Journal of Network Intelligence
Taiwan Ubiquitous Information
 ISSN 2414-8105 (Online)
Volume 3, Number 2, May 2018

L1-L2 Norms Based Target Representation for Visual Tracking

Yuanyun Wang, Jun Wang, Chengzhi Deng*, Huasheng Zhu, Shengqian Wang

[1]Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing, Nanchang Institute of Technology, Nanchang 330099, China
[2]School of Information Engineering, Nanchang Institute of Technology, Nanchang 330099, China
wangyy_abc@163.com, dengchengzhi@126.com
*Corresponding author:wangjun012778@126.com

Received Jan. 2018, Revised Feb. 2018

Abstract. *It is a challenging task to develop a robust appearance model due to complicated appearance variations such as illumination variation, partial occlusion, motion blur and background clutters. Some existing appearance models are often built upon a linear combination of a set of templates with least square methods. With such kind of representation, visual tracking is not robust when significant appearance variations are in presence. In this work, we propose a novel target representation for visual tracking. A target candidate is represented by a linear combination of a set of target templates from previous frames with ℓ_1-norm to characterize the coding residual. In the meantime, ℓ_2-norm is used to regularize the coding coefficient. The proposed target appearance model is robust to partial occlusion. A novel likelihood evaluation function is proposed based on the reconstruction residual and coding coefficient. Experimental results on challenging video sequences in comparison with state-of-the-art algorithms demonstrate the effectiveness and robustness of the proposed tracking algorithm.*
Keywords: Visual tracking; Particle filter; Non-sparse representation; Appearance model.

1. **Introduction.** Visual tracking is a well-known issue in computer vision with a variety of tasks such as vehicle navigation, security and surveillance, human-computer interaction, etc. Despite much progress has been made in recent decades [1], it remains a challenging task due to factors such as partial occlusions, illumination variations, background clutters and out-of-plane rotation. In visual tracking, a target candidate is usually manually selected in the first frame. Thereby, a tracking algorithm is required to associate the tracked target in the rest video frames. The appearance model is primarily important in a tracking algorithm, which represents a target candidate and is used to evaluate the observation likelihood of a target candidate belonging to a target in the current frame. A good appearance model should be robust to significant appearance variations.

A target is usually represented by a set of dictionary templates. Each target candidate is described by a linear combination of pre-defined templates. The observation likelihood is evaluated based on the distance between a target candidate and the corresponding templates. Recently, sparse representation techniques [2] have been applied to visual tracking [3, 4], where a target candidate (i.e., an image patch) is sparsely represented by a set of templates with online update. These tracking algorithms use ℓ_1-norm sparsity constraint on coding coefficients and achieve robustness to appearance variations caused by partial occlusions and outliers. However, the expensive computation is a drawback

due to solving ℓ_1-norm minimization problems. Moreover, in these tracking algorithms, a number of templates are used to learn and represent target candidates. The computational cost grows with the number of templates.

Inspired by collaborative representation based face recognition [5], in this work, we propose a novel appearance model based on a set of target templates by using ℓ_1-norm to characterize the coding residual (i.e., the reconstruction error between a target candidate and the templates) to achieve robustness to occlusions and outliers. In the meantime, by exploiting non-sparse ℓ_2-norm to regularize coding coefficients the computational cost is less expensive than using ℓ_1-norm to regularize coding coefficients. Moreover, we propose a novel likelihood evaluation function based on the reconstruction residual and the estimated template coefficient, which makes tracking algorithm more stable. Finally, we propose a robust tracking algorithm based on the proposed appearance model and the observation likelihood in a particle filter framework. Experimental results against several state-of-the-art tracking algorithms demonstrate that the proposed tracking achieves superior tracking results.

The remainder of this paper is organized as follows. Section 2 summarizes the related work. Section 3 presents the proposed visual tracking algorithm, which includes ℓ_1 and ℓ_2-norms based target representation, a novel observation likelihood and the template update scheme. Section 4 evaluates experimental results of the proposed tracking algorithm in comparison with state-of-the-art algorithms on some challenging video sequences. Section 5 concludes this work.

2. **Related work.** Visual tracking is an important issue in computer vision with numerous applications. Generally speaking, based on the types of appearance models adopted, visual tracking algorithms can be categorized as either generative [6, 7, 8, 9] or discriminative [10, 11, 12, 13, 14, 15]. Here, we briefly review some representative tracking algorithms related to our work.

Generative tracking algorithms typically learn an appearance model to represent a target candidate and localize the target by searching for the image region that has the minimum reconstruction residual to the target model in the current frame. Adam *et al.* [9] divides a target candidate into multiple non-overlapping image patches. The Earth Mover's Distance (EMD) is used to measure the similarity between a patch and the corresponding patch in the target model. The proposed algorithm represents a target candidate by a histogram which takes into account the spatial distribution of the pixel intensities. The tracking algorithm can alleviate the drift problem because of fixed target templates. However, it is not robust to cluttered background and drastic illumination variations. In [16], representative samples are used as target templates by undergoing the principle component analysis. A target candidate is represented by a set of basis vectors. The proposed tracking algorithm is robust to pose and illumination variations. In [17], a local subspace collaborative tracking algorithm is proposed, which uses multiple linear and nonlinear subspaces learned to model target appearances.

