ICIC Express Letters
Part B: Applications
Volume **7**, Number **4**, April **2016**

ICIC International ©2016 ISSN 2185-2766
pp. **789–795**

ROBUST VISUAL TRACKING WITH IMPROVED COLLABORATIVE REPRESENTATION

Chenhua Liu[1], Jun Kong[1,2], Min Jiang[1] and Shengwei Tian[2]

[1]Key Laboratory of Advanced Process Control for Light Industry, Ministry of Education
Jiangnan University
No. 1800, Lihu Avenue, Wuxi 214122, P. R. China
liuchenhua1991@163.com

[2]College of Electrical Engineering
Xinjiang University
No. 14, Shengli Road, Urumqi 830047, P. R. China
kongjun@jiangnan.edu.cn

Received September 2015; accepted December 2015

Abstract. *Visual tracking is a challenging problem for the appearance changes caused by extrinsic and intrinsic factors. In this paper, a robust object tracking algorithm exploiting both collaborative representation and coding residual is proposed within the Bayesian inference framework. To solve the constrained convex optimization problem, we propose an effective numerical algorithm for the minimization problem based on the Augmented Lagrange Multiplier (ALM) method, which guarantees the object representation to be solved efficiently. Extensive experimental results on several challenging sequences confirm the effectiveness of the approach, which significantly outperforms competitive trackers in terms of accuracy measures including the overlap ratio and center location error, respectively.*

Keywords: Visual tracking, Collaborative representation, Bayesian inference, Augmented Lagrange Multiplier

1. **Introduction.** As one of the underlying issues in computer vision, object tracking plays a significant role due to potential applications in critical tasks such as image compression, video surveillance, and activity analysis. While much progress has been made in the past decades, designing a robust visual tracking system is still a challenging problem due to multitudinous challenges including background clutter, fast motion, varying illumination, and occlusion.

Recently, more attention has been paid to the sparse representation for object tracking [1-4]. In [1], Mei and Ling present an L_1 tracker based on sparse representation of target templates and trivial templates. The tracking task is aimed to search the most possible patch with sparse representation and the error term is treated as arbitrary but sparse noise handled with trivial templates. Some other L_1 based tracking methods have been proposed from different views to improve the effects of the tracking. In [2], Jia et al. present an L_1 tracker based on the structural local sparse appearance model that integrates local and global information of an observed image through an alignment pooling method. In [4], Zhuang et al. propose a discriminative sparse similarity map obtained from a multi-task reverse sparse coding approach with Laplacian term for visual tracking.

Some latest researches in face recognition and visual tracking show that methods with collaborative representation can also have a well performance compared with the sparse representation [5-8]. In [5], Yang et al. reveal that it is the collaborative representation, not the sparse representation, that truly improves the accuracy of face recognition. Inspired by the collaborative representation in [5] and subspace learning in [6], Xiao et al. further propose an L_2-regularized based tracking method to powerfully use all of the

orthogonal Principal Component Analysis (PCA) basis vectors in subspace for object representation [8]. To handle the appearance changes in the process of tracking, the square noise templates are introduced to represent the corrupted object. However, the weak sparse projection coefficient of L_2-regularized may deteriorate the ambiguity of square noise templates that reconstruct both the foreground and background. Moreover, the residual modeled by Gaussian cannot well tolerate the outliers (e.g., occlusion) for the weak sparsity of L_2-norm.

Motivated by the above-mentioned work and success of outliers handling in face recognition [9], we propose a robust tracking algorithm based on collaborative representation of PCA basis vectors. Different from the object representation in [8], we use the L_1-norm to measure the coding residual for robustness to outliers. To solve the minimization problem of object representation, the Augmented Lagrange Multiplier (ALM) method [10] is adopted which can guarantee the representation model to be solved effectively.

The paper is organized as follows. Section 2 presents the proposed object representation. Section 3 presents the tracking framework. Section 4 conducts extensive experiments to demonstrate the performance of proposed method. Section 5 places the conclusion.

2. Object Representation.

In this section, we propose an improved collaborative representation model for robust visual tracking and an effective numerical algorithm to solve the proposed appearance model.

