International Journal of Hybrid Information Technology
Vol.9, No.4 (2016), pp. 375-382
http://dx.doi.org/10.14257/ijhit.2016.9.4.32

Robust and Fast Tracking via Joint Collaborative Representation

Fei Zhou, Guizong Zhang, Xinyue Fan, Dandan Yi

Chongqing Key Laboratory of Optical Communication and Networks, Chongqing University of Posts and Telecommunications
zhoufei@cqupt.edu.cn, zhangguizong23@sina.com,
fanxy@cqupt.edu.cn,dandanxinhai@163.com

Abstract

In this paper, we present a robust and fast tracking method based on joint collaborative representation. Traditional sparse coding based tracking methods code the candidates as a sparse linear combination of a series of object and trivial templates and perform time consuming L1 regularizations. In contrast to these methods, this paper adopts the L2-regularized least square models to reduce the computational complexity. The tracked object can be represented by the linear combination of a series of object templates, and also can be represented by candidate samples in the current frame. We propose a joint objective function to handle the tracking process. In addition, we introduce an effective update scheme to deal with the change of target appearance over time. Experiments on several challenging image sequences show that our proposed tracking method is robust and efficient.

Keywords*: Object tracking, L2-regularized least square, collaborative representation.*

1. Introduction

Visual tracking has been widely applied in real-world application such as automatic surveillance, human computer interaction, intelligent transportation etc. Although many tracking algorithms have been proposed and tested over the past few decades, many challenges still remain. Occlusion, illumination variation, poses change, background clutter are the main challenges for visual tracking.

Current tracking algorithms can be categorized into methods based on filters [1], classifiers [2], statistics features [3], sparse representation [4-8] and so on. Ever since the first time sparse representation was introduced into visual tracking by Mei [8], various efficient and effective trackers have been proposed. The main ideas of these trackers are similar which use a series of object and trivial templates to represent the tracked objects. The object templates are used to describe the valid part of the object and trivial templates are adopted to handle with outliers (e.g., occlusion). In [4], Xiao et al. proposed a robust and fast tracking algorithm, in which object tracking is performed by solving L2-regularized least square problems in a Bayesian inference framework. Besides, the appearance of the tracked target is modeled with several PCA (Principal Components Analysis) basis vectors and square templates. The tracker not only utilizes the strength of subspace representation but also take partial occlusion into consideration. Exploiting multiple views of various types of image features can significantly improve tracking performance because of their complementary characteristics. To this end, Hong et al. [5] proposed to use four types of complementary features, i.e. intensity, color histogram, HOG (Histogram of Oriented Gradient) and LBP (Local Binary Patterns) in the appearance representation, and then cast tracking as a multi-task multi-view sparse learning problem.

In this paper, we propose a robust and fast tracking method based on joint collaborative representation which works well in handling high computational complexity, partial

ISSN: 1738-9968 IJHIT

occlusion and other challenging factors. There are several key differences between former papers and our study. First, we adopt the L2-regularized least square method to solve the proposed representation model. Second, we model target appearance with PCA basis vectors, and account for occlusion with trivial templates. Third, we construct another representation model to improve the robustness of our tracker according to the premise that the tracked target can not only be sparsely represented by a series of object templates but also can be sparsely represented by candidates in the current frame. Finally, to account for the appearance variations, we proposed an effective update scheme.

The remaining of this paper is organized as follows: Section 2 briefly introduces the motivation of this work. Section 3 provides some discussion of the proposed joint collaborative representation model; In section 4, we analysis the template update strategy. Extensive experiments are conducted in Section 5. Finally, we conclude the paper in Section 6.

2. Motivation of This Work

2.1. Sparse Representation or Collaborative Representation

Over the past few decades, the sparse representation methods have been widely applied in pattern recognition and computer vision. In [9], Wright et al. used sparse representation for robust face recognition for the first time. According to it, Mei and Ling [8] presented a L1-tracker which aims at determining the most likely patch with a sparse coding of object templates and modeling partial occlusion by trivial templates.

