J Opt (April 2024) 53(2):1129–1136
https://doi.org/10.1007/s12596-023-01293-9

RESEARCH ARTICLE

Discriminative learning of online appearance modeling methods for visual tracking

Zhongming Liao[1,2] · **Xiuhong Xu**[3] · **Zhaosheng Xu**[4] · **Azlan Ismail**[1,5]

Received: 6 April 2023 / Accepted: 27 June 2023 / Published online: 22 July 2023

Abstract Appearance variations are a challenging issue in visual tracking systems. Typically, appearance modeling is used to deal with the challenge of representing and detecting objects in these systems. Appearance modeling is generally structured of parts such as visual target representation and online learning update modeling. Various online learning methods have been proposed to perform the task of object representation and update the model. The discriminative online learning model, as the main focus of the study, is investigated in this paper. Correspondingly, describe current procedures fully, highlighting their benefits and drawbacks. This study aims to give in-depth research into methodologies based on discriminative online learning. A critical review of current approaches' benefits and drawbacks is covered. The finding of this research is investigation of discriminative online learning methods for appearance modeling in visual tracking systems. It provides a comprehensive analysis of current approaches, evaluating their benefits and drawbacks, and comparing their performance to identify the most effective approach for addressing appearance variations in object tracking. The approaches are evaluated, and performance comparisons are made to identify the most effective approach to discriminative online learning for appearance modeling.

Keywords Online learning modeling · Discriminative models · Visual representation · Visual tracking

Introduction

One of the most challenging computer vision issues is visual tracking. Applications include motion analysis, video surveillance, and advanced diver assistance systems [1, 2]. Visual tracking refers to the difficulty of determining a target's motion from a set of images. Due to developments in computer technology, it has emerged as one of the most discussed topics in computer vision [1]. Many applications include surveillance, autonomous navigation, and medical diagnosis [2, 3].

Because of the unavoidable appearance changes in an open environment, traditional tracking techniques that start with fixed models of the target typically fail [4]. The object's fundamental qualities, such as non-rigid construction and posture changes, and the environment's qualities, such as shifting changes, camera motion, view-point, camera scale, and occlusion, can be attributed to these discrepancies. Adaptive techniques that gradually alter a target's representation over time have been developed in order to successfully handle these changes [4].

Researchers consider visual representation and appearance modeling to deal with visual tracking difficulties. Visual representation refers to how information is communicated through images, graphs, charts, and other visual

✉ Xiuhong Xu
xxh258639049@163.com

[1] Faculty of Computer and Mathematical Sciences, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

[2] Academic Affairs Office, Xinyu College, Xinyu 338004, Jiangxi, China

[3] College of Photovoltaic Power Generation, Jiangxi New Energy Technology Vocational College, Xinyu 338004, Jiangxi, China

[4] School of Mathematics and Computer Science, Xinyu College, Xinyu 338004, Jiangxi, China

[5] Institute for Big Data Analytics and Artificial Intelligence (IBDAAI), Kompleks Al-Khawarizmi, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia

media. Visual representations are essential for communicating complex ideas and data quickly and efficiently, as they allow people to process large amounts of information more effectively than text alone. One common type of visual representation is the infographic, which combines images and text to present information visually appealingly. Other examples include diagrams, maps, and flowcharts. Effective visual representations are designed to be easy to read and understand, and they often use color, size, and other design elements to convey important information [5].

Appearance modeling is a technique used to create realistic digital representations of objects, people, or environments. Appearance modeling involves capturing data about the appearance of an object or environment, such as its texture, color, and lighting, and using that data to create a 3D model. Appearance modeling is used in various applications, from video game development to product design. One common use of appearance modeling is in the creation of virtual try-on technology, which allows customers to see how clothing or other products will look on them before making a purchase. Appearance modeling can also create realistic simulations of real-world environments, such as cityscapes or natural landscapes, for use in movies or video games [6, 7].

