Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 404978, 7 pages
http://dx.doi.org/10.1155/2013/404978

Research Article

Visual Object Tracking Based on 2DPCA and ML

Ming-Xin Jiang,[1,2] Min Li,[1] and Hong-Yu Wang[2]

[1] *School of Information & Communication Engineering, Dalian Nationalities University, Dalian 116600, China*
[2] *School of Information & Communication Engineering, Dalian University of Technology, Dalian 116600, China*

Correspondence should be addressed to Min Li; limin@dlnu.edu.cn

Received 7 March 2013; Accepted 23 May 2013

Academic Editor: Yudong Zhang

We present a novel visual object tracking algorithm based on two-dimensional principal component analysis (2DPCA) and maximum likelihood estimation (MLE). Firstly, we introduce regularization into the 2DPCA reconstruction and develop an iterative algorithm to represent an object by 2DPCA bases. Secondly, the model of sparsity constrained MLE is established. Abnormal pixels in the samples will be assigned with low weights to reduce their effects on the tracking algorithm. The object tracking results are obtained by using Bayesian maximum a posteriori (MAP) probability estimation. Finally, to further reduce tracking drift, we employ a template update strategy which combines incremental subspace learning and the error matrix. This strategy adapts the template to the appearance change of the target and reduces the influence of the occluded target template as well. Compared with other popular methods, our method reduces the computational complexity and is very robust to abnormal changes. Both qualitative and quantitative evaluations on challenging image sequences demonstrate that the proposed tracking algorithm achieves more favorable performance than several state-of-the-art methods.

1. Introduction

As one of the fundamental problems of computer vision, visual tracking plays a critical role in advanced vision-based applications (e.g., visual surveillance, human-computer interaction, augmented reality, intelligent transportation, and context-based video compression) [1–3]. However, building a robust model-free tracker is still a challenging issue due to the difficulty arising from the appearance variability of an object of interest, which includes intrinsic appearance variability (e.g., pose variation and shape deformation) and extrinsic factors (illumination changes, camera motion, occlusions, etc.).

Typically, a complete tracking system can be divided into three main components: (1) an appearance observation model, which evaluates the likelihood of a candidate state belonging to the object model, (2) a motion model, which aims to model the states of an object over time (such as Kalman filtering and particle filtering), and (3) a search strategy for finding the most likely states in the current frame (e.g., mean shift and sliding window). In this paper, we are devoted to developing a robust appearance model.

Due to the power of subspace representation, subspace-based trackers (e.g., [4, 5]) are robust to in-plane rotation, scale change, illumination variation, and pose change. However, they are sensitive to partial occlusion caused by their underlying assumption that the error term is Gaussian distributed with small variances. This assumption does not hold for object representation when partial occlusion occurs as the noise term cannot be modeled with small variances.

An effective tracking algorithm (called L1 tracker) based on sparse representation within a particle filter framework is developed in [6]. The L1 tracker represents the tracked target by using a set of target templates and trivial templates. The target templates depict a subspace on the tracked object and the trivial templates aim to model the occlusion effectively. However, the use of trivial templates increases the number of templates significantly, which make the computational complexity of L1 tracker too high to satisfy real applications.

In [7], the authors also presented a sparse coding-based tracker by combing sparse coding and Kalman filtering and fusing the color and gradient features. To account for the variations of the tracked object during the tracking

processing, they use a template update strategy by replacing a random template of the original template library with the last tracking result. However, this simple update manner can easily introduce tracking errors when abnormal changes occur, which may cause tracking drift.

Motivated by aforementioned discussions, we propose an object tracking algorithm based on 2DPCA and MLE. Firstly, we introduce regularization into the 2DPCA reconstruction and develop an iterative algorithm to represent an object by 2DPCA bases. Secondly, the model of sparsity constrained MLE is established. Abnormal pixels in the samples will be assigned with low weights to reduce their affects on the tracking algorithm. The object tracking results are obtained by using Bayesian maximum a posteriori probability (MAP) estimation. Finally, to further reduce tracking drift, we employ a template update strategy which combines incremental subspace learning and the error matrix. This strategy adapts the template to the appearance change of the target and reduces the influence of the occluded target template as well. The experimental results show that our algorithm can achieve stable and robust performance especially when occlusion, rotation, scaling, or illumination variation occurs.