For a set of M training templates $D=[d_1,d_2,\mathrm{L}\ ,d_M]\in R^{d\times M}\ (d?\ \ M)$ where d is the dimension of the training templates. A candidate in the current frame $y\in R^d$ can be sparsely represented by

$$y=Dx+e=[D\ I]\begin{bmatrix}x\\e\end{bmatrix}=D'c \tag{1}$$

where $x\in R^M$ is the sparse coefficient, I corresponds to the trivial template set, e is a sparse error. Object tracking in [14] is formulated as solving an L1-regularized least square problem which is known to typically yield sparse solutions,

$$c_{opt}=\arg\min_c(\|y-D'c\|_2^2+\lambda\|c\|_1) \tag{2}$$

where λ is the regularization parameter to balance the relationship between the reconstruction term and the L1-norm constraint.

In [7] and [11], the working mechanism of sparse representation was challenged and a very simple and effective FR scheme, namely collaborative representation, that is L2-regularized least square method. There is

$$c_{opt}=\arg\min_c(\|y-D'c\|_2^2+\lambda\|c\|_2^2) \tag{3}$$

The solution of Eq.(3) can be easily and analytically obtained by solving the extremum problems of the expression:

$$-D'^{\mathrm{T}}\ y+D'^{\mathrm{T}}\ D'\mathrm{c}+\lambda\,\mathrm{Ic}=0 \tag{4}$$

So the optimal solution can be derived as

$$c_{opt}=(\mathrm{D}'^{\mathrm{T}}\mathrm{D}'+\lambda I)^{-1}\ \mathrm{D}'^{T}\ y \tag{5}$$

Let $P = (\mathrm{D'^{T}D'} + \lambda I)^{-1}\mathrm{D'}$ as a projection matrix. Once a query sample y comes, we can simply project y onto P. Experiments on face recognition and visual tracking showed that methods based on collaborative representation have very competitive recognition or tracking accuracy but with significantly lower complexity.

2.2. Object Templates or PCA Basis Vectors

Templates in Eq.(2) and (3) collected by random sampling are highly correlated, so they are not enough to represent appearance variations of the tracked target. Several tracking methods in [10, 11] have demonstrated that PCA subspace representation with online update is effective in dealing with appearance variations such as rotation, scale, illumination variation and pose change. So we have:

$$y = Ua + e \tag{6}$$

where y indicates the observation vector, a denotes the corresponding coefficient, $U \in R^{d \times k}$ is a matrix of PCA basis vectors (k is the number of PCA basis), e is the error term.

However, it has also been shown that the PCA subspace representation based methods are sensitive to occlusion. To exploit the strength of both PCA basis vectors and trivial templates, we construct the representation model by

$$y = Ua + e = [\mathrm{U\ I}]\begin{bmatrix} a \\ e \end{bmatrix} \tag{7}$$

Compare with Eq.(6), e is the error term that contains partial occlusion. Compare with Eq. (1), a needn't to be sparse. So, we present the objective function by combining sparse representation and PCA subspace learning.

$$\min \left\| y - Ua - e \right\|_2^2 + \lambda \left\| \begin{matrix} a \\ e \end{matrix} \right\|_2^2 \tag{8}$$

The computational complexity of which is $O(\mathrm{dk} + \mathrm{dn})$ (n is the number of trivial templates) [7] while the computational complexity of L1 tracker is $O(\mathrm{m}(\mathrm{dk} + \mathrm{d}^2))$ (m is the number of iterations to achieve convergence which ranges from dozens to hundreds). Consequently, this improved method is much quicker than L1 tracker.

3. Joint Collaborative Representation Model

In [12], Wang et al. proposed that the tracked object can be sparsely represented by both the object templates and candidate samples in the current frame.