As mentioned earlier, one of the primary challenges limiting the efficiency of real-time visual tracking algorithms is the absence of appropriate appearance models [1, 8]. Due to the fixed models, conventional template-matching tracking techniques are unable to adjust to appearance changes. As a result, dynamic templates built on online learning are utilized to depict how a target's look varies due to adjustments to posture and lighting. The tracking framework incorporates the online learning method to update the target's appearance model flexibly in response to changes in appearance [9].

Visual tracking techniques are often divided into two classes: generative and discriminative. The generative tracking techniques develop a model to depict the target object's appearance. Finding the item whose look is closest to the modeled appearance is how the tracking problem is then stated. Instead, discriminative tracking approaches, which outperform generative tracking techniques regarding accuracy and speed, seek to distinguish the target item from the background [1, 10].

Therefore, this study focuses on a discriminative online learning-based method for appearance modeling. An extensive investigation of the methods that have emerged over the years is presented, allowing the reader to appreciate the historical development of this field. The main contributions of this study are listed as follows:

1. Review several online learning-based discriminative techniques for modeling physical appearance.
2. Outlining a critical examination of current methodologies for discriminatory online learning-based approaches and discussing their benefits and drawbacks.
3. Addressing the effectiveness of online learning strategies for appearance modeling techniques for visual tracking.

The remainder of this paper consists of "Introduction" section presents an introduction. The "Review of online learning modeling methods" section discusses appearance modeling involving visual representation and online learning. The discriminative online learning-based methods investigate in this section. Results and discussion are presented in "Discussions and analysis" section. Finally, this study concludes in "Conclusion" section.

Review of online learning modeling methods

Because of target appearance variation, online learning modeling is the most important module in visual tracking. This module generally consists of two components [6]: visual target representation and online learning modeling of the model. Visual target representation focuses on leveraging various visual elements to represent the target in pictures. Online training focuses on creating a model of the target and updating it under conditions of appearance variation in order to recognize the target in subsequent frames. Because it is the first stage in online learning modeling, visual representation is only briefly discussed in the following parts. As was already said, this work uses discriminative-based techniques to model online learning.

Visual target representation

Visual tracking is required to represent a target in the image to describe the target using visual features and track the target in the following images. Because the description's effectiveness substantially impacts the entire tracking process, visual target representation is a crucial duty in the appearance modeling module. Moreover, such a description is not always given to the tracker beforehand; thus, it could need to be created in real-time using both past knowledge and unknown information or an online creating model [11]. As a result, choosing the right characteristics for visual target representation and description is essential for visual tracking. This study only introduces the visual representation and briefly describes it because it is essential for any appearance modeling method.

Discriminative-based appearance models

In dynamic and long processing of visual tracking, having a good target description for its representation in the scene is not enough to deal with target appearance changes. In this case, the target representation is not adaptive to appearance variation conditions. These appearance variations can arise from illumination changes, pose variations, geometrical transformations of the target, etc. To handle such variations, generating a target model is required, and the target model is needed to be incrementally updated to be adjusted and adapted to the new circumstances [6, 12]. This study focuses on discriminative online learning for appearance modeling. Figure 1 shows online discriminative-based appearance modeling methods.

Discriminative-based appearance models deal with a binary classification to classify the foreground and background regions. They intend to classify the target and non-target regions discriminately. They adopted highly discriminative and informative features for visual target tracking. These models can predict a scene's target and non-target regions using online learning classification functions. This online learning procedure gradually modifies visual elements to anticipate the monitored object in a complex background.

Boosting-based discriminative appearance models

Due to their successful performance in discriminative learning capacities, the BDAMs are currently widely used in visual target tracking [6]. To be more precise, the self-learning boosting-based models train a classifier first using the data from previous frames before using the learned classifier to assess potential target regions at the current frame [6]. In the following sections, we will discuss the categories in detail.

