Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **2**, March **2018**

ℓ_1-Regularized Hull Representation for Visual Tracking

Jun Wang[1,2], Yuanyun Wang[1,2], Ke Wang[3], and Chengzhi Deng*[1,2]

[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

[3]School of Electrical and Automation Engineering,
East China JiaoTong University, Nanchang 330013, China

wangjun012778@126.com.
*Corresponding author: dengchengzhi@126.com

Received March, 2017; revised December, 2017

Abstract. *Due to various factors such as partial occlusions, fast motion and illumination variations, developing an effective and efficient appearance model is a challenging task. In this paper, we propose a simple and effective tracking algorithm with an appearance model based on ℓ_1-regularized hull representation with target templates. ℓ_1-regularized affine combinations can cover target appearances which do not appear in target templates. ℓ_1 constraint enables the tracking algorithm to robustly deal with partial occlusions and outliers. A novel likelihood function is introduced, which is derived from the reconstruction residual between a target candidate and the target templates and target template coefficients. Experimental results on several challenging video sequences against state-of-the-art tracking algorithms demonstrate that the proposed tracking algorithm is robust to partial occlusions, illumination variations, background clutters, etc.*

Keywords: Visual tracking, Regularized hull, Sparse representation, Particle filter

1. **Introduction.** Visual tracking is a fundamental research problem in computer vision with a variety of applications such as vehicle navigation, human-computer interaction, video surveillance, etc. The goal of visual tracking is to locate a tracked target across a video sequence. In recent decades, much progress has been made in visual tracking [1]. However, it is a challenging tasks to design a robust target appearance model due to the influence of factors such as illumination variation, fast motion, partial occlusions, background clutters and out-of-plane rotation.

Generally speaking, tracking algorithms can be classified as either generative [5, 6, 7, 8, 9, 10, 12, 13], or discriminative [14, 17, 18, 20] based on types of observation. Generative tracking algorithms typically consider tracking problem as searching for an image region that has the minimal reconstruct residual to the tracked target in the current frame. In [2], a target candidate is divided into multiple non-overlapping image patches, which are represented by intensity histograms. The tracking algorithm in [2] can alleviate the drift problem because the fixed target template are used, however it is not robust to dynamic scene variations. Kwon *et al.*[3] use multiple basic appearance models to adapt significant appearance variations, and use multiple basic motion models to cover motion variations. The algorithm [3] is robust to complicated appearance variations. He *et al.*[4] represent a target by a locality sensitive histogram, which is robust to drastic illumination variations. Wang *et al.*[11] propose a Least Soft-thresold Squares (LSS) regression to represent a target candidate and compute the reconstruction error by the LSS distance.

In a discriminative tracking algorithm, visual tracking is formulated as a binary classification problem and learn a classifier to distinguish the tracked target from its surrounding background. In [16], Zhang *et al.* construct a very sparse measurement matrix to extract feature for target appearance. A naive Bayes classifier is learnt to distinguish a target from surrounding background. Babenko *et al.* [15] propose a discriminative tracking algorithm by introducing multiple instance learning to update the classifiers. In [19], a tracking algorithm based on detection is proposed to track a target in a long-term. In [21], Wang *et al.* introduce a sequential training method for CNN into visual tracking.

Recently, sparse representations are successfully used in visual tracking [22, 23, 26]. Sparse based target representations are robust to partial occlusions and outliers. The L1 tracking algorithm [22] combines target templates and trivial templates to represent target candidates. In [24], local patches of a target candidate are sparsely represented by the corresponding patches in the dictionary templates. Based on both holistic templates and local representations, Zhong *et al.*[25] propose a sparsity-based collaborative appearance model. Zhang *et al.*[27] use a sparse and discriminative hashing method for visual tracking, and introduce sparsity into the hash coefficient vectors for select the discriminative feature. In [28], by exploiting temporal consistency, a robust appearance model with low-rank constraints is proposed. In [29], a sparse tracker is proposed based on circulant target templates, which sample particles by using circular shifts of target templates.

Recently, convex and affine hull models based on image sets are proposed [30] and used to face recognition and image set classification. In [13], ℓ_2-regularized affine hull representation is proposed.

Motivated by the above-mentioned work, we propose an ℓ_1-regularized hull representation based tracking algorithm. A target candidate is approximated by an ℓ_1-regularized affine combinations upon target templates. The proposed appearance representation has the advantages of both of affine hulls (i.e., covering the unknown target appearances that do not appear in target templates) and sparse representation (i.e., it is robust to partial occlusions and outliers). We also present an effective function to evaluate observation likelihoods of a target candidate belonging to the tracked target. Preliminary results of this work are presented in [31]. Numerous experiments on challenging video sequences against state-of-the-art tracking algorithms demonstrate the effectiveness and robustness of the proposed appearance model and the tracking algorithm.

2. **Particle filter for visual tracking.** In a particle filter framework, the target state and the corresponding observation are denoted as $\mathbf{s}_t$ and $\mathbf{y}_t$ at frame t, respectively. The tracking problem is formulated as an estimation of the posterior distribution $p(\mathbf{s}_t|\mathbf{y}_{1:t})$, where $\mathbf{y}_{1:t} = \{\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_t\}$ are observations from previous t frames. In the first frame, a set of target candidates $\mathbf{X}_1 = \{\mathbf{x}_1^1, \mathbf{x}_1^2, \cdots, \mathbf{x}_1^m\}$ are sampled by extracting image regions surrounding $\mathbf{s}_1$ with importance weights $w_1^i = \frac{1}{m}$. These particles are propagated according to the motion model $p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)$. Based on particles $\mathbf{X}_t = \{\mathbf{x}_t^1, \mathbf{x}_t^2, \cdots, \mathbf{x}_t^m\}$ with weight w_t^i, $\mathbf{s}_t$ is estimated as

$$\hat{\mathbf{s}}_t = \sum_{i=1}^{m} w_t^i \mathbf{x}_t^i. \tag{1}$$

In the tracking process, the state $\mathbf{x}_t^i$ is often assumed to relate to $\mathbf{x}_{t-1}^i$. As a result, w_t^i is updated as

$$w_t^i = w_{t-1}^i p(\mathbf{y}_t^i|\mathbf{x}_t^i), \tag{2}$$

as shown in Eqn. (8), $p(\mathbf{y}_t^i|\mathbf{x}_t^i)$ is the observation likelihood of particle $\mathbf{x}_t^i$.

