
PETS01D1Human1	5.98	**0.93**	**0.87**	**5.09**	1.68	0.97	6.12	1.16	1.14	5.95	1.39	0.91

each method, and this is easy to test and judge the sensitivity of corruption. Besides, each run of the experiment used the same starting position, initial frame, frame number of video sequences, and software environment.

5.1. Qualitative Comparison. The test sequences, PETS01D1Human1, show that a person went to the extreme left side from the lower right corner of the screen, which telephone poles will cause short-term shelter to it. Tracking results are shown in Figure 5; the image frames are # 49, # 88, # 155, # 257, # 324, and # 405. All methods can effectively track the target, and the tests show that in the circumstances of the same light intensity, the same camera angle, and slight shelter, all methods can effectively track the target. It also indicates that the proposed method in the paper and the contrastive method are effective target tracking algorithms.

In the Car4 sequence, when the cars pass through the bridge and the shade, intensity of illumination altered obviously. Tracking results are shown in Figure 6; and the image frames are #41, #792, #234, #309, #416, and #602. When the cars go through the bridge, MIL will be ineffective significantly, but will not lose target; IVT will also be ineffective, but it can snap back. The method in this paper and L, compared with MIL and IVT, can locate the target accurately.

In the David Indoor sequence, the degeneration is include twice illumination change, expression change, and partial occlusion. The track result was shown in Figure 7, and the image frames are #15, #158, #299, #309, #319, and #419. The method in this paper can locate the target accurately; contrastively, the result of L1 is ineffective. The reason is that the target gray histogram was changed by light intensity, that affects the feature of image gray histogram; the methods of MIL and IVT may be more sensitive to the effects.

The OneLeaveShopReenter2cor sequence shows a woman walking through a corridor, when a man walks by, which lead to large occlusions Clothes with similar color are the occluder. The track result was shown in Figure 8; the image frames are #47, #198, #266, #374, #380, and #489. The method in this paper and L1 can locate the target accurately. When occlusion happened, MIL put the occluder as target and missed the target; The target is similar with the occluded,and then the IVT is difficult to discriminate object and occluded.

In conclusion, both the method in this paper and L1 can locate the target accurately. And they have strong robustness for occlusions, pose changes, significant illumination variations, and so forth.

5.2. Quantitative Comparison. We evaluate the tracking performance by position error. The position Error is approximated by the distance between the central position of the tracking result and the manually labeled ground truth. Table 1 shows the statistical data of position error which includes maximum, mean and standard deviation. Figure 9 shows the errors of all four trackers.

From previous comparison results, we can see that proposed method can track the target more accurately in video sequence OneLeaveShopReenter2cor, David Indoor, and Car4 than other methods. The max, mean, and standard deviation of position errors are smaller than IVT and MIT. Therefore, in complex environment, our method has a better robustness. Comparing with L1, the result of tracking to sequence OneLeaveShopReenter2cor and Car4 shows that L1 has higher stability in the scene where illumination did not change significantly. However, the standard deviation of position error of L1 tracker in those sequences is smaller than proposed method, that L1 update capability is less than proposed method, when grayscale histogram distribution changed greatly. The dictionary in L1 is constructed by target template, so robustness of learned dictionary is better than it.

6. Conclusion

In this paper, a tracking algorithm was proposed based on sparse representation and dictionary learning. Based on biased discriminant analysis, we proposed an effective Incremental learning algorithm to construct overcompleted dictionary. Positive and negative samples are obtained during tracking process and are used for updating discriminant dictionary by biased discriminant analysis. Then we proposed sparse representation global and local classification for set of central points and set of local states. Compared to the state-of-the-art tracking methods, the proposed algorithm improves the discriminating performance of completed dictionary and the adaptive ability of appearance model. It has a strong robustness to illumination changes, perspective changes, and targets rotation itself.

References

[1] T. Bai and Y. F. Li, "Robust visual tracking with structured sparse representation appearance model," *Pattern Recognition*, vol. 45, pp. 2390–2404, 2012.

[2] F. Chen, Q. Wang, S. Wang, W. Zhang, and W. Xu, "Object tracking via appearance modeling and sparse representation," *Image and Vision Computing*, vol. 29, pp. 787–796, 2011.

[3] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 25, no. 5, pp. 564–577, 2003.

[4] D. A. Ross, J. Lim, R. S. Lin, and M. H. Yang, "Incremental learning for robust visual tracking," *International Journal of Computer Vision*, vol. 77, no. 1–3, pp. 125–141, 2008.

[5] B. Babenko, S. Belongie, and M. H. Yang, "Visual tracking with online multiple instance learning," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPR '09)*, pp. 983–990, June 2009.

[6] Z. Yin and R. T. Collins, "Object tracking and detection after occlusion via numerical hybrid local and global mode-seeking," in *Proceedings of the 26th IEEE Conference on Computer Vision and Pattern Recognition (CVPR '08)*, pp. 1–8, June 2008.

[7] S. Avidan, "Ensemble tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 29, pp. 261–271, 2007.

[8] X. Mei and H. Ling, "Robust visual tracking using L1 minimization," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '09)*, pp. 1436–1443, 2009.

[9] Z. Han, J. Jiao, B. Zhang, Q. Ye, and J. Liu, "Visual object tracking via sample-based Adaptive Sparse Representation (AdaSR)," *Pattern Recognition*, vol. 44, no. 9, pp. 2170–2183, 2011.

[10] B. Liu, L. Yang, J. Huang, P. Meer, L. Gong, and C. Kulikowski, "Robust and fast collaborative tracking with two stage sparse optimization," *Lecture Notes in Computer Science*, vol. 6314, no. 4, pp. 624–637, 2010.

[11] X. Mei, H. Ling, Y. Wu, E. Blasch, and L. Bai, "Minimum error bounded efficient L1 tracker with occlusion detection," in *Proceedings of the IEEE Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1257–1264, 2011.

[12] J. A. Tropp and S. J. Wright, "Computational methods for sparse solution of linear inverse problems," *Proceedings of the IEEE*, vol. 98, pp. 948–958, 2010.

[13] Q. Zhang and B. Li, "Discriminative K-SVD for dictionary learning in face recognition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 2691–2698, June 2010.

[14] J. Mairal, F. Bach, J. Ponce, and G. Sapiro, "Online dictionary learning for sparse coding," in *Proceedings of the 26th International Conference On Machine Learning (ICML '09)*, pp. 689–696, June 2009.

[15] M. Yang, L. Zhang, X. Feng, and D. Zhang, "Fisher discrimination dictionary learning for sparse representation," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 543–550, 2011.

[16] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[17] J. Wen, X. Gao, Y. Yuan, D. Tao, and J. Li, "Incremental tensor biased discriminant analysis: a new color-based visual tracking method," *Neurocomputing*, vol. 73, no. 4–6, pp. 827–839, 2010.