IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received December 5, 2019, accepted December 11, 2019, date of publication December 16, 2019,
date of current version December 31, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2959942

Progress and Outlook of Visual Tracking: Bibliographic Analysis and Perspective

YATING LIU[1,2], **KUNFENG WANG**[3], **(Senior Member, IEEE), XUESONG LI**[1,2], **TIANXIANG BAI**[1,2], **AND FEI-YUE WANG**[1], **(Fellow, IEEE)**

[1] State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, Beijing 100190, China
[2] University of Chinese Academy of Sciences, Beijing 100049, China
[3] College of Information Science and Technology, Beijing University of Chemical Technology, Beijing 100029, China

Corresponding author: Kunfeng Wang (wangkf@mail.buct.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2018YFC1704400, and in part by the National Natural Science Foundation of China under Grant U1811463.

ABSTRACT Benefitting from continuous progress in computer architecture and computer vision algorithms, the visual tracking field has earned its rapid development in recent years. This paper surveys this interesting field through bibliographic analysis on the Web-of-Science literature from 1990 to 2019. Specifically, statistical analysis methods are used to obtain the most productive authors and countries/regions, the most cited papers, and so on. In order to realize an in-depth analysis, the co-authors, co-keywords and keyword-author co-occurrence networks are built to intuitively exhibit the evolution of research hotspots and the collaboration patterns among world-wide researchers. Brief introductions of the topics that occur frequently in co-keywords networks are provided as well. Furthermore, existing challenges and future research directions within the visual tracking field are discussed, revealing that tracking-by-detection and deep learning will continue receiving much attention. In addition, the parallel vision approach should be adopted for training and evaluating visual tracking models in a virtual-real interaction manner.

INDEX TERMS Visual tracking, bibliographic analysis, collaboration patterns, research hotspots, parallel vision.

I. INTRODUCTION

Visual tracking is a process that uses computer vision methods to derive the locations and motion trajectories of moving objects from a video. Usually, the locations of the same object are often spatially close in adjacent frames, which allows us to attach it with a trajectory and predict its future position. After obtaining the motion state and trajectory of the target, the knowledge can be used for motion analysis, action recognition and scene understanding. Furthermore, this field has a wide range of applications such as intelligent monitoring, human-computer interaction, visual navigation, and medical diagnosis [1]. For example, in intelligent monitoring, visual tracking has been widely used in many communities and public areas to improve social security and convenience services [2], [3]; in the military field, visual tracking is used to implement vision-based navigation control in military installations [4]. In short, the visual tracking technology has considerable application prospects in both civil and military fields.

To solve the visual tracking problem, multiple disciplines including pattern recognition, machine learning, image processing, and computer vision are closely integrated, so that collaborations are demanded from interdisciplinary fields. The visual tracking issue can be characterized in various ways, e.g., single-target tracking or multi-target tracking, rigid object tracking or non-rigid object tracking, stationary-camera or moving-camera tracking, single-camera or multi-camera tracking, and single-scene or cross-scene tracking [5]. Fig. 1 shows the different kinds of tracking diagrams. As we know, visual tracking is a complex issue, making it difficult to grasp the progress of the entire field comprehensively. Therefore, we present a survey on visual tracking from the point of view of bibliographic analysis. After collecting 11832 literatures published in the past 30 years from Web of Science Collection with our query, we adopt the bibliographic analysis and collaboration pattern analysis methods to acquire in-depth information. We expand Anderson's method [6] and conduct our analysis in three levels of

The associate editor coordinating the review of this manuscript and approving it for publication was Habib Ullah.