Kwon *et al.* [7] learn a set of basic appearance models and basic motion models to cover complicated appearance variations and motion variations, respectively. The tracking algorithm is robust to motion variations. Wang *et al.* [6] use the least soft-thresold squares (LSS) distance to measure the similarity between a target candidate and the target templates, which is robust to partial occlusion and illumination variations. Xiao *et al.*[18] use the ℓ_2-regularized least square to model target appearances, which provides a fast tracking performance. In [19, 20], correlation filters are used to model appearance models. In [21, 22], correlation filters based trackers are proposed and achieve robust performance.

Recently, sparse linear representations have been introduced to target representations. In L1 algorithm [3], a target candidate is sparsely represented by using both target templates and trivial templates. The target templates are used to represent target appearance, and trivial templates are used to describe outliers or occlusions. The L1 algorithm is robust to partial occlusions. However, it is time-consuming in solving ℓ_1 minimization problem, which limits the tracking performance in real time. Jia *et al.* [23] propose a structural local sparse appearance model, where a target candidate is sparsely represented by using the partial information and spatial information via a alignment-pooling method. Taking advantage of generative and discriminative models, Zhong *et al.* [4] propose a sparsity-based collaborative appearance model based on both holistic templates and local representations. Recently, Zhang *et al.* [24] propose structural spare tracking algorithm by exploiting the spatial layout structure among the local patches inside each target candidate. In [25], a target candidate is represented by sparse combinations of particles by exploiting underlying low-rank constraints.

Discriminative tracking algorithms consider visual tracking as a binary classification problem, in which a classifier is learnt to distinguish a target from the around background. Avidan [10] proposes an ensemble tracking algorithm by combining a set of weak classifiers into a strong classifier and computes the confidence value for each pixel. The target is located by a vote confidence map. Bai *et al.*[11] consider the contribution of confidences as a weight vector and combine a set of weak classifiers into a strong classifiers. Babenko *et al.* [15] introduce the multiple instance learning framework into visual tracking where positive and negative bags are considered as training samples. Kalal *et al.* [14] formulate visual tracking in a tracking-learning-detecting framework. In [14], a bootstrapping classifier is learnt and used to select potential samples for updating unlabeled data with positive and negative constraints. Hare *et al.*[12] propose a tracking-by-detecting algorithm based on an online structured output support vector machine (SVM). Ning *et al.* [26] learn linear structured SVM and explicit feature map to track object. In [27, 28, 29], the features based on deep convolutional neural networks are learnt.

3. The proposed visual tracking algorithm.

In this section, we describe ℓ_1-ℓ_2 norms based target representation and a likelihood evaluation based on the reconstruction residual and the coding coefficient. Based on the target representation and the likelihood evaluation, we outline the proposed tracking algorithm in a particle filter framework [30].

3.1. ℓ_1-ℓ_2 norms based target representation.

During tracking, m particles (i.e., target candidates) are sampled at the t-th frame, the state of a particle is denoted as $\mathbf{x}_t^i, i = 1, 2, \cdots, m$. The corresponding observation of $\mathbf{x}_t^i$ is denoted as $\mathbf{y}_t^i$ at frame t. The state of the located target at frame t is denoted as $\hat{\mathbf{x}}_t$, and the corresponding observation is denoted as $\hat{\mathbf{y}}_t$.

In visual tracking, the observation $\mathbf{y}_t^i$ of a target candidate is often represented by a linear combination of target templates

$$\mathbf{y}_t^i \approx \mathbf{d}_1\alpha_1 + \mathbf{d}_2\alpha_2 + \cdots + \mathbf{d}_n\alpha_n, \tag{1}$$

where $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_n]$ is a set of target templates, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]^T \in R^n$ is the corresponding template coefficient vector.

Different from sparse linear representations in [3, 4, 23], in the proposed tracking algorithm, the observation $\mathbf{y}_t^i$ of a target candidate is approximated in the form of non-sparse combinations of a set of target templates by solving

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{y}_t^i - \mathbf{D}\boldsymbol{\alpha}\|_1 + \lambda\|\boldsymbol{\alpha}\|_2^2 \,, \tag{2}$$

where $\mathbf{D}$ is a set of templates, $\|\cdot\|_1$ and $\|\cdot\|_2$ denote the ℓ_1 and ℓ_2 norms, respectively. In [3], $\mathbf{D}$ includes target templates and trivial templates, which are used to represent target candidates and to handle partial occlusions, respectively. In Eqn. (2), $\mathbf{D}$ only includes a set of target templates.

Let $\mathbf{e} = \mathbf{y}_t^i - \mathbf{D}\boldsymbol{\alpha}$. Eqn. (2) can be-written as

$$\hat{\boldsymbol{\alpha}} = \arg\min_{\boldsymbol{\alpha}} \|\mathbf{e}\|_1 + \lambda\|\boldsymbol{\alpha}\|_2^2 \ , s.t.\ \mathbf{y}_t^i = \mathbf{D}\boldsymbol{\alpha} + \mathbf{e}. \tag{3}$$

In Eqn. (3), $\hat{\boldsymbol{\alpha}}$ can be efficiently computed by Augmented Lagrange Multiplier method [5]. The corresponding augmented Lagrangian function is as follows

$$L_\mu(\mathbf{e}, \boldsymbol{\alpha}, \mathbf{z}) = \|\mathbf{e}\|_1 + \lambda\|\boldsymbol{\alpha}\|_2^2 + \langle \mathbf{z}, \mathbf{F}\rangle + \frac{\mu}{2}\|\mathbf{F}\|_2^2, \tag{4}$$

where $\mathbf{F} = \mathbf{y}_t^i - \mathbf{D}\boldsymbol{\alpha} - \mathbf{e}$, μ is a constant that penalises large reconstruction error. $\mathbf{z}$ is a Lagrange multiplier vector. $\mathbf{z}$ and $\boldsymbol{\alpha}$ are iteratively estimated in Eqn. (4).

