ROBUST VISUAL TRACKING BASED ON SPARSE PCA-L1

Yuanyuan Zhang[1] and Fuxiang Wang[2]

[1]National Key Lab of CNS/ATM, School of Electronic and Information
Engineering, Beihang University, Beijing, China
buaa_zyy@163.com
[2]National Key Lab of CNS/ATM, School of Electronic and Information
Engineering, Beihang University, Beijing, China
wangfx@buaa.edu.cn

ABSTRACT

Recently, visual tracking based on sparse principle component analysis has drawn much research attention. As we all know, principle component analysis (PCA) is widely used in data processing and dimensionality reduction. But PCA is difficult to interpret in practical application and all those principal components are linear combinations of all variables. In our paper, a novel visual tracking method based on sparse principal component analysis and L1 tracking is introduced, which we named the method SPCA-L1 tracking. We firstly introduce trivial templates of L1 tracking method, which are used to describe noise, into PCA appearance model. Then we use lasso model to achieve sparse coefficients. Then we update the eigenbasis and mean incrementally to make the method robust when solving different kinds of changes of the target. Numerous experiments, where the targets undergo large changes in pose, scale and illumination, demonstrate the effectiveness and robustness of the proposed method.

KEYWORDS

Visual tracking, sparse principal component analysis, particle filter

1. INTRODUCTION

Visual tracking is a very important part in computer vision field. The applications of surveillance, vehicle navigation, medical diagnostic, virtual reality and human computer interface are all based on visual tracking. And it is a hard problem because there are many different and varying circumstances that have to be considered in tracking algorithm, such as illumination variation, occlusion, pose changes, and background clutters.

There are two major categories of tracking methods now, discriminative and generative methods, used in current tracking techniques. Discriminative online learning methods [1] treat object tracking as a classification problem. Generative online learning methods are adopted to track an object by searching for region most similar to the target model. There are three main tracking strategies in generative online learning area. The first strategy is about tracking method based on templates, in which dynamic and multi-feature templates are very important to the tracking. The

David C. Wyld et al. (Eds) : CSEN, SIPR, NCWC - 2016
pp. 125–135, 2016. © CS & IT-CSCP 2016 DOI : 10.5121/csit.2016.61011

second one is method based on sparse representation, in which most of the appearance information about the target is represented by a linear combination of only a few basis vectors. And the third one is method based on subspace analysis, which we are studying about. It includes the methods based on Non-negative matrix factorization (NMF), methods based on Kernel, and the ones based on principal component analysis (PCA).

Recently, the PCA draws more and more attention. There are many works following PCA. Ross et al. [4] proposed an adaptive probabilistic tracking approach to update the models of a target by means of incremental eigenbasis updates. Kwon and Lee [10] apply sparse PCA to formulate tracking. Under the subspace assumption, Sui and Zhang[11] propose a locally structured Gaussian Process and cast tracking as a regression problem. These methods based on PCA provide a compact representation of the target, which is efficient in feature extraction and removing redundant information. At the same time, the probabilistic model facilitates efficient computation.

But subspaces learning is essentially sensitive to occlusion. For this reason, some methods for occlusion handling need to be used with subspace learning together. Wang et al. [12] use the subspace model to represent the target and impose sparsity on the residual errors to deal with occlusions.

In this paper, under the framework of particle filter, we proposed a new tracking method based on sparse principal component analysis, which can work more robustly, especially with the occlusion. Our contributions are as follows:

1) Reconstruct new target models with PCA subspace learning and trivial templates, so we can, at the same time, describe the candidates and noise, the trivial templates can deal with the noise, like occlusion, during the tracking;

2) Use lasso model to get the sparse coefficient of principle components;

3) Incrementally learn and update the low-dimensional subspace representation, including correctly update the eigenbasis.

The remaining part of this paper is organized as follows. In Section 2, we review some relevant approaches that motivated our work. The details of our method are described in Section 3. Experimental results are reported in Section 4. We conclude this paper in Section 5.

