Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2013, Article ID 352634, 7 pages
http://dx.doi.org/10.1155/2013/352634

*Research Article*

# Robust Online Object Tracking Based on Feature Grouping and 2DPCA

**Ming-Xin Jiang, Jun-Xing Zhang, and Min Li**

*College of Information & Communication Engineering, Dalian Nationalities University, Dalian 116600, China*

Correspondence should be addressed to Jun-Xing Zhang; zhangjunxing@dlnu.edu.cn

Received 20 March 2013; Revised 7 May 2013; Accepted 13 May 2013

Academic Editor: Shangbo Zhou



We present an online object tracking algorithm based on feature grouping and two-dimensional principal component analysis (2DPCA). Firstly, we introduce regularization into the 2DPCA reconstruction and develop an iterative algorithm to represent an object by 2DPCA bases. Secondly, the object templates are grouped into a more discriminative image and a less discriminative image by computing the variance of the pixels in multiple frames. Then, the projection matrix is learned according to the more discriminative image and the less discriminative image, and the samples are projected. The object tracking results are obtained using Bayesian maximum a posteriori probability estimation. Finally, we employ a template update strategy which combines incremental subspace learning and the error matrix to reduce tracking drift. 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

Online object tracking is a fundamental problem in many computer vision applications such as surveillance, driver assistance systems, and human-computer interactions [1–4]. Although researchers have made great progresses in this area, object tracking remains a challenging problem due to the difficulty arising from the appearance variability of an object. Intrinsic and extrinsic changes inevitably cause large appearance variation. Due to the nature of the tracking problem, an effective appearance model is of prime importance for the success of a tracking algorithm [5–8].

In recent years, as a popular dimensionality reduction and feature extraction technique, linear subspace learning has been successfully used in robust visual tracking. Supervised discriminative methods for classification and regression have also been exploited to solve visual tracking problems. For example, Avidan developed a tracking algorithm that employs the support vector machine (SVM) classifier within an optic flow framework [9]. Along similar lines, Williams et al. developed a method in which an SVM-based regressor was used for tracking [10]. As a result of training the regressor on in-plane image motion, this method is not effective in tracking objects with out-of-plane movements. In [11], the author combines the sparse coding and Kalman filtering together and chooses the color histogram and gradient histogram as features to track the object. The template updating strategy of the algorithm is to replace a random template of the original template library with the last tracking result. This updating strategy can easily introduce tracking errors when abnormal changes happen to the object, which would result in failure of tracking. A method by casting object tracking as a sparse approximation problem in a particle filter framework is proposed in [12]. The author solves the problem of object occlusion through the introduction of trivial template which, however, extends the number of templates greatly and increases the computational complexity. In that case, the practical value of algorithm is greatly reduced.

Motivated by the abovementioned discussions, we propose an online object tracking algorithm based on feature grouping and 2DPCA. Firstly, we introduce regularization into the 2DPCA reconstruction and develop an iterative


(a) (b) (c) (d) (e) (f)

Figure 1: Tracking results of test video "Lemming."

algorithm to represent an object by 2DPCA bases. Secondly, the object templates are grouped into a more discriminative image and a less discriminative image by computing the variance of the pixels in multiple frames. Then, the projection matrix is learned according to the more discriminative image and the less discriminative image, and the samples are projected using the projection matrix. The object tracking results are obtained using Bayesian maximum a posteriori probability (MAP) estimation (Figures 1, 2, and 3). 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 proposed method has strong robustness to abnormal changes.

## 2. Robust Online Object Tracking Based on Feature Grouping and 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 [13]). 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.

Given a series of image matrices $\mathbf{Y} = [\mathbf{Y_1}; \mathbf{Y_2}; \ldots; \mathbf{Y_d}]$, 2DPCA aims to obtain an orthogonal left-projection matrix


(a) (b) (c) (d) (e) (f)

Figure 2: Tracking results of test video "Car4."