2.1. Improved collaborative representation.

Given an orthogonal PCA subspace $\boldsymbol{D} \in R^{d\times m}$, where d and m respectively represent the feature dimension and the number of basis vectors, the target region $\boldsymbol{y} \in R^{d\times 1}$ can be represented by an image subspace with projection coefficient $\boldsymbol{c} \in R^{m\times 1}$. To collaboratively represent object, we use L_2-norm to regulate $\boldsymbol{c}$, and then we have the following minimization problem:

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}}\{\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}\|_2^2 + \lambda\|\boldsymbol{c}\|_2^2\} \tag{1}$$

where λ is a regularization parameter.

However, when outliers occur in sequence, using L_2-norm to measure the representation fidelity is less robust than L_1-norm for the L_1-norm could tolerate the outliers [9], and then we have:

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}}\{\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}\|_1 + \lambda\|\boldsymbol{c}\|_2^2\} \tag{2}$$

Let $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}$, and we can rewrite Equation (2) as

$$\boldsymbol{c}^* = \arg\min_{\boldsymbol{c}}\{\|\boldsymbol{e}\|_1 + \lambda\|\boldsymbol{c}\|_2^2\} \quad \text{s.t.} \quad \boldsymbol{y} = \boldsymbol{D}\boldsymbol{c} + \boldsymbol{e} \tag{3}$$

2.2. Effective numerical method for solving Equation (3).

Equation (3) is a constrained convex optimization problem which can be efficiently solved through ALM operation. The corresponding ALM function is:

$$L_\tau(\boldsymbol{e}, \boldsymbol{c}, \boldsymbol{\gamma}) = \|\boldsymbol{e}\|_1 + \lambda\|\boldsymbol{c}\|_2^2 + \langle\boldsymbol{\gamma}, \boldsymbol{y} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{e}\rangle + \frac{\tau}{2}\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{e}\|_2^2 \tag{4}$$

where $\langle\cdot,\cdot\rangle$ denotes the inner product operator, $\boldsymbol{\gamma}$ is a vector of Lagrange multiplier, τ is a constant that determines the penalty for large representation error, and $\{\tau_k\}$ is a monotonically increasing positive sequence. The ALM method iteratively estimates the optimal solutions and the Lagrange multiplier by minimizing the augmented Lagrangian function:

$$(\boldsymbol{e_{k+1}}, \boldsymbol{c_{k+1}}) = \arg\min_{\boldsymbol{e},\boldsymbol{c}} L_{\tau_k}(\boldsymbol{e}, \boldsymbol{c}, \boldsymbol{\gamma_k}) \tag{5}$$

$$\boldsymbol{\gamma_{k+1}} = \boldsymbol{\gamma_k} + \tau_k(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{e}) \tag{6}$$

$$\tau_{k+1} = \rho\tau_k \tag{7}$$

where ρ is a constant that ensures $\{\tau_k\}$ monotonically increases. The minimization in Equation (5) could be implemented by alternatively and iteratively updating the two unknowns $\boldsymbol{c}$ and $\boldsymbol{e}$ as follows:

$$\begin{cases} \boldsymbol{c_{k+1}} = \arg\min\limits_{\boldsymbol{c}} L_{\tau_k}(\boldsymbol{e_k}, \boldsymbol{c}, \boldsymbol{\gamma_k}) \\ \boldsymbol{e_{k+1}} = \arg\min\limits_{\boldsymbol{e}} L_{\tau_k}(\boldsymbol{e}, \boldsymbol{c_{k+1}}, \boldsymbol{\gamma_k}) \end{cases} \tag{8}$$

We could have a closed-form solution:

$$\begin{cases} \boldsymbol{c_{k+1}} = (\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D} + 2\lambda/\tau_k)^{-1}\boldsymbol{D}^{\mathrm{T}}\left(\boldsymbol{y} - \boldsymbol{e_k} + \boldsymbol{\gamma_k}/\tau_k\right) \\ \boldsymbol{e_{k+1}} = S_{1/\tau_k}\left[\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c_{k+1}} - \boldsymbol{e_{k+1}}\right] \end{cases} \tag{9}$$

where $S_\theta(x)$ is the soft thresholding operator, which defines as sgn(x)max(|x|-θ). Let $\boldsymbol{P_k} = (\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D} + 2\lambda/\tau_k)^{-1}\boldsymbol{D}^{\mathrm{T}}$, $\boldsymbol{P_k}$ is independent from $\boldsymbol{y}$, therefore, we can pre-calculate it as a set of projection matrices for all the candidates in each frame. Once a candidate image patch $\boldsymbol{y}$ comes, we can simply project $\boldsymbol{y}$ onto $\boldsymbol{P_k}$ via $\boldsymbol{P_k}\boldsymbol{y}$ in the first stage of ALM, which makes the calculation quickly. The entire algorithm for solving Equation (3) is summarized in Algorithm 1.