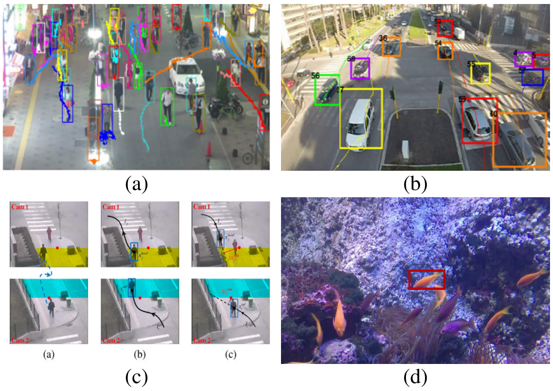


FIGURE 1. Visual tracking diagrams. (a) shows multi-people tracking. (b) shows multi-car tracking. (c) shows multi-camera multi-object tracking. (d) shows single-object tracking.

data summarization, and discuss the results from 10 different viewpoints, such as keywords, co-authors and author-keywords. Through this analysis, we are able to find research hotspots, active authors and institutions, and potential fields. We also come up with some thoughts and perspectives on future research trends of visual tracking, especially on how tracking could be integrated with the parallel vision approach [7], [8].

The main contributions of this paper are as follows:

1. After retrieving the visual tracking papers from the Web-of-Science database, we present some basic statistical analysis results, such as the most productive authors and institutions, to get a basic understanding of the past 30 years' development of visual tracking.

2. We investigate the trends of keyword changes, research hotspots and research methods, as well as the co-occurrence relationship among authors and keywords, based on a division of the past 30 years into three decades. This investigation benefits a better understanding of how the visual tracking field has been evolving since 1990.

3. Based on the results of thorough bibliographic analysis, we propose some in-depth thoughts and perspectives on potential hotspots in future research of the visual tracking topic.

The remainder of this paper is organized as follows. Section II introduces related works on visual tracking survey. Section III describes our analysis methods. Analysis results of the most cited papers, the most productive authors, the co-occurrence of keywords and other information acquired from the analyzed papers are presented in Section IV. The collaboration patterns of authors and co-keywords networks are described in Section V. Section VI elaborates the research trends in the visual tracking field, especially by extending existing methods via virtual-real interaction. Finally, the conclusion is drawn in Section VII.

II. RELATED WORKS

In the past decades, visual tracking has witnessed its remarkable progress. Accordingly, some literatures in academic journals and conferences have reviewed the existing works on visual tracking from different aspects.

A. VISUAL TRACKING REVIEWS

Some researchers make comprehensive review and analysis of visual tracking algorithms. For example, Yilmaz *et al.* [9] review the state-of-the-art visual tracking algorithms, which are classified into different categories based on the object and motion representations used. They also predict new research trends. Arnold *et al.* [10] systematically evaluate visual tracking algorithms on 315 video fragments, where 19 most cited trackers are compared. Other papers [11], [12] summarize the difficulties in tracking objects and discuss important issues related to tracking, including the use of appropriate image features, selection of motion models, and detection of objects.

Some researchers focus mainly on reviewing the application of visual tracking. Hou and Han [13] discuss the application of visual tracking technology in video surveillance, image compression, and three-dimensional reconstruction. In addition, they categorize the visual tracking methods into bottom-up and top-down methods. Emanuele and Konstantinos [14] introduce the technical requirements and related technical algorithms of visual tracking. In particular, they present an overview of 28 recent papers on subsea video tracking and related motion analysis problems, arguably capturing the state-of-the-art subsea video tracking and suggesting useful research directions for the subsea video processing community. Hansen and Ji [15] review the progress and latest technologies of video-based eye detection and tracking. The method of gaze estimation is also investigated and compared according to their geometric properties and the reported accuracies. Buch *et al.* [16] conduct a comprehensive review of automatic video analysis for urban surveillance.

In addition, some literatures only discuss a small branch of visual tracking methods. Gu *et al.* [17] introduce the mean-shift-based framework and analyze its shortcomings. Chen [18] summarizes the development, conundrum, and application of Kalman filter for robotic visual perception. Li *et al.* [19] review the methods and applications of 2D appearance models of object tracking methods. Zhang *et al.* [20] review the progress of appearance-model-learning-based tracking methods, including target feature description and three categories of target appearance models.