3. **Proposed tracking algorithm.**

In this section, we propose a novel tracking algorithm where a target candidate is represented by ℓ_1-regularized affine combinations of a set of target templates. The observation likelihood of a target candidate is derived from both the reconstruction error and the ℓ_1-regularized template coefficients.

3.1. **Target Representations.**

In sparse representation based visual tracking algorithms, the sparse representation of a target candidate $\mathbf{y}$ is represented by sparse combinations of target templates and trivial templates as [22, 24, 25]:

$$\min \|\mathbf{y} - \mathbf{B}\mathbf{c}\|_2^2 + \lambda\|\mathbf{c}\|_1, \tag{3}$$

where $\mathbf{B}$ is templates, $\|\cdot\|_2$ and $\|\cdot\|_1$ denote the ℓ_2 and ℓ_1 norms respectively. In [22], $\mathbf{B}$ includes target templates $\mathbf{D}$ and trivial templates $\mathbf{I}$, which are used to represent the target candidates and to represent partial occlusions, respectively.

The sparse appearance model is robust to partial occlusion and outliers. However, When significant appearance variations appear, the target representation based on sparse constrain is sensitive to illumination variations, severe occlusions and background clutters.

Inspired by hull representation based face recognition and sparse representation techniques, in the proposed target representation, a target candidate $\mathbf{y}$ is represented by ℓ_1-regularized affine combinations of target templates in a template dictionary $\mathbf{D}$ as:

$$\min_{\boldsymbol{\alpha}} \|\mathbf{y}_t^i - \mathbf{D}\boldsymbol{\alpha}\|_2^2, \ s.t. \ \|\boldsymbol{\alpha}\|_1 \leq \delta, \ \sum_{j=1}^{n} \alpha_j = 1, \tag{4}$$

where $\sum(\cdot)$ is the affine constraint [30], $\mathbf{D} = [\mathbf{d}_1, \mathbf{d}_2, \cdots, \mathbf{d}_n]$ is a set of target templates, $\boldsymbol{\alpha} = [\alpha_1, \alpha_2, \cdots, \alpha_n]^T \in R^n$ is the template coefficient vector which is to be estimated. The affine combinations of target templates can cover unknown target appearances that do not appear in the template set. With the estimated $\hat{\boldsymbol{\alpha}}$, the target candidate $\mathbf{y}$ is approximately represented as $\mathbf{D}\hat{\boldsymbol{\alpha}}$. Using the Lagrangian function, Eqn. (4) can be rewritten as follow

$$F(\boldsymbol{\alpha}, \lambda_2) = \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_1 + \lambda_2(\mathbf{e}\boldsymbol{\alpha} - 1) + \lambda_3\|\mathbf{e}\boldsymbol{\alpha} - 1\|_2^2, \tag{5}$$

where λ_1 is a positive parameter to balance the reconstruction error and the regularizer, λ_2 is the Lagrange multiplier, and λ_3 is a penalty parameter. $\mathbf{e}$ is a row vector whose elements are 1. Then, $\boldsymbol{\alpha}$ and λ_2 are optimized alternatively. The iteration processing of minimizing $\boldsymbol{\alpha}$ goes as

$$\begin{aligned} \boldsymbol{\alpha}^{(t+1)} &= \arg\min_{\boldsymbol{\alpha}} F(\boldsymbol{\alpha}, \lambda_2^{(t)}) \\ &= \arg\min_{\boldsymbol{\alpha}} \|\mathbf{y} - \mathbf{D}\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_1 + \frac{\lambda_3}{2}\|\mathbf{e}\boldsymbol{\alpha} - 1 + \frac{\lambda_2^{(t)}}{\lambda_3}\|_2^2, \\ &= \arg\min_{\boldsymbol{\alpha}} \|\tilde{\mathbf{Y}} - \tilde{\mathbf{D}}\boldsymbol{\alpha}\|_2^2 + \lambda_1\|\boldsymbol{\alpha}\|_1, \end{aligned} \tag{6}$$

where $\tilde{\mathbf{Y}} = [\mathbf{y}; (\frac{\lambda_3}{2})^{\frac{1}{2}}(1 - \frac{\lambda_2^{(t)}}{\lambda_3})]$, $\tilde{\mathbf{D}} = [\mathbf{D}; (\frac{\lambda_3}{2})^{\frac{1}{2}}\mathbf{e}]$. The solution of $\boldsymbol{\alpha}^{(t+1)}$ can be obtained by solving ℓ_1-minimization optimization problem such as LASSO.

When $\boldsymbol{\alpha}^{(t+1)}$ is obtained, $\lambda_2^{(t+1)}$ is updated as follows

$$\lambda_2^{(t+1)} = \lambda_2^t + \lambda_3(\mathbf{e}\boldsymbol{\alpha}^{(t+1)} - 1). \tag{7}$$

In the proposed target appearance, a target candidate is represented by an affine hull of target templates with sparse constraint. The representation $\mathbf{D}\hat{\boldsymbol{\alpha}}$ of $\mathbf{y}$ can be considered as a point in the affine subspace upon target templates, which is the closest point to $\mathbf{y}$.