2. RELATED WORK

In this section, we will briefly introduce the tracking method IVT (Incremental Visual Tracking) [4, 5] and L1 minimization visual tracking [7]. Both of the methods are in particle framework. The IVT is a traditional tracking method based on incremental subspace learning. It learns and updates the low-dimensional subspace representation of the targets. In order to estimate the locations of the target in the new coming frames, IVT predicts the candidates by using a sampling algorithm with likelihood estimates instead of gradient descent. It performs well with the variation of the target and surrounding illumination. But when the target is occluded by other things, it may drift away and miss the target in the end. And L1 minimization is a method by casting tracking as a sparse approximation problem. It introduces a set of trivial templates to deal with the occlusion problem. The trivial templates are used to capture the noise and occlusion. The L1 tracking is robust with the noise and partly occlusion. But the tracking speed is slowly and it is not real-time.

Our work is motivated by the IVT method and L1 minimization method. We use PCA to reconstruct the target templates, so the templates are orthometric with each other and need smaller storage space. At the same time, we introduce the trivial templates to handle the noise. And we choose the candidate with smallest residual as the new target. Then we update the subspace by incrementally updating the eigenbasis.

3. OUR WORK

In this section, the details of our tracking method based on sparse PCA-L1 will be introduced. Our tracking method is conducted within the particle filtering framework. The main parts of our method are as follows.

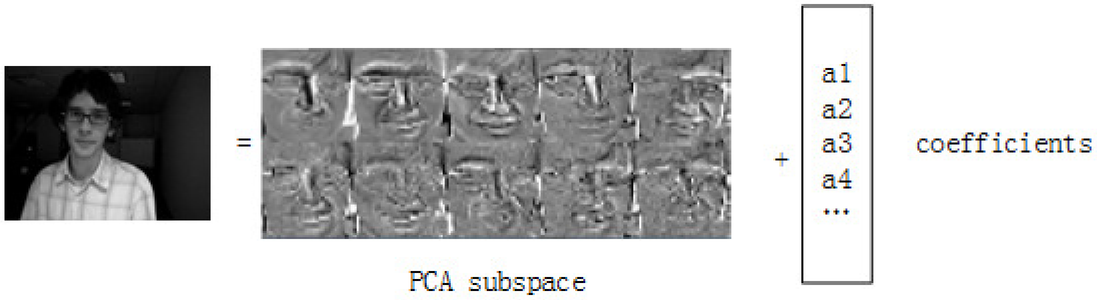
A. New target models subspace and trivial templates

We want to reconstruct new target models involving PCA subspace and trivial templates, so we can describe the candidates and noise at the same time. The trivial templates can deal with the noise, like occlusion, during the tracking at the same time. The Fig.1 shows the difference between the models of L1 tracking, IVT tracking and our tracking method, Sparse PCA-L1. In Fig.1 (a), we can see the subspace representation of the target object in IVT tracking. Fig.1 (b) shows the models of L1 minimization tracking, which includes trivial templates to capture the occlusion. Our tracking models are represented in Fig.1 (c). At initialization, we manually select the first target template from the first frame; then apply zero mean unit norm normalization, it is the first template. The rest templates are created by moving one pixel in four possible directions at the corner points of the first template in the first frame. These are the basic models. Then we use PCA to learn those basic models, making them orthotropic to each other. Meanwhile we introduce trivial templates to represent the occlusion. Thus our models need less space than L1 tracking models and more robust than IVT when the target is occluded.

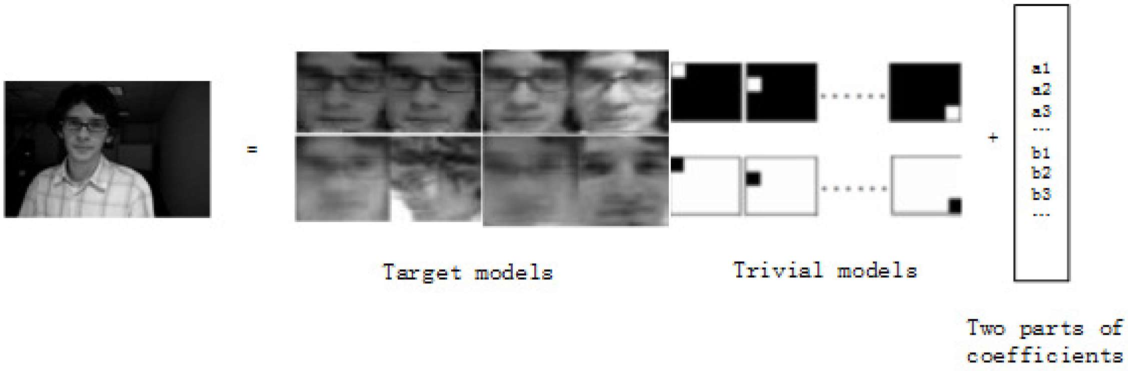
We use $T = [t_1, t_2..t_n] \in R^{d\times n}(d \gg n)$ to represent the new template set, and use $I = [i_1, ...i_d] \in R^{d\times d}$ to represent the trivial template set. So the candidate can be describe by the following formulation:

$$y = (T, I)\begin{pmatrix} a_T \\ a_I \end{pmatrix} + e \tag{1}$$

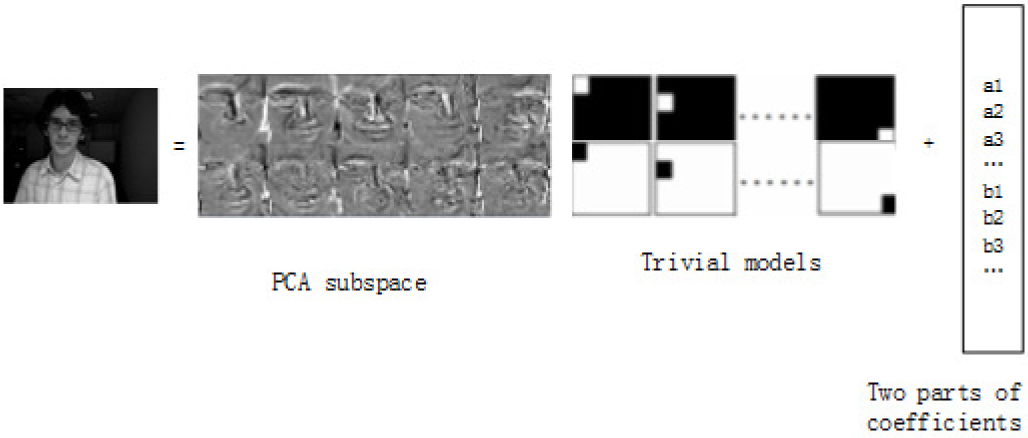
Where y is the observed candidate, a_T is the coefficient of the main templates, a_x is the coefficient of the trivial templates and e is the representation residual.



(a) IVT target models



(b) L1 minimization tracking target models



(c) SPCA-L1 target models

By using the new models, we can collect the most information of the target and candidates. Meanwhile when the target is covered, the trivial templates can represent the noise and make our tracking robust. It will be proved in the experiment part following.

B. Lasso model for sparse representation

As we introduced in the former part, after learning the basic templates, we use subspace learning method PCA to reduce the dimension of template set. Then we need to get the sparse coefficients a in formation (1).

We know that the formation has more than one solution. But we want to get the sparse solution, so we exploit the compressibility in the transform domain by solving the problem as an l_1 regularized least squares problem, which is known to typically yield sparse solutions. As the coefficients include the ones of trivial templates, so the problem in our tracking method is transformed to the following formation:

$$\hat{a} = \arg\min_{a}\{\frac{1}{2}\| y - Ba \|_2^2 + \lambda \| a \|_1 + \frac{d}{2}\| a_I \|_2^2\} \qquad s.t. \qquad a \geq 0 \tag{2}$$

Where $B = [T, I]$ in formation (1). $\| . \|_1$ and $\| . \|_2$ denote the l_1 and l_2 norms respectively.

At last, we choose the candidate which has the smallest residual as the new target.

$$identity(y) = argmin\{\| e \|_2^2\} = argmin\{\| y - Ba \|_2^2\} \tag{3}$$

C. Online Tracking models and incrementally update of the mean and eigenbasis

The appearance of a target object often change drastically due to different kinds of factors. Therefore, we need to adapt the appearance online to reflect these changes. The appearance model we have chosen, a eigenbasis, is typically learned off-line from a set of training images, $d \times n$ matrix, $T = \{t_1, ..., t_n\}$, by taking computing the eigenvectors U. Then we present our approach for drawing particles in the motion parameter space and predicting the most likely candidate with the help of the learned appearance model.