In the proposed target appearance model, we use ℓ_1-norm measure the coding residual for the robustness to partial occlusions or outliers. ℓ_2-norm regularization on α brings much less computational cost than ℓ_1-norm regularization. In the mean, the ℓ_2-norm is used on the coding coefficient, which prevents any single template from taking a dominant role in representing a target candidate. The proposed appearance model can obtain a more stable target representation.

As shown in the following experiment section, non-sparse ℓ_2-norm regularization on coding coefficient can enhance the discrimination of target representation. The proposed appearance model turns out to robust to illumination variations, partial occlusions, etc.

3.2. **Likelihood evaluation.** The likelihood evaluation is a key issue in visual tracking, which reflects the similarity between a target candidate and the corresponding target models. Based on the estimated $\hat{\boldsymbol{\alpha}}$ in Eqn. (2), the reconstruction residual between a target candidate $\mathbf{x}_t^i$ and the templates $\mathbf{D}$ is measured

$$d(\mathbf{y}_t^i, \mathbf{D}\hat{\boldsymbol{\alpha}}) = (\mathbf{y}_t^i - \mathbf{D}\hat{\boldsymbol{\alpha}})^T(\mathbf{y}_t^i - \mathbf{D}\hat{\boldsymbol{\alpha}}) \tag{5}$$

where $\mathbf{D}$ is dictionary templates, $\hat{\boldsymbol{\alpha}}$ is the coefficient vector estimated by Eqn. (2).

The regularization coefficient reflects the importance of target templates in representing a target candidate. In the proposed appearance model, it is introduced into the observation likelihood. The proposed likelihood evaluation is computed based on the reconstruction residual with the coding coefficient. Based on Eqn. (2), the observation likelihood is computed as

$$p(\mathbf{y}_t^i|\mathbf{x}_t^i) \propto exp\left\{-\eta(d(\mathbf{y}_t^i, \mathbf{D}\hat{\boldsymbol{\alpha}})) - \lambda\|\hat{\alpha}\|_1\right\}, \tag{6}$$

where $d(\mathbf{y}_t^i, \mathbf{D}\hat{\boldsymbol{\alpha}}))$ is the reconstruction residual between a target candidate $\mathbf{x}_t^i$ and target templates $\mathbf{D}$, η is the standard deviation of the Gaussian function.

3.3. **Target location and template update.** The proposed tracking algorithm is proposed in a particle filter framework. A particle (i.e., a target candidate) corresponds to an image patch in terms of motion state variables in each frame image, which is described by a motion state variable including 2D position (x, y) and target scale s.

In a particle filter framework, visual tracking is formulated as an estimation of the posterior distribution $p(\mathbf{x}_t|\mathbf{y}_{1:t})$, where $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_t\}$ are observations from the first frame to frame t. The goal of visual tracking is to estimate the target state $\mathbf{x}_t$ recursively

$$p(\mathbf{x}_t|\mathbf{y}_{1:t}) \propto p(\mathbf{y}_t|\mathbf{x}_t) \int p(\mathbf{x}_t|\mathbf{x}_{t-1}) p(\mathbf{x}_{t-1}|\mathbf{y}_{1:t-1}) d\mathbf{x}_{t-1}, \tag{7}$$

Algorithm 1: Proposed tracking algorithm

Input: the t-th frame, target state $\hat{\mathbf{x}}_{t-1}$ at the $(t-1)$-th frame.

1 Sample m target candidates $\{\mathbf{x}_t^i\}_{i=1}^m$ according to $\hat{\mathbf{x}}_{t-1}$ at the $(t-1)$-th frame and motion model by a Gaussian function with Eqn. (8).

2 Extract the feature and obtain $\mathbf{y}_t^i$ for $\{\mathbf{x}_t^i\}_{i=1}^m$.

3 Compute the observation likelihood $p(\mathbf{y}_t^i|\mathbf{x}_t^i)$ for each particle with Eqns. (2), (6) and target template $\mathbf{D}_{t-1}$.

4 Evaluate target state $\hat{\mathbf{x}}_t$ with Eqn. (9).

5 Update $\mathbf{D}_t$ according to the update scheme in Section 3.3.

Output: Target state $\hat{\mathbf{x}}_t$ at t-th frame.

where $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ denotes the motion model. In this work, $p(\mathbf{x}_t|\mathbf{x}_{t-1})$ is modelled by a Gaussian distribution

$$p(\mathbf{x}_t|\mathbf{x}_{t-1}) = \mathcal{N}(\mathbf{x}_t; \mathbf{x}_{t-1}, \Sigma), \tag{8}$$

where $\Sigma = diag(x, y, s)$ is a diagonal covariance matrix whose elements are the parameters of the target 2D position and target scale.