As shown in experimental results, the target representation based on ℓ_1-regularized hull is robust to illumination variations and background clutters.

3.2. **Likelihood Evaluation.** The observation likelihood of a target candidate reflects the probability that a target candidate belongs to the tracked target. Different from existing visual tracking algorithms, in the proposed algorithm, the likelihood evaluation of a candidate $\mathbf{y}$ is derived from both the reconstruction error and the template coefficients as

$$p(\mathbf{y}|\mathbf{x}) \propto exp\left\{-\gamma(d(\mathbf{y}, \mathbf{D}\hat{\boldsymbol{\alpha}})) + \eta\|\hat{\alpha}\|_1\right\}, \tag{8}$$

where $d(\mathbf{y}, \mathbf{D}\hat{\boldsymbol{\alpha}})$ is the reconstruction residual between the target candidate $\mathbf{y}$ and the target templates $\mathbf{D}$, γ is the standard deviation of the Gaussian. By introducing the template coefficients $\hat{\alpha}$ into the observation function, the more stable observation likelihood is obtained. The term $\|\hat{\alpha}\|_1$ in Eqn. (8) introduces the discriminative information for the evaluation function.

The reconstruction error between a target candidate $\mathbf{y}$ and the corresponding templates $\mathbf{D}$ is computed as

$$d(\mathbf{y}, \mathbf{D}\hat{\boldsymbol{\alpha}}) = (\mathbf{y} - \mathbf{D}\hat{\boldsymbol{\alpha}})^T(\mathbf{y} - \mathbf{D}\hat{\boldsymbol{\alpha}}), \tag{9}$$

where $\mathbf{D}$ is target templates, $\hat{\boldsymbol{\alpha}}$ is the coefficient vector estimated by Eqn. (4).

3.3. **Template Update.** In order to adapt to target appearance variations, a template update scheme is necessary. In our work, in the first frame, the tracked target is selected as a target template. The other target templates are initialized by perturbing a few pixels within a radius (3 pixels in our algorithm) around the target center location. The template that has the least template coefficient in estimated vector $\hat{\boldsymbol{\alpha}}$ in Eqn. (4) is swapped out, in the meantime, the current tracking result is added to the template set as a new target template.

Based on the proposed target representation, the likelihood evaluation and the template update, we present the proposed tracking algorithm which is outlined in Algorithm 1. For all the video sequences, we manually select the initial target locations and initialize a set of particles in a particle filter framework.

Algorithm 1: Proposed tracking algorithm

```
Select a set of image patches as targe templates $\mathbf{D}_1 = [\mathbf{d}_1, \cdots, \mathbf{d}_n]$ according to
  state $\mathbf{s}_1$ in the first frame $F_1$, sample $m$ particles $\{\mathbf{x}_1^i\}_{i=1}^m$ with equal weights.
Input: t-th video frame.
Resample $m$ particles $\{\mathbf{x}_t^i\}_{i=1}^m$ according to $p(\mathbf{x}_t^i|\mathbf{x}_{t-1}^i)$.
Crop the corresponding image patches $\{\mathbf{y}_t^i\}_{i=1}^m$ according to $\{\mathbf{x}_t^i\}_{i=1}^m$.
for $i = 1$ to $m$ do
    Evaluate the observation likelihood $p(\mathbf{y}_t^i|\mathbf{x}_t^i)$ using Eqn. (8).
    Update particle weight $w_t^i$ via Eqn. (2).
end
Obtain state $\hat{\mathbf{s}}_t$ with Eqn. (1).
Extract $\mathbf{y}_t$ according to $\hat{\mathbf{s}}_t$, and estimate template coefficient $\hat{\boldsymbol{\alpha}}$ via Eqn. (4).
Update $\mathbf{D}_t$ according to the update scheme in Section 3.3.
Return $\hat{\mathbf{s}}_t$.
```

Table 1. The main attributes of the twelve 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
Basketball	725	576×432	34×81	RGB	√	√		√	√	√
Bolt	350	640×360	26×61	RGB		√	√		√	√
CarDark	393	320×240	29×23	RGB	√			√		
CouponBook	327	320×240	62×98	RGB	√				√	
David2	537	320×240	27×34	Gray		√	√			
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				√		
MountainBike	228	640×360	67×56	RGB	√	√	√			
Singer2	366	624×352	67×122	RGB	√	√	√	√		√
Sylvester	1345	320×240	51×61	Gray		√	√	√		

4. **Experiments.** The proposed tracking algorithm is evaluated with 9 state-of-the-art tracking algorithms. These state-of-the-art tracking algorithms include: Struck[17], VTS[6], Frag[2], SCM[25], FCT[16], OAB[14], L1[22], LSST[11] and PCOM[12]. For fairness, we use the source codes or the binary codes provided by the authors, and initialize all the evaluated algorithms with default parameters in our experiments.

In our experiments, 12 challenging video sequences from a recent benchmark [1] are used to evaluate the tracking performance of the ten tracking algorithms. Table 1 summarized the main challenging aspects of the video sequences.

The proposed tracking algorithm is implemented in MATLAB. All the evaluated tracking algorithms are tested on a PC with Intel(R) Core(TM) i5-2400 3.10GHZ and 8GB memory. The number of particles is set to 400. The histograms of sparse coding (HSC) [32] is used as the feature descriptor. The value of γ in Eqn. (8) is set to 20. The values of λ_1, λ_2 and λ_3 are initialized as 0.001, 0.05 and 0.1, respectively. In our experiments, the size of the dictionary templates is 25 for maintaining the effectiveness and diversity of the dictionary templates. The average processing time of the proposed tracking algorithm is 0.72 s per frame.