B. BIBLIOGRAPHIC ANALYSIS WORKS

Scientific bibliographic analysis is a powerful bibliometric approach, which can study the conceptual structure of a specific research area. It is regarded as an interdisciplinary space between disciplines, fields, professional and personal documents. It focuses on surveying and delineating research areas to determine the structure and evolution. In the past years, some researchers use network analysis methods to analyze data in different areas. Wang *et al.* [21] investigate the productivity and collaboration patterns in the publications of IEEE Transactions on Intelligent Transportation Systems (TITS) between 2010 and 2013. Xu *et al.* [22] identify the most productive authors, institutions, and countries/regions in IEEE TITS from 2000 to 2015. Three networks including co-authorship network, keyword co-occurrence network and author co-keyword network are generated to analyze the collaboration patterns among authors in the ITS field. Moral-Munoz *et al.* [23] introduce the H-Classics metric which is based on the H-index to analyze the ITS field, and provide more insights about the scientific structure of ITS research. Emrouznejad and Yang [24] analyze the scholarly literature in Data Envelopment Analysis (DEA) from 1978 to 2016. The statistics based on different aspects, current studies and future trends are provided in that field.

In this paper, the bibliographic analysis approach is used to survey the visual tracking field. A deeper understanding of the field is gained through Web of Science Collection over the last 30 years, including the number of literatures and highly cited authors, as well as the evolution of keywords.

III. METHOD FOR LITERATURE ANALYSIS

In this section, we generally elaborate our method in three steps to present an in-depth analysis of the visual tracking literature. We focus on the relationship among authors and high-frequency keywords, which helps to obtain the research status and predict research trends in this field.

A. DATA COLLECTION

In order to capture influential and authoritative literature, we extract data from Web of Science Collection, using the following "advanced search" query: "TS = (Object tracking and Computer Vision) AND TI = (Tracking or Tracklet or Tracklets)". The collected documents include proceedings, articles and reviews from 1990 to 2019. The research papers from 1990 to 2019 are collected from numerous journals and conferences (e.g., IEEE TPAMI and CVPR), and 11832 papers are obtained consequently.

B. DATA ANALYSIS

Three levels of data summarization are introduced in this part, including basic analysis, co-authors analysis and authors-keywords analysis.

On the first level, the basic information of these papers are extracted, including the most cited papers, and the most productive authors, institutions and countries/regions. In addition, we divide the past 30 years into three time intervals with 10 years each, which will facilitates the investigation of high-frequency keywords at different time periods. Besides, the changes of tracking methods in three decades are summarized to get a better understanding of the visual tracking research.

On the second level, bibliographic analysis and collaboration pattern analysis are conducted to discover relevant cooperation relationship among researchers and research groups. The cooperation between researchers is demonstrated in the social network so that research communities can be discovered from clusters of co-authors.

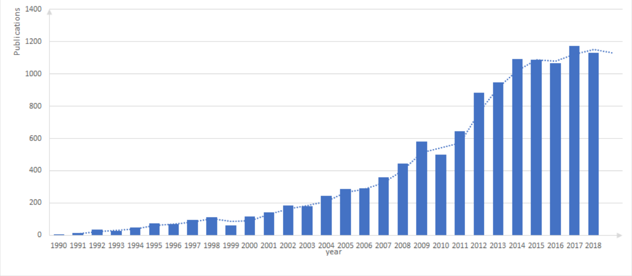


FIGURE 2. The numbers of collected literatures per year from 1990 to 2018.

On the third level, we generate author-keyword and keyword-author networks to get a deeper understanding of the research stages in visual tracking field. This time, we obtain the relationship between authors and keywords occurring together. As a result, we are able to find those researchers who are involved in a specific topic, and those topics in which a specific group of authors is engaged.

Furthermore, the average degree, the average shortest path length, the diameter, and the clustering coefficient [25] are calculated for each cluster to better understand the cooperation frequency.

C. VISUALIZATION

With the aim of disclosing co-occurrence relationships intuitively, we exploit the visualization tool Gephi [26] to build our network graphs.

IV. BIBLIOGRAPHIC ANALYSIS

In this section, we summarize the most productive authors, institutions, and countries/regions during the past 30 years. Furthermore, the highly cited papers are identified and the frequently occurring keywords are investigated, in order to get a basic understanding of the research progress in the visual tracking field.