Maximizing the posterior in Eqn. (9) is equal to maximizing the observation likelihood $p(\mathbf{y}_t|\mathbf{x}_t)$. In visual tracking, the posterior probability $p(\mathbf{x}_t|\mathbf{y}_{1:t})$ in Eqn. (7) is approximated by a set of particles with corresponding importance weights $\{w_t^i\}_{i=1}^m$, where $w_t^i \propto p(\mathbf{y}_t|\mathbf{x}_t^i)$. The optimal state of the target at frame t is defined by maximizing the posterior probability estimation

$$\hat{\mathbf{x}}_t = \arg \max_{\{\mathbf{x}_t^i\}_{i=1}^m} p(\mathbf{y}_t|\mathbf{x}_t^i) p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i). \tag{9}$$

Since target appearances vary significantly during the tracking process, target templates should be updated in order to maintain the effectiveness. Fixed target templates are not sufficient to handle recent appearance variations. Target templates are updated to capture appearance variations. In the first frame, the first target template is manually selected. The remaining target templates are selected by perturbating a few pixels (4 pixels in our experiments) with a radius around the corner of the first target template. The first target template is the most informative information, so it is always remained in target templates. During tracking, the recent tracking result reflects appearance variations. At the current frame, the tracking result is added into target templates. In the mean, the oldest target template is swapped out.

3.4. **The complete tracking algorithm.** We outline the proposed tracking algorithm in Alg. 1 by integrating the ℓ_1-ℓ_2 norms based target representation, the regularized observation likelihood and the template updating in a particle filter framework. At the t-th frame, a number of particles are generated according to the motion model and extract the corresponding image patch feature (i.e., steps 1-2 in Alg. 1). Then, the observation likelihood of each particle is computed according to Eqns. (2) and (6) and target template $\mathbf{D}_{t-1}$ obtained at frame $t-1$ (i.e., step 3 in Alg. 1). With evaluated observation likelihood of each particle, the tracked target is located. The current tracking result is added into the target templates and the oldest target template is replaced.

4. **Experiments.** In this section, we evaluate the proposed tracking algorithm on six challenging video sequences against seven state-of-the-art algorithms. These tracking algorithms include: Struck [12], VTS[8], SCM [4], FCT [13], DLS [26], LSST[6] and PCOM[31]. We use the source codes or binary codes provided by the authors. These competing algorithms demonstrate excellent tracking performances in a recent evaluation

[1] except for DLS, LSST and PCOM which are recently proposed. All the evaluated algorithms are implemented in MATLAB on a PC with Inter(R) Core(TM) i5-2400 3.10GHZ and 8GB memory. The number of particles is set to 300. Histograms of Sparse Codes (HSC) [32] is extracted as feature descriptors. The average processing time of the proposed tracking algorithm is 0.75 s per frame. The size of the target templates is set to 25.

Table 1 summarizes the main attributes of the video sequences. In the *Bolt* sequence, the target moves fast and rotates in-plane and out-of-plane. In the mean, the target is occluded by himself and the other athletes. In the *CouponBook* sequence, the appearance of the target change drastically and the other distracters which has a similar appearance as the tracked target. There are drastic illumination variations in the $Fish$ and Man sequence. In the $Football$ and $Football1$ sequences, the targets are occluded by the other similar objects in these sequences with a complicated background and rotate in plane and out-of-plane.

4.1. **Quantitative Evaluation.** The tracking results are presented using four evaluation measures: average center location error, success rate and average overlap rate and precision[1].

Table 2 show the average center location errors (in pixels) of the eight tracking algorithms on the six sequences. Fig. 1 also shows the precision plots in terms of location error threshold for these evaluated algorithms. From Table 2 and Fig. 1, we can see that the proposed tracking algorithm achieves the best or the second best tracking results on five sequences against the other tracking algorithms. In the meantime, DLS and the proposed tracking algorithm achieve robust tracking performance over all these sequences.

Table 3 shows success rates of the eight tracking algorithms on the six video sequences. We also present the success plots of the tracking algorithms in Fig. 2. From these quantitative evaluations, we can see that the proposed algorithm achieves favorable results in most of the video sequences against the evaluated tracking algorithms. Table 4 show the average overlap rates of the tracking algorithms on the six sequences. As seen from Table 4, the proposed algorithm produces the best or the second best tracking results on five video sequences.

4.2. **Qualitative Evaluation.** Fig. 3 show some tracking results of all the evaluated algorithms. A detailed analysis on the tracking results is discussed based on the main challenging factors in each video sequence.

Illumination variation and **Occlusion**: In the $Fish$ and Man sequences, the targets undergo drastic illumination variations. Struck learns the appearance variations and

TABLE 1. The main attributes of the six video sequences. Target size: the initial target size in the first frame; BC: background clutter; OPR: out-of-plane rotation; IPR: in-plane rotation; IV: illumination variation; Occ: occlusion; Def: deformation.