$\mathbf{U}$, an orthogonal right-projection matrix $\mathbf{V}$, and the projection coefficients $\mathbf{A} = [\mathbf{A_1}; \mathbf{A_2}; \ldots; \mathbf{A_d}]$ by solving the following objective function:

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

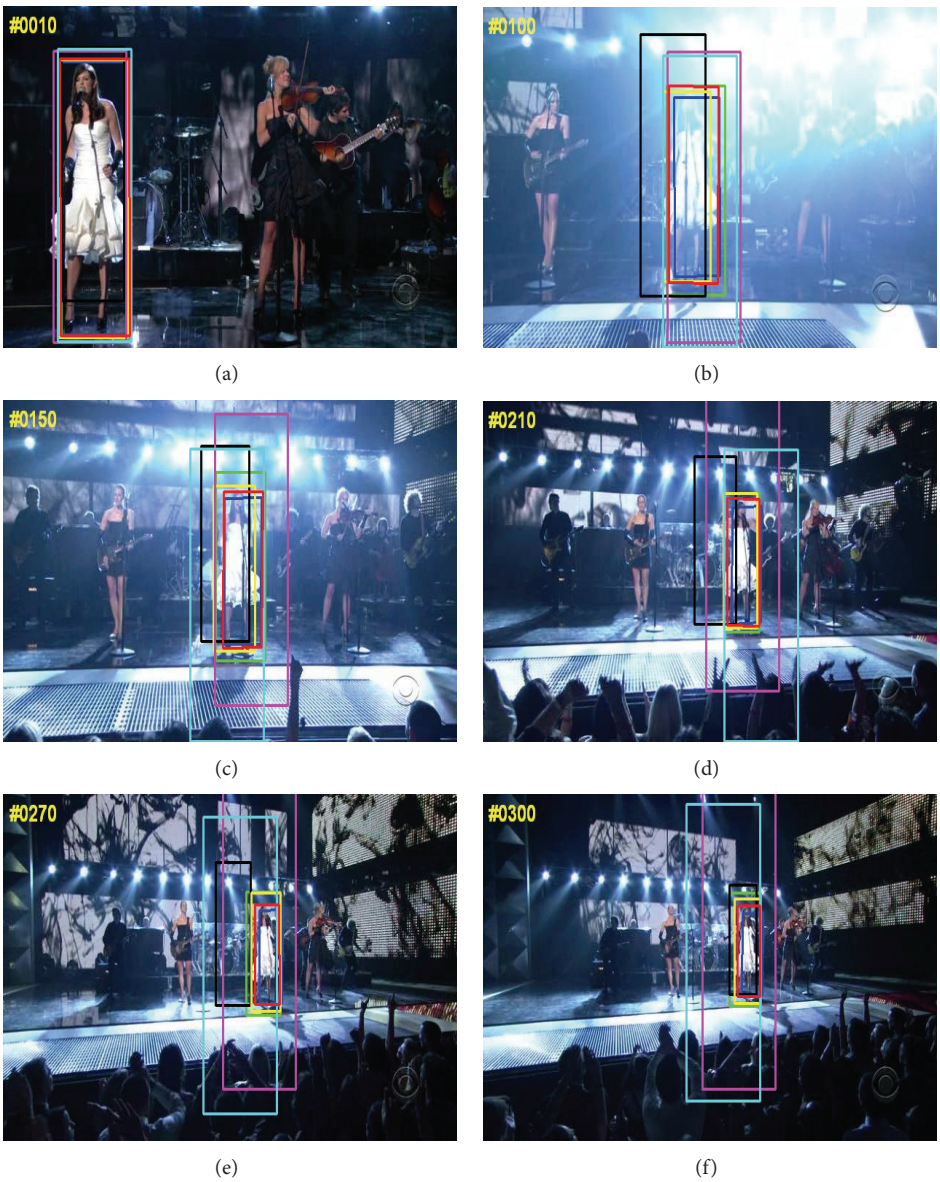
Then the coefficient $\mathbf{A}_i$ can be approximated by $\mathbf{A}_i \approx \mathbf{U}'\mathbf{Y}_i\mathbf{V}$. We note that the underlying assumption of (6) is that the error term is Gaussian distributed with small variances. This assumption is not able to deal with partial occlusion as the error 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

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




(a) (b) (c) (d) (e) (f)

Figure 3: Tracking results of test video "Singer1."

