Drift-free tracking via the construction of an effective dictionary

Li Zhao, Pengcheng Huang, Fei Liu, Hui Huang and Huiling Chen

Abstract
Template dictionary construction is an important issue in sparse representation (SP)-based tracking algorithms. In this article, a drift-free visual tracking algorithm is proposed via the construction of an effective template dictionary. The constructed dictionary is composed of three categories of atoms (templates): nonpolluted atoms, variational atoms, and noise atoms. Moreover, the linear combinations of nonpolluted atoms are also added to the dictionary for the diversity of atoms. All the atoms are selectively updated to capture appearance changes and alleviate the model drifting problem. A bidirectional tracking process is used and each process is optimized by two-step SP, which greatly reduces the computational burden. Compared with other related works, the constructed dictionary and tracking algorithm are both robust and efficient.

Keywords
Dictionary construction, bidirectional, model drifting

Date received: 27 March 2020; accepted: 5 May 2020

Topic Area: Robot Vision for Surveillance and Security
Topic Editor: Antonio Fernandez-Caballero
Associate Editor: Jianhua Zhang

Introduction
In many industry scenarios, such as intelligent traffic systems, industrial robots, and smart security surveillance systems, visual sensors have become increasingly more common due to their low cost and nonintrusiveness. Object tracking is an important problem in visual sensor-based intelligent systems and has been studied extensively in recent decades. Technically, object tracking involves locating a specified region in a video sequence and has significant potential applications in various fields, including visual surveillance, intelligent transportation systems, human–computer interaction, and intelligent driving.

A variety of tracking algorithms have been investigated in the existing literature. From the perspective of appearance modeling, there are two kinds of tracking algorithms: generative and discriminative. For generative models, the modeling of the object using Gaussian mixture models is both effective and computationally efficient. Zhou et al. assume that the observations are composed of different components and modeled the object by utilizing an adaptive mixture of Gaussians. Taking the spatial distribution of the object into consideration, Yu and Wu propose a spatial-appearance model that captures the properties of both local appearance changes and global spatial changes to fit nonrigid appearance variations. In the literature, a spatial-color mixture of Gaussians model is introduced to model the object. It considers not only the common similarity measure based on color histograms but also the spatial layout of the colors.

College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou, China

Corresponding authors:
Li Zhao and Hui Huang, College of Computer Science and Artificial Intelligence, Wenzhou University, Wenzhou 325000, China.
Email: lizhao@wzu.edu.cn; huanghui@wzu.edu.cn
These generative model-based tracking algorithms all use pixelwise object representation and are apt to be corrupted by noise. To consider the object appearance as a whole, linear subspace learning, which represents the object as a vector, has been widely applied to visual tracking. Based on the subspace constancy assumption, Black and Jepson\textsuperscript{2} propose a subspace learning algorithm for template tracking. However, this algorithm does not work well if the appearance of the object gradually changes. The model proposed in the literature\textsuperscript{10,11} learns and presents the object appearance by low-dimensional subspace in an incremental way, and it can therefore efficiently capture nonrigid appearance variations and recover all motion parameters efficiently. In the literature,\textsuperscript{12} the observation model is decomposed into multiple basic observation models that are constructed by the sparse principal component analysis of a set of feature templates. To utilize more spatial layout information, higher-order subspace learning algorithms are proposed.\textsuperscript{13–16} In the literature,\textsuperscript{13,14} an online tensor decomposition framework is introduced for object tracking. It can adapt to the appearance changes of a target by gradually learning a low-order tensor eigenspace representation. Due to the fact that the appearance variations are highly nonlinear, nonlinear manifold learning methods have been proposed. In the research,\textsuperscript{17} Porikli et al. utilize a covariance matrix\textsuperscript{18} descriptor to capture the spatial correlation information in the appearance of object.