The sample mean of the training images is $\bar{t} = \frac{1}{n}\sum_{i=1}^{n} t_i$, and the sample covariance matrix is $\frac{1}{n-1}\sum_{i=1}^{n}(t_i - \bar{t})(t_i - \bar{t})^T$,and its eigenvector is describe as U. Equivalently one can obtain U by computing the singular value decomposition $T = U\Sigma V^T$ of the centered data matrix $[(t_1 - \bar{t})...(t_n - \bar{t})]$, with columns equal to the respective training images minus their mean. When the additional m images, $d \times m$ matrix, $W = \{t_{n+1}, ...t_{n+m}\}$ comes, we retrain the eigenbasis. This update could be performed by computing the singular value decomposition $[T \quad W] = U'\Sigma' V'^T$ of the centered data matrix $[(t_1 - \overline{t'})...(t_{n+m} - \overline{t'})]$, where $\overline{t}'$ is the average of the entire $n + m$ training images.

We describe a new method based on the Sequential Karhunen Loeve (SKL) algorithm. Letting $\tilde{W}$ be the component of W orthogonal to U, we can express the concatenation of T and W in a partitioned form as follows:

$$[T \quad W] = [U \quad \tilde{W}][\begin{matrix} \Sigma & U^T W \\ 0 & \tilde{W}^T W \end{matrix}][\begin{matrix} V^T & 0 \\ 0 & I \end{matrix}] \tag{4}$$

Let $R=\begin{bmatrix}\Sigma & U^TW\\ 0 & \tilde{W}^TW\end{bmatrix}$, which is a square matrix of size $k+m$,where k is the number of singular values in Σ . The SVD of R, $R=\tilde{U}\tilde{\Sigma}\tilde{V}^T$, can be computed in constant time regardless of n , the initial number of data. Then the SVD of $[T \quad W]$ can be expressed as follows:

$$[T \quad W]=([U \quad \tilde{W}]\tilde{U})\tilde{\Sigma}(\tilde{V}^T\begin{bmatrix}V^T & 0\\ 0 & I\end{bmatrix}) \tag{5}$$

By using this algorithm, we can calculate the updating process more efficiently. The following table (1) shows the storage space and computational complexity change before and after the optimization. It is clear that both storage space and computational complexity are reduced, and it is helpful to track more quickly.

Table1: Comparison before and after the optimization

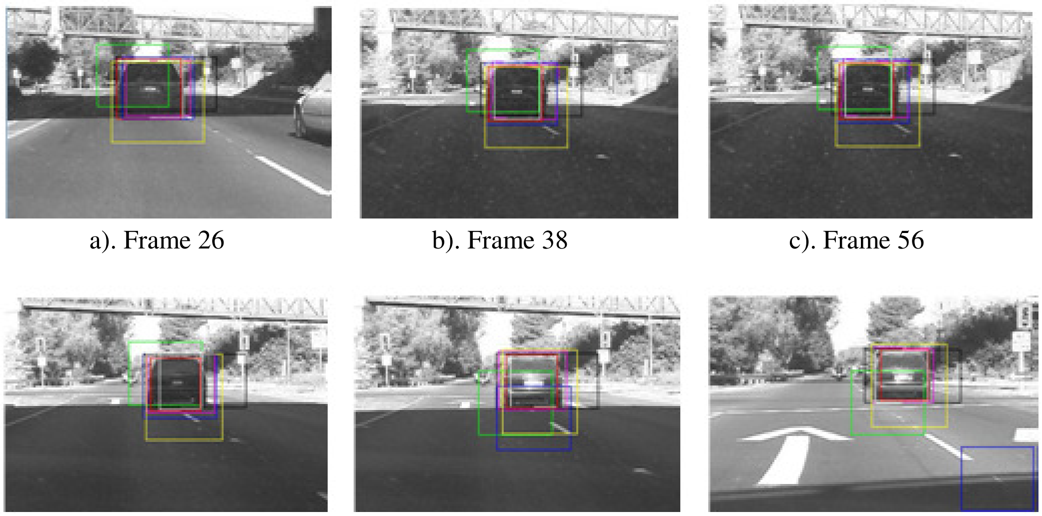
	Update object	Storage space	Computational complexity
Before	$[T \quad W]=U'\Sigma'V'^T$	$O(m(n+q)^2)$	$O(m(n+q)^2)$
After	$[T \quad W]=[U \quad \tilde{W}]\begin{bmatrix}\Sigma & U^TW\\ 0 & \tilde{W}^TW\end{bmatrix}\begin{bmatrix}V^T & 0\\ 0 & I\end{bmatrix}$	$O(m(k+q))$	$O(mq^2)$

