

Research Article

Visual Tracking Based on Discriminative Compressed Features

Wei Liu ¹ and Hui Wang²

¹Department of Modern Education Technology, Ludong University, Yantai, China

²Lab, CNCERT/CC, Yumin Road No. 3A, Beijing 100029, China

Correspondence should be addressed to Wei Liu; ldulw@sina.com

Received 3 April 2018; Revised 13 June 2018; Accepted 11 July 2018; Published 1 August 2018

Academic Editor: Lei Zhang

Copyright © 2018 Wei Liu and Hui Wang. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Visual tracking is a challenging research topic in the field of computer vision with many potential applications. A large number of tracking methods have been proposed and achieved designed tracking performance. However, the current state-of-the-art tracking methods still can not meet the requirements of real-world applications. One of the main challenges is to design a good appearance model to describe the target's appearance. In this paper, we propose a novel visual tracking method, which uses compressed features to model target's appearances and then uses SVM to distinguish the target from its background. The compressed features were obtained by the zero-tree coding on multiscale wavelet coefficients extracted from an image, which have both the low dimensionality and discriminate ability and therefore ensure to achieve better tracking results. The experimental comparisons with several state-of-the-art methods demonstrate the superiority of the proposed method.

1. Introduction

Visual tracking aims at locating the target of interest from an image sequence, which is one of the most activated research topics in the field of computer vision with many potential applications such as video surveillance, human-computer interaction, navigation, and automatic driving. It has attracted increasing interest in the past few decades [1–16]. However, due to a variety of challenging factors such as illumination changes, pose deformation, and occlusion, the performance of visual tracking is still far away from requirements in practical applications. The main difficulty is that it is not easy to design a good appearance modeling method, which is not only good at distinguishing the target from its background but also being robust to the above-mentioned appearance changes. Finding a good appearance modeling is a challenging problem in many visual applications such as image classification [17–19] and video recognition [20–22].

In the literature, there are a variety of visual tracking methods with focus on developing effective appearance modeling methods. Most of these methods can be classified into two groups: generative methods and discriminative methods. The former learns generative features from samples that only contain the target, whose purpose is to represent the target as

accurate as possible. The latter learns discriminative features from samples including both the target and its background, which usually involves solving an optimization function. To achieve better tracking performance, discriminative methods attracted more attention.

In this paper, to overcome the challenges caused by low contrast, illuminative changes, and scale changes, we propose a novel tracking method using discriminative compressed features, which is real-time and able to process multiple scales of the target. The main idea of the proposed method is that it combines compressive sensing and multiscale texture transformation to extract compressed texture features and then uses SVM to classify the target from its background. The compressed features have both the low dimensionality and discriminate ability and therefore ensure to achieve better tracking results. The experimental comparisons with several state-of-the-art methods demonstrate the superiority of the proposed method.

The rest of this paper is organized as follows. In Section 2, we review the work closely related to our proposed approach. Section 3 gives a detailed description of the proposed tracking method. Experimental results are reported and analyzed in Section 6. We conclude this paper in Section 6.

2. Related Work

In the past decades, there are many tracking methods that have been proposed, which can be roughly divided into generative methods and discriminative methods. The former focuses on modeling the appearance of the tracked target and then finds the candidate that is the most similar to the target template as the tracking result. The representative methods include those trackers based on sparse representation [23–29]. In [29], sparse coding is used to extract features from sampled patches. The local sparse features are then pooled into a global representation. In [28], an online learning sparse representation is proposed for visual tracking to handle occlusion. In [25], a joint sparse representation framework is used to combine multi-cue features for visual tracking. Since features from different cues describe the tracked target from different aspects, more robust tracking results can be obtained when multi-cue features are used. In [23], a biologically inspired appearance model is proposed to model target appearance, which is also based on features extracted using sparse coding.