While for discriminative trackers, many models are proposed to select different discriminative features for tracking. The support vector tracker\textsuperscript{19} uses an offline learned support vector machine as the classifier and embeds it into an optical flow-based tracker. In the literature,\textsuperscript{20} a discriminative classification rule is learned to distinguish between the object and background. These algorithms require a large hand-labeled data set for training, and the support vector machine classifier is not updated once trained. To adapt to the object appearance changes, discriminative trackers have been extended to include online learning. Collins et al.\textsuperscript{21} classify tracking as tracked object and background. A variance ratio is used to measure feature discriminability and select the best color space feature from a feature pool for tracking. Avdan\textsuperscript{22} labels pixels by a combination of some weak classifiers and constructs a probability map to represent the probabilities that particular pixels belong to the tracking object or its background. However, the pixel-based features (color and gradient) have very limited discriminative power, especially if the background shares a similar color with the object. To overcome this disadvantage, Grabner et al.\textsuperscript{23,24} select discriminative local tracking features from a large feature pool by online boosting. The weak classifiers discriminate the object from the background to obtain corresponding features. Zhang et al.\textsuperscript{25} propose a graph embedding-based subspace learning method, which can simultaneously learn the subspace of the object and its local discriminative subspace against the background. Li et al.\textsuperscript{26} propose a novel correlation filter-based tracker, which is robust to background clutters and scale variations of the target. Zhou et al.\textsuperscript{27} exploit appearance and the background context to design a robust correlation filter-based tracker.

In recent years, $l_1$-norm constrained sparse representation (SR) has attracted increasingly more attention and applied to object tracking.\textsuperscript{28–34} Despite the great success of SP in the field of tracking, less research is focused on how to establish an effective visual tracking template in the dictionary. SR requires an overcomplete template dictionary, so the linear combination of these templates can be used to approximate the estimation of new samples with very sparse coefficients. For an online video sequence, the preliminary construction of an overcomplete dictionary beforehand is difficult or even impossible. Therefore, a dictionary with a small number of atoms is collected when tracking is started, and the atoms are updated online during the tracking. However, there are two limitations to this method: (1) the atoms are far from complete and (2) the atoms are gradually contaminated by the tracking errors, ultimately resulting in drift problems. In contrast, because all video frames are available, overcomplete atoms can be built for offline video sequences. Thus, the construction of a dictionary with overcomplete atoms is a key problem that remains to be solved, and some methods have been proposed in the signal processing field to address this problem.\textsuperscript{35–38} In the literature,\textsuperscript{35} Aharon et al. propose a method alternates between a process of updating the dictionary atoms and sparse-coding the examples based on the current dictionary. Yaghoobi et al.\textsuperscript{36} introduce diverse constraints to spread the dictionary learning problem and use optimized methods to solve it. In the literature,\textsuperscript{37} an iterative algorithm based on the least-squares cost is proposed to construct dictionaries. Nevertheless, the labeling of so much sample data in a video is time-consuming.

From the preceding analysis of existing research, an effective dictionary construction method for offline video tracking is proposed in the present study. The major characteristics of the proposed method are summarized as follows:

- Three categories of atoms are constructed: nonpoluted atoms and their linear combination, variational atoms, and noise atoms. All the atoms are selectively updated to capture the appearance changes and alleviate the model-drifting problem.
- The algorithm adopts a two-step method, which effectively solves the optimization problem via two sets of SP and reduces the huge computational burden.
- From the perspective of control theory, the presented pursuit algorithm combines the key frame constraint and bitracking constraint, which make the open-loop essence of the tracking problem well-posed.
To this end, equation (1) is reformulated as

\[ m \approx V a = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n \]  

where \( a = (a_1, a_2, \cdots, a_n)^T \) denotes the vector of coefficient parameter. If the number of atoms \( n \) is large, \( m \in \mathbb{R}^d \) is usually sparse.

However, the image samples \( m \) in visual applications are inevitably degenerated because of noise or occlusion. To this end, equation (1) is reformulated as

\[ m = V a + e = a_1 v_1 + a_2 v_2 + \cdots + a_n v_n + e \]  

where \( e \) denotes the noise vector. Once a pixel in \( m \) is polluted, the corresponding entry of \( e \) is nonzero. Nevertheless, for different image candidates, the locations of the damaged pixels are different and unknown. Additionally, the magnitudes of errors are random. By combining these noise atoms, the error vector \( e \) is expressed as

\[ e = 1 e = e_1 i_1 + e_2 i_2 + \cdots + e_k i_k \]  

where \( i_k \) is a vector, whose \( k \)‘th item is 1 and other items are 0. Based on equation (3), equation (2) can be reformulated as follows

\[ m = [V I] \begin{bmatrix} a \\ e \end{bmatrix} = Bp \]  

Figure 1. The flowchart of the proposed tracking algorithm.