4. RESULTS

In this section, we present the experiment results achieved by applying our method. To demonstrate the performance of our method, we track the classical videos in the visual tracking, involving 1) target pose change 2) different light conditions 3) scale change and so on. We do the experiment on the computer: Inter(R) Core(TM) i5-4590 CPU, 3.30GHz, 4GB. And we use Matlab to do the simulation. All the results are as follow: the red color stands for our method, the pink is IVT, the yellow one represents L1 tracking, the black is APG, the blue is CT, and the green is MIL. We pay especially attention to the comparison of our method with IVT, which based on subspace.

A. Intuitive comparison

As the Fig.2 shows, the first car4 video shows a car passing beneath a bridge with sudden illumination change. Our algorithm can track the car efficiently. While the MIL drifts a little when the car undergoes illumination, and CT drifts and at last loses the target.

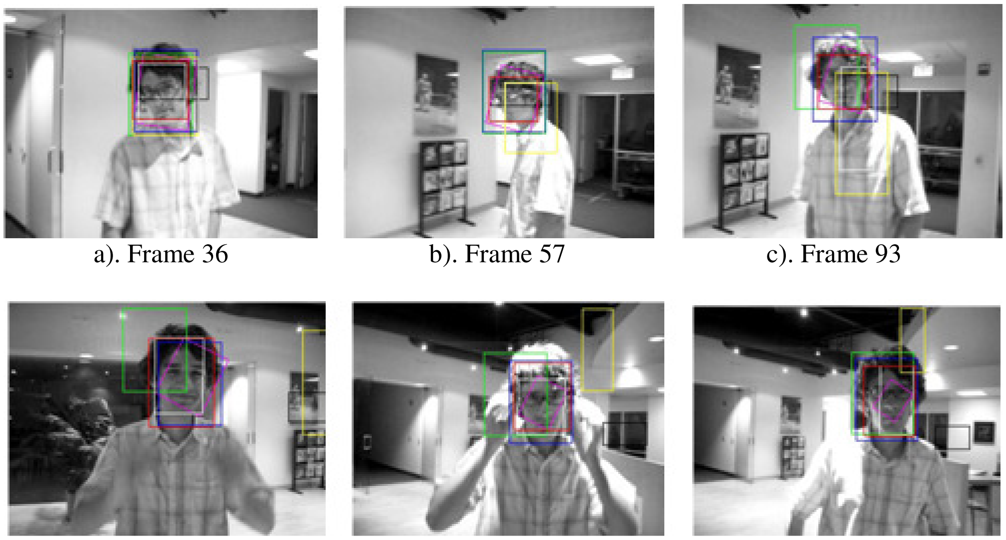


a). Frame 26 b). Frame 38 c). Frame 56

d). Frame 71 e). Frame 76 f). Frame 96

Fig.2 Car4 tracking results

In Fig.3, the second David indoor video involves light change and target pose change, recorded at 15 frames per second with a moving digital camera. The man moves from a dark area to a bright area, the back to the dark area, who undergoes lighting and pose changes. Notice that there is a large scale variation in the target. It's clear that our algorithm is able to track the target throughout the sequences, no matter of light change or scale variation. The MIL method and L1 tracking experience drift, and the L1 even loses the target at #311.