4.1. **Quantitative evaluation.** Table 2 presents the average center location errors (in pixels) for all the tracking algorithms on the 12 video sequences. Fig. 1 shows the precision plots for the evaluated algorithms on the 12 video sequences. From Table 2 and Fig. 1, it can be seen that the proposed tracking algorithm achieves the best tracking results in 7 out of the 12 video sequences. Struck achieves the best tracking results on the *CarDark*, *David*2 and *man* sequences. LSST performs well on the *Fish*, *CarDark* and *David*2 sequences. Additionally, the proposed tracking algorithm obtains the smallest average center location error over all the 12 video sequences.

Table 3 presents the success rates for the evaluated algorithms on the 12 sequences. Fig. 2 shows the success rate plots for all the tracking algorithms. As seen from Table 3 and Fig. 2, the proposed tracking performs well against the state-of-the-art algorithms. It achieves the best tracking results in 9 out of the 12 video sequences. The proposed tracking algorithm achieves the highest average success rate over the 12 sequences. Besides, Struck obtains the best or the second best results in 5 video sequences, and achieves the second highest average success rate over the 12 sequences.

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

Sequence	Struck	VTS	Frag	SCM	FCT	OAB	L1	LSST	PCOM	Ours
Basketball	176.7	**7.0**	29.0	13.3	89.9	136.1	77.8	19.9	22.8	**11.5**
Bolt	387.8	369.8	150.5	203.2	267.9	**31.2**	118.4	376.4	363.3	**6.3**
CarDark	**1.0**	17.7	30.7	10.0	46.5	2.8	32.7	**1.6**	2.7	4.6
CouponBook	15.0	65.1	56.2	**6.0**	18.6	24.9	66.3	8.0	8.3	**4.9**
David2	**1.6**	55.1	4.5	5.1	14.6	28.8	56.4	**1.8**	**1.8**	4.3
Fish	**3.9**	43.6	24.7	8.3	19.6	39.8	36.4	**2.9**	11.8	7.2
Football	15.3	115.3	14.6	**6.9**	15.8	19.4	68.4	13.2	54.2	**4.4**
Football1	**7.0**	7.5	11.9	10.4	23.7	36.7	59.3	8.6	23.4	**5.4**
Man	**2.3**	22.7	44.6	2.9	16.5	2.8	2.6	**2.4**	2.5	2.5
MountainBike	12.9	17.0	34.0	10.1	155.0	12.6	141.8	131.2	**8.0**	**7.2**
Singer2	174.7	101.9	35.9	172.2	22.9	170.5	145.8	**14.2**	188.7	**12.1**
Sylvester	11.7	22.0	22.7	7.9	**7.7**	11.9	31.0	67.5	62.3	**4.8**
Average	67.5	70.4	38.3	**38.0**	58.2	43.1	69.7	54.0	62.5	**6.3**

Table 3. Success rates (%). The best two results are shown in red color and blue color, respectively.

Sequence	Struck	VTS	Frag	SCM	FCT	OAB	L1	LSST	PCOM	Ours
Basketball	11.5	**86.1**	51.7	59.0	23.3	1.2	24.0	11.3	23.5	**72.0**
Bolt	1.4	2.9	3.7	**14.3**	0.9	2.3	1.4	0.9	0.9	**92.9**
CarDark	**100**	68.5	45.8	61.8	12.2	89.8	64.9	**100**	**100**	98.5
CouponBook	**100**	39.4	40.9	**100**	98.5	57.6	39.4	97.0	**100**	**100**
David2	**100**	36.1	89.8	80.1	48.6	36.1	27.9	**100**	**100**	**100**
Fish	**100**	35.9	47.3	86.6	54.0	23.5	20.2	**100**	96.4	**100**
Football	69.3	41.4	72.9	**88.7**	55.3	68.5	16.3	62.7	53.9	**93.1**
Football1	**89.2**	58.1	43.2	39.2	6.8	44.6	12.2	51.4	44.6	**96.0**
Man	99.3	22.4	21.0	98.5	13.4	99.3	98.5	**100**	**100**	99.3
MountainBike	81.6	86.0	70.6	95.2	39.5	81.1	25.9	44.3	**99.6**	**100**
Singer2	3.6	36.1	45.9	3.0	71.0	3.6	4.1	**74.9**	3.6	**97.8**
Sylvester	80.3	69.7	50.3	86.6	**91.0**	71.2	55.5	30.3	45.2	**100**
Average	**69.7**	48.5	48.6	67.8	42.9	48.2	32.5	64.4	64.0	**95.8**

Table 4 presents the average overlap rates on the 12 video sequences. As seen from Table 4, the proposed tracking achieves the best tracking results on 8 video sequences and the highest average overlap rate over all the sequences. Struck obtains favorable tracking results among all the other algorithms.

4.2. **Qualitative evaluation.** Next, a detailed analysis on the 12 video sequences for all the evaluated tracking algorithms are given. Fig. 3 represents some tracking results.