Sequence	Frames	Image size	Target size	Color	BC	OPR	IPR	IV	Occ	Def
Bolt	350	640×360	26×61	RGB		√	√		√	√
CouponBook	327	320×240	62×98	RGB	√				√	
Fish	476	320×240	60×88	Gray				√		
Football	362	624×352	39×50	Gray	√	√	√		√	
Football1	74	352×288	26×43	RGB	√	√	√			
Man	134	241×193	26×40	RGB				√		

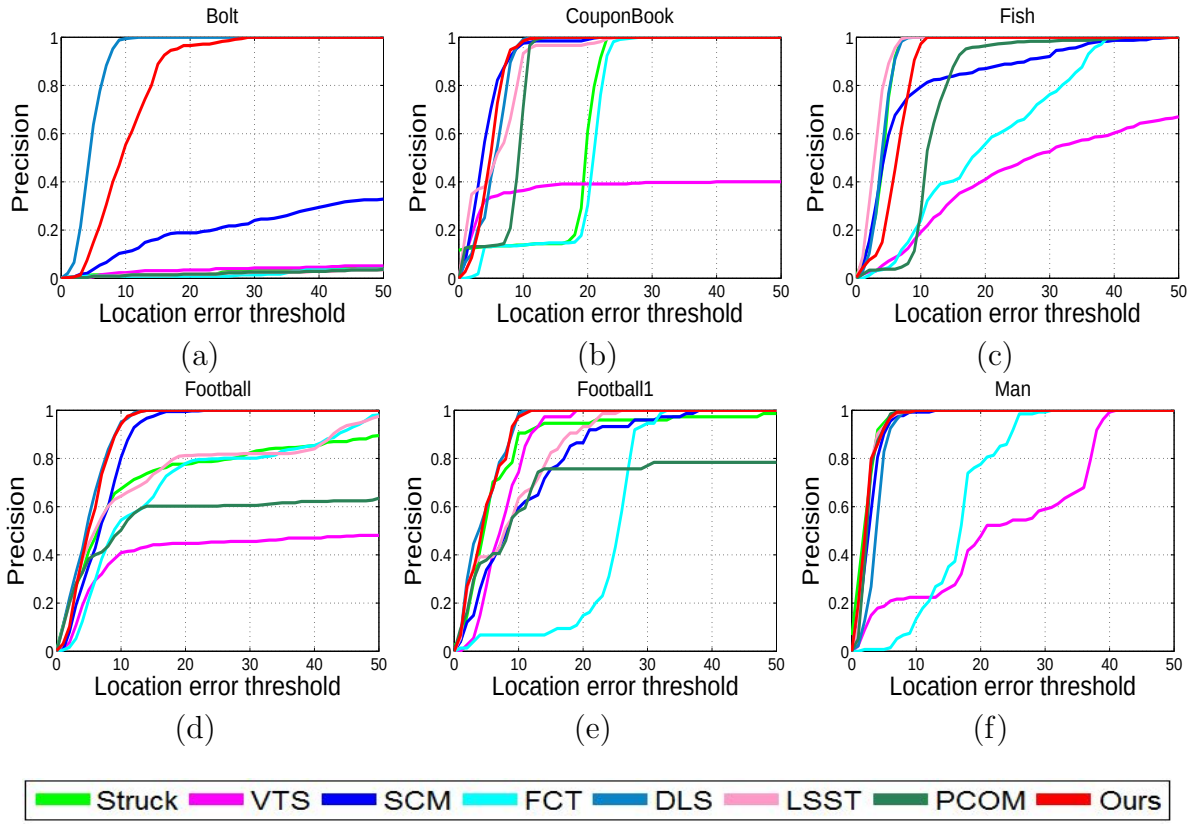

FIGURE 1. Precision plots in terms of location error threshold.

achieves robust tracking results in the *Fish* sequence. PCOM uses the probability continuous outlier model to alleviate the influence of illumination variations. The proposed tracking algorithm performs well in these sequences via the ℓ_1-norm constraint on the coding residual. DLS achieves robust tracking performance via learn structured SVM classifier. In the *Football* sequence, the target is occluded by other distracters which are similar to the target in appearance. FCT, Struck and PCOM drift to track the target when the target undergoes rotation and is occluded. In contrast, DLS and the proposed tracking algorithm achieve more accurate tracking results.

Background clutters and **deformation**: The target is partially occluded and influenced by the other coupon book In the *CouponBook* sequence. Struck, SCM, DLS, PCOM and the proposed tracking algorithm track the target accurately throughout the video sequences. In the *Football* sequence, the targets are in a cluttered background.

TABLE 2. Average center location errors (in pixels). The best two results are shown in red and blue colors, respectively.

Sequence	Struck	VTS	SCM	FCT	DLS	LSST	PCOM	Ours
Bolt	387.8	369.8	203.2	267.9	**4.5**	376.4	363.3	**9.9**
CouponBook	15.0	65.1	6.0	18.6	**5.4**	8.0	8.3	**4.9**
Fish	**3.9**	43.6	8.3	19.6	**3.9**	**2.9**	11.8	6.1
Football	15.3	115.3	6.9	15.8	**4.8**	13.2	54.2	**4.7**
Football1	7.0	7.5	10.4	23.7	**4.4**	8.6	23.4	**3.9**
man	**2.3**	22.7	2.9	16.5	3.8	**2.4**	2.5	**2.4**
Average	71.9	104.0	39.6	60.3	**4.5**	68.6	77.2	**5.3**

TABLE 3. Success rate (in percentage). The best two results are shown in red and blue colors, respectively.