2. Visual Object Tracking Model Based on 2DPCA and MLE: The Theory of 2DPCA

2.1. The Theory of 2DPCA. Principal component analysis (PCA) is a well-established linear dimension-reduction technique, which has been widely used in many areas (such as face recognition [8]). It finds the projection directions along which the reconstruction error to the original data is minimum and projects the original data into a lower dimensional space spanned by those directions corresponding to the top eigenvalues. Recent studies demonstrate that two-dimensional principal component analysis (2DPCA) could achieve performance comparable to PCA with less computational cost [9, 10].

Given a series of image matrices $\mathbf{Y} = [Y_1 \ Y_2 \ \cdots \ Y_d]$, 2DPCA aims to obtain an orthogonal left-projection matrix $\mathbf{U}$, an orthogonal right-projection matrix $\mathbf{V}$, and the projection coefficients $\mathbf{A} = [A_1 \ A_2 \ \cdots \ A_d]$ by solving the following objective function:

$$\min_{\mathbf{U},\mathbf{V},\mathbf{A}_i} \frac{1}{d}\sum_{i=1}^{d}\left\|\mathbf{Y}_i - \mathbf{U}\mathbf{A}_i\mathbf{V}'\right\|_F^2. \tag{1}$$

Then the coefficient $\mathbf{A}_i$ can be approximated by $\mathbf{A}_i \approx \mathbf{U}'\mathbf{Y}_i\mathbf{V}$. We note that the underlying assumption of (1) is that the error term is Gaussian distributed with small variances. This assumption is not able to deal with partial occlusion as the error term cannot be modeled with small variances when occlusion occurs. In this paper, we propose an object tracking algorithm by using 2DPCA basis matrices and an additional MLE error matrix $\mathbf{Y} \approx \mathbf{U}\mathbf{A}\mathbf{V}' + \mathbf{e}$.

Let the objective function be

$$L(\mathbf{A},\mathbf{E}) = \frac{1}{2}\left\|\mathbf{Y} - \mathbf{U}\mathbf{A}\mathbf{V}' - \mathbf{E}\right\|_F^2 + \lambda\|\mathbf{e}\|_1; \tag{2}$$

the problem is

$$\begin{aligned} \min_{\mathbf{A},\mathbf{E}} \quad & L(\mathbf{A},\mathbf{E}) \\ \text{s.t.} \quad & \mathbf{U}'\mathbf{U} = \mathbf{I}; \ \ \mathbf{V}'\mathbf{V} = \mathbf{I}, \end{aligned} \tag{3}$$

where $\mathbf{Y}$ denotes an observation matrix, $\mathbf{A}$ indicates its corresponding projection coefficient, and λ is a regularization parameter. $\mathbf{e}$ describes the error matrix.

2.2. MLE Model. The basic idea of sparse coding is to use the templates in a given dictionary $\mathbf{T}$ to represent a testing sample y (as $y \approx T\alpha$), where α is sparse coding coefficient vector. Traditionally, the sparsity can be measured by L0-norm and the L0-norm minimization is an NP-hard problem. Fortunately, [11] proves that when the solution is sparse enough, L0-norm minimization is equivalent to the L1-norm minimization.

Therefore, the sparse coding problem can be defined as [12, 13]

$$\begin{aligned} \min_{\alpha} \quad & \|\alpha\|_1 \\ \text{s.t.} \quad & \left\|y - T\alpha\right\|_2^2 \le \varepsilon, \end{aligned} \tag{4}$$

where $\varepsilon > 0$ is a very small constant. This model shows two constraints in sparse coding: one is that $\min_\alpha \|\alpha\|_1$ constrains the sparsity of represented signal; the other is that $\|y - T\alpha\|_2^2 \le \varepsilon$ constrains the accuracy of the represented signal [14–17].