Background Clutters: As shown in Fig. 3, VTD is able to track the target in the *Football*1 sequence due to the use of multiple basic appearance models. Because using of the partial and spatial information, ASLA achieves favorable tracking results in the *CarDark* sequence. In the *CouponBook* sequence, when the drastic appearance variation

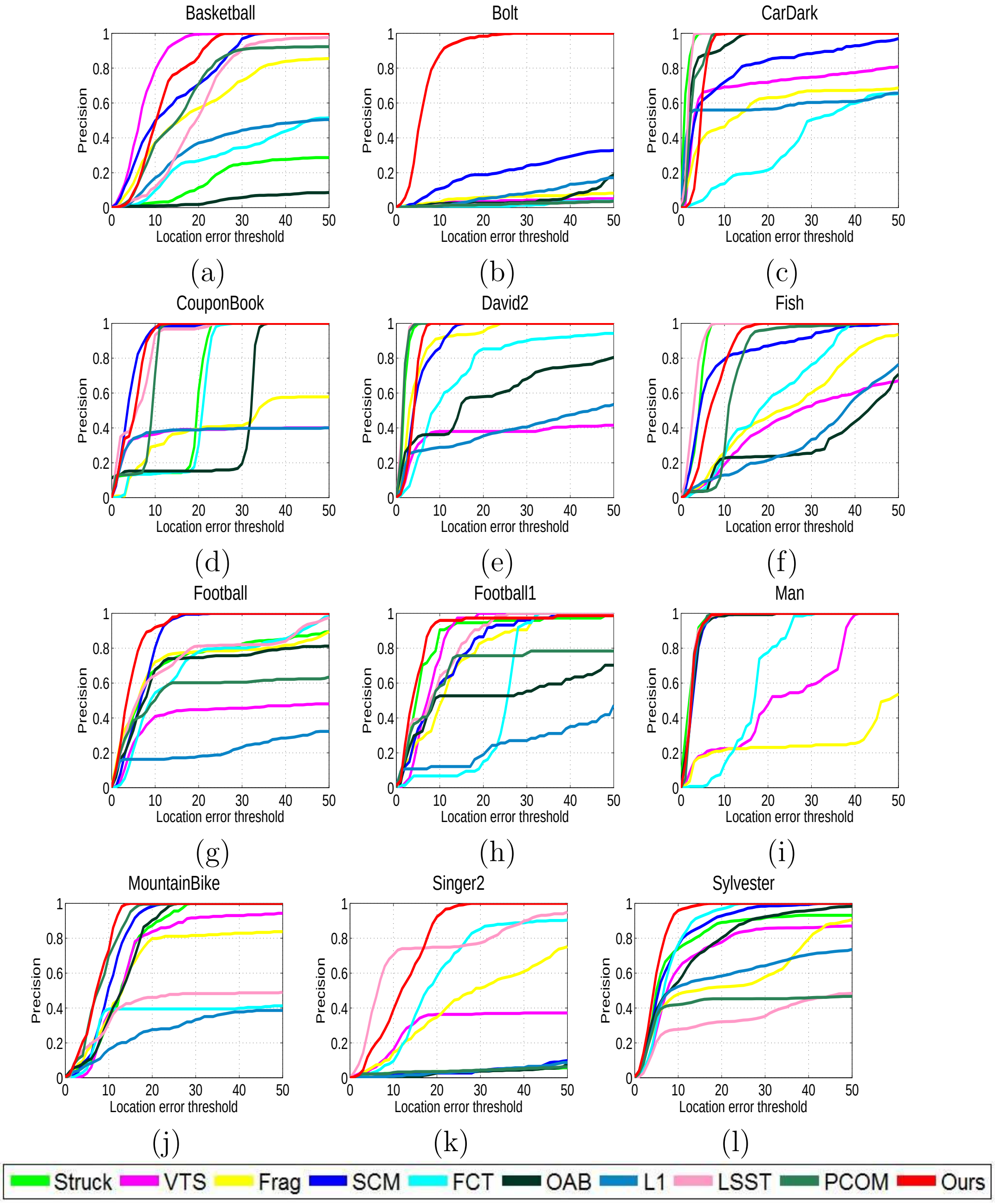
TABLE 4. Average overlap rates (%). The best two results are shown in red color and blue color, respectively.

Sequence	Struck	VTS	Frag	SCM	FCT	OAB	L1	LSST	PCOM	Ours
Basketball	12.9	**67.2**	49.2	55.3	23.0	2.5	44.6	24.1	22.8	**57.1**
Bolt	1.7	2.3	3.3	**12.9**	1.4	2.2	3.5	1.0	1.0	**69.6**
CarDark	**89.0**	56.0	42.7	54.5	13.9	79.7	56.5	**86.3**	80.5	70.4
CouponBook	70.2	35.5	37.1	**82.3**	64.8	57.1	35.2	80.2	80.5	**86.2**
David2	**85.7**	25.7	72.3	65.6	44.3	39.2	26.3	72.8	**82.3**	73.9
Fish	**84.3**	34.4	48.9	74.0	54.3	31.6	28.6	**80.9**	65.4	79.7
Football	55.7	30.8	56.2	**60.3**	47.5	52.1	16.2	53.0	42.4	**70.7**
Football1	**66.0**	53.2	48.4	45.4	16.9	37.8	13.1	53.9	48.2	**70.8**
Man	81.9	27.4	17.5	71.9	26.4	80.0	65.3	70.0	**82.3**	**82.7**
MountainBike	62.2	60.5	53.7	67.3	30.9	62.6	23.4	36.4	**73.1**	**74.4**
Singer2	4.2	27.5	44.9	5.3	56.9	4.3	6.0	**66.4**	4.4	**67.3**
Sylvester	66.0	57.6	46.4	67.8	**67.6**	61.4	46.4	27.7	35.9	**75.0**
Average	**56.6**	39.8	43.4	55.2	37.3	42.5	30.4	54.4	51.6	**73.1**

occurs, L1, Frag and VTS drift away from the target and track the other distracter until the end of the sequence. SCM achieves robust tracking performance due to the sparsity-based discriminative classifier, which distinguishes the target from a cluttered background in these sequences. The proposed tracking algorithm can performs well in these sequences. This is attributed to that affine combinations cover unknown appearances that do not appear in the template set.