Self-learning single-instance boosting-based models,

Various applications for computer vision and target detection and tracking are based on online boosting models [13]. To make a clear description of self-learning single-instance models, we conduct a table and categorize these models as shown in Table 1.

Co-learning single-instance BDAMs As stated in [6], the "model drift" issue affects self-learning boosting-based models. Additional models based on semi-supervised learning methods are used for visual object tracking to address this issue. Grabner et al. [14] proposed a semi-supervised algorithm that uses online boosting, as illustrated in Fig. 2.

Multi-instance boosting-based models Multiple instance learning is utilized for target tracking in order to address the underlying uncertainty of object localization, as shown in Fig. 3. Multi-instance boosting-based models can be broadly categorized into classes for self- and co-learning. Table 2 includes information about these categories in detail.

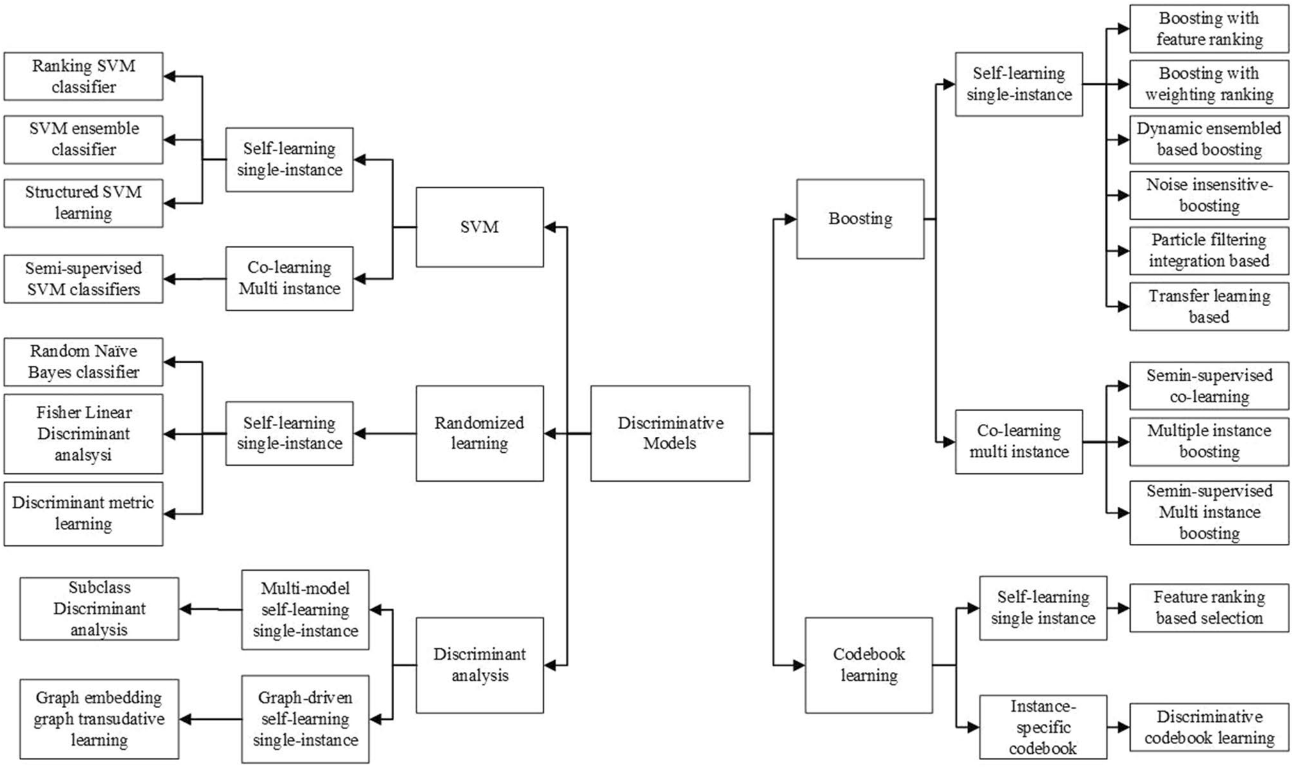
Ranking SVM classifier
SVM ensemble classifier
Structured SVM learning
Semi-supervised SVM classifiers
Self-learning single-instance
Co-learning Multi instance
SVM
Random Naïve Bayes classifier
Fisher Linear Discriminant analsysi
Discriminant metric learning
Self-learning single-instance
Randomized learning
Discriminative Models
Subclass Discriminant analysis
Multi-model self-learning single-instance
Graph embedding graph transudative learning
Graph-driven self-learning single-instance
Discriminant analysis
Boosting
Self-learning single-instance
Boosting with feature ranking
Boosting with weighting ranking
Dynamic ensembled based boosting
Noise insensitive-boosting
Particle filtering integration based
Transfer learning based
Co-learning multi instance
Semin-supervised co-learning
Multiple instance boosting
Semin-supervised Multi instance boosting
Codebook learning
Self-learning single instance
Feature ranking based selection
Instance-specific codebook
Discriminative codebook learning