Algorithm 1 Effective ALM method for solving (3)

set $\boldsymbol{e_1} = \boldsymbol{c_1} = \boldsymbol{\gamma_1} = 0$, $\tau_1 = 10$
Input: The PCA subspace $\boldsymbol{D}$, the candidate sample $\boldsymbol{y}$
for $k = 1, 2, \ldots,$ until both the $\boldsymbol{c}$ and $\boldsymbol{e}$ are convergent to optimal state **do**
  $\boldsymbol{c_{k+1}} = (\boldsymbol{D}^{\mathrm{T}}\boldsymbol{D} + 2\lambda/\tau_k)^{-1}\boldsymbol{D}^{\mathrm{T}}\left(\boldsymbol{y} - \boldsymbol{e_k} + \boldsymbol{\gamma_k}/\tau_k\right)$
  $\boldsymbol{e_{k+1}} = S_{1/\tau_k}\left[\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c_{k+1}} - \boldsymbol{e_{k+1}}\right]$
  $\boldsymbol{\gamma_{k+1}} = \boldsymbol{\gamma_k} + \tau_k\left(\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c} - \boldsymbol{e}\right)$
  $\tau_{k+1} = \rho\tau_k$
end for
Output: The optimal $\boldsymbol{c^*}$ and $\boldsymbol{e^*}$

3. **Tracking Framework.** The object tracking task can be casted as a Bayesian inference problem in the Markov model. Given a series of observed samples $\boldsymbol{y_{1:t}} = \{\boldsymbol{y_1}, \boldsymbol{y_2}, \ldots, \boldsymbol{y_t}\}$, the purpose is to recursively estimate the hidden state variable:

$$p(\boldsymbol{x_t}|\boldsymbol{y_{1:t}}) \propto p(\boldsymbol{y_t}|\boldsymbol{x_t}) \int p(\boldsymbol{x_t}|\boldsymbol{x_{t-1}}) p(\boldsymbol{x_{t-1}}|\boldsymbol{y_{1:t-1}}) d\boldsymbol{x_{t-1}} \tag{10}$$

where $\boldsymbol{x_t}$ is the object state, $\boldsymbol{y_t}$ is the observation at time t, $p(\boldsymbol{x_t}|\boldsymbol{x_{t-1}})$ denotes the motion model between two continuous object states while $p(\boldsymbol{y_t}|\boldsymbol{x_t})$ indicates the observation model which is applied to computing the likelihood of candidates. The entire tracking procedure is summarized in Figure 1. At the outset, the state of the target is manually initialized. Then, the candidate samples can be obtained from the motion model. Once the state parameters of residual and projection coefficient are obtained from Algorithm 1, we can evaluate the likelihood of each candidate state. Finally, the samples are cumulated to update the subspace for handling the change of target object. The whole tracking procedure will keep running until the target state of last frame is obtained.

Motion Model: Let $\boldsymbol{x_t} = \{l_x, l_y, \theta, s, \alpha, \phi\}$, where l_x, l_y, θ, s, α, ϕ indicate x, y translations, rotation angle, scale, aspect ratio, and skew respectively. These affine parameters are supposed to be independent and modeled by six scalar Gaussian distributions. We use the random walk to formulate the state transition, i.e., $p(\boldsymbol{x_t}|\boldsymbol{x_{t-1}}) = N\left(\boldsymbol{x_t}; \boldsymbol{x_{t-1}}, \boldsymbol{\Psi}\right)$, where $\boldsymbol{\Psi}$ is a diagonal covariance matrix of the affine parameters.

Observation Model: The residual $\boldsymbol{e}$ can be considered as the error for the reconstruction. However, the small Gaussian noises still exist in real situations. Thus, the observation likelihood can be measured as:

$$p(\boldsymbol{y}|\boldsymbol{x}) = \exp\left(-\sigma E(\boldsymbol{c^*}, \boldsymbol{e^*})\right) \tag{11}$$

where $E(\boldsymbol{c}^*, \boldsymbol{e}^*) = (1/2)\|\boldsymbol{y} - \boldsymbol{D}\boldsymbol{c}^* - \boldsymbol{e}^*\|_2^2 + \lambda\|\boldsymbol{e}^*\|_1$, and σ is a constant controlling the shape of the Gaussian kernel. The former part of $E(\boldsymbol{c}^*, \boldsymbol{e}^*)$ accounts for the small Gaussian noises and the latter part aims to penalize the primary residual pixels.