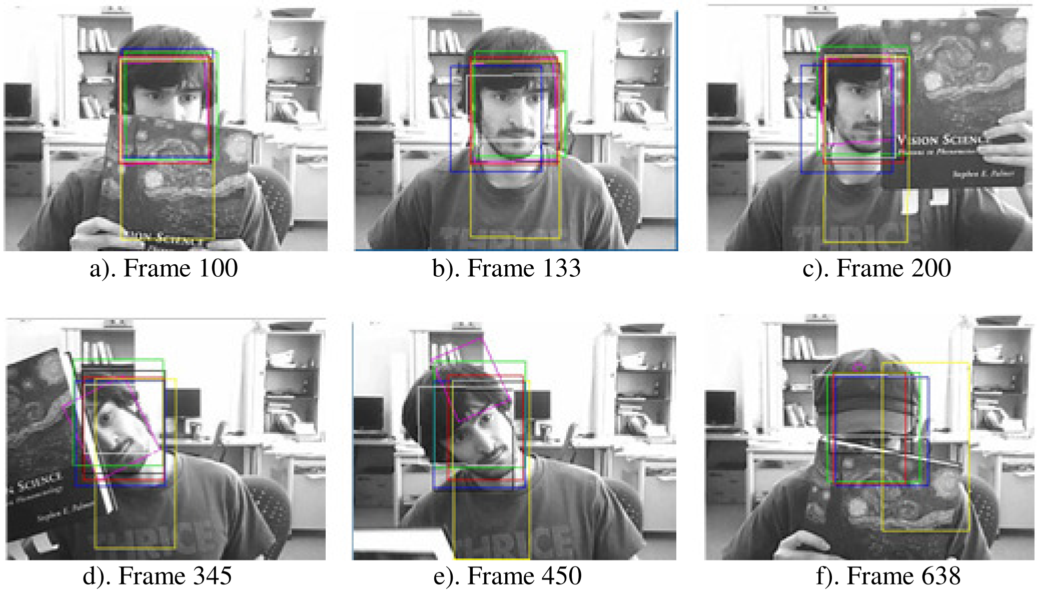


a). Frame 36 b). Frame 57 c). Frame 93

d). Frame 189 e). Frame 300 f). Frame 311

Fig.3 David indoor tracking results

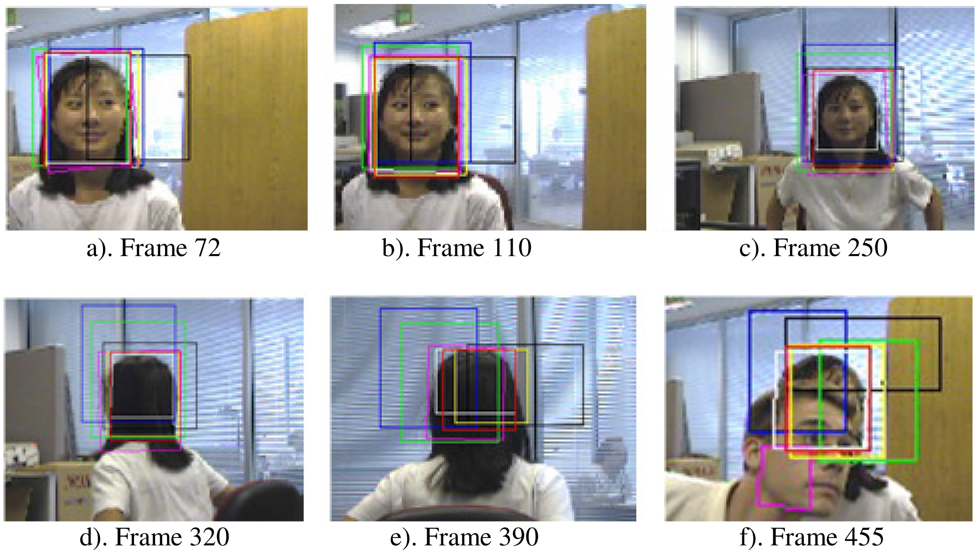
About Fig.4, the third video occluded face2 involves occlusion. In the video, the target is occluded by a book. We can find that when the man's face is covered by a book and a hat, L1 tracking and CT drifts away from the target. And when the man tilted his head, the IVT is not robust. It is the same condition in Fig.3 when David turns left at #189, and since then the tracking rectangle is not identical to the target. During the process, our method perform well and deal with occlusion problem perfectly.


a). Frame 100 b). Frame 133 c). Frame 200

d). Frame 345 e). Frame 450 f). Frame 638

Fig.4 Occluded face2 tracking results

In Fig.5, It is a video of a girl, who turns her head around and later is covered by a man's face. MIL and APG perform a little bad, especially when the man covers the girl's face. And the IVT mistakes the man as the target instead of the girl because of occlusion. And our method locks the target even she turns around. And just drifts a little when the man appears.