A. ANNUAL NUMBERS OF LITERATURES

The results extracted from the WOS are shown below. We get 11832 papers contributed by 27055 authors who are associated with 6541 institutions in 104 countries/regions, and for each paper the following information is obtained: authors, affiliations, title, year of publication, citations, sources, abstract and keywords. Fig. 2 shows the numbers of collected literatures from 1990 to 2018.

It can be found that during the years from 1990 to 1999, the field of visual tracking did not catch much attention, as the total number of core papers during that period is 531. From 2000 to 2009, a stable increase can be found and even the publication count 581 in 2009 exceeded the total number of publications 531 from 1990 to 1999. Since 2010, the visual tracking field has experienced a rapid development as reflected by the numbers of publications and citations, and even more than 1000 papers were published yearly from 2014 to 2018.

TABLE 1. The most productive authors from 1990 to 2019.

Index	Author	Number of Papers
1	Ming-Hsuan Yang	67
2	Huchuan Lu	51
3	Weiming Hu	40
4	Jie Yang	36
5	Dong Wang	33
6	Fatih Porikli	30
7	Hongxun Yao	29
8	Junliang Xing	26
9	Xiaoqin Zhang	24
10	Luc Van Gool	22
11	Jenq-Neng Hwang	22
12	Tianzhu Zhang	22
13	Shengping Zhang	21
14	Guillaume-Alexandre Bilodeau	21
15	Michael Felsberg	21

TABLE 2. The most productive institutions from 1990 to 2019.

Index	Institution	Number of Papers
1	Chinese Academy of Sciences	475
2	University of California System	201
3	Beijing Institute of Technology	193
4	Centre National DE LA Recherche Scientifique	172
5	Shanghai Jiao Tong University	155
6	Tsinghua University	152
7	Harbin Institute of Technology	147
8	INRIA	132
9	Beihang University	107
10	National University of Defense Technology China	105

TABLE 3. The most productive countries/regions from 1990 to 2019.

Index	Countries/Regions	Number of Papers
1	China	4366
2	USA	2153
3	Germany	727
4	Japan	708
5	South Korea	682
6	UK	602
7	France	601
8	Canada	397
9	Australia	380
10	India	357

The annual publication in the past 30 years has met a rapid growth with some small ups and downs. From 1995, the field ran into its first rise that mainly attribute to the development of classification methods, including SVM and Adaboost, and image descriptors, such as SIFT. From 2005 to 2009, the progress in conditional random field and other machine learning methods promoted the second development of visual tracking. Since 2010, the field of visual tracking papers has ushered a remarkable growth with deep learning methods.

B. THE MOST PRODUCTIVE AUTHORS

Table 1 shows the 15 most productive authors from 1990 to 2019. It is apparent that Ming-Hsuan Yang from USA is the most productive author with 67 publications. He is followed by Huchuan Lu and Weiming Hu from China who have published 51 and 40 papers, respectively.

They engage in varied research directions, indicating that the research in visual tracking is wide and diverse. For example, Ming-Hsuan Yang is interested in face detection [27], collaborative model [28] and hierarchical convolution [29]. Huchuan Lu investigates salient object detection [30], person re-identification (Re-ID) [31] and attention networks [32]. Weiming Hu studies hidden Markov model [33], graph learning [34] and sparse representation [35].

C. THE MOST PRODUCTIVE INSTITUTIONS

Table 2 shows the statistical result of the most productive institutions. As is shown, 7 Chinese institutions are listed in the top-10 most productive institutions. Chinese Academy of Sciences ranks the first with 475 published papers in the dataset which accounts for 3.99% of the total data. University of California System, Centre National DE LA Recherche Scientifique, and INRIA are three institutions that are outside China but listed in the top-10 most productive institutions, accounting for 1.688%, 1.445%, and 1.109% respectively. These facts indicate that Chinese institutions have made great contributions to the development of visual tracking.