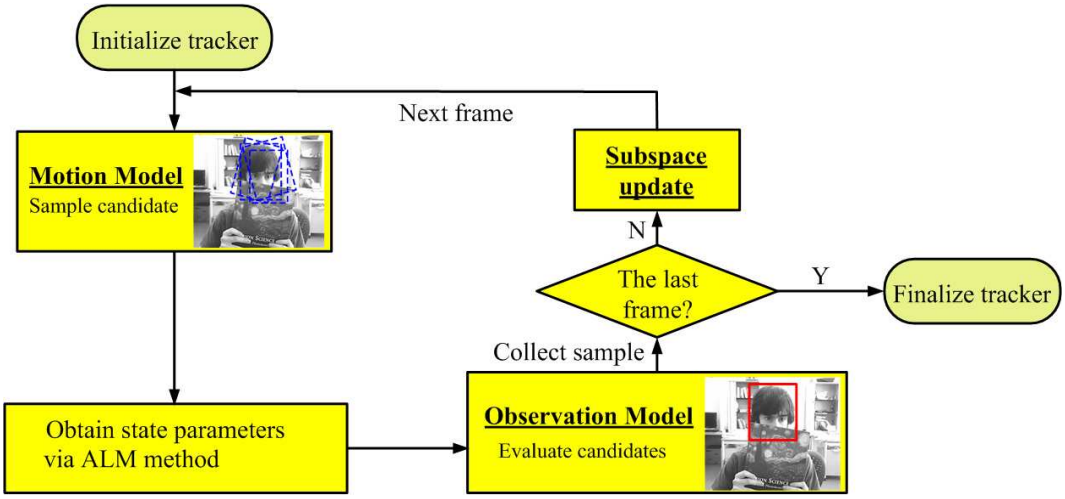

FIGURE 1. The illustration of the proposed tracking framework

Subspace update: The columns of subspace $\boldsymbol{D}$ are PCA basis vectors. To learn the appearance of the target object while tracking processes, it is necessary to incrementally update the subspace when new observations arrive. In this paper, we adopt the incremental PCA method [6] to update the subspace when enough samples are cumulated. As the residual term can be used to identify outliers, the sample used to update the subspace can be extracted as:

$$y_i = \begin{cases} y_i & |e_i| = 0 \\ \mu_i & \text{otherwise} \end{cases} \tag{12}$$

where y_i, μ_i and e_i are the i-th element of $\boldsymbol{y}$, $\boldsymbol{\mu}$, and $\boldsymbol{e}$ respectively, and $\boldsymbol{\mu}$ is the mean vector in [6].

4. **Experiments.** The proposed tracker is implemented in MATLAB and runs at 6 frames per second on a 3.5GHz CPU with 8GB memory. We empirically set $\lambda = 5e^{-2}$, $\rho = 1.5$, $\sigma = 20$. The location of the target in the first frame is manually labeled. Each observation is normalized to 32×32 pixels, and 16 PCA basis vectors are used for the subspace in all the experiments. 600 particles are adopted and our tracker is incrementally updated when 5 samples are cumulated.

To prove the effectiveness of the proposed algorithm, we use eight challenging image sequences containing different challenging factors (e.g., illumination change, severe occlusion, and background clutter) and compare our method with six competitive methods: Incremental Visual Tracking (IVT) [6], Probability Continuous Outlier Model (PCOM) [7], Adaptive Structural Local Appearance (ASLA) [2], Sparsity based Collaborative Model (SCM) [3], Discriminative Sparse Similarity Tracking (DSST) [4], and L2-regularized Least Square (L2-RLS) [8]. For a fair evaluation, we run these codes with the same bounding box in the first frame.

4.1. **Quantitative evaluation.** Quantitative evaluation is aimed to fairly evaluate the ability of tracking methods. We evaluate the aforementioned algorithms by computing their average overlap rates and center errors. It should be noted that a bigger overlap rate or a smaller center error means a more proper result. Given the result of each frame and corresponding ground truth, we can get the overlap rate by the PASCAL VOC [11] criterion. The results are listed in Table 1 and Table 2.