Fig. 1 Representative of online discriminative-based appearance modeling methods

Table 1 Review of self-learning single-instance boosting-based models

Models	References	Descriptions	Pros	Cons
Conventional models	[13, 18, 19]	Traditional boosting algorithm and single instance learning	Simple and effective	It may not perform well on noisy data
Non-noise sensitive	[20]	Incorporates self-learning mechanism	Robust to noisy data	It may take longer to converge than conventional models
Transfer learning-based	[21]	Leverages knowledge from related tasks or domains	Effective when the target task has limited data or is in a different domain	It may not perform well when the source task is not closely related to the target task

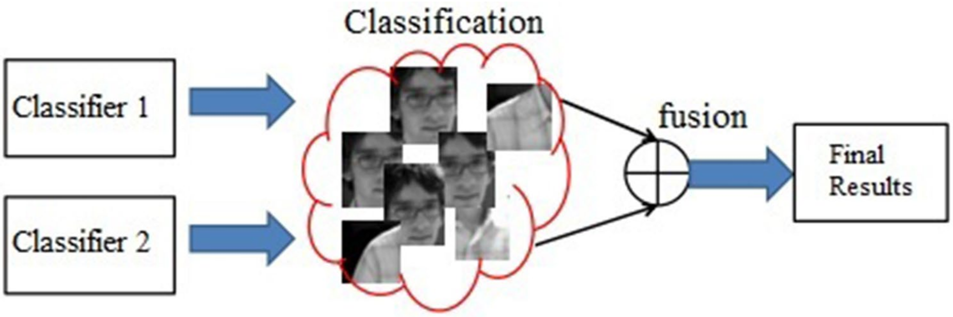

Fig. 2 A sample of a typical co-learning problem

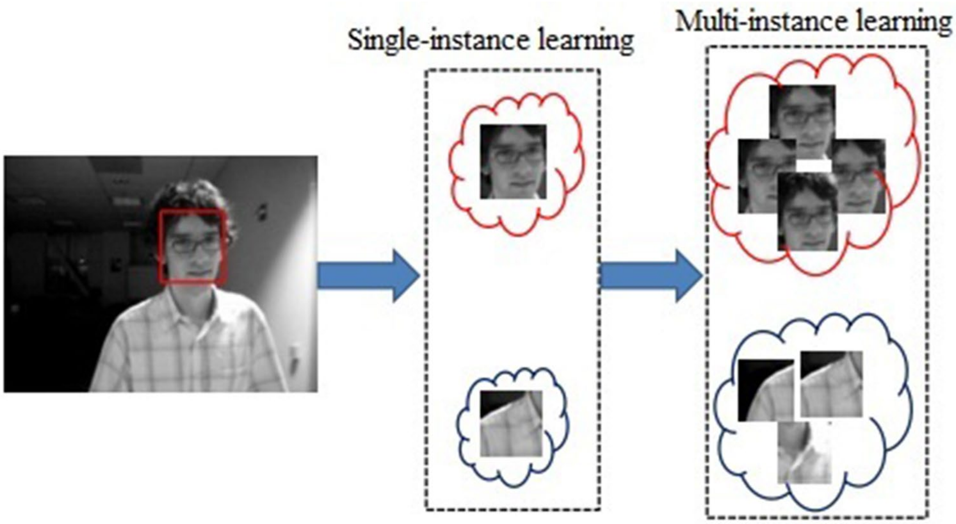

Fig. 3 Illustration of single-instance multi-instance learning

Table 2 Review of multi-instance boosting-based models

Models	References	Descriptions	Pros	Cons
Self-learning multi-instance-based	[22, 23]	Learns from instances and bags, effective on noisy data	It can take longer to converge and may not handle varying bag sizes well	It may take longer to converge than conventional models
Co-learning multi-instance boosting-based	[24]	Learns from multiple views of the data, effective in multi-sensor or multi-modality systems	It may require more computational resources and may not handle highly correlated instances well	It may require more computational resources than other models It may not perform well when the instances within each bag are highly correlated

SVM-based discriminative appearance models (SDAMs)

SDAMs aim to develop discriminative SVM classifiers with a margin of error to increase inter-class separability. SDAMs have a high discriminative capacity since they can find and retain instructive samples as support vectors for object/non-object classification. Designing resilient SDAMs requires careful kernel selection and efficient kernel computation. SDAMs are commonly based on self-learning SDAMs and co-learning SDAMs, according to the applied learning methods. SVM-based discriminative appearance models are shown in Table 3.

Randomized learning-based models