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

*2.2. Feature Grouping.* In object tracking problem, $\mathbf{T} = [\mathbf{t_1}, \mathbf{t_2}, \ldots, \mathbf{t_n}] \in \mathbb{R}^{d\times n}$ consists of $n$ templates and forms the object template library, where $\mathbf{t}_i \in \mathbb{R}^d$, $d \gg n$. In our experiments, we make $n = 10$ and the object template size $32 \times 32$; that is, $\mathbf{d} = 1024$. The result image block in current frame is denoted as $\mathbf{y}$, $\mathbf{y} \in \mathbb{R}^d$, which has the same size as object template $\mathbf{t}_i$.

We would like to decompose each template $\mathbf{t}_i$, $i = 1, 2, \ldots, n$, into a more discriminative image $\mathbf{t}_i^a$ and a less discriminative image $\mathbf{t}_i^b$. First, we rewrite $\mathbf{t}_i$ as $t_{i,j}$, $j = 1, 2, \ldots d$, and compute the variance of each pixel in each template. if the pixel $t_{i,j}$ has a smaller variance, we can think that this pixel is more stable. Based on this heuristic, we can decompose those templates into a more discriminative image and a less discriminative image.

Denote by $\bar{t}_j$ the mean of the templates $t_{i,j}$. The variance of the $j$th pixel in template $\mathbf{t}_{i,j}$ is

$$\sigma_j = \frac{1}{d}\sum_{j=1}^{d}\left(t_{i,j} - \bar{t}_j\right). \tag{4}$$

We want to put those pixels having smaller $\sigma_j$ into a more discriminative image and the remaining into a less discriminative image $\mathbf{t}_i^b$.

After decomposing, each template can be written as $\mathbf{t}_i = \mathbf{t}_i^a + \mathbf{t}_i^b$, $i = 1, 2, \ldots, n$, and we have $\mathbf{T} = \mathbf{T}_a + \mathbf{T}_b$. However, if we learn the projection matrix $\mathbf{P}$ only based on

$\mathbf{T}_a$, the result cannot be very satisfying because $\mathbf{T}_b$ is also useful for determining the projection direction. To effectively exploit the information in both $\mathbf{T}_a$ and $\mathbf{T}_b$, we propose to learn a subspace where the energy in $\mathbf{T}_a$ is well preserved and the energy in $\mathbf{T}_b$ is suppressed. Denote by $\bar{\mathbf{t}}$, $\bar{\mathbf{t}}^a$, and $\bar{\mathbf{t}}^b$ the mean vectors of $\mathbf{T}$, $\mathbf{T}_a$, and $\mathbf{T}_b$, respectively, and let $\bar{\mathbf{t}}_i = \mathbf{t}_i - \bar{\mathbf{t}}$, $\bar{\mathbf{t}}_i^a = \mathbf{t}_i^a - \bar{\mathbf{t}}^a$, and $\bar{\mathbf{t}}_i^b = \mathbf{t}_i^b - \bar{\mathbf{t}}^b$ be the centralized image vectors. Accordingly we have the centralized datasets $\overline{\mathbf{T}}$, $\overline{\mathbf{T}}_a$, and $\overline{\mathbf{T}}_b$. Clearly, we have $\bar{\mathbf{t}} = \bar{\mathbf{t}}^a + \bar{\mathbf{t}}^b$, $\bar{\mathbf{t}}_i = \bar{\mathbf{t}}_i^a + \bar{\mathbf{t}}_i^b$, and $\mathbf{P}\bar{\mathbf{t}}_i = \mathbf{P}\bar{\mathbf{t}}_i^a + \mathbf{P}\bar{\mathbf{t}}_i^b$. After projection, the average energy of $\mathbf{P}\bar{\mathbf{t}}_i^a$ is