Illumination variation: In the *Fish*, *Man*, *Sylvester*, *Singer*2 and *CarDark* sequences, the targets undergo drastic illumination variation. Especially, in the *CarDark* sequences, the contrast between the target and the background is low. ASLA captures the appearance variation due to the illumination variation by exploiting the partial and spatial information in the *Fish*, *CarDark* and *Sylvester* sequences. Frag can track the target when there is no drastic illumination variation, however it drifts away from the target when it undergoes drastic illumination variation. L1 only achieves robust tracking results in the *man* sequence. SCM is robust to illumination variation by holistic templates and local representations in the *Man* and *CarDark* sequences. In SCM, the sparsity-based discriminative classifier model selects discriminative features and introduces backgrounds as the negative templates to obtain accurate confidence values. In our tracking proposed, the proposed appearance model takes advantage of the affine combinations and sparse representation and achieves favorable performance in these sequences.

In-plane and **out-of-plane rotations**: The *David*2, *Bolt*, *Sylvester* and *MountainBike* sequences consist of both in-plane and out-of-plane rotations. The proposed tracking algorithm performs well in the *Bolt* sequence and accurately track the target until the end of the sequence, but all the other tracking algorithms only track the target in the first 50 frames. Struck and PCOM are robust and accurate in the *David*2 sequence. Frag uses fixed target templates, so it keeps away from the target when the target appearances undergo out-of-plane rotation in the *Sylvester* sequence. FCT learns a representative feature to track the target and achieve robust tracking results. In the *MountainBike* sequence, L1 and FCT lose the target when it rotates and drift away from the target until the end of the sequence. In the *Bolt* and *Singer*2 sequences, SCM and L1 lose the tracked targets when severe rotations happen. The proposed tracking algorithm use hull representation with ℓ_1-constraint to model target appearances. The proposed tracker



(a) (b) (c)

(d) (e) (f)

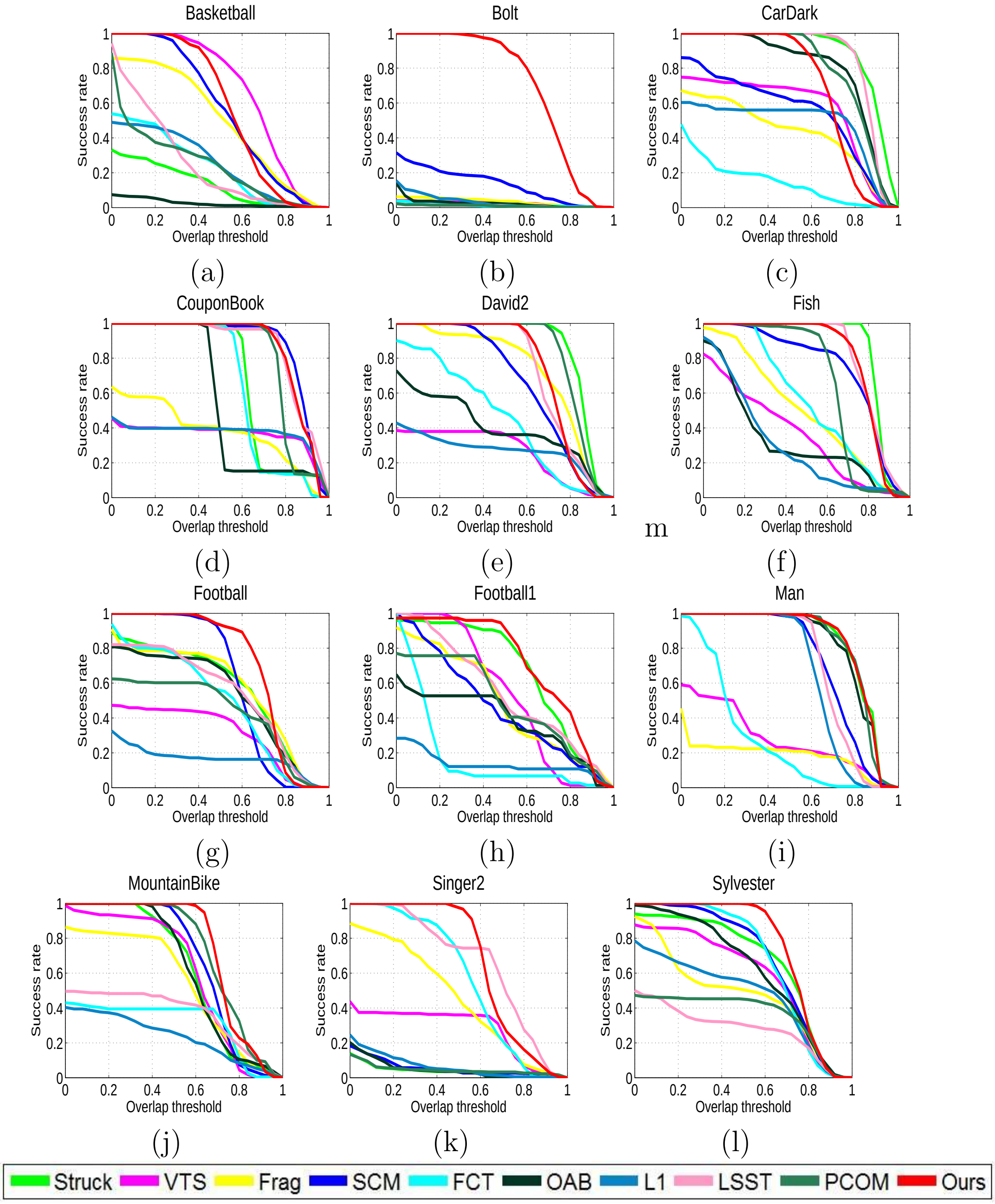
(g) (h) (i)

(j) (k) (l)

Figure 1. Precision plots in terms of location error threshold (in pixels).

select the closest affine combination as the current target representation. The proposed algorithm can successfully track the target in these sequences.