The remainder of the article is organized as follows. The “SP-based tracker” section discusses some work of the SP framework in the tracking context. We present the detail of the proposed tracker in “Proposed tracking system” section. Some results are shown in “Experimental results” section, and conclusions are made in “Conclusion” section.

Motivation

Before introducing the motivation of this article, a short review of traditional SP-based tracker is first presented to make this article self-contained.

SP-based tracker

In the applications of object tracking, it is assumed that the manifold of the object lies in a linear subspace for a short period of time. The assumption is rational because the appearances of the object are similar among the consecutive period of time. The assumption is rational because the manifold of the object lies in a linear subspace for a short period of time. In the applications of object tracking, it is assumed that the object changes, it can be represented by some atoms.

Inspired and guided by the key frame-based trackers, this article proposes the construction of a valid and large template dictionary via the use of a key frame-based algorithm. Because the goal is to collect as many representative atoms as possible, the objects in several key frames are
manually selected as the nonpolluted atoms; however, these atoms are not sufficient. To enlarge the atom set, three categories of atoms are introduced, namely (1) the given nonpolluted atoms and their linear combinations, (2) some variational atoms that are used to adapt to the appearance changes, and (3) some noise atoms that deal with occlusion and noise. The noise atom was defined in “SP-based tracker” section, and the other two categories of atoms are introduced in the following subsections.

**Nonpolluted atoms.** The upper portion of Figure 2 shows that in the user-specified $k$ key frames, the target area to be traced is manually marked. For the $j$th frame, the target region of disturbed 0–2 pixels is to generate some new image regions. The mismatch problem can be alleviated via intensive sampling around the target area. The cropped regions are then adopted as the nonpolluted atoms $D_{ni} = \{d_{n1i}, d_{n2i}, \ldots, d_{nji}\}$, where $d_{nji}$ denotes for the $j$th non-polluted atom. In this way, for all key frames, the nonpolluted atom set is constructed as $D_n = \{D_{n1}, D_{n2}, \ldots, D_{nk}\}$.

**Variational atoms.** To effectively capture the changes in the appearance of the target object, the variational atoms can be initialized by the linear combination of the randomly selected nonpolluted atoms in the two corresponding frames (see the second line of Figure 2). More precisely, consider the subsequence between the first and second key frames. Let $D_v = \{d_{v1}, d_{v2}, \ldots, d_{vj}\}$ denote the variational atom set. The atoms for the changed atom group are as follows

$$d_{ij} = \alpha d_{n1}^2 + (1 - \alpha)d_{n2}^2$$

where $\alpha$ is a random weight that is uniformly generated in the interval $[0, 1]$. In addition, with the use of the atom updating strategy (see “Atom updating” section), the variations of the object appearance can be effectively captured by these atoms. By adding the variational atoms and noise atoms, the complete dictionary is constructed as $\mathcal{D} = [D_n, D_v, D_i]$.

**Two-step SP optimization**

According to the preceding section, the number of atoms inside the dictionary is large, and solving equation (5) is therefore time-consuming. To solve this problem, two-step SP optimization is proposed.

The notations used in this article are first reviewed. The dictionary contains the nonpolluted atoms $D_n = \{D_{n1}, D_{n2}, \ldots, D_{nk}\}$, the variational atoms $D_v$, and the noise atoms $D_i$. For example, when the tracker is conducted between the key frames 1 and 2, atoms $D_{n1}$ and $D_{n2}$ are first selected, and the variational atoms $D_v$ are then generated by equation (6) based on $D_{n1}$ and $D_{n2}$. For the remaining non-polluted atoms $D_n' = \{D_{n1}', \ldots, D_{nk}'\}$, a selection algorithm is used as follows.