The discriminative methods learn a binary classifier, which is then used to classify a candidate as the target or background [5, 8, 14, 16, 30–34]. In [30], Yakut and Kehtarnavaz proposed to track ice-hockey pucks by combining three pieces of information in ice-hockey video frames using an adaptive gray-level thresholding method. In [31], Topkaya et al. proposed a multiple object tracking method using tracklet clustering, which first obtains short yet reliable tracklets and then clusters the tracklets over time based on color and spatial and temporal attributes. In [32], Wang and Zhao proposed an adaptive appearance model called Principal Component-Canonical Correlation Analysis (P3CA) to extract discriminative features for object tracking. In [14], Qi et al. propose a CNN based tracking method, which uses correlation filters to construct six weak trackers on outputs of six CNN layers. These weak trackers are then adaptively combined by a Normal Hedge algorithm. In [34], a further improved method is proposed which uses a SNT to compute the loss of each weak tracker, which achieves better tracking performance.

3. Discriminative Compressed Features

3.1. Multiscale Wavelet Transformation. Multiscale wavelet is a kind of wavelet which consists of more than two scale functions. It preserves the local properties of time-frequency domains while overcoming the drawbacks of a single wavelet and therefore has more properties of different frequencies. In this paper, we choose the GHM multiscale wavelet [35], which can be obtained by recursively calculating as follows:

$$v_{j,k} = \sum_m G_{m-2k} v_{j-1,m} \quad (1)$$

$$w_{j,k} = \sum_m H_{m-2k} v_{j-1,m} \quad (2)$$

where $v_{j,k}$ and $w_{j,k}$ are low-frequency coefficients and high-frequency coefficients of the j th scale of the input signal, respectively. $v_{j-1,m}$ denotes the low-frequency coefficients of

the $(j-1)$ th scale; k and m are the indices of the current scales, which are dependent on the input image. The multiwavelet filters are defined as

$$G_0 = \begin{bmatrix} \frac{3}{5\sqrt{2}} & \frac{4}{5} \\ -\frac{1}{20} & -\frac{3}{10\sqrt{2}} \end{bmatrix} \quad (3)$$

$$G_1 = \begin{bmatrix} \frac{3}{5\sqrt{2}} & 0 \\ \frac{9}{20} & \frac{1}{\sqrt{2}} \end{bmatrix}$$

$$G_2 = \begin{bmatrix} 0 & 0 \\ \frac{9}{20} & -\frac{3}{10\sqrt{2}} \end{bmatrix} \quad (4)$$

$$G_3 = \begin{bmatrix} 0 & 0 \\ -\frac{1}{20} & 0 \end{bmatrix}$$

$$H_0 = \frac{1}{10} \begin{bmatrix} -\frac{1}{2} & -\frac{3}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & 3 \end{bmatrix} \quad (5)$$

$$H_1 = \frac{1}{10} \begin{bmatrix} \frac{9}{2} & -\frac{10}{\sqrt{2}} \\ -\frac{9}{\sqrt{2}} & 0 \end{bmatrix}$$

$$H_2 = \frac{1}{10} \begin{bmatrix} \frac{9}{2} & -\frac{3}{\sqrt{2}} \\ \frac{9}{\sqrt{2}} & -3 \end{bmatrix} \quad (6)$$

$$H_3 = \frac{1}{10} \begin{bmatrix} -\frac{1}{2} & 0 \\ -\frac{1}{\sqrt{2}} & 0 \end{bmatrix}$$

3.2. Compressed Multiscale Features. It is easy to obtain low-frequency components and high-frequency components after the signals are filtered by wavelet transformation. In general, most energy of the signal is in the low-frequency components. In contrast, high-frequency components of the signal reflect the details of the input image. Therefore, the simplest way of compressing the input image is to set the high-frequency coefficients to be zero when reconstructing the input image using wavelet transformation. The other option is to set the high-frequency coefficients of some local regions to be zero or to set the high-frequency coefficients based on a threshold, which will cause severe loss of image details, blurred images after compression, or loss of image information.

Wavelet transformation is able to composite the input image at different scales. More importantly, the subimage at each resolution has different frequency properties and different orientation selections. Therefore, it can be used to

encode different information of the input image at different scales.