Occlusion and **deformation**: The target in the *Basketball* sequence undergoes occlusion and non-rigid deformation. Overall, VTS, SCM and the proposed tracking algorithm perform well on this sequence. In the *Football* sequence, the target is occluded by multiple similar objects. SCM takes advantage of generative and discriminative models, and distinguishes the target from cluttered background. VTS and L1 fail to track the target, and track an incorrect object when the target is occluded by a similar object. In the *Bolt* sequence, the target undergoes partial occlusion and deformation. L1 and SCM fail to track the target after the 50th frame. As shown in Fig. 3(b), the proposed tracking performs robustly to those appearance variations. This attributes to two reasons: (1) the template set include representative target appearances from previous frames. The template set



(a) (b) (c)

(d) (e) (f)

(g) (h) (i)

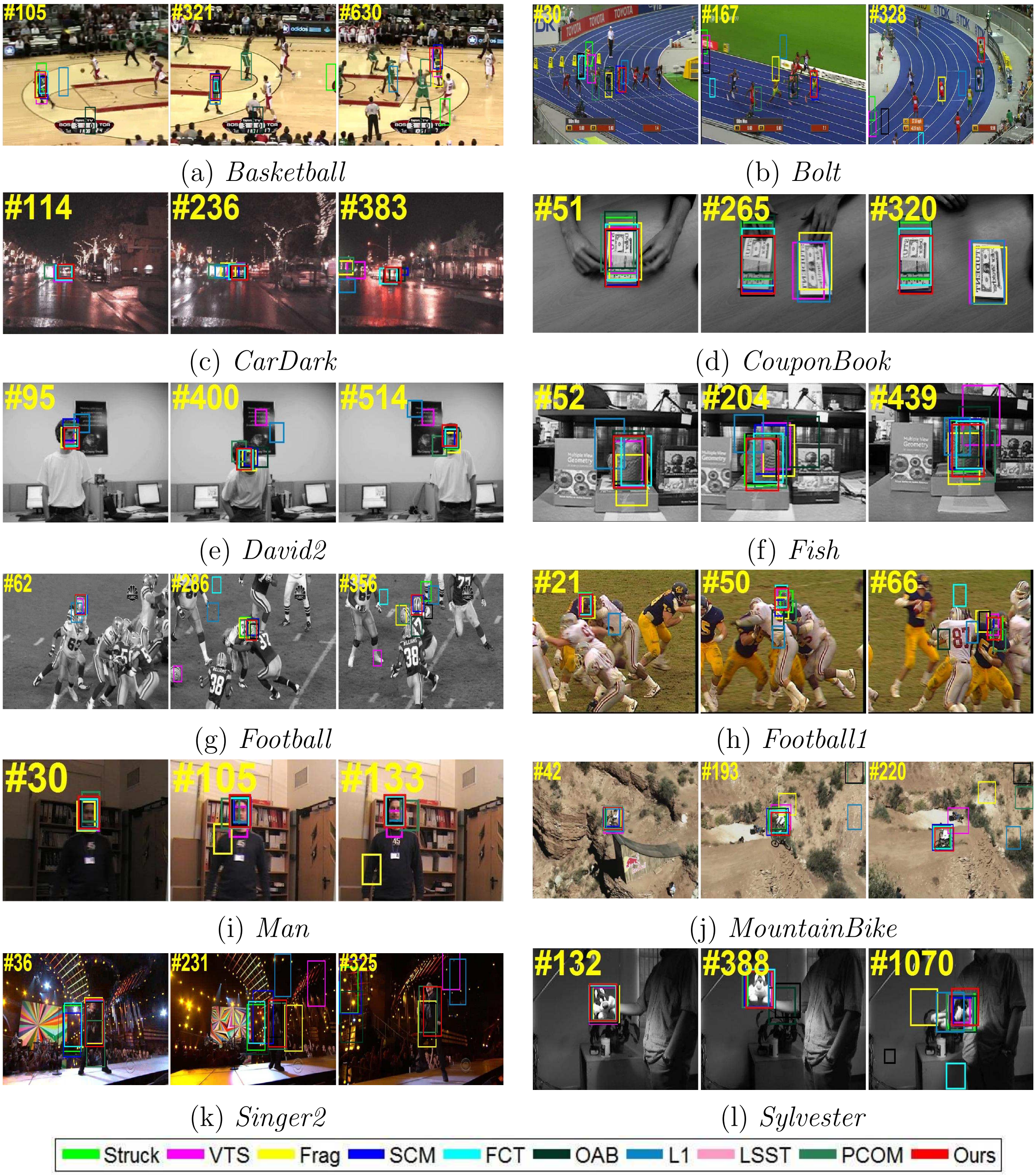
(j) (k) (l)

Figure 2. Success plots in terms of overlap threshold.

maintains the diversity and effectiveness of target templates; (2) the ℓ_1-regularized affine combinations of target templates can cover unknown target appearances. The other tracking algorithms fail to track the target after the 50th frame in the *Bolt* sequence.

From the quantitative comparisons and qualitative analysis, it can be seen that the proposed algorithm is robust to background clutters, in-plane and out-of plane rotations, illumination variations.

5. **Conclusion.** We have presented a simple yet efficient visual tracking algorithm based on ℓ_1-regularized hull representations. A target candidate is represented by an ℓ_1-regularized affine combination of target templates, which can cover unknown target appearances. The proposed appearance model exploits the advantages of both affine hull representation and sparse constraint in visual tracking. The novel observation likelihood function introduces



(a) *Basketball* (b) *Bolt*

(c) *CarDark* (d) *CouponBook*

(e) *David2* (f) *Fish*

(g) *Football* (h) *Football1*

(i) *Man* (j) *MountainBike*

(k) *Singer2* (l) *Sylvester*

Figure 3. The tracking results on the 12 sequences.

discriminative information and helps to predict the target location accurately. Comprehensive experiments show that the proposed target representation is robust to illumination variation, background clutters, rotations and partial occlusions. Both quantitative and qualitative comparisons demonstrate the effectiveness and robustness against state-of-the-art tracking algorithms.