Building a randomized or diversified classifier ensemble to create a model for target appearance is possible with randomized learning-based techniques based on random input and feature selection. They are effective in real-time systems and have little computational cost. They can be expanded to address issues with multi-class learning. Additionally, they make it possible to leverage parallel processing on multi-core and GPU-based platforms, which can be improved with random learning-based techniques to cut down on processing time significantly. Because of their arbitrary feature selection, they suffer from tracking performance in various scenes and target more look variants. Several randomized learning-based models are put forth in visual trackings, such as online random forests [15] and random naive Bayes classifiers [16]. Table 4 presents some existing models under randomized learning-based models.

Discriminant analysis-based models

Discriminant analysis-based models are an algorithm used in visual tracking to handle appearance variations in the target object being tracked. Appearance variations occur due to lighting conditions, pose, scale, and occlusion changes. The basic idea behind these models is to create a discriminative function that distinguishes the target object from the background and other objects in the scene. This function is learned based on training data consisting of samples of the target object in different appearance variations [6]. The following section discusses these branches in detail.

Conventional discriminant analysis models Table 5 presents the review of conventional models of discriminant analysis models.

Graph-driven discriminant analysis models Recent discriminant analysis models utilize graph-based learning techniques for discriminant analysis for visual target track-

Table 3 Review of SVM-based discriminative appearance models

Models	References	Descriptions	Pros	Cons
Self-learning-based	[25, 26]	The self-learning for the purpose of classifying objects and non-objects, SDAMs must create SVM classifiers	It can improve the accuracy of the classifier over time Can handle noise and variability in the data	It may be sensitive to the quality of the initial training set It may require additional computational resources for each iteration
Co-learning-based	[27, 28]	Incorporates multiple sources of information, can handle noise and variability	It can improve the accuracy of the classifier by incorporating multiple sources of information Can handle noise and variability in the data	It may require additional computational resources for training multiple classifiers It may not perform well if the different views of the data are highly correlated