$$E_a = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{P}\bar{\mathbf{t}}_i^a\right\|_2^2 = \frac{1}{n}\sum_{i=1}^{n}\left(\mathbf{P}\bar{\mathbf{t}}_i^a\right)^T\left(\mathbf{P}\bar{\mathbf{t}}_i^a\right) = \operatorname{tr}\left\{\mathbf{P}\left(\frac{1}{n}\overline{\mathbf{T}}_a\overline{\mathbf{T}}_a^T\right)\mathbf{P}^T\right\} = \operatorname{tr}\left\{\mathbf{P}\mathbf{S}_\mathbf{a}\mathbf{P}^T\right\}, \tag{5}$$

where $\mathbf{S}_\mathbf{a} = (1/n)\overline{\mathbf{T}}_a\overline{\mathbf{T}}_a^T$ is the total scatter matrix of $\mathbf{T}_a$ and "tr" is the matrix trace operator.

Similarly, the average energy of $\mathbf{P}\bar{\mathbf{t}}_i^b$ is

$$E_b = \frac{1}{n}\sum_{i=1}^{n}\left\|\mathbf{P}\bar{\mathbf{t}}_i^b\right\|_2^2 = \operatorname{tr}\left\{\mathbf{P}\mathbf{S}_b\mathbf{P}^T\right\}, \tag{6}$$

where $\mathbf{S}_b = (1/n)\overline{\mathbf{T}}_b\overline{\mathbf{T}}_b^T$ is the total scatter matrix of $\mathbf{T}_b$.

To preserve the $\mathbf{T}_a$ while suppressing the $\mathbf{T}_b$, we seek for a projection matrix $\mathbf{P}$ to maximize the energy $E_a$ while minimizing the energy $E_b$ by solving the following optimization problem:

$$J_P = \arg\max_P \frac{E_a}{E_b} = \arg\max_P \frac{\operatorname{tr}\left\{\mathbf{P}\mathbf{S}_\mathbf{a}\mathbf{P}^T\right\}}{\operatorname{tr}\left\{\mathbf{P}\mathbf{S}_b\mathbf{P}^T\right\}}. \tag{7}$$

An equivalent form of (4) is

$$J_P = \arg\max_P \frac{E_a}{E_b} = \arg\max_P \operatorname{tr}\left(\mathbf{P}\mathbf{S}_\mathbf{a}\mathbf{P}^T\right) \quad \text{s.t. } \mathbf{P}\mathbf{S}_\mathbf{a}\mathbf{P}^T = \mathbf{I}. \tag{8}$$

Apparently, the desired $\mathbf{P}$ can be computed by using generalized eigenvalue decomposition; that is, matrix $\mathbf{P}$ is composed of the generalized eigenvectors of $\mathbf{S}_\mathbf{a}w = \lambda\mathbf{S}_b w$ corresponding to the $p$ largest eigenvalues.

## 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,

$$p\left(x_t \mid Y_t\right) \propto p\left(y_t \mid x_t\right)\int p\left(x_t \mid x_{t-1}\right)p\left(x_{t-1} \mid Y_{t-1}\right)dx_{t-1}, \tag{9}$$

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 $t$th frame through maximum a posteriori probability estimation; that is,

$$\hat{x}_t = \arg\max_{x_t^l} p\left(x_t^l \mid Y_t\right), \quad l = 1, 2, \ldots, N, \tag{10}$$

where $x_t^l$ stands for the $l$th sample of state variable $x_t$ in $t$th 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 $t$th frame, $\theta_t$ stands for the rotation angle, $S_t$ represents the scale change, $\alpha_t$ stands for the aspect ratio, and $\phi_t$ stands for the direction of tilt.