a). Frame 72 b). Frame 110 c). Frame 250

d). Frame 320 e). Frame 390 f). Frame 455

Fig.5 Girl tracking results

B. Accuracy comparison

To evaluate the accuracy of our tracking algorithm, we consider the center position errors of each tracking method. As we can see, the center position errors of CT and MIL in sequences ‘Car4, David indoor and Girl’ increase at or after 100 frames. The L1 tracking method does a bad job in David indoor. Meanwhile, the IVT performs badly when the target is occluded because the center position errors increase in Occluded face2. But in these 4 videos, our tracking method does better than the other 5 methods.

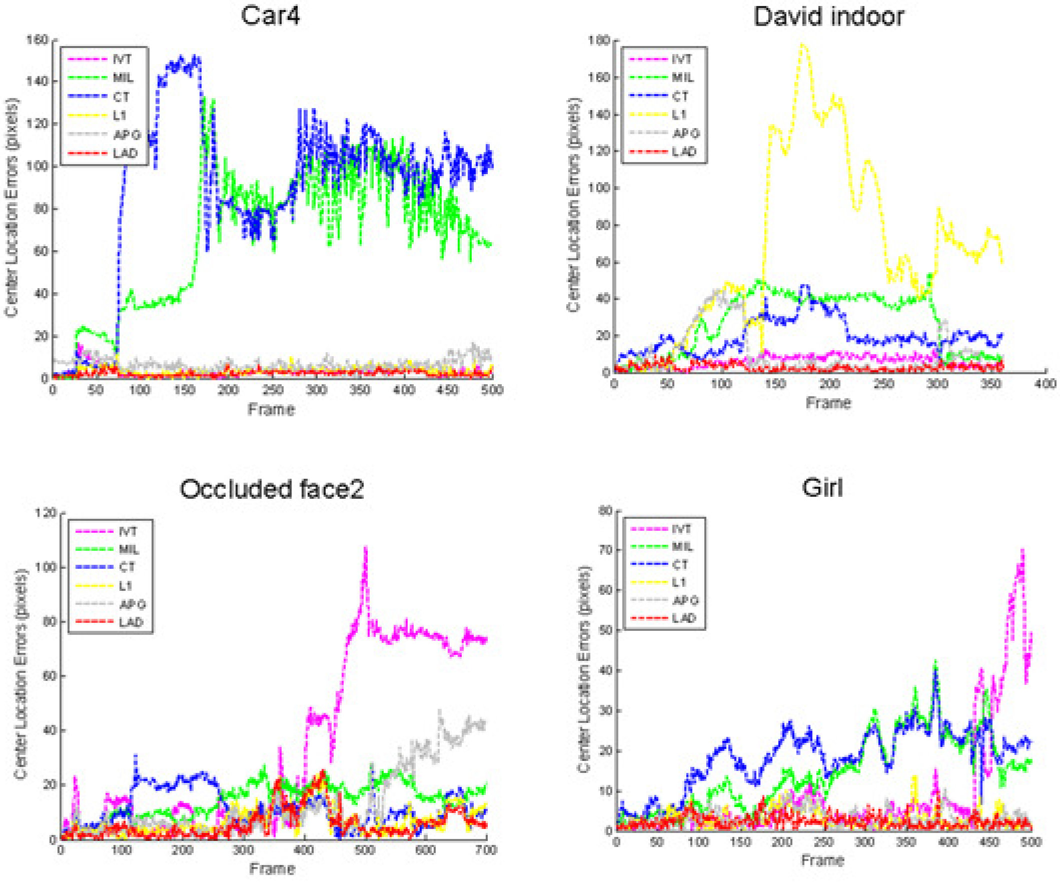


Fig.6 Center position errors of four challenging sequences

Besides, we calculate the average center position errors of the above results. As showing in Table 2, the red one represents the smallest average center position errors, the blue is the second best. So our tracking enhances a little in accuracy than IVT and L1 tracking methods.