Table 4 Review of randomized learning-based appearance models

Models	References	Descriptions	Pros	Cons
Online random forests	[15]	By using Random Naive Bayes classifiers, this approach creates a varied classifier ensemble	Robust to noise and outliers in the data Can handle large datasets with incremental learning It can be parallelized for efficient computation	Requires careful selection of hyperparameters It may be sensitive to the quality of the data sampling
Random naive Bayes classifiers	[16]	This technique, called MIForests, builds randomized trees using multiple-instance learning	Can handle high-dimensional data Fast training and prediction times Can handle missing data	Assumes independence between features, which may not always be true May be sensitive to the quality of the data sampling

Table 5 Review of conventional models of discriminant analysis models

Models	References	Descriptions	Pros and Cons	Cons
Uni-modal-based	[29–31]	The data for the object class are distributed according to a uni-modal Gaussian curve	Multi-modal distributions in the object and background classes are difficult for uni-modal-based models to handle well	It performs badly when the target and background have multi distributions
Multi-modal-based	[32]	These models create a sample model using a combination of Gaussian distributions from the object and backdrop classes	It is helpful in varying occlusion appearances	Having a high computing cost

ing. Typically, these graph-driven models are categorized into graph-embedding-based and graph-transductive-based methods. Table 6 presents the review of graph-driven discriminant analysis models.

Codebook learning-based models

These models rely on the concept of a codebook, which is a dictionary of visual patterns learned from the target object in different appearance variations. The basic idea behind these models is to represent the target object as a bag of visual words, where each word corresponds to a visual pattern in the codebook. The bag of visual words is then used as a feature vector to track the target object.

The codebook is learned from training data consisting of samples of the target object in different appearance variations. The codebook can be learned using unsupervised learning techniques such as k-means clustering or supervised learning techniques such as support vector machines (SVMs) [17].

Discussions and analysis

The discriminative-based appearance models primarily focus on how to match the data from various target classes. The key challenge with these models is determining if the provided model is properly defined. Incremental learning, defined as a model update of the target visual representation during the tracking process, is introduced to make the model more effective. The foreground and background may be efficiently separated with this technique. However, due to the background regions' resemblance to the target class, they continue to experience appearance fluctuation and distractions.

Discriminative-based appearance models deal with a binary classification to classify the foreground and background regions. They are used to separate the target and non-target regions in an image. They adopted highly discriminative and informative visual features for target tracking. Visual features are incrementally updated in this online learning process to represent the target in the complicated background. Thus, they can achieve effective and efficient predictive performances [6].

In conclusion, generative approaches solely consider the object's appearance and ignore background information. The discriminative approaches, in contrast, calculate a border area that separates the item from its surroundings by considering information about both the object and the backdrop [35]. The number of monitored objects and the average position inaccuracy are used to evaluate the tracking outcome in the visual tracking community objectively [36,

Table 6 Review of graph-driven discriminant analysis models

Models	References	Descriptions	Pros	Cons
Graph embedding-based	[33]	In this method, the graph is represented as a matrix of pairwise similarities between nodes, which is then embedded into a low-dimensional space using techniques such as spectral embedding or convolutional graph networks	Can handle non-linear and complex relationships between data points It can capture the global structure of the data Robust to noisy or incomplete data	It is limited to small or medium-sized datasets due to computational complexity It may require a careful selection of hyperparameters
Graph transductive learning-based	[34]	In this method, the graph is represented as a matrix of pairwise similarities between nodes, and a set of labeled nodes is provided as input	Can handle large datasets with limited labeled data Robust to noisy or incomplete data It can capture the global structure of the data	It may require a careful selection of hyperparameters It may be sensitive to the quality of the graph representation

37]. As a result, if there are enough examples, discriminative approaches perform better than generative ones.

Conclusion

This work concentrated on online learning modeling, a vital process for appearance modeling that is mostly employed for visual tracking. This thesis discusses appearance modeling, one of the key components of visual tracking systems. One of the key components of visual tracking systems is appearance modeling, which is covered. Statistical modeling and visual representation are the two main components of appearance modeling. These two are thoroughly covered in this paper because they substantially impact the outcome of moving object identification, which is crucial for visual tracking systems. This work emphasizes generative online learning-based methods and thorough reviews utilizing highly regarded and peer-reviewed literature.