D. THE MOST PRODUCTIVE COUNTRIES/REGIONS

Table 3 shows the most productive countries/regions. China ranks first among all countries as its institutions rank highly in the most productive institutions. Specifically, China has 4366 research papers included in the WOS dataset. Other productive countries are USA with 2152 papers, Germany with 727 papers, and so on.

E. LEADING PUBLISHERS

In this part, we calculate the numbers and proportions of articles focusing on tracking in different publishers, as shown in Table 4. The top 5 journals/conferences which have published the most papers on visual tracking are: Lecture Notes in Computer Science, Proceedings of SPIE, IEEE International Conference on Image Processing, IEEE Conference on Computer Vision and Pattern Recognition, and International Conference on Pattern Recognition. Among them, Lecture Notes in Computer Science publishes 795 papers on visual tracking, with a percentage of 6.678%.

F. THE MOST CITED PAPERS

In this part, we use Web of Science to dig out papers with the highest citation numbers. As suggested by [36], we remove the bias and normalize the citation counts using Eq. (1),

$$C_{norm} = \frac{CC}{\frac{D_{\text{diff}}}{365} + \frac{IN}{NI}}, \tag{1}$$

where C_{norm} denotes the normalized citation count of an article, and CC indicates the original citation count. D_{diff} denotes the number of days between the present date and the 1st of January of the next year after publication. IN and NI

TABLE 4. Proportion of literatures by the publishers.

Index	Publisher	Number of Papers	Percentage (%)
1	Lecture Notes in Computer Science	795	6.678
2	Proceedings of The Society of Photo Optical Instrumentation Engineers	558	4.687
3	IEEE International Conference on Image Processing	307	2.579
4	IEEE Conference on Computer Vision and Pattern Recognition	195	1.638
5	International Conference on Pattern Recognition	176	1.478
6	Computer Engineering and Applications	143	1.201
7	IEEE Transactions on Pattern Analysis and Machine Intelligence	138	1.786
8	IEEE International Conference on Robotics and Automation	122	1.025
9	IEEE Transactions on Image Processing	122	1.025
10	Multimedia Tools and Applications	119	1.006

TABLE 5. The normalized most cited papers.

Index	Title	Normalized citation counts	Countries	Authors
1	Online Object Tracking: A Benchmark [37]	238	USA	Yi Wu; Jongwoo Lim; Ming-Hsuan Yang
2	Tracking-Learning-Detection [38]	223	UK	Zdenek Kalal; Krystian Mikolajczyk; Jiri Matas
3	Kernel-Based Object Tracking [39]	190	USA	Dorin Comaniciu; Visvanathan Ramesh; Peter Meer
4	Object tracking: A survey [9]	185	USA	Alper Yilmaz ; Omar Javed; Mubarak Shah
5	KinectFusion: Real-Time Dense Surface Mapping and Tracking [40]	176	UK	Richard A. Newcombe; Shahram Izadi; Otmar Hilliges; David Molyneaux; David Kim; Andrew J. Davison; Pushmeet Kohli; Jamie Shotton; Steve Hodges; Andrew Fitzgibbon
6	Incremental Learning for Robust Visual Tracking [41]	172	USA	David A. Ross; Jongwoo Lim; Ruei-Sung Lin; Ming-Hsuan Yang
7	Robust Object Tracking with Online Multiple Instance Learning [42]	161	USA	Boris Babenko; Ming-Hsuan Yang; Serge Belongie
8	Object Tracking: Benchmark [43]	154	USA	Yi Wu; Jongwoo Lim; Ming-Hsuan Yang
9	CONDENSATION - Conditional Density Propagation for Visual Tracking [44]	145	UK	Michael Isard; Andrew Blake
10	Visual Tracking: An Experimental Survey [10]	132	USA	Arnold W. M. Smeulders; Dung M. Chu; Rita Cucchiara; Simone Calderara; Afshin Dehghan; Mubarak Shah