The analysis of the two constraint terms mentioned earlier is as follows. For object tracking, the accuracy constraint is more important than the sparsity one, especially when occlusion, rotation, scaling, or illumination variation happens to the object. In that case, considering some possible abnormal changes, whether the model can accurately describe the object or not will directly determine the success or failure of tracking algorithm. Most of current algorithms are presented under the assumption that the sparse coding residual $\mathbf{e} = y - T\widehat{\alpha}$ follows the Gaussian distribution. In practice, however, this assumption is limited when abnormal changes happen which will inevitably lead to the failure of tracking algorithm.

In sparsity constraints, though L1-norm minimization is more efficient than the L0-norm minimization, the fact is that the L1-norm minimization programming is still very time consuming. Object tracking algorithms are different from face recognition algorithms in that face recognition algorithms do not demand fast processing speed in a sample training process, while in object tracking, slow processing speed will directly affect the practical value of the object tracking algorithm. In that case, the introduction of L1-norm minimization into the field of object tracking would greatly reduce the performance of tracking algorithms.

We note that the tracking accuracy and speed are two important aspects for evaluating the performance of object tracking algorithms. Therefore, in this paper, we develop an MLE-based model that improves the traditional sparse coding model from the two aspects and then apply it to achieve an effective and efficient tracker.

In the field of object tracking, accuracy is the most important issue. Hence, at first, we need to improve the accuracy constraint term in the traditional sparse coding model.

When the reconstruction error $\mathbf{e} = y - T\widehat{\alpha}$ follows the Gaussian distribution, the traditional sparse coding solution can be written as

$$\widehat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\{ \left\| (y - T\alpha) \right\|_2^2 + \lambda \|\alpha\|_1 \right\}, \tag{5}$$

where λ is a regularization parameter. For object tracking, the dictionary $\mathbf{T} = [t_1, t_2, \ldots, t_n] \in R^{d \times n}$ consists of n templates and forms the object template library.consider $t_i \in R^d$, $d \gg n$. In our experiments, we make $n = 20$ and the object template size 32×32; that is, $d = 1024$. The result image block in current frame is denoted as y, $y \in \mathbb{R}^d$. $\alpha = (\alpha_1, \alpha_2, \ldots, \alpha_n)^t \in R^d$ denotes the coefficient vector of sparse coding. Equation (5) is obviously a minimum variance estimation problem with sparsity constraint. When object's reconstruction error $\mathbf{e} = y - T\widehat{\alpha}$ follows the Gaussian distribution, the solution of (5) is the maximum likelihood estimation.

However, in practical applications, when the object suffers from occlusion, rotation change, scale change, or illumination variation, the reconstruction errors $\mathbf{e}$ of abnormal pixels will not follow the Gaussian distribution. In that case, these algorithms may not track the object accurately. Therefore, we need to build a more adaptive object representing model.

First, we rewrite the dictionary $\mathbf{T}$ as $\mathbf{T} = [r_1; r_2; \ldots r_d]$, where row vector $r_j \in R^n$, $j = 1, 2, \ldots, d$, is the jth row of $\mathbf{T}$. Meanwhile, we rewrite tracking result image block $\mathbf{y}$ as $\mathbf{y} = [y_1; y_2; \ldots; y_d]$, where y_j, $j = 1, 2, \ldots, d$, is the jth pixel of $\mathbf{y}$. In that case, the reconstruction error $\mathbf{e} = y - T\widehat{\alpha} = [e_1; e_2; \ldots; e_d]$, where $e_j = y_j - r_j\widehat{\alpha}$, $j = 1, 2, \ldots, d$, is the jth pixel's reconstruction error.