**Algorithm 1.** Step I: selection algorithm.

**Input:** non-polluted atoms $D_n' = \{D_{n1}', \ldots, D_{nk}'\}$ and a candidate sample $y$.

1. Compute the distance between $y$ and every atom $d_i$ in $D_n'$, next get a nearest neighbor subset $I$ in $D_n'$: $D_n'' = \{D_n'|d_i \in D_n'\}$ and $i \in I$).
2. Handle the non-negative least squares problem:

$$\beta^* = \arg \min_{\beta} ||D_n'' \beta - y||_2, \; \text{s.t.} \; \beta \geq 0$$

3. The non-polluted templates are selected from $D_n''$ as follows:

$$D_n^* = \{D_n''|d_i \in D_n'' \text{ and } \beta_i^* \geq 0\}$$

**Output:** the selected atoms $D_n^*$.

Accordingly, most of the atoms in $D_n'$ will not be used in the optimization process, resulting in a significant acceleration. All the atoms used in the final dictionary are $\mathcal{D} = \{D_{n1}, D_{n2}, D_v, D_i\}$. The optimization problem in step II is conducted, as given by equation (5).
Bitracking procedure

To take both forward and backward sequential information into consideration, the tracker is managed by a bitracking procedure. As shown in Figure 3, the tracking process is not conducted in the chronological order. The tracking process of the left part of the figure is responsible for obtaining forward sequential information and that of the right part of the figure enables the capture of backward sequence information. The tracker begins from frame 1; the first tracking step is from frame 1 to frame 3, the second tracking step is from frame 2 to frame 4, the third tracking step is from frame 3 to frame 5, and so on. The tracking procedure stops until the end of the forward and backward loops. In this way, the object in every frame has two tracking results. To obtain more reliable results, the frame with the smallest difference between the two results is selected as the intersection of the bitracking process.

Atom updating

In the majority of tracking applications, the changes of both the target and the environment must be handled by the tracker simultaneously. If the atoms used in the tracker are updated frequently, the atom will be gradually polluted by tracking errors, leading to the model-drifting problem. Therefore, it is necessary to design an appropriate updating strategy for tracking. In the proposed model, nonpolluted atoms and variational atoms are updated online in different manners.

- The nonpolluted atoms are manually labeled and generated before the tracker begins. During the tracking procedure, the nonpolluted atoms are not updated.
- The variational atoms are generated, as given by equation (6), and updated in the same manner, as in the literature.28

Discussion

This section discusses the reasons for the effectiveness of the proposed dictionary and tracking process.

From the perspective of control theory, tracking is essentially an open-loop problem; there is no feedback in the tracking process. Therefore, tracking errors inevitably accumulate, leading to the model-drifting problem. To alleviate this problem, some special constraints must be introduced. Traditionally, there are two kinds of constraints, the first of which is the key frame-based constraint.41,42 In this work, the manually labeled ground truth in the key frames acts as special feedback. The optimization is conducted on the whole sequence, which minimizes the tracking errors in the key frames. The second kind of constraint is the bidirectional tracking constraint,43,44 which leads to a new minimization criterion that combines both the forward and backward tracking errors. Both types of constraints can improve the robustness and accuracy of tracking; however, they are time-consuming and therefore cannot fulfill the real-time requirement.

In the present work, the object templates in the key frames can be naturally incorporated into the SP framework and are used for the tracking of the whole sequence. Additionally, the optimization process is solved efficiently by the algorithm in “Two-step SP optimization” section.

Iterative extension

As stated in “Discussion” section, tracking is an open-loop problem that is inevitably corrupted by image noise. Although the labeled templates from key frames provide a constraint for the drifting problem, the tracker still cannot successfully track any object in arbitrary video sequences.

The proposed tracking framework enables an iterative way to refine tracking results when the performance of the tracker is not satisfactory. When the tracker deviates from the true position of the object and can never be recovered again, the tracking process is paused, the most representative frames are selected, and the image region of the object is extracted. By randomly linear combination of the regions in the two corresponding frames, their offsprings are generated. They are then added into the template dictionary, and the tracking process is restarted. From the theoretical analysis in control domain, it is found that the interactive process together with the key frame-based constraints essentially forms a feedback to the tracking process and thus can greatly improve the tracking performance in theory.

Experimental results

Two comparative experiments, each involving different dictionaries and tracking processes, are first presented to verify the asserted contributions of this work. Next, to confirm the performance of the proposed tracker, several traditional tracking algorithms are compared, and an iterative tracking example is presented. The average pixel error is adopted to measure the tracking accuracies of the different methods.