TABLE 1. Average overlap rate. The best result is shown in bold font.

Sequence	IVT	ASLA	SCM	PCOM	DSST	L2-RLS	Ours
Occlusion2	0.73	0.70	0.82	**0.83**	0.60	0.78	**0.83**
DavidOutdoor	0.52	0.46	0.38	0.57	0.13	**0.75**	0.74
DavidIndoorNew	0.44	0.42	0.51	**0.76**	0.60	0.23	**0.76**
Singer1	0.47	0.82	0.84	0.60	0.70	0.24	**0.86**
Boy	0.19	**0.79**	0.53	0.31	0.78	**0.79**	**0.79**
Jumping	0.62	0.67	0.73	0.68	0.61	0.73	**0.74**
Stone	0.12	0.51	**0.62**	0.43	0.10	0.37	0.60
Deer	0.24	0.63	0.61	0.55	0.63	0.60	0.69
Average	0.416	0.625	0.630	0.591	0.519	0.561	**0.751**
Speed(fps)	32	9	0.5	20	4	10	6

TABLE 2. Average center error (pixel). The best result is shown in bold font.

Sequence	IVT	ASLA	SCM	PCOM	DSST	L2-RLS	Ours
Occlusion2	7.8	6.9	4.4	4.5	11.9	5.5	**3.4**
DavidOutdoor	52.4	86.5	67.1	51.2	209.8	**6.0**	7.4
DavidIndoorNew	35.9	32.4	17.7	3.8	11.9	132.0	**3.4**
Singer1	11.9	3.8	3.2	10.8	12.8	72.8	**2.7**
Boy	177.2	**2.8**	51.8	146.7	3.2	2.9	3.0
Jumping	6.4	5.2	3.9	4.9	6.8	3.8	**3.7**
Stone	115.1	3.7	**2.6**	29.3	56.6	25.7	3.1
Deer	135.2	**5.9**	10.1	14.9	8.8	9.4	6.1
Average	67.7	18.4	20.1	33.3	40.2	32.3	**4.1**

4.2. **Qualitative evaluation.**

Severe Occlusion: We test two sequences (*Occlusion2*, *DavidOutdoor*) with occlusion. The IVT tracker does not take the occlusion into consideration for the object representation and this tracker is less effective for the sequence with severe occlusion. Overall, the L2-RLS and our tracker can perform better than other trackers. As for the ambiguity of the L_2-regularized coefficient to square template, we adopt the L_1-norm to measure the residual. Therefore, our tracker is more robust to occlusion compared with L2-RLS tracker.

Illumination Change: Figure 2(b) presents the tracking results in the sequences (*Singer1*, *DavidIndoorNew*) with drastic illumination change. The L2-RLS tracker is less effective in both of the two sequences for the weak tolerance to outliers. As our tracker applies the L_1-norm to tolerating the outliers and PCA basis vectors to modelling the subspace respectively, our tracker is more robust to the illumination change.

Motion Blur: Figure 2(c) shows results from two challenging sequences (*Boy*, *Jumping*) with abrupt motion. It is a challenging task to estimate the locations of the target when abrupt motion occurs. Moreover, the imprecise prediction of location will cause the tracked target to be inaccurate and degenerate the subspace or template dictionary. It can be seen that our tracker and L2-RLS tracker perform well in both of the two sequences for the powerful ability to collaboratively represent the object.