Assume that $e_1, e_2, \ldots, e_n$ are independently and identically distributed according to a certain probability density function $p_\theta(e_j)$, where θ denotes the parameter set that characterizes the distribution. Then the likelihood function would be $L_\theta(e_1, e_2, \ldots, e_d) = \prod_{j=1}^n p_\theta(e_j)$, and MLE aims to maximize this likelihood function or, equivalently, minimize the objective function: $-\ln L_\theta(e_1, e_2 \ldots, e_d) = \sum_{j=1}^n f_\theta(e_j)$, where $f_\theta(e_j) = -\ln p_\theta(e_j)$, to simplify the computation.

Taking into account the sparsity constraint of α, the MLE of α can be formulated as the following minimization:

$$\widehat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\{ \sum_{j=1}^{d} f_\theta\left(e_j\right) + \lambda \|\alpha\|_1 \right\}. \tag{6}$$

According to [6], formula (6) can be converted into weighted sparse coding problem

$$\widehat{\alpha} = \underset{\alpha}{\operatorname{argmin}} \left\{ \left\| W^{1/2} (y - T\alpha) \right\|_2^2 + \lambda \|\alpha\|_1 \right\}, \tag{7}$$

where W is a diagonal matrix with diagonal elements as follows:

$$W_{j,j} = \frac{\exp\left(\mu\sigma - \mu e_j^2\right)}{\left(1 + \exp\left(\mu\sigma - \mu e_j^2\right)\right)}, \tag{8}$$

which also stands for the jth pixel's weight value. μ and σ are positive constants. If we make $W_{j,j} = 2$, then the model would be the traditional sparse coding problem. Hence, we can see that formula (7) is more adaptive than (3).

In this study, we choose it as the weight function

$$W_{j,j} = \frac{1}{1 + 1/\exp\left(-\beta e_j\right)}, \tag{9}$$

where β is a scale factor (we choose $\beta = 10$ in our experiments). The physical meaning of $W_{j,j}$ is to allocate smaller weights to those pixels with bigger residuals (probably abnormal pixels) and allocate bigger weights to pixels with smaller residuals. By setting a reasonable weight threshold, we can get rid of those abnormal pixels lower than the threshold and do further sparse coding. In that case, we can effectively reduce the effect of abnormal pixels and therefore achieve good performance during the tracking processing.

From (9), we can see that the weight value $W_{j,j}$ is bounded between 0 and 1 which makes sure that even the pixels with very small residuals would not have too large weight values. This would guarantee the stability of the algorithm.

3. Bayesian MAP Estimation

We can regard object tracking as a hidden state variables' Bayesian MAP estimation problem in the Hidden Markov model; that is, with a set of observed samples $Y_t = \{y_1, y_2, \ldots, y_t\}$, we can estimate the hidden state variable x_t using Bayesian MAP theory.

According to the Bayesian theory,

$$p\left(x_t \mid Y_t\right) \propto p\left(y_t \mid x_t\right) \int p\left(x_t \mid x_{t-1}\right) p\left(x_{t-1} \mid Y_{t-1}\right) dx_{t-1}, \tag{10}$$

where $p(x_t \mid x_{t-1})$ stands for a state transition model for two consecutive frames and $p(y_t \mid x_t)$ stands for an observation likelihood model. We can obtain the object's best state in tth frame through maximum posterior probability estimation; that is,

$$\widehat{x_t} = \underset{x_t^l}{\operatorname{argmax}}\, p\left(x_t^l \mid Y_t\right), \quad l = 1, 2, \ldots, N, \tag{11}$$