It is widely thought of the fact that the targets in a video sequence are redundant in both spatial and frequency domains. The former indicates the adjacent pixels have spatial correlation. The latter indicates that the adjacent frequencies of a pixel have some kinds of correlation. On the other hand, the statistical features of image signals indicate that large coefficients always exist in low-frequency regions and therefore small bits can be assigned to those small coefficients or they will not be transmitted at all. It will cause high compression rates and very small information loss.

The compression method based on multiscale wavelet transformation applies the zero-tree coding to compression of high spectral images. The principle behind this method is that it exploits the structure correlation of high spectral images to construct only one effective (shared) image and then further determine the positions of nonzeros of multiscale wavelet coefficients. The shared image is obtained by combining multiscale frequency coefficients and therefore removes spatial redundancy and frequency redundancy with the purpose of improving compression efficiency.

The one-dimensional wavelet transformation filters the input signal by low-pass filtering and high-pass filtering and then obtains low-frequency components and high-frequency components by downsampling. According to Mallat algorithm, two-dimensional wavelet transformation can be implemented by several one-dimensional wavelet transformation and obtain low-frequency and high-frequency components, respectively. Given an input image with m rows and n columns, the process of 2D wavelet transformation is that it first decomposes the input image along its each row using 1D wavelet transformation, which will obtain L and H two parts. The second step is to decompose the L and H parts along its column using 1D wavelet transformation. With these two steps, the input image will get four parts (LL, HL, LH, and HH). The second level, third level, or higher level's wavelet transformation can be obtained by using such a process on the former level. Therefore, the wavelet transformation is an iterative process.

To meet the real-time requirements, the dimensionality of appearance features should not be too high. To meet this requirement, in this paper, we adopt compressive sensing to reduce the dimensionality of high-dimensional appearance features. Let $u \in \mathbf{R}^D$ be the wavelet features and Γ be a random matrix computed using the same method as in [26]. The compressed features $v \in \mathbf{R}^d$ can be computed as $v = \Gamma u$.

4. Discriminative SVM Tracking

SVM is for classic binary pattern classification since it was proposed by Vapnik in 1995. In this paper, we use SVM as our tracking model.

4.1. SVM Tracking. To classify the target from its background, our tracking method tries to find a hyperplane in the D -dimensional compressed feature space to distinguish the features of the target and its background.

To achieve this aim, the optimization objective is to maximize the classifier's margin in the feature space. In other words, we need to meet the following conditions:

$$x_i \cdot w + b \geq 0 \quad \text{if } y_i = +1 \quad (7)$$

$$x_i \cdot w + b \leq 0 \quad \text{if } y_i = -1 \quad (8)$$

where y_i is the class label of the i th sample. For example, if the sample is target, $y_i = +1$. Otherwise, if the sample is background, $y_i = -1$.

Given training samples and their corresponding labels, we first extract compressed features from each sample using the method introduced in Section 3. The features with their labels can then be fed to SVM to train SVM's parameters. In the tracking stage, for each target candidate, we can also extract the compressed features using the same method as like in the training stage. Then we can feed the extracted features to SVM to predicate its label. If the features are classified as +1, it is considered as the potential target. Otherwise, it is not considered as the potential target. The final target is selected as the potential target candidate with the largest probability.

4.2. Model Update. To make the proposed tracker adapt to target appearance changes over time, the tracker needs to be updated online. To this aim, we update the model using the collected positive and negative samples. In particular, we collect a set of positive and negative samples at time t . Using the proposed appearance model, we can extract the compressed features for all positive and negative samples. Then the SVM model can be updated as

$$u_t^1 = \lambda u_t^1 + (1 - \lambda) u_1^1 \quad (9)$$

$$u_t^0 = \lambda u_t^0 + (1 - \lambda) u_1^0 \quad (10)$$

$$\delta_t^1 = \sqrt{\lambda (\delta_t^1)^2 + (1 - \lambda) (\delta_1^1)^2 + \lambda (1 - \lambda) (u_t^1 - u_1^1)^2} \quad (11)$$

$$\delta_t^0 = \sqrt{\lambda (\delta_t^0)^2 + (1 - \lambda) (\delta_1^0)^2 + \lambda (1 - \lambda) (u_t^0 - u_1^0)^2} \quad (12)$$

where λ denotes the learning rate, which controls the speed of model updating.