Sequence	Struck	VTS	SCM	FCT	DLS	LSST	PCOM	Ours
Bolt	1.4	2.9	14.3	0.9	**99.7**	0.9	0.9	**80.3**
CouponBook	**100**	39.4	**100**	98.5	**100**	97.0	**100**	**100**
Fish	**100**	35.9	86.6	54.0	**100**	**100**	96.4	**100**
Football	69.3	41.4	88.7	55.3	**96.4**	62.7	53.9	**97.5**
Football1	**89.2**	58.1	39.2	6.8	**89.2**	51.4	44.6	**98.7**
man	99.3	22.4	98.5	13.4	**100**	**100**	**100**	**100**
Average	76.5	33.4	71.2	38.1	**97.6**	68.7	66.0	**96.1**

TABLE 4. Average overlap rate (in percentage). The best two results are shown in red and blue colors, respectively.

Sequence	Struck	VTS	SCM	FCT	DLS	LSST	PCOM	Ours
Bolt	1.7	2.3	12.9	1.4	**75.9**	1.0	1.0	**61.8**
CouponBook	70.2	35.5	**82.3**	64.8	83.1	80.2	80.5	**86.0**
Fish	**84.3**	34.4	74.0	54.3	**83.3**	80.9	65.4	78.6
Football	55.7	30.8	60.3	47.5	**71.4**	53.0	42.4	**70.1**
Football1	66.0	53.2	45.4	16.9	**70.3**	53.9	48.2	**73.3**
man	81.9	27.4	71.9	26.4	67.5	70.0	**82.3**	**82.2**
Average	60.0	30.6	57.8	35.2	**75.2**	56.5	53.3	**75.3**

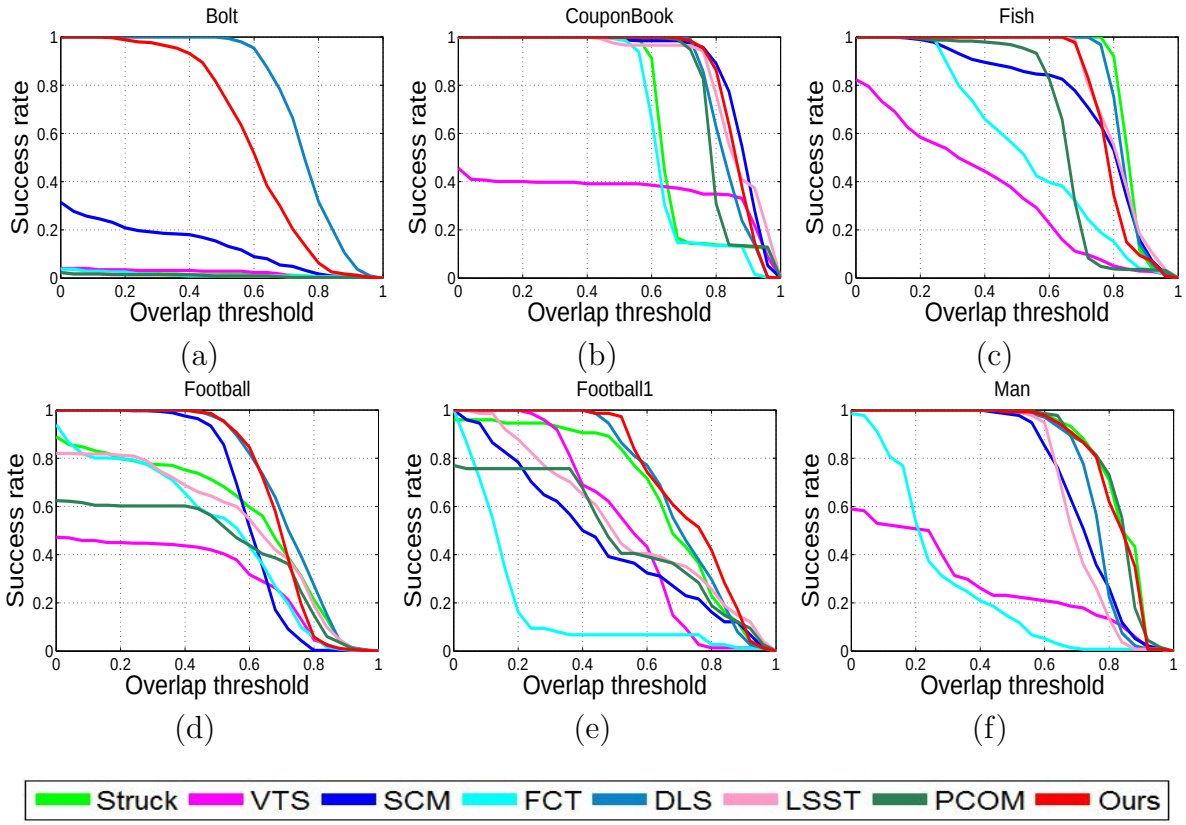

FIGURE 2. Success plots in terms of overlap threshold.