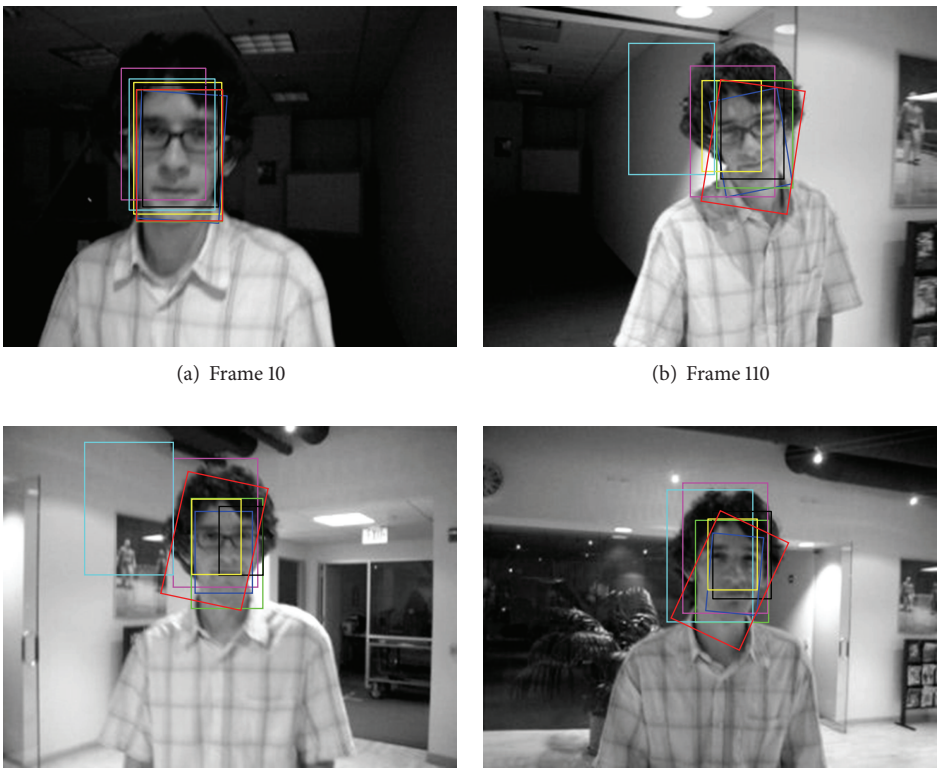
where x_t^l stands for the lth sample of state variable x_t in tth frame. In this paper, we choose $N = 400$.

3.1. State Transition Model. We choose object's motion affine transformation parameters as state variable $x_t = \{x_t, y_t, \theta_t, S_t, \alpha_t, \phi_t\}$, where x_t and y_t, respectively, represent the x-direction and y-direction translation of the object in tth

(a) Frame 170 (b) Frame 340

(c) Frame 530 (d) Frame 650

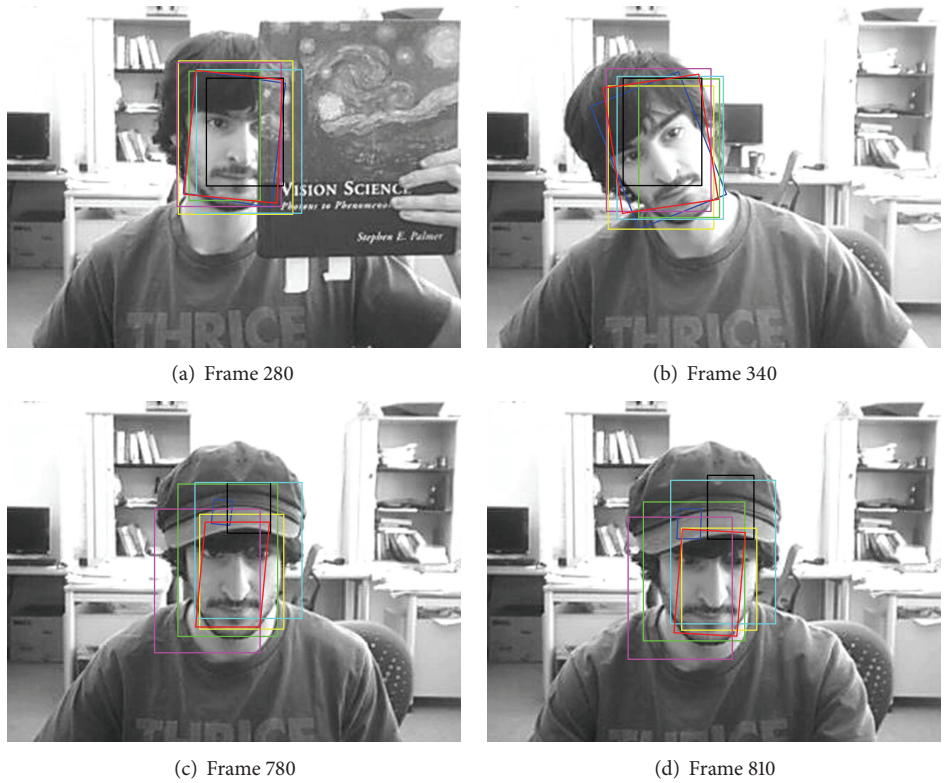
Figure 1: Tacking results of test video "Car4."