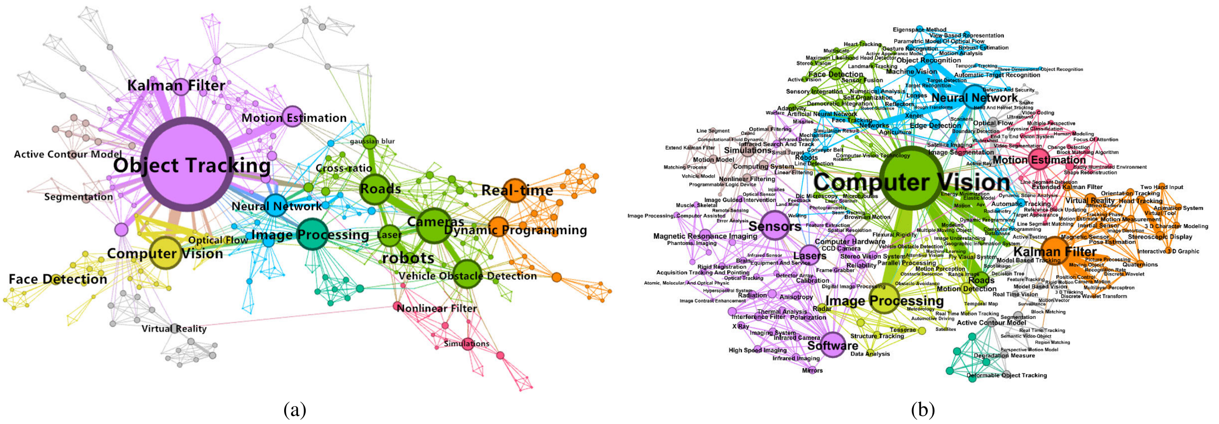


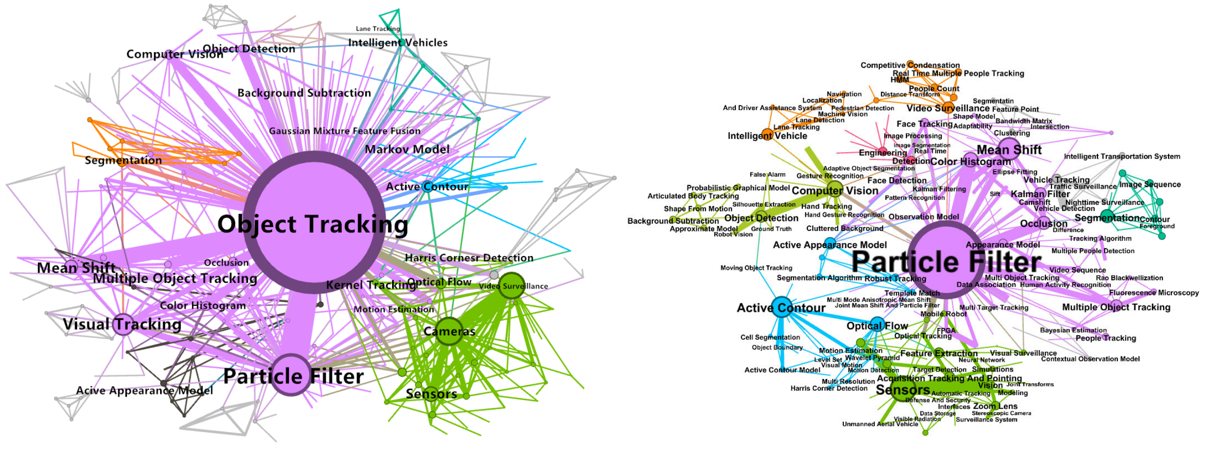
FIGURE 3. The co-occurrence keywords in the 1990s. (a) shows the co-occurrence of keywords in the 1990s and (b) shows the co-occurrence of filtered keywords possessing the same meanings in this period.

indicates the issue number of the article and the number of issues of the journal in the publication year.

The refined top-10 most cited papers are listed in Table 5, with the citation counts ranging from 238 to 132. Obviously, more than half of these papers are from USA, while the co-authors of them are from many countries. This reflects the leading position of USA in original research.