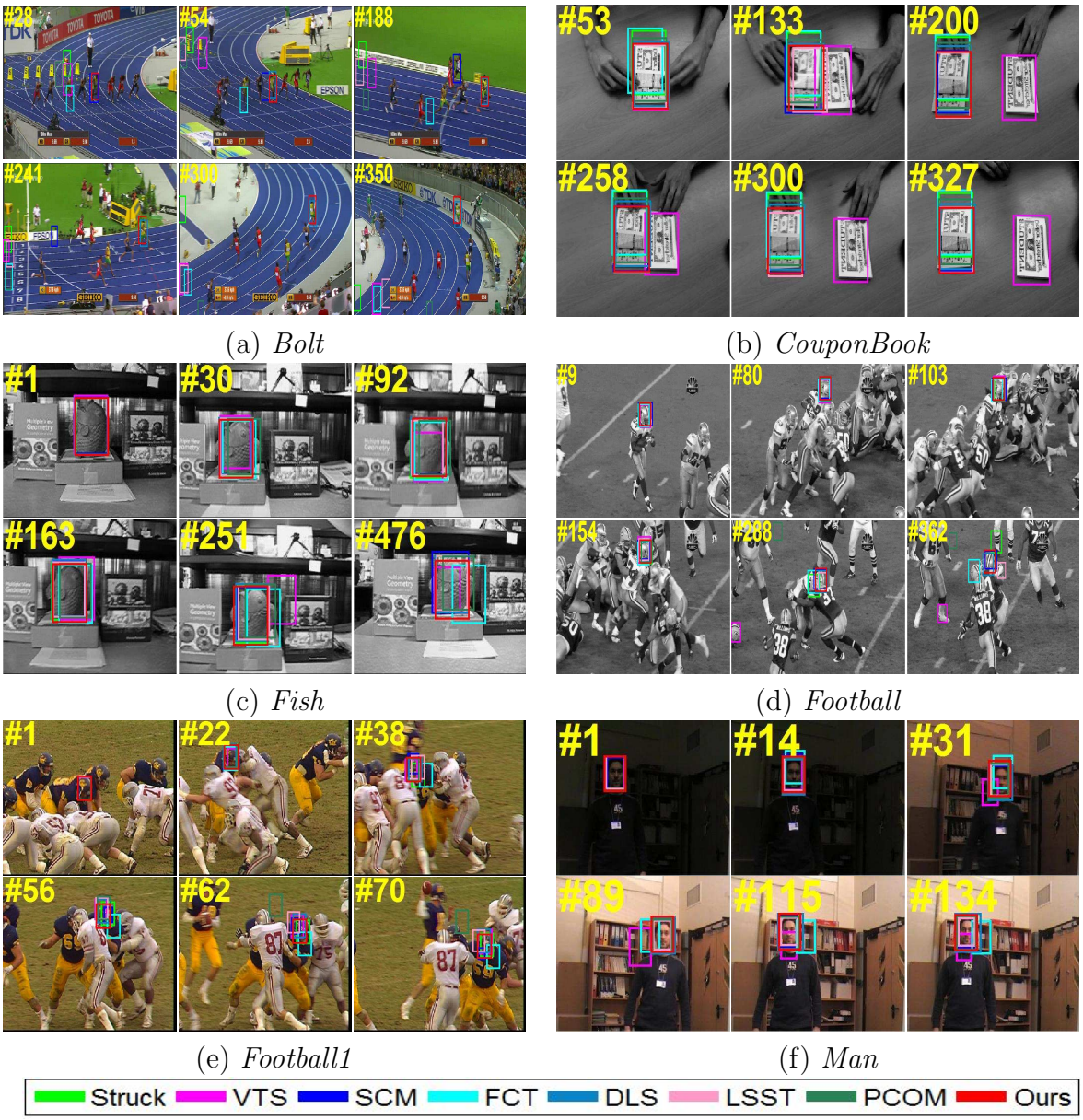

FIGURE 3. The tracking results obtained on the 6 video sequences.

DLS and the proposed algorithm achieve good tracking performance in dealing with appearance variations by background clutters. In the *Football*1 sequence, FCT and PCOM fail to track the target after the 43rd frame due to the influence of cluttered background and rotations. The proposed algorithm can track the target successfully.

In-plane and **Out-of-plane rotations**: The *Bolt*, *Football* and *Football*1 sequences are influenced by both in-plane and out-of-plane rotations. In the *Bolt* sequence, the proposed tracking algorithm and DLS achieve favorable tracking performances in the whole sequence, while all the other tracking algorithm only tracking the target in the first 50 frames. DLS and the proposed tracking algorithm can accurately track the target in the *Football* and *Football*1 sequences.

5. **Conclusion.** We have presented a simple but effective tracking algorithm, in which a target is represented by a linear combinations of a set of templates from previous frames with ℓ_1-norm to characterize the representation residual. In the meantime, ℓ_2-norm is

used to regularize the coding coefficient. As shown, this target representation is robust to partial occlusions. Moreover, a novel likelihood evaluation method is proposed to obtain a stable likelihood evaluating. Experimental results demonstrate the favorable performance in comparison with some state-of-the-art algorithms. The proposed algorithm turns out to be robust to partial occlusions, drastic illumination variations and rotations.

Acknowledgment. This work is supported by the National Natural Science Foundation of China (No:61661033 and 61461032).

REFERENCES

[1] Y. Wu, J. Lim, and M. Yang, Online Object Tracking: A Benchmark, *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2411-2418.

[2] J. Wright, A. Yang, A. Ganesh, S. Sastry, and Y. Ma, Robust face recognition via sparse representation, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31 (2) (2009), pp. 210-227.

[3] X. Mei, H. Ling, Robust visual tracking and vehicle classification via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (11) (2011), pp. 2259-2272.