Different dictionaries

In this section, the proposed method is compared with the classic $L_1$ tracker.28 To make a fair comparison, the constraints of the key frames and the bitracking procedure are not used.
As shown in Figure 4, when the object undergoes large changes in pose and illumination, the $L_1$ tracker is unable to follow the object quickly. There are two reasons for these results: (1) The atoms adopted in the $L_1$ tracker are inadequate, so they cannot capture the appearance changes, thus leading to tracking failure. (2) The template updated in the $L_1$ tracker introduces errors into the template, and thus, the template deviates from the tracking target, leading to the drifting problem. In our method, the nonpolluted templates are chosen from the key frames to construct the dictionary, which has significant effect on avoiding target drifting. Figure 5 shows that, as compared with the $L_1$ tracker, the proposed method achieves superior tracking performance.

**Different tracking procedures**

To demonstrate bitracking process outperforming the normal tracking procedure, the proposed algorithm is tested on two sequences via the use of the two tracking procedures, respectively.
As shown in Figure 6, the normal tracking process cannot track the sudden motion of the target, although many nonpolluted templates are used. In contrast, the bitracking process obtains accurate results, as the motion of the object is estimated from both tracking directions, resulting in the improvement of the robustness of the tracking algorithm.

As shown in Figure 7, the bitracking process achieves better performance than the normal tracking process.

Comparison with state-of-the-art methods

The proposed algorithm (normal tracking process and bitracking process) is compared with several state-of-the-art tracking algorithms, namely the $L_1$ tracker,\textsuperscript{28} incremental visual tracking (IVT),\textsuperscript{10} semi-supervised online boosting (SSOB),\textsuperscript{24} and multiple instance learning (MIL).\textsuperscript{45} Furthermore, these methods are tested on multiple video
sequences that contain illumination changes, occlusion, background interference, and posture changes.

In the first experiment, the proposed algorithm is compared with SSOB and MIL. As shown in Figure 8, MIL fails to track the object in the second image and remaining sequences, because the employed Haar feature in MIL lost its discrimination due to the large change in illumination. SSOB loses track of the object in the fourth image. As shown in Figure 9, the average error of the proposed algorithm is notably much lower than those of SSOB and MIL. These results demonstrate that the proposed algorithm exhibited an accurate tracking performance superior to those of SSOB and MIL.

In the second experiment, all the tracking algorithms, including the normal tracking process, bitracking process, $L_1$ tracker, MIL, IVT, and SSOB, are compared. Figure 10 shows that most of the algorithms involved in the experiment achieve good results, excluding SSOB and MIL in the car sequence. The reason for this is that SSOB and MIL are susceptible to illumination changes and deformations, and thus tracking failure occurs. As presented in Table 1, the proposed method achieved the best performance in terms of the tracking speed. These results demonstrate that the proposed method not only exhibits accurate tracking performance but is also characterized by a greatly reduced time consumption.

## Conclusion

This article proposes a drift-free visual tracking algorithm based on a constructed template dictionary. A set of templates, namely some nonpolluted templates, their offspring, one stable template, and variable templates, is used in the dictionary. To accommodate changes and prevent the model-drifting problem, these templates are selectively updated. In addition, the tracking process is bidirectional, which results in improved tracking performance of the proposed algorithm. The effectiveness of the proposed tracking algorithm is proven by several comparison experiments.

## Declaration of conflicting interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

## Funding

The author(s) disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported in part by the National Natural Science Foundation of China [Grant No. 61922064], in part by the Zhejiang Provincial Natural Science Foundation [Grant Nos LR17F030001 and LQ19F020005], in part by the Project of Science and Technology Plans of Wenzhou City [Grant Nos C20170008, G20150017, and ZG2017016].