In practice, we randomly sample particles from a diagonalized Gaussian distribution $p(x_t \mid x_{t-1}) = \mathrm{N}(x_t; x_{t-1}, \Sigma)$ to generate M candidate particles $\{\mathrm{x}_t^1, \mathrm{x}_t^2, \mathrm{L}\ , \mathrm{x}_t^M\}$, where x_t indicates the objects state in the t-th frame, and build a candidate dictionary $\mathrm{A}_t = [\mathrm{z}_t^1, \mathrm{z}_t^2, \mathrm{L}\ , \mathrm{z}_t^M]$. To construct a robust tracker, we present a novel objective function to improve our method.

$$J(\mathrm{y}_t, a_t, b_t) = (\left\| \mathrm{y}_t - U_t a_t - e_t \right\|_2^2 + \lambda_1 \left\| \begin{matrix} a_t \\ e_t \end{matrix} \right\|_2^2) + \mu(\left\| \mathrm{y}_t - A_t b_t \right\|_2^2 + \lambda_2 \left\| b_t \right\|_2^2) \tag{9}$$

where μ is a constant parameter to balance the relationship between the PCA subspace representation model and candidate representation model.

With the non-negativity constraints ($a_t \geq 0, b_t \geq 0$), the optimal representation coefficients can be solved by,

$$\{\hat{y}_t, \hat{a}_t, \hat{b}_t\} = \underset{y_t, a_t, b_t}{\operatorname{argmin}} (\|y_t - U_t a_t - e_t\|_2^2 + \lambda_1 \left\| \begin{matrix} a_t \\ e_t \end{matrix} \right\|_2^2) + \mu(\|y_t - A_t b_t\|_2^2 + \lambda_2 \|b_t\|_2^2) \tag{10}$$

Here y_t, a_t, b_t are unknown. We adopt an iterative method to seek the optimal solution of the optimization problem (i.e., Eq.(10)). The detail iterative algorithm can be summed up in the following two steps:

a. Fix y_t, solve a_t and b_t. In this case, Eq. (10) will be divided into two L2-regularization problems, which can be easily and quickly solved by the procedure mentioned above.

b. Fix a_t and b_t, solve y_t. In this case, Eq. (10) will be rewritten as

$$\hat{y}_t = \underset{y_t}{\operatorname{argmin}} (\|y_t - U_t a_t - e_t\|_2^2 + \mu \|y_t - A_t b_t\|_2^2) \tag{11}$$

Eq.(11) can be obtained by solving the extremum problems. The optimization can be solved efficiently by repeating step a and b until the difference of objective function (i.e., $J(y_t, a_t, b_t)$) between two iterations or the number of iterations meet a stopping criterion. Figure 1 shows the basic idea of our tracking algorithm. We now provide the algorithm for optimizing the joint collaborative representation below:

Input: $\hat{y}_{t-1}$, U_t, A_t, λ_1, λ_2, μ.

Output: $\hat{y}_t = y_{t,i}$, $\hat{a}_t = a_{t,i}$, $\hat{b}_t = b_{t,i}$.

1. Initialize $y_{t,0} = \hat{y}_{t-1}, i = 0$.
2. While not converged do
3. $$a_{t,i+1} = \underset{a_t}{\operatorname{argmin}} (\|y_{t,i} - U_t a_t - e_t\|_2^2 + \lambda_1 \left\| \begin{matrix} a_t \\ e_t \end{matrix} \right\|_2^2), a_t \geq 0$$
4. $$b_{t,i+1} = \underset{b_t}{\operatorname{argmin}} \|y_{t,i} - A_t b_t\|_2^2 + \lambda_2 \|b_t\|_2^2, b_t \geq 0$$
5. $$y_{t,i+1} = \underset{y_t}{\operatorname{argmin}} (\|y_t - U_t a_{t,i+1} - e_t\|_2^2 + \mu \|y_t - A_t b_{t,i+1}\|_2^2)$$
6. $i = i + 1$
7. end

When y_t, a_t, b_t are obtained, the optimal state $\hat{x}_t$ can be approximated by,

$$\hat{x}_t = \sum_{i=1}^{M} \hat{b}_t^i x_t^i \tag{12}$$

where $\hat{b}_t^i$ represents the weight distribution calculated by normalization.