Additionally, a critical analysis was completed to discuss the benefits and drawbacks of the current approaches. To further examine appearance modeling in visual tracking, the fusion of generative and discriminative online learning can be investigated for appearance modeling. Moreover, a deep learning-based approach can be implemented to improve online learning performance.

References

1. L. Gao, B. Liu, P. Fu, M. Xu, J. Li, Visual tracking via dynamic saliency discriminative correlation filter. Appl. Intell. **52**(6), 5897–5911 (2022)
2. X. Gao, J. Szep, P. Satam, S. Hariri, S. Ram, J.J. Rodriguez, 2020 Spatio-temporal processing for automatic vehicle detection in wide-area aerial video. IEEE Access **8**, 199562–199572 (2020)
3. M. Ang, E. Sundararajan, K. Ng, A. Aghamohammadi, T. Lim, Investigation of threading building blocks framework on real time visual object tracking algorithm. Appl. Mech. Mater. **666**, 240–244 (2014)
4. A. Aghamohammadi, M.C. Ang, E.A. Sundararajan, N.K. Weng, M. Mogharrebi, S.Y. Banihashem, A parallel spatiotemporal saliency and discriminative online learning method for visual target tracking in aerial videos. PLoS ONE **13**(2), e0192246 (2018)
5. Y. Cao, G. Fu, J. Yang, Y. Cao, M.Y. Yang, Accurate salient object detection via dense recurrent connections and residual-based hierarchical feature integration. Signal Process. Image Commun. **78**, 103–112 (2019)
6. P.R. Vadamala, A.F. Aklak, Discriminative appearance model with template spatial adjustment for visual object tracking. Soft Comput. **27**, 1–14 (2023)
7. B. Gao, M.W. Spratling, Explaining away results in more robust visual tracking. Vis. Comput. **39**, 1–15 (2022)
8. X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, A.V.D. Hengel, A survey of appearance models in visual object tracking. ACM Trans. Intell. Syst. Technol. (TIST) **4**(4), 58 (2013)
9. Y. Li, S. Wang, Q. Tian, X. Ding, A survey of recent advances in visual feature detection. Neurocomputing **149**, 736–751 (2015)
10. Y. Liu, H. Yan, W. Zhang, M. Li, L. Liu, An adaptive spatiotemporal correlation filtering visual tracking method. PLoS ONE **18**(1), e0279240 (2023)
11. Q. Liu, X. Zhao, Z. Hou, Survey of single-target visual tracking methods based on online learning. IET Comput. Vis. **8**(5), 419–428 (2014)
12. L. Chen, Y. Liu, A robust spatial-temporal correlation filter tracker for efficient UAV visual tracking. Appl. Intell. **53**, 1–16 (2022)
13. H. Yang, L. Shao, F. Zheng, L. Wang, Z. Song, Recent advances and trends in visual tracking: a review. Neurocomputing **74**(18), 3823–3831 (2011)
14. A.W. Smeulders, D.M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, M. Shah, Visual tracking: an experimental survey. IEEE Trans. Pattern Anal. Mach. Intell. **36**(7), 1442–1468 (2014)
15. H. Grabner, M. Grabner, H. Bischof, Real-time tracking via online boosting, in *BMVC*, vol. 6 (2006)
16. H. Grabner, C. Leistner, H. Bischof, Semi-supervised on-line boosting for robust tracking, in *Computer Vision–ECCV 2008* (Springer, 2008), pp. 234–247
17. A. Saffari, C. Leistner, J. Santner, M. Godec, H. Bischof, On-line random forests, in *2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)* (IEEE, 2009), pp. 1393–1400
18. C. Leistner, A. Saffari, H. Bischof, Miforests: multiple-instance learning with randomized trees, in *Computer Vision–ECCV 2010* (Springer, 2010), pp. 29–42
19. J. Gall, N. Razavi, L. Gool, On-line adaption of class-specific codebooks for instance tracking (2010)
20. S. Avidan, Ensemble tracking. IEEE Trans. Pattern Anal. Mach. Intell. **29**(2), 261–271 (2007)
21. X. Liu, T. Yu, Gradient feature selection for online boosting, in *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007* (IEEE, 2007), pp. 1–8
22. C. Leistner, A. Saffari, P.M. Roth, H. Bischof, On robustness of on-line boosting-a competitive study, in *2009 IEEE 12th International Conference on Computer Vision Workshops (ICCV Workshops)* (IEEE, 2009), pp. 1362–1369
23. S. Wu, Y. Zhu, Q. Zhang, A new robust visual tracking algorithm based on transfer adaptive boosting. Math. Methods Appl. Sci. **35**(17), 2133–2140 (2012)
24. B. Babenko, M.-H. Yang, S. Belongie, Visual tracking with online multiple instance learning, in *IEEE Conference on Computer Vision and Pattern Recognition, 2009. CVPR 2009* (IEEE, 2009), pp. 983–990
25. X. Li, C. Shen, Q. Shi, A. Dick, A. Van den Hengel, Non-sparse linear representations for visual tracking with online reservoir metric learning, in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2012), pp. 1760–1767
26. X. Li, A. Dick, C. Shen, A. van den Hengel, H. Wang, Incremental learning of 3D-DCT compact representations for robust visual tracking. IEEE Trans. Pattern Anal. Mach. Intell. **35**(4), 863–881 (2013)
27. S. Avidan, Support vector tracking. IEEE Trans. Pattern Anal. Mach. Intell. **26**(8), 1064–1072 (2004)
28. R. Yao, Q. Shi, C. Shen, Y. Zhang, A. van den Hengel, Robust tracking with weighted online structured learning, in *Computer Vision—ECCV 2012* (Springer, 2012), pp. 158–172
29. F. Tang, S. Brennan, Q. Zhao, H. Tao, Co-tracking using semi-supervised support vector machines, in *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007* (IEEE, 2007), pp. 1–8
30. F. Yang, H. Lu, Y.-W. Chen, Robust tracking based on boosted color soft segmentation and ICA-R, in *2010 17th IEEE International Conference on Image Processing (ICIP)* (IEEE, 2010), pp. 3917–3920