$$u^1 = \frac{1}{\alpha} \sum_{k=1}^{\alpha} v_{1,l}^k \quad (13)$$

$$u^0 = \frac{1}{\alpha} \sum_{k=1}^{\alpha} v_{0,l}^k \quad (14)$$

$$\delta^1 = \sqrt{\frac{1}{\alpha} \sum_{k=1}^{\alpha} (v_{1,l}^{(k)} - u^1)^2} \quad (15)$$

$$\delta^0 = \sqrt{\frac{1}{\alpha} \sum_{k=1}^{\alpha} (v_{0,l}^{(k)} - u^0)^2} \quad (16)$$

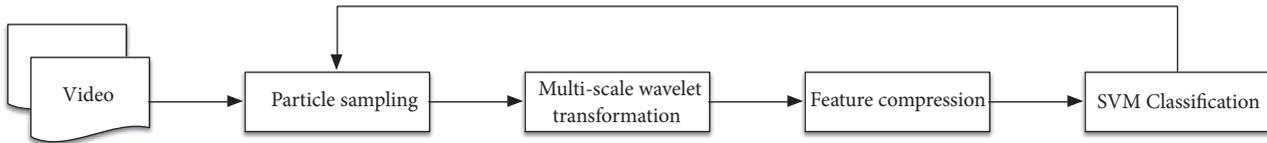


FIGURE 1: The flowchart of the proposed tracking method.

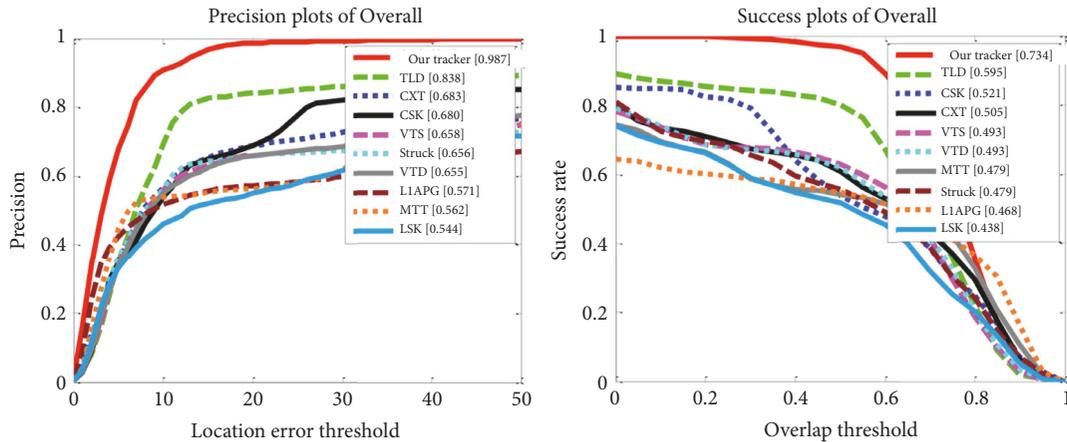


FIGURE 2: Overall precision plots and success plots on the test sequences.

5. Experiment Results

The target tracking is implemented in a particle filter framework. Several sequences from the OTB100 dataset have been chosen to evaluate the proposed tracking method. At the first frame, the target is initialized manually. Of course, the target can be initialized by a detector when the method is applied in real systems. After the target is initialized, a set of particles are sampled around the target. Whether each particle is considered as the target or not is based on the output of SVM scoring. In the next frame, the particles are sampled using the tracking result in the last frame as mean and a predefined covariance. The process is repeated frame by frame. The flowchart of the proposed tracking method is shown in Figure 1.

To test the performance of the proposed method, we compared the proposed method to several state-of-the-art trackers including TLD [36], CXT [1], Struck [37], LIAPG [38], and MTT [39]. By quantitatively and qualitatively analyzing the experimental results, we demonstrate the outstanding performance of the proposed method.