## ORCID iD

Li Zhao  
https://orcid.org/0000-0001-5787-2705

## References

1. Zhang X, Li C, Tong X, et al. Human pose estimation and tracking via parsing a tree structure based human model. *IEEE Trans Syst Man Cybern* 2014; 44(5): 580–592.
2. Zhou S, Chellappa R, and Moghaddam B. Visual tracking and recognition using appearance-adaptive models in particles filters. *IEEE Trans Image Process* 2004; 13(11): 1434–1456.
3. Zhang X, Li W, Ye X, et al. Robust hand tracking via novel multi-cue integration. *Neurocomputing* 2015; 157: 296–305.
4. Xu J, Park S, and Zhang X. A bio-inspired motion sensitive model and its application to estimating human gaze positions under classified driving conditions. *Neurocomputing* 2019; 345(14): 23–35.
5. Yu T and Wu Y. Differential tracking based on spatial appearance model (SAM). In: Proceedings of IEEE conference on computer vision and pattern recognition, New York, 17–22 June 2006, pp. 720–727.

6. Zhang X, Hu W, Bao H, et al. Robust head tracking based on multiple cues fusion in the kernel-bayesian framework. *IEEE Trans Circuits Syst Video Technol* 2013; 23(7): 1197–1208.

7. Wang H, Suter D, Schindler K, et al. Adaptive object tracking based on an effective appearance filter. *IEEE Trans Pattern Anal Mach Intell* 2007; 29(9): 1661–1667.

8. Zhang X, Hu W, Xie N, et al. A robust tracking system for low frame rate video. *Int J Compu Vis* 2015; 115(3): 279–304.

9. Black MJ and Jepson AD. Eigen tracking: robust matching and tracking of articulated objects using a view-based representation. *Int J Comput Vision* 2004; 26(1): 63–84.

10. Ross DA, Lim J, Lin RS, et al. Incremental learning for robust visual tracking. *Int J Comput Vision* 2008; 77: 125–141.

11. Zhang X, Hu W, Qu W, et al. Multiple object tracking via species-based particle swarm optimization. *IEEE Trans Circuits Syst Video Technol* 2010; 20(11): 1590–1602.

12. Kwon J and Lee KM. Visual tracking decomposition. In: *Proceedings of IEEE conference on computer vision and pattern recognition*, San Francisco, 13–18 June 2010, pp. 1269–1276.

13. Hu W, Li X, Zhang X, et al. Incremental tensor subspace learning and its applications to foreground segmentation and tracking. *Int J Comput Vis* 2011; 91(3): 303–327.

14. Zhang X, Shi X, Hu W, et al. Visual tracking via dynamic tensor analysis with mean update. *Neurocomputing* 2011; 74(17): 3277–3285.

15. Wu Y, Cheng J, Wang J, et al. Real-time visual tracking via incremental covariance tensor learning. In: *Proceedings of IEEE international conference on computer vision*. Kyoto, 29 September–2 October 2009, pp.1631–1638.

16. Zhang X, Wang D, Zhou Z, et al. Robust low-rank tensor recovery with rectification and alignment. *IEEE Trans Pattern Anal Mach Intell* 2020. DOI: 10.1109/TPAMI.2019.2929043.

17. Porikli F, Tuzel O, and Meer P. Covariance tracking using model update based on Lie algebra. In: *Proceedings of international conference on computer vision and pattern recognition*. New York, 17–22 June 2006, pp. 728–735.

18. Tuzel O, Porikli F, and Meer P. Region covariance, a fast descriptor for detection and classification. In: *Proceedings of european conference on computer vision*, New York, NY, USA, 2006, pp. 589–600.

19. Avidan S. “Support vector tracking”. *IEEE Trans Pattern Anal Mach Intell* 2004; 26(8): 1064–1072.

20. Shen C, Kim J, and Wang H. Generalized kernel-based visual tracking. *IEEE Trans Circuits Syst Video Technol* 2010; 20(1): 119–130.

21. Collins R, Liu Y, and Leordeanu M. Online selection of discriminative tracking features. *IEEE Trans Pattern Anal Mach Intell* 2005; 25: 1631–1643.

22. Avidan S. Ensemble tracking. *IEEE Trans Pattern Anal Mach Intell* 2007; 29(2): 261–271.

23. Grabner H, Grabner M, and Bischof H. Real-time tracking via on-line boosting. In: *Proceedings of british machine vision conference*, 2006, pp. 46–56. Springer.

24. Grabner H, Grabner M, and Bischof H. Semi-supervised online boosting for robust tracking. In: *Proceedings of european conference on computer vision*, France, Marseille, 10–12 October, pp. 234–247.