(a) Frame 10 (b) Frame 110

(c) Frame 220 (d) Frame 310

Figure 2: Tacking results of test video "David."



(a) Frame 280 (b) Frame 340

(c) Frame 780 (d) Frame 810

Figure 3: Tacking results of test video "Faceocc2."

frame, θ_t stands for the rotation angle, S_t represents the scale change, α_t stands for the aspect ratio, and ϕ_t stands for the direction of tilt.

We assume that the state transition model follows the Gaussian distribution; that is,

$$p(x_t \mid x_{t-1}) = N(x_t; x_{t-1}, \Psi), \tag{12}$$

where Ψ is a diagonal matrix whose diagonal elements are motion affine parameter's variation $\sigma_x^2, \sigma_y^2, \sigma_\theta^2, \sigma_S^2, \sigma_\alpha^2, \sigma_\phi^2$.

3.2. Observation Likelihood Model. We use object's reconstruction error to build observation likelihood model; that is,

$$p(y_t \mid x_t) = \prod_{j=1,2,\ldots,d} \mathbb{N}\left(e_j^t, \mu, \sigma^2\right), \tag{13}$$

where $\mathbb{N}(\cdot)$ means Gaussian distribution, μ and σ^2, respectively, represent the mean and variation of Gaussian distribution, d stands for the number of pixels of an object template, and $e_j^t = \| y_j^t - \Phi_j^t \widehat{\gamma}_j^t \|_2$ stands for the reconstruction error of jth pixel of object templates in tth frame.

3.3. Templates Updating. To consider that the appearance of the target may change during the tracking processing, it is necessary to dynamically update the template library.

In this paper, we use a method named "Half Updating Strategy" to update the templates. We take the tracking results of first n frames as the initial templates, and from $(n+1)$th frame on, we use the algorithm mentioned earlier to obtain and save the tracking results. During this process, if the result image block has equal to 50% abnormal pixels, then we do not update the tracker. When we have $n/2$ tracking results, that is, half of the number of initial templates, we replace the first $n/2$ templates in original template library with the newly accumulated $n/2$ tracking results. Then a "Half Updating" is finished.

Table 1: The description of test videos.

Name of test videos	Number of frames	Video description
Car4	659	Illumination variation and scale variation
David	462	Illumination variation, in-plane rotation, and off-plane rotation
Faceocc2	819	Partly occlusion, in-plane rotation, and off-plane rotation

4. Experimental Results and Analysis

In order to evaluate the performance of our tracker, we conduct experiments on three challenging image sequences (Table 1 and Figures 1, 2, and 3). These sequences cover most challenging situations in object tracking: occlusion,

motion blur, in-plane and out-of-plane rotation, large illumination change, scale variation, and complex background. For comparison, we run six state-of-the-art algorithms with the same initial position of the target. These algorithms are the Frag tracking [18], IVT tracking [19], MIL tracking [20], L1 tracking [6], PN tracking [21], and VTD tracking [22] methods. We present some representative results in this section.