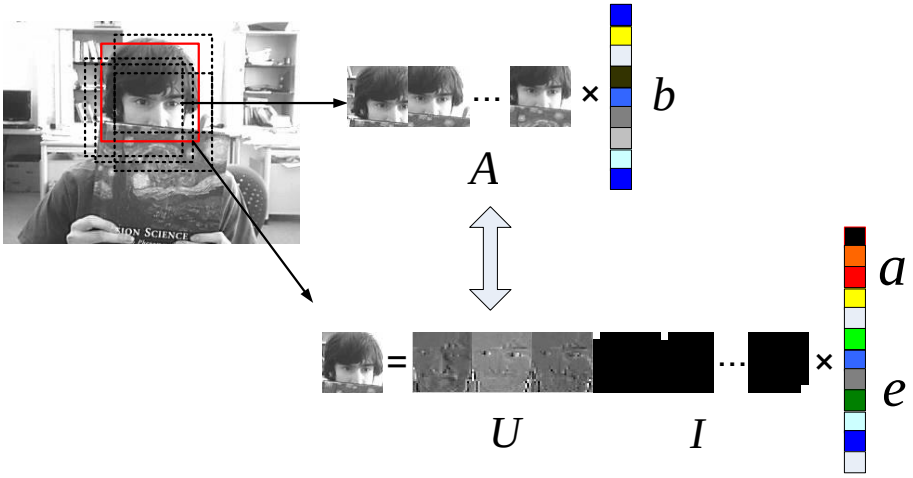


Figure 1. The basic Idea of our Tracking Algorithm

4. Update Framework

To handle appearance change of a target object for visual tracking, we proposed a novel update mechanism. As we know, a_t represents the importance of different PCA basis vectors, error term e_t reflects the possibility of partial occlusion or misalignment. When a new optimal candidate is obtained, we can obtain the occlusion ratio η_t ($\eta_t = N_{nonzero} / N_{total}$). If $\eta_t < \varepsilon$, it means that this sample is reasonable and we directly update the model with this sample; If $\eta_t > \varepsilon$, it means that a significant part of the target object is occluded and we discard it. Then we update the observation model by recalculate the PCA basis vectors.

5. Experiments

In this section, we present extensive experimental results to validate the effectiveness and efficiency of our proposed tracker. We evaluate our proposed method on six sequences. In this paper, each tracked object patch is resized to 32×32 pixels. The number of PCA basis vectors is 10. Square templates are used as trivial templates [4]. In each frame, the number of particles is 300 and the radium is 4. We set $\lambda_1=\lambda_2=0.001$, $\mu=1$, and compare the proposed algorithms with four state-of-the-art trackers: Visual Tracking Decomposition (VTD) tracker [13], incremental visual tracking (IVT) [14], L1 tracker [8], Multiple Instance Learning (MIL) [15].

To evaluate the aforementioned trackers, we use their average center errors as the criterion for accuracy measure in this paper. Fig. 2 presents the tracking results on different video sequences with illumination variation, pose change and background clutter.

Table 1. Average Center Errors of Tracking Algorithms (FPS: frame per seconds)

Sequence	VTD	IVT	L1	MIL	OURS
FaceOcc1	11.1	9.2	6.5	32.3	5.0
FaceOcc2	10.4	10.2	11.1	14.1	6.2
Car4	12.3	2.9	4.1	60.1	4.5
Car11	27.1	2.1	33.3	16.2	4.1
Deer	11.9	127.2	171.5	66.5	10.3
Panda	94.8	169.8	94.0	103.4	3.0
Average	27.9	53.6	53.4	48.8	5.5
FPS	4	32	0.5	32	8