Acknowledgment. This work is partially supported by the Jiangxi Science and Technology Research Project of Education within the Department of China (No:GJJ151127), by the National Natural Science Foundation of China (No:61661033 and 61362036), by the Jiangxi Natural Science Foundation of China (No:20161BAB202041 and 20161BAB202040), and by the Open Research Fund of Jiangxi Province Key Laboratory of Water Information Cooperative Sensing and Intelligent Processing (No:2016WICSIP019). The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers.

REFERENCES

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

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

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

[4] S. He, Q. Yang, R. Lau, J. Wang, and M. Yang, Visual tracking via locality sensitive histograms, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2427–2434, 2013.

[5] J. Wang, H. Wang and Y. Yan, Robust visual tracking by metric learning with weighted histogram representations, *Neurocomputing*, vol.153, no.1, pp. 77–88, 2015.

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

[7] J. Wang, Y. Wang, and H. Wang, Adaptive appearance modeling with point-to-set metric learning for visual tracking, *IEEE Trans. on Circuits and Systems for Video Technology*, 27(9), 2017, pp. 1987-2000.

[8] D. J. Guo, Z. M. Lu and H. Luo, Multi-Channel Adaptive Mixture Background Model for Real-time Tracking, *Journal of Information Hiding and Multimedia Signal Processing*, vol.7, no.1, pp. 216–221, 2016.

[9] X. Li, C. Shen, Q. Shi, A. Dick, and A. Hengel, Non-sparse linear representations for visual tracking with online reservoir metric learning, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1760–1767, 2012.

[10] D. Ross, J. Lim, R. Lin, and M. Yang, Incremental learning for robust visual tracking, *Int. J. Comput. Vision*, vol.77, no.1, pp. 125–141, 2008.

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

[12] D. Wang, and H. Lu, Visual Tracking via probability continuous outlier model, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3478–3485, 2014.

[13] J. Wang, H. Wang, and W. Zhao, Affine hull based target representation for visual tracking, *J. Vis. Commun. Image R.*, vol.30, no.1, pp. 266–276, 2015.

[14] H. Grabner, M. Grabner, H. Bischof, Real-Time Tracking via On-line Boosting, *Proc. of British Machine Vision Conference*, pp. 47-56, 2006.

[15] B. Babenko, M.-H. Yang, S. Belongie, Robust object tracking with online multiple instance learning, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.33, no.8 pp. 1619–1632, 2011.

[16] K.Zhang, L.Zhang, and M.Yang, Fast Compressive Tracking, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.36, no.10, pp.2002–2015, 2014.

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

[18] M. Danelljan, F.S. Khan, M. Felsberg, and J. van de Weijer, Adaptive color attributes for real-time visual tracking, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1090–1097, 2014.

[19] Z. Kalal, K. Mikolajczyk, and J. Matas, Tracking-learning-detection, *IEEE Trans. Pattern Anal. Mach. Intell.* vol.34, no.7, pp. 1409–1422, 2012.

[20] J. Zhang, S. Ma, and S. Sclaroff, MEEM: Robust tracking via multiple experts using entropy minimization, *Proc. of European Conference on Computer Vision*, pp. 188–203, 2014.

[21] L. Wang, W. OuYang, X. Wang, H. Lu, STCT: Sequentially Training Convolutional Networks for Visual Tracking, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 150–158, 2016.

[22] X. Mei, H. Ling, Robust visual tracking and vehicle classification via sparse representation, *IEEE Trans. Pattern Anal. Mach. Intell.*, vol.33, no.11, pp. 2259–2272, 2011.

[23] C. Bao, Y. Wu, H. Ling, and H. Ji, Real time robust l1 tracker using accelerated proximal gradient approach, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1830–1837, 2012.

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

[25] W. Zhong, H. Lu, and M. Yang, Robust object tracking via sparse collaborative appearance model, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1838–1845, 2012.

[26] T. Zhang, S. Liu, C. Xu, S. Yan and B. Ghanem, Structure sparse tracking, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 150–158, 2015.

[27] L. Zhang, H. Lu, D. Du and L. Liu, Sparse hashing tracking, *IEEE Trans. on Image Processing*, vol.25, no.2, pp. 840–849, 2016.

[28] T. Zhang, S. Liu, N. Ahuja, M.H. Yang and B. Ghanem, Robust visual tracking via consistent low-rank sparse learning, *International Journal of Computer Vision*, vol.111, no.2, pp. 171–190, 2015.

[29] T. Zhang, A. Bibi, and G. Bernard, In Defense of Sparse Tracking: Circulant Sparse Tracker, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3880–3888, 2016. ?I

[30] P. Zhu, W. Zuo, L. Zhang, S. Shiu, and D. Zhang, Image Set based Collaborative Representation for Face Recognition, *IEEE Trans. INF. FOREN. SEC.*, vol.9, no.7, pp. 1120–1132, 2014.

[31] J. Wang, Y. Wang, C. Deng, S. Wang, et al., Sparse affine hull for visual tracking, *Proc. of International Conference of Digital Home*, pp. 85–88, 2016.

[32] X. Ren, and D. Ramanan, Histograms of sparse codes for object detection, *Proc. of IEEE Conference on Computer Vision and Pattern Recognition*, pp. 3246–3253, 2013.