31. R.-S. Lin, M.-H. Yang, S.E. Levinson, Object tracking using incremental fisher discriminant analysis, in *Proceedings of the 17th International Conference on Pattern Recognition, 2004. ICPR 2004* (IEEE, 2004), pp. 757–760
32. N. Jiang, W. Liu, Y. Wu, Order determination and sparsity-regularized metric learning adaptive visual tracking, in *2012 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (IEEE, 2012), pp. 1956–1963
33. X. Wang, G. Hua, T.X. Han, Discriminative tracking by metric learning, in *Computer Vision–ECCV 2010* (Springer, 2010), pp. 200–214
34. Z. Xu, P. Shi, X. Xu, Adaptive subclass discriminant analysis color space learning for visual tracking, in *Advances in Multimedia Information Processing—PCM 2008* (Springer, 2008), pp. 902–905
35. X. Zhang, W. Hu, S. Maybank, X. Li, Graph based discriminative learning for robust and efficient object tracking, in *IEEE 11th International Conference on Computer Vision, 2007. ICCV 2007* (IEEE, 2007), pp. 1–8
36. Y. Zha, Y. Yang, D. Bi, Graph-based transductive learning for robust visual tracking. Pattern Recogn. **43**(1), 187–196 (2010)
37. M.Y. Abbass, K.-C. Kwon, N. Kim, S.A. Abdelwahab, F.E.A. El-Samie, A.A. Khalaf, A survey on online learning for visual tracking. Vis. Comput. **37**, 993–1014 (2021)

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.