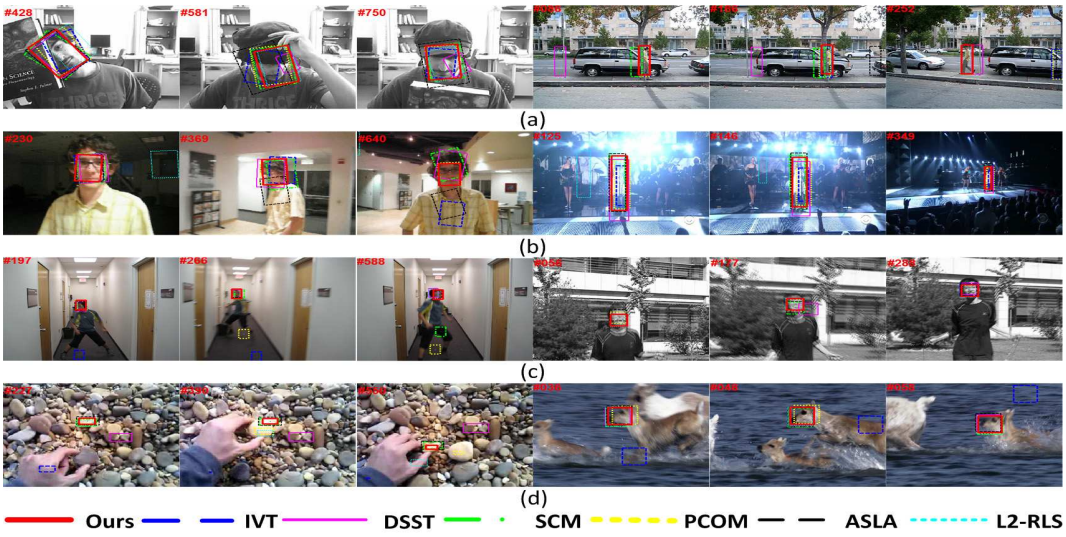

FIGURE 2. Sample tracking results on eight challenging sequences: (a) *Occlusion2* and *DavidOutdoor* with occlusion; (b) *DavidIndoorNew* and *Singer1* with illumination change; (c) *Boy* and *Jumping* with fast motion; (d) *Stone* and *Deer* with background clutter

Background Clutter: Figure 2(d) shows the tracking results in the *Stone* and *Deer* with complex background. Moreover, the *Stone* sequence contains partial occlusion and the *Deer* sequence contains motion blur, respectively. As the proposed tracker can effectively handle the outliers in the process of tracking, our tracker can be more effective in these two sequences.

5. **Conclusions.** This paper presents a robust visual tracking method based on the improved L_2-regularized collaborative representation. Different from the traditional supposition of Gaussian to the coding residual, we use the L_1-norm to measure the coding residual for the tolerating of the outliers. Moreover, an effective ALM based numerical algorithm is applied to solving the minimization problem of object representation. Extensive experimental results validate the proposed method can achieve more favorable performance than several competitive methods. In the future, we plan to integrate multiple visual cues (e.g., color) into our object representation for more effective tracking.

Acknowledgment. This work is partially supported by Xinjiang Uygur Autonomous Regions University Science and Research Key Project (XJEDU2012I08), National Natural Science Foundation of China (61362030, 61201429), and Technology Research Project of The Ministry of Public Security of China (2014JSYJB007).

REFERENCES

[1] X. Mei and H. Ling, Robust visual tracking using L_1 minimization, *IEEE International Conference on Computer Vision*, pp.1436-1443, 2009.
[2] X. Jia, H. Lu and M. Yang, Visual tracking via adaptive structural local sparse appearance model, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1822-1829, 2012.
[3] W. Zhong, H. Lu and M. Yang, Robust object tracking via sparsity based collaborative model, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.1838-1845, 2012.
[4] B. Zhuang, H. Lu, Z. Xiao and D. Wang, Visual tracking via discriminative sparse similarity map, *IEEE Trans. Image Processing*, vol.23, no.4, pp.1872-1881, 2014.
[5] L. Zhang, M. Yang and X. Feng, Sparse representation or collaborative representation: Which helps face recognition?, *IEEE International Conference on Computer Vision*, pp.471-478, 2011.
[6] D. Ross, J. Lim, R. Lin and M. Yang, Incremental learning for robust visual tracking, *International Journal of Computer Vision*, vol.77, no.1, pp.125-141, 2008.

[7] D. Wang, H. Lu and M. Yang, Visual tracking via probability continuous outlier model, *IEEE Conference on Computer Vision and Pattern Recognition*, pp.3478-3485, 2014.

[8] Z. Xiao, H. Lu and D. Wang, L2-RLS based object tracking, *IEEE Trans. Circuits and Systems for Video Technology*, vol.24, no.8, pp.1301-1309, 2014.

[9] M. Yang, X. Feng, Y. Ma and D. Zhang, Collaborative representation based classification for face recognition, *Technical Report*, 2012.

[10] Z. Lin, M. Chen, L. Wu and Y. Ma, The augmented lagrange multiplier method for exact recovery of corrupted low-rank matrices, *Technical Report*, University of Illinois at Urbana-Champaign, 2009.

[11] M. Everingham, L. Van Gool, C. K. Williams, J. Winn and A. Zisserman, The pascal visual object classes (VOC) challenge, *International Journal of Computer Vision*, vol.88, no.2, pp.303-338, 2010.