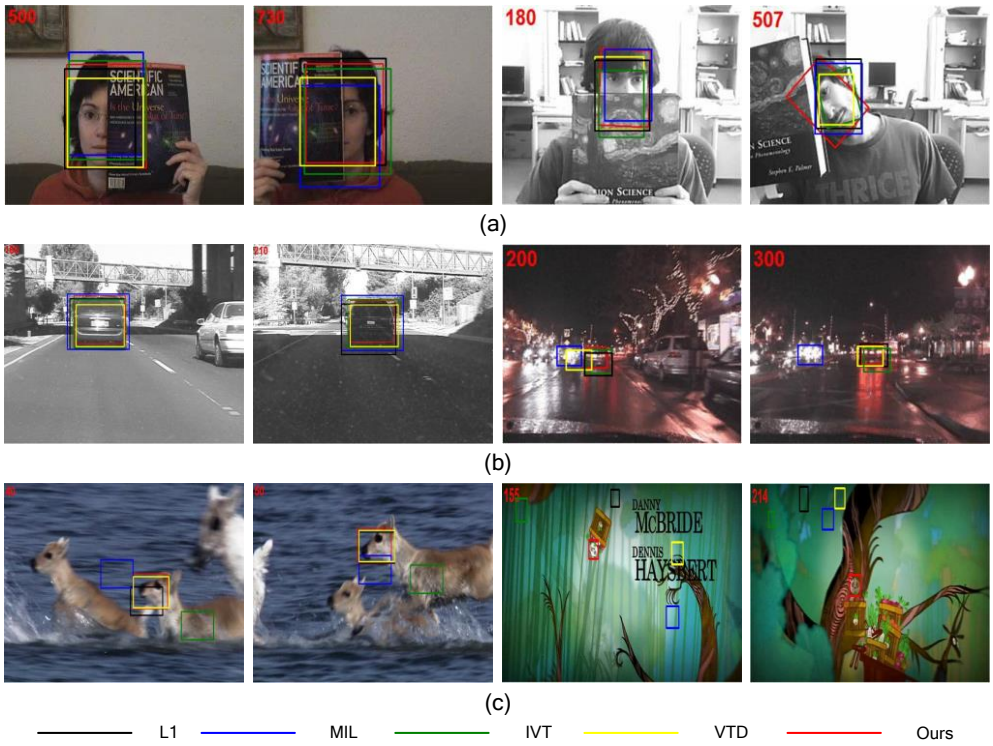


Figure 2. Qualitative Evaluation of Different Trackers (a) FaceOcc1 and FaceOcc2 with Partial Occlusion. (b) Car4 and Car11 with Illumination Variation. (c) Deer and Panda with Rotation and Pose Variation

Partial occlusion: Figure 2(a) demonstrates that the proposed method performs robust to partial occlusion. While the VTD, IVT, L1 and MIL tracking methods are less effective. These improvements benefit from the application of trivial templates and update framework in our tracking method.

Illumination variation: Figure 2(b) shows tracking results from different video sequences to evaluate whether our tracker is able to tackle drastic illumination variation. The IVT tracking method and the proposed tracker perform well because the PCA subspace is robust to illumination change. The MIL tracker gets lost in tracking process in this case whereas VTD and L1 algorithms perform slightly better.

Rotation and pose variation: Figure 2(c) presents the tracking results where the objects suffer rotation and pose variation. The VTD tracking method gets lost when the tracked target rotates, while our tracker performs well throughout this sequence.

The average center location errors are shown in Table 1, which demonstrates that our tracker achieves better performance than other state-of-the-art trackers. This improvement attributes to the complementary of PCA subspace representation model and candidate representation model. In addition, the update scheme in our tracking method uses the latest tracking result to update the PCA basis vectors, which improve the robustness to appearance variation. Our tracker is slower than the IVT and MIL trackers and generally more stable and accurate. Besides, our tracker is more effective and much faster than the L1 tracker.

6. Conclusions

In this paper, we proposed a novel robust and fast tracking method based on joint collaborative representation. The tracked object can not only be collaboratively represented by a series of PCA basis vectors, but also can be collaboratively represented

by candidate samples. We derived the objective function to integrate PCA subspace representation model and candidate representation model. Then we adopted an effective update scheme to handle with the appearance variation of the tracked object. Experiments on several video sequences demonstrate the effectiveness and efficiency of our tracking method. Our future work will focus on how to extend our tracking model for multiple object tracking.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61471077, 61301126), the Fundamental and Frontier Research Project of Chongqing (cstc2013jcyjA40034, cstc2013jcyjA40041), the Science and Technology Project of Chongqing Municipal Education Commission (KJ1400413).