Table 2: Average center position errors

Trackers Sequences	LAD (pixels)	IVT	MIL	CT	L1	APG
Car4	2.8	3.4	67.5	90.2	3.6	6.7
David indoor	3.3	6.1	27.0	7.5	66.7	10.2
Occluded face2	5.9	33.7	14.3	11.1	6.9	13.0
Girl	2.9	9.3	14.6	18.3	3.0	3.2

4. CONCLUSION

The tracking method sparse PCA-L1 is a robust tracking method. Because we add the trivial template and PCA subspace together, the tracking method does a good job when the target undergo pose, illumination and appearance change and occlusion. We mix the advantages that PCA subspace can represent most information of the target and trivial template can describe the noise like occlusion and so on. At the same time, we update the eigenbasis and mean accurately and efficiently. So when the target changes, we can gradually form the new models according to the eigenbasis.

REFERENCES

[1] Z. Kalal, J. Matas, and K. Mikolajczyk. P-N learning: Bootstrapping binary classifiers by structural constraints. In CVPR, 2010.3,7.

[2] Grabner,H., Leistner,C., Bischof,H.: 'Semi-supervised online boosting for robust tracking'. Proc. European Conf. Computer Vision (ECCV), Marseille, France, October 2008,pp.234-247

[3] Sankarayanan, K., Davis, J.W.: 'One class multiple instance learning and applications to target tracking'. Proc. IEEE Conf. Computer Vision(ACCV), Daejeon, Korea, November 2012, pp.1-14

[4] Ross,D. , Lim, J. , Yang, M.H : 'Adaptive probabilistic visual tracking with incremental subspace update'. Proc. European Conf. Computer Vision(ECCV), Prague, Czech Republic, May 2004, pp.470-482

[5] Lim, J., Ross, D., Lin, R.S., Yang, M.H.: 'Incremental learning for visual tracking', in Weiss, Y., Bottou, L., (Eds.), 'Advances in neural information processing systems' (MTI Press, 2005), pp.793-800

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

[7] Mei, X., Ling, H.B.: 'Robust visual tracking using L1 minimization'. Proc. IEEE Int. Conf. Computer Vision(ICCV), Kyoto, Japan, September 2009, pp.1436-1443w

[8] Zou,H., and Hastie, T.(2005),'Regularization and Variable selection via the Elastic Net' Journal of the Royal Statistical Society, Series B,67,301-320.

[9] Jolliffe,I.T., Tredafilov, N.T., and Uddin, M.(2003),'A Modified Principle Component Technique Based on the Lasso' Journal of Computational and Graphical Statistics, 12,531-547.

[10] J. Kwon and K. Lee. Visual tracking decomposition. In CVPR,2010.2,3,7.

[11] Y. Sui and L. Zhang.' Visual Tracking via Locally Structured Gaussian Process Regression'. IEEE Signal Processing Letters,22(9):1331-1335,2015.3,7

[12] D. Wang, H. Lu and M. Yang. 'Online object tracking with sparse prototypes'. IEEE Transactions on Image Processing(TIP), 22(I):314-325,2013.2,3

[13] T. Zhang, B. Ghanem, S. Liu, and N. Ahuji, 'Low rank sparse learning for robust visual tracking'. In ECCV,2012.3,6,7.

[14] S. Hare, A. Saffari, and P. Torr. Struck:' Structured output tracking with kernels'. In ICCV,2011.3.

[15] Yao Sui, Yafei Tang, Li Zhang. 'Discriminative Low Rank Tracking'. ICCV 2015.3002-3010.

[16] A. d' Aspremont, L. EI Ghaoui, M.Jordan, and G.Lanckriet.' A direct formulation for sparse PCA using semidefinite programming'. SIAM Review ,46(3),2007.

AUTHORS

Yuanyuan Zhang received B.S. degree in Electronic science and technology from Nanjing University of Aeronautics and Astronautics, Nanjing, China, in 2014. Now she is studying the M.S. degree in School of Electronics and Information Engineering of Beihang University, Beijing, China. Her current research interests lie in the areas of image processing and computer visual tracking.

Fuxiang Wang received the B.S. degree in 1999 and the Ph.D. degree in 2007, all from Beihang University, Beijing, China. He is currently a lecture with the School of Electronics and Information Engineering, Beihang University, Beijing, China. His current research interests lie in the areas of blind separation and their applications.