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

$$p\left(x_t \mid x_{t-1}\right) = N\left(x_t; x_{t-1}, \Psi\right), \tag{11}$$

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

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

$$p\left(y_t \mid x_t\right) = \prod_{j=1,2,\ldots,d} \mathbb{N}\left(e_j^t, \mu, \sigma^2\right), \tag{12}$$

where $\mathbb{N}(\cdot)$ means Gaussian distribution, $\mu$ and $\sigma^2$, respectively, represent the mean and variation of Gaussian distribution, $d$ stands for the number of pixels of an object template, and $e_j^t$ stands for the reconstruction error of $j$th pixel of object templates in $t$th frame.

## 4. Experimental Results and Analysis

In order to show the robustness of the object tracking algorithm based on projection discussed in this paper, we choose several sets of public test videos taken under different environments to test the performance of our algorithm (Figures 4, 5, and 6). Given the limited space, in this section we only list three of them to show the tracking results and error curves. Different abnormal changes have happened to the moving objects in chosen test videos such as occlusion, rotation, illumination variation, or scale variation (Table 1).

The implementation of the algorithm is based on Windows operating system. The configuration of the computer is AMD Athlon (TM) X2 Dual Core QL-62 2.00 GHz, 1.74 GB memory. In order to evaluate the performance of the algorithm, we choose six currently most representative and classic tracking algorithms to do the comparison. The six classic algorithms are L1 Tracker [8], IVT Tracker [14], PN

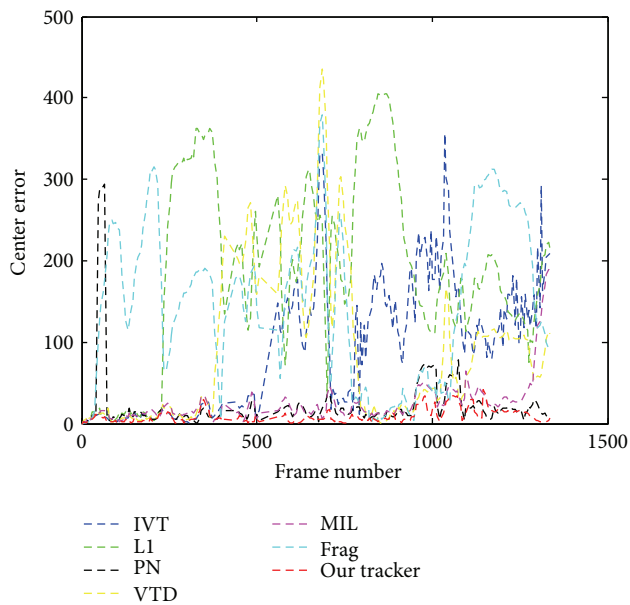



Figure 4: Quantitative evaluation in terms of center location error for test video "Lemming."

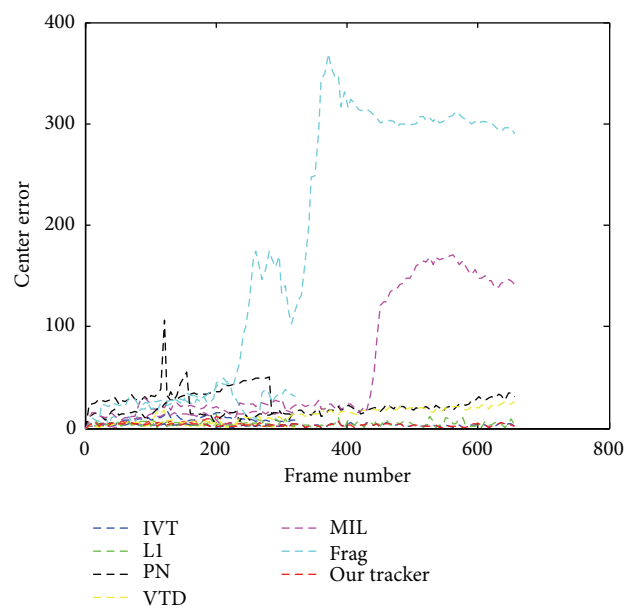



Figure 5: Quantitative evaluation in terms of center location error for test video "Car4."