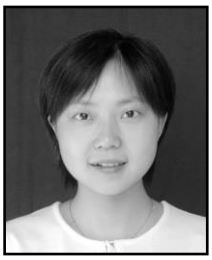
References

[1] K. Okuma, A. Taleghani, N. de Freitas, J. J. Little and D. G. Lowe, "A boosted particle filter: Multi-target detection and tracking", Proceeding of the 8th ECCV, **(2004)**; Prague, Czech Republic.
[2] Z. Kalal, K. Mikolajczyk and J. Matas, "Tracking-learning-detection", IEEE Trans. Patt. Anal. Mach. Intell., vol. 34, no. 7, **(2012)**, pp. 1409-1422.
[3] H. Grabner, C. Leistner and H. Bischof, "Semi-supervised on-line boosting for robust tracking", Proceeding of the 10th ECCV, **(2008)**; Marseille, France.
[4] X. Ziyang, L. Huchuan and W. Dong, "L2-RLS based object tracking", IEEE Trans. Circuits Syst.,vol. 24, no. 8, **(2013)**, pp. 1301-1309.
[5] H. Zhibin and M. Xue, "Tracking via robust multi-task multi-view joint sparse representation", Proceeding of ICCV, **(2013)**.
[6] J. Xu, L. Huchuan and Y. Ming-Hsuan, "Visual tracking via adaptive structural local sparse appearance model", Proceeding of CVPR, **(2012)**.
[7] D. Zhang, Y. Meng and F. Xiangchu, "Sparse representation or collaborative representation: which helps face recognition?", Proceeding of ICCV, **(2011)**.
[8] X. Mei and H. Ling, "Robust visual tracking using l1 minimization", Proceeding of ICCV, **(2009)**.
[9] J. Wright and A. Yang, "Robust face recognition via sparse representation", IEEE Trans. Patt. Anal.Mach. Intell., vol. 31, no. 2, **(2009)**, pp. 210-227.
[10] T. Wang, I. Gu and P. Shi, "Object tracking using incremental 2d-pca learning and ml estimation", Proceeding of ICASSP, **(2007)**.
[11] D. Wang, H. Lu and Y. Chen, "Incremental mpca for color object tracking", Proceeding of ICPR **(2010)**.
[12] D. Wang, H. Lu and C. Bo, "Online visual tracking via two view sparse representation", IEEE Signal Processing Letters, vol. 21, no. 9, **(2014)**, pp. 1031-1034.
[13] J. Kwon and K. M. Lee, "Visual tracking decomposition", Proceeding of CVPR, **(2010)**.
[14] D. Ross, J. Lim, R. Lin and M. Yang, "Incremental learning for robust visual tracking", Proceeding of IJCV, **(2008)**.
[15] B. Babenko, M. Yang and S. Belongie, "Robust object tracking with online multiple instance learning", IEEE Trans. Patt. Anal. Mach. Intell., vol. 33, no. 8, **(2011)**, pp. 1619-1632.

Authors

Fei Zhou. He received the PhD degree from University of Electronic Science and Technology of China, in 2006. Now, He is an associated professor of ChongQing University of Posts and Telecommunications. He authored/coauthored about 20 technical articles. His research interests include signal processing, image processing and cognitive sensing.

Guizong Zhang. He is pursuing the M.S. degree in Information and Communication Engineering, Chongqing University of Posts and Telecommunications (CQUPT), China. His current research interests include image processing, machine vision and artificial intelligence.



Xinyue Fan. She received the M.S. degree from University of Electronic Science and Technology of China in 2005. Her research interests include geolocation, wireless communication systems, and communication signal processing.

Dandan Yi. She is pursuing the M.S. degree in Information and Communication Engineering, Chongqing University of Posts and Telecommunications (CQUPT), China. His current research interests include image processing and wireless communication.