25. Zhang X, Hu W, Chen S, et al. Graph-embedding-based learning for robust object tracking. *IEEE Trans Ind Electronics* 2014; 61(2): 1072–1084.

26. Li J, Zhou X, Chan S, et al. Robust object tracking via large margin and scale-adaptive correlation filter. *IEEE Access* 2017; 6: 12642–12655.

27. Zhou X, Li J, Chen S, et al. Multiple perspective object tracking via context-aware correlation filter. *IEEE Access* 2018; 6(1): 43262–43273.

28. Mei X and Ling H. Robust visual tracking using $L_1$ minimization. In: *Proceedings of international conference on computer vision*, Kyoto, Japan, 29 September–2 October 2009, pp. 1–8.

29. Liu B, Yang V, Huang J, et al. Robust and fast collaborative tracking with two stage sparse optimization. In: *Proceedings of european conference on computer vision*, 2010, pp. 624–637. IEEE.

30. Li H, Shen C, and Shi Q. Real-time visual tracking using compressive sensing. In: *Proceedings of international conference on computer vision and pattern recognition*, Providence, RI, USA, 20–25 June 2011, pp. 1305–1312. IEEE.

31. Zhang X, Li W, Hu W, et al. Block covariance based tracker with a subtle template dictionary. *Pattern Recogn* 2013; 46: 1750–1761.

32. Jia X, Lu H, and Yang M. Visual tracking via adaptive structural local sparse appearance model. In: *Proceedings of IEEE conference on computer vision and pattern recognition*. Providence, RI, USA, 16–21 June 2012, pp. 1822–1829. IEEE.

33. Zhang T, Ghanem B, Liu S, et al. Robust visual tracking via multi-task sparse learning. In: *Proceedings of IEEE conference on computer vision and pattern recognition*, Providence, RI, USA, 16–21 June 2012, pp. 2042–2049. IEEE.

34. Chan S, Zhou JX, and Li S. Adaptive compressive tracking based on locality sensitive histograms. *Pattern Recogn* 2017; 72: 517–531.

35. Aharon M, Elad M, and Bruckstein A. K-SVD: an algorithm for designing overcomplete dictionaries for sparse representation. *IEEE Trans Signal Process* 2006; 54(11): 4311–4322.

36. Yaghnoobi M, Blumsenas T, and Davies ME. Dictionary learning for sparse approximations with the majorization method. *IEEE Trans Signal Process* 2009; 57(6): 2178–2191.

37. Skretting K and Engan K. Recursive least squares dictionary learning algorithm. *IEEE Trans Signal Process* 2010; 58(4): 2121–2130.

38. Zhang X, Liu Q, Wang D, et al. Self-taught semisupervised dictionary learning with nonnegative constraint. *IEEE Trans Ind Inf* 2020; 16(1): 532–543.
39. Donoho D and Tsaig Y. Fast solution of $L_1$-norm minimization problems when the solution may be sparse. Preprint. 2006. http://www.stanford.edu/tsaig/research.html

40. Agarwala A, Hertzmann A, Salesin D, et al. Keyframe-based tracking for rotoscoping and animation. *ACM Trans Graph* 2005; 24(3): 584–591.

41. Wei Y, Sun J, Tang X, et al. Interactive offline tracking for color objects. In: *Proceedings of international conference on computer vision*, Rio de Janeiro, Brazil, 14–21 October 2007, pp. 1–8. IEEE.

42. Wei X and Chai J. Interactive tracking of 2D generic objects with spacetime optimization. In: *Proceedings of european conference on computer vision*, Marseille, France, 12–18 October 2008, pp. 657–670.

43. Sun J, Zhang W, Tang X, et al. Bidirectional tracking using trajectory segment analysis. In: *Proceedings of international conference on computer vision*, Nice, France, France, 13–16 October 2005, pp. 717–724.

44. Wu H, Chellappa R, Sankaranarayanan A, et al. Robust visual tracking using the time-reversibility constraint. In: *Proceedings of international conference on computer vision*, Rio de Janeiro, Brazil, 14–21 October 2007, pp. 1–8. IEEE.

45. Babenko B, Yang MH, and Belongie S. Robust object tracking with online multiple instance learning. *IEEE Trans Pattern Anal Mach Intell* 2010; 33: 1619–1632.