Two frame based metrics widely used in tracking performance evaluation are (1) center location error, which is defined as the Euclidean distance between the central location of the tracked target and the manually labeled ground-truthed position; (2) bounding box overlap which is the ratio of the areas of the intersection and the union of the bounding box indicating the tracked subject and the ground-truthed bounding box. To measure the overall performance of a tracker on a test sequence, success rate and precision

score are adopted. The former is computed as the percentage of image frames, which have a bounding box overlap larger than a given threshold. The latter is the percentage of image frames, which have a central position error less than a given threshold. In each case, when multiple thresholds are used, a curve is drawn to show how success rates or precision scores are affected by different thresholds. These curves are, namely, success plot and precision plot, respectively. In practical evaluations, we average the curves of a tracker over all the sequences, which have the same challenge and show a curve for each challenge item rather than a test sequence. In addition, we use the area under curve (AUC) of the success plot to quantitatively measure the overall performance of a tracker on a challenge item.

5.1. Quantitative Comparison. The overall precision plots and success plots are shown in Figure 2, from which we can see that the proposed method outperforms other methods in terms of the overall precision plots and success plots.

5.2. Qualitative Comparison. To further show the superiority of the proposed method, we show several examples of tracking results on Figures 3 and 4. As we can see from Figure 3, the proposed tracker outperforms other trackers on several representative frames on two sequences. More tracking results are shown in Figure 4, from which we can see that the proposed tracker also achieves the best tracking performance.

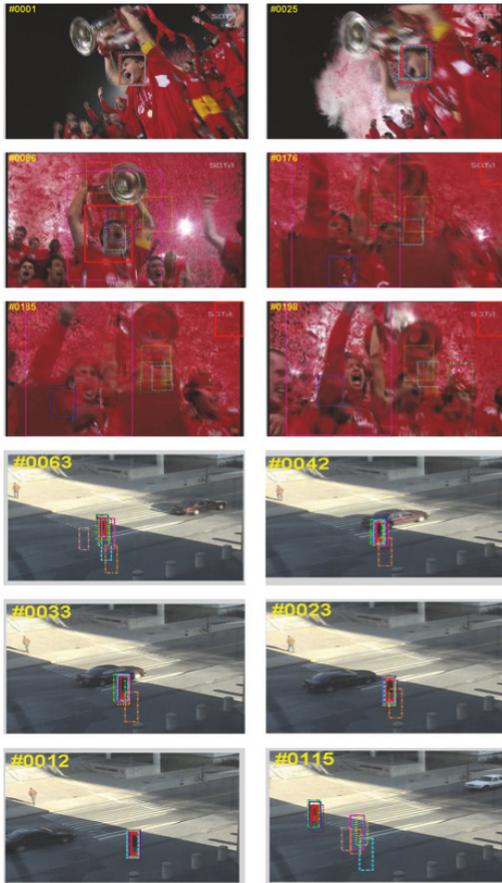


FIGURE 3: Examples of tracking results on representative frames of two sequences.

6. Conclusion

In this paper, we propose to use compressed features to model the tracked target's appearance and then use SVM to perform tracking. The experimental results indicate the proposed method outperforms several state-of-the-art methods. The advantages of the proposed method are twofold: (1) It is good at handling scale changes of the target over time because the used features are obtained by multiscale wavelet transformation. (2) The speed of the proposed method can achieve real-time because the dimensionality of the used features was reduced by compressed sensing techniques.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.



FIGURE 4: Examples of tracking results on representative frames of other four sequences.

Acknowledgments

The research is supported by Project of Shandong Province Higher Educational Science and Technology Program (no. J14LN64).

References

- [1] T. B. Dinh, N. Vo, and G. Medioni, "Context tracker: exploring supporters and distracters in unconstrained environments," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR '11)*, pp. 1177–1184, June 2011.
- [2] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "Exploiting the circulant structure of tracking-by-detection with kernels," in *Proceedings of the European Conference on Computer Vision*, pp. 702–715, 2012.
- [3] J. Kwon and K. M. Lee, "Tracking by sampling trackers," in *Proceedings of the 2011 IEEE International Conference on Computer Vision (ICCV '11)*, pp. 1195–1202, November 2011.
- [4] J. Han and P. H. N. De With, "Real-time multiple people tracking for automatic group-behavior evaluation in delivery simulation training," *Multimedia Tools and Applications*, vol. 51, no. 3, pp. 913–933, 2011.
- [5] Z. Han, Q. Ye, and J. Jiao, "Combined feature evaluation for adaptive visual object tracking," *Computer Vision and Image Understanding*, vol. 115, no. 1, pp. 69–80, 2011.