5. Conclusions/Outlook

This paper presents a robust tracking algorithm via 2DPCA and MLE. In this work, we represent the tracked object by using 2DPCA bases and an MLE error matrix. With the proposed model, we can remove the abnormal pixels and thus reduce the effect of abnormal pixels on tracking algorithms. We take the object's reconstruction error into the Bayesian maximum posterior probability estimation framework and design a stable and robust tracker. Then, we explicitly take partial occlusion and misalignment into account for appearance model update and object tracking. Experiments on challenging video clips show that our tracking algorithm performs better than several state-of-the-art algorithms. Our future work will be the generalization of our representation model into other related fields.

Acknowledgments

This research described in this paper was supported by the Fundamental Research Funds for the Central Universities (DC110321, DC120101132, and DC120101131). This work was supported by Project of Liaoning Provincial Department of Education (L2012476, and L2010094). This work was supported by National Natural Science Foundation of China (61172058).

References

[1] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: a survey," *ACM Computing Surveys*, vol. 38, no. 4, pp. 229–240, 2006.

[2] M. X. Jiang, Z. J. Shao, and H. Y. Wang, "Real-time object tracking algorithm with cameras mounted on moving platforms," *International Journal of Image and Graphics*, vol. 12, no. 3, Article ID 1250020, 2012.

[3] M. Jiang, M. Li, and H. Wang, "A robust combined algorithim of object tracking based on moving object detection," in *Proceedings of the International Conference on Intelligent Control and Information Processing (ICICIP '10)*, pp. 619–622, Dalian, China, July 2010.

[4] L. Zhang, P. Zhu, Q. Hu, and D. Zhang, "A linear subspace learning approach via sparse coding," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 755–761, Barcelona, Spain, November 2011.

[5] M. X. Jiang, M. Li, and H. Y. Wang, "Object tracking algorithm based on projection matrix," *Journal of Convergence Information Technology*, vol. 7, pp. 209–217, 2012.

[6] X. Mei and H. Ling, "Robust visual tracking and vehicle classification via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 11, pp. 2259–2272, 2011.

[7] 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.

[8] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[9] J. H. Yin, C. Y. Fu, and J. K. Hu, "Using incremental subspace and contour template for object tracking," *Journal of Network and Computer Applications*, vol. 35, pp. 1740–1748, 2012.

[10] D. Wang, H. Lu, and X. Li, "Two dimensional principal components of natural images and its application," *Neurocomputing*, vol. 74, no. 17, pp. 2745–2753, 2011.

[11] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ1-norm solution is also the sparsest solution," *Communications on Pure and Applied Mathematics*, vol. 59, no. 6, pp. 797–829, 2006.

[12] 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.

[13] A. Wagner, J. Wright, A. Ganesh, Z. Zhou, H. Mobahi, and Y. Ma, "Toward a practical face recognition system: robust alignment and illumination by sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 2, pp. 372–386, 2012.

[14] Y. Meng, L. Zhang, J. Yang, and D. Zhang, "Robust sparse coding for face recognition," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 625–632, Colorado Springs, Colo, USA, June 2011.

[15] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: which helps face recognition?" in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 471–478, Barcelona, Spain, November 2011.

[16] C. Chiang, C. Duan, S. Lai, and S. Chang, "Learning component-level sparse representation using histogram information for image classification," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 1519–1526, Barcelona, Spain, November 2011.

[17] W. Dong, X. Li, L. Zhang, and G. Shi, "Sparsity-based image denoising via dictionary learning and structural clustering," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 457–464, Barcelona, Spain, November 2011.

[18] A. Adam, E. Rivlin, and I. Shimshoni, "Robust fragments-based tracking using the integral histogram," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '06)*, pp. 798–805, New York, USA, June 2006.

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

[20] B. Babenko, M. H. Yang, and S. Belongie, "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, Miami, Fla, USA, June 2009.

[21] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: bootstrapping binary classifiers by structural constraints," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 49–56, San Francisco, Calif, USA, June 2010.

[22] J. Kwon and K. M. Lee, "Visual tracking decomposition," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR '10)*, pp. 1269–1276, San Francisco, Calif, USA, June 2010.