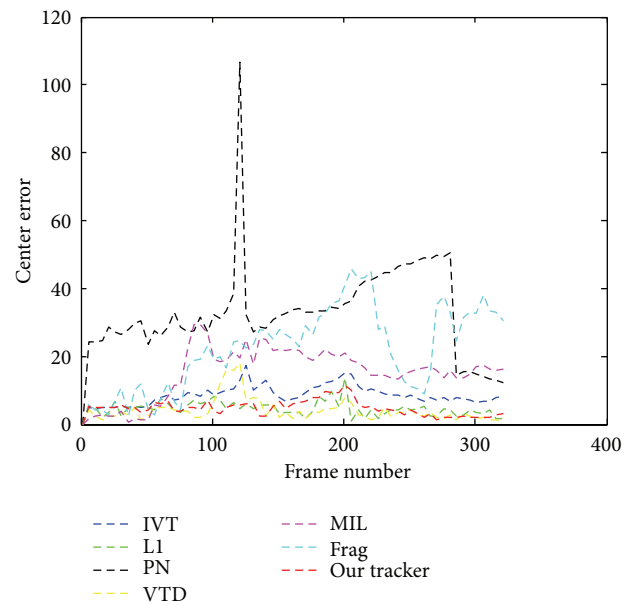



Figure 6: Quantitative evaluation in terms of center location error for test video "Singer1."

Table 1: The description of test videos.

| Name of test videos | Number of frames | Video description |
|---|---|---|
| Lemming | 1336 | Out-of-plane rotation, scale change, occlusion, and background cluster |
| Car4 | 659 | Illumination variation and scale variation |
| Singer1 | 351 | Illumination variation and scale change |

Table 2: The average processing speed of our method.

| Name of test videos | The average processing speed (Frame/S) |
|---|---|
| Lemming | 32.57 |
| Car4 | 35.13 |
| Singer1 | 29.98 |

Tracker [15], VTD Tracker [16], MIL Tracker [17], and Frag Tracker [18].

The average processing speeds of our method for different test videos are listed in Table 2. The statistical data is obtained using Opencv2.2. Table 2 shows that the proposed algorithm can meet the real-time performance.

## 5. Conclusions

This paper presents a robust tracking algorithm via feature grouping and 2DPCA. In this work, we represent the tracked object by using 2DPCA bases and a feature grouping. With the proposed model, we can reduce the effect of abnormal pixels on tracking algorithms. We obtained the tracking result using Bayesian maximum a posteriori probability estimation framework and designed 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) and 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, Article ID 1250020, 12 pages, 2012.

[3] M. X. Kristan, S. Kovačič, A. Leonardis, and J. Perš, "A two-stage dynamic model for visual tracking," *IEEE Transactions on Systems, Man, and Cybernetics B*, vol. 40, no. 6, pp. 1505–1520, 2010.

[4] M. X. Jiang, M. Li, and H. Y. 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, August 2010.

[5] F. Su, G. Fang, and N. M. Kwok, "Adaptive colour feature identification in image for object tracking," *Mathematical Problems in Engineering*, vol. 2012, Article ID 509597, 18 pages, 2012.

[6] P. Han, J. Du, and M. Fang, "Spatial object tracking using an enhanced mean shift method based on perceptual spatial-space generation model," *Journal of Applied Mathematics*, vol. 2013, Article ID 420286, 13 pages, 2013.

[7] A. M. Sarhan, A. I. Saleh, and R. K. Elsadek, "A reliable event-driven strategy for real-time multiple object tracking using static cameras," *Advances in Multimedia*, vol. 2011, Article ID 976463, 20 pages, 2011.

[8] 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.

[9] S. Avidan, "Support vector tracking," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 8, pp. 1064–1072, 2004.

[10] O. Williams, A. Blake, and R. Cipolla, "A sparse probabilistic learning algorithm for real-time tracking," in *Proceedings of the 8th IEEE International Conference on Computer Vision*, pp. 353–360, October 2003.

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

[12] 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.

[13] 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.

[14] 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.

[15] 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.

[16] 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.

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

[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, NY, USA, June 2006.