- [6] Z. Han, J. Jiao, B. Zhang, Q. Ye, and J. Liu, "Visual object tracking via sample-based Adaptive Sparse Representation (AdaSR)," *Pattern Recognition*, vol. 44, no. 9, pp. 2170–2183, 2011.
- [7] J. Han, E. J. Pauwels, P. M. De Zeeuw, and P. H. N. De With, "Employing a RGB-D sensor for real-time tracking of humans across multiple re-entries in a smart environment," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 2, pp. 255–263, 2012.
- [8] S. Gao, Z. Han, C. Li, Q. Ye, and J. Jiao, "Real-Time Multipedestrian Tracking in Traffic Scenes via an RGB-D-Based Layered Graph Model," *IEEE Transactions on Intelligent Transportation Systems*, vol. 16, no. 5, pp. 2814–2825, 2015.
- [9] L. Zhang, W. Wu, T. Chen, N. Strobel, and D. Comanicu, "Robust object tracking using semi-supervised appearance dictionary learning," *Pattern Recognition Letters*, vol. 62, pp. 17–23, 2015.
- [10] S. Zhang, H. Zhou, H. Yao, Y. Zhang, K. Wang, and J. Zhang, "Adaptive NormalHedge for robust visual tracking," *Signal Processing*, vol. 110, pp. 132–142, 2015.
- [11] S. Zhang, S. Kasiviswanathan, P. C. Yuen, and M. Harandi, "Online dictionary learning on symmetric positive definite manifolds with vision applications," in *Proceedings of the AAAI Conference on Artificial Intelligence*, pp. 3165–3173, January 2015.
- [12] Z. He, X. Li, X. You, D. Tao, and Y. Y. Tang, "Connected component model for multi-object tracking," *IEEE Transactions on Image Processing*, vol. 25, no. 8, pp. 3698–3711, 2016.
- [13] X. Li, Q. Liu, Z. He, H. Wang, C. Zhang, and W.-S. Chen, "A multi-view model for visual tracking via correlation filters," *Knowledge-Based Systems*, vol. 113, pp. 88–99, 2016.
- [14] Y. Qi, S. Zhang, L. Qin et al., "Hedged deep tracking," in *Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2016*, pp. 4303–4311, July 2016.
- [15] Z. He, S. Yi, Y.-M. Cheung, X. You, and Y. Y. Tang, "Robust Object Tracking via Key Patch Sparse Representation," *IEEE Transactions on Cybernetics*, vol. 47, no. 2, pp. 354–364, 2017.
- [16] R. Shi, J. Zhang, Z. Xie, J. Gao, and X. Zheng, "Robust tracking with per-exemplar support vector machine," *IET Computer Vision*, vol. 9, no. 5, pp. 699–710, 2015.
- [17] P. Wilf, S. Zhang, S. Chikkerur, S. A. Little, S. L. Wing, and T. Serre, "Computer vision cracks the leaf code," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 113, no. 12, pp. 3305–3310, 2016.
- [18] L. Liu, Z. Lin, L. Shao, F. Shen, G. Ding, and J. Han, "Sequential discrete hashing for scalable cross-modality similarity retrieval," *IEEE Transactions on Image Processing*, vol. 26, no. 1, pp. 107–118, 2017.
- [19] Y. Guo, G. Ding, L. Liu, J. Han, and L. Shao, "Learning to hash with optimized anchor embedding for scalable retrieval," *IEEE Transactions on Image Processing*, vol. 26, no. 3, pp. 1344–1354, 2017.
- [20] S. Zhang, H. Yao, X. Sun et al., "Action recognition based on overcomplete independent components analysis," *Information Sciences*, vol. 281, pp. 635–647, 2014.
- [21] F. Jiang, S. Zhang, S. Wu, Y. Gao, and D. Zhao, "Multi-layered gesture recognition with Kinect," *Journal of Machine Learning Research (JMLR)*, vol. 16, pp. 227–254, 2015.
- [22] K. Chen, G. Ding, and J. Han, "Attribute-based supervised deep learning model for action recognition," *Frontiers of Computer Science*, vol. 11, no. 2, pp. 219–229, 2017.
- [23] S. Zhang, X. Lan, H. Yao, H. Zhou, D. Tao, and X. Li, "A biologically inspired appearance model for robust visual tracking," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 28, no. 10, pp. 2357–2370, 2017.
- [24] S. Zhang, X. Lan, Y. Qi, and P. C. Yuen, "Robust Visual Tracking via Basis Matching," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 27, no. 3, pp. 421–430, 2017.
- [25] X. Lan, S. Zhang, and P. C. Yuen, "Robust joint discriminative feature learning for visual tracking," in *Proceedings of the 25th International Joint Conference on Artificial Intelligence*, pp. 3403–3410, July 2016.
- [26] S. Zhang, H. Zhou, F. Jiang, and X. Li, "Robust visual tracking using structurally random projection and weighted least squares," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 25, no. 11, pp. 1749–1760, 2015.
- [27] S. Zhang, H. Yao, X. Sun, and X. Lu, "Sparse coding based visual tracking: review and experimental comparison," *Pattern Recognition*, vol. 46, no. 7, pp. 1772–1788, 2013.
- [28] S. H. Zhang, H. Yao, H. Zhou, X. Sun, and S. H. Liu, "Robust visual tracking based on online learning sparse representation," *Neurocomputing*, vol. 100, pp. 31–40, 2013.
- [29] S. Zhang, H. Yao, X. Sun, and S. Liu, "Robust visual tracking using an effective appearance model based on sparse coding," *ACM Transactions on Intelligent Systems and Technology*, vol. 3, no. 3, pp. 43:1–43:18, 2012.
- [30] M. Yakut and N. Kehtarnavaz, "Ice-hockey puck detection and tracking for video highlighting," *Signal, Image and Video Processing*, vol. 10, no. 3, pp. 527–533, 2016.
- [31] I. S. Topkaya, H. Erdogan, and F. Porikli, "Tracklet clustering for robust multiple object tracking using distance dependent Chinese restaurant processes," *Signal, Image and Video Processing*, vol. 10, no. 5, pp. 795–802, 2016.
- [32] Y. Wang and Q. Zhao, "Robust object tracking via online Principal Component–Canonical Correlation Analysis (P3CA)," *Signal, Image and Video Processing*, vol. 9, no. 1, pp. 159–174, 2015.
- [33] D. Shan and C. Zhang, "Visual tracking using IPCA and sparse representation," *Signal, Image and Video Processing*, vol. 9, no. 4, pp. 913–921, 2015.
- [34] Y. Qi, S. Zhang, L. Lei Qin et al., "Hedging Deep Features for Visual Tracking," in *Proceedings of the IEEE Transactions on Pattern Analysis and Machine Intelligence (IEEE T-PAMI)*, 2018.
- [35] J. Sembiring, A. S. Sabzevary, and K. Akizuki, "Stochastic process on multiwavelet," *IFAC Proceedings Volumes*, vol. 35, no. 1, pp. 211–215, 2002.
- [36] Z. Kalal, J. Matas, and K. Mikolajczyk, "P-N learning: bootstrapping binary classifiers by structural constraints," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 49–56, June 2010.
- [37] S. Hare, A. Saffari, and P. H. S. Torr, "Struck: structured output tracking with kernels," in *Proceedings of the IEEE International Conference on Computer Vision (ICCV '11)*, pp. 263–270, IEEE, Barcelona, Spain, November 2011.
- [38] C. Bao, Y. Wu, H. Ling, and H. Ji, "Real Time Robust L1 Tracker Using Accelerated Proximal Gradient Approach," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1830–1837, June 2012.
- [39] T. Zhang, B. Ghanem, S. Liu, and N. Ahuja, "Robust visual tracking via multi-task sparse learning," in *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition*, 2012.