[4] W. Zhong, H. Lu, and M. Yang, Robust object tracking via sparsity-based collaborative model, *IEEE Conference on Computer Vision and Pattern Recognition,* 2012, pp. 1838-1845.

[5] L. Zhang, M. Yang, X. Feng, Y. Ma, and D. Zhang, Collaborative Representation based Classification for Face Recognition, Technical report. arXiv: 1204.2358.

[6] D. Wang, H. Lu, and M. Yang, Least Soft-thresold Squares Tracking, *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 2371-2378

[7] J. Kwon, and K. Lee, Visual tracking decomposition, *IEEE Conference on Computer Vision and Pattern Recognition*, 2010, pp. 1269-1276.

[8] J. Kwon, and K. Lee, Tracking by sampling trackers, *IEEE International Conference on Computer Vision*, 2011, pp. 1195-1202.

[9] A. Adam, E. Rivlin, and I. Shimshoni, Robust fragments-based tracking using the integral histogram, IEEE Conference on Computer Vision and Pattern Recognition, 2006, pp. 798-805.

[10] S. Avidan, Ensemble tracking, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 29(2) (2007), pp.261-271.

[11] Q. Bai, Z. Wu, S. Sclaroff, M. Betke, and C. Monnier, Randomized Ensemble Tracking, *IEEE International Conference on Computer Vision*, 2013, pp. 2040-2047.

[12] S. Hare, A. Saffari, P. H. Torr, Struck: Structured output tracking with kernels, *IEEE International Conference on Computer Vision*, 2011, pp. 263-270.

[13] K.Zhang, L.Zhang, and M.Yang, Fast Compressive Tracking, *IEEE Trans. Pattern Anal. Mach. Intell.* 36(10)(2014), pp.2002-2015.

[14] Z. Kalal, K. Mikolajczyk, and J. Matas, Tracking-learning-detection, *IEEE Trans. Pattern Anal. Mach. Intell.* 34 (7) (2012), pp. 1409-1422.

[15] B. Babenko, M.-H. Yang, S. Belongie, Robust object tracking with online multiple instance learning, *IEEE Trans. Pattern Anal. Mach. Intell.* 33 (8) (2011), pp. 1619-1632.

[16] D. Ross, J. Lim, R. Lin, and M. Yang, Incremental learning for robust visual tracking, *International Journal of Computer Vision,* 77 (1) (2008), pp. 125-141.

[17] C. Ma, X. Yang, C. Zhang, et al., Long-term correlation tracking, *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 5388-5396.

[18] Z. Xiao, H. Lu, D. Wang, L2-RLS Based Object Tracking, *IEEE Trans. Circuits Syst. Video Technol.,* 2014, 24(8)(2014), pp. 1301-1308.

[19] T. Liu, G. Wang, Q. Yang, Real-time part-based visual tracking via adaptive correlation filters, *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 4902-4912.

[20] M. Danelljan, G. Hager, F. Khan, et al., Learning Spatially Regularized Correlation Filters for Visual Tracking, *IEEE International Conference on Computer Vision*, 2015, pp, 4310-4318.

[21] M. Mueller, N. Smith, B. Ghanem, Context-Aware Correlation Filter Tracking, *IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 1387-1395.

[22] T. Zhang, C. Xu, M. Yang, Multi-task Correlation Particle Filter for Robust Visual Tracking, *IEEE Conference on Computer Vision and Pattern Recognition,* 2017, pp. 4819-4827.

[23] X. Jia, H. Lu, and M. Yang, Visual tracking via adaptive structural local sparse appearance model, *IEEE Conference on Computer Vision and Pattern Recognition*, 2012,1822-1829.

[24] T. Zhang, S. Liu, C. Xu, S. Yan and B. Ghanem, Structure sparse tracking, *IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 150-158.
[25] T. Zhang, S. Liu, N. Ahuja, M. H. Yang, B. Ghanem, Robust visual tracking via consistent low-rank sparse learning, *Int. J. Comput. Vision*, 111(2)(2015), 171-190.
[26] J. Ning, J. Yang, S. Jiang, L. Zhang and M-H Yang, Visual Tracking via Dual Linear Structured SVM and Explicit Feature Map, *IEEE Conference on Computer Vision and Pattern Recognition,* 2016, pp. 4266-4274.
[27] C. Ma, J.B Huang, X.K Yang and M-H. Yang, Hierarchical Convolutional Features for Visual Tracking, *IEEE International Conference on Computer Vision*, 2015, pp. 3074-3082.
[28] C. Huang, S. Lucey, D. Ramanan, Learning Policies for Adaptive Tracking with Deep Feature Cascades, IEEE International Conference on Computer Vision, 2017, pp. 1-8.
[29] Y. Qi, S. Zhang, L. Qin, et al., Hedged Deep Tracking, *IEEE Conference on Computer Vision and Pattern Recognition,* 2016, pp. 4303-4311.
[30] A. Doucet, D. Freitas, and N. Gordon. Sequential Monte Carlo Methods In Practice. *Springer, New York,* 2001.
[31] D. Wang, and H. Lu, Visual Tracking via probability continuous outlier model, *IEEE Conference on Computer Vision and Pattern Recognition,* 2014, pp. 3478-3485.
[32] X. Ren, and D. Ramanan, Histograms of sparse codes for object detection, *IEEE Conference on Computer Vision and Pattern Recognition*, 2013, pp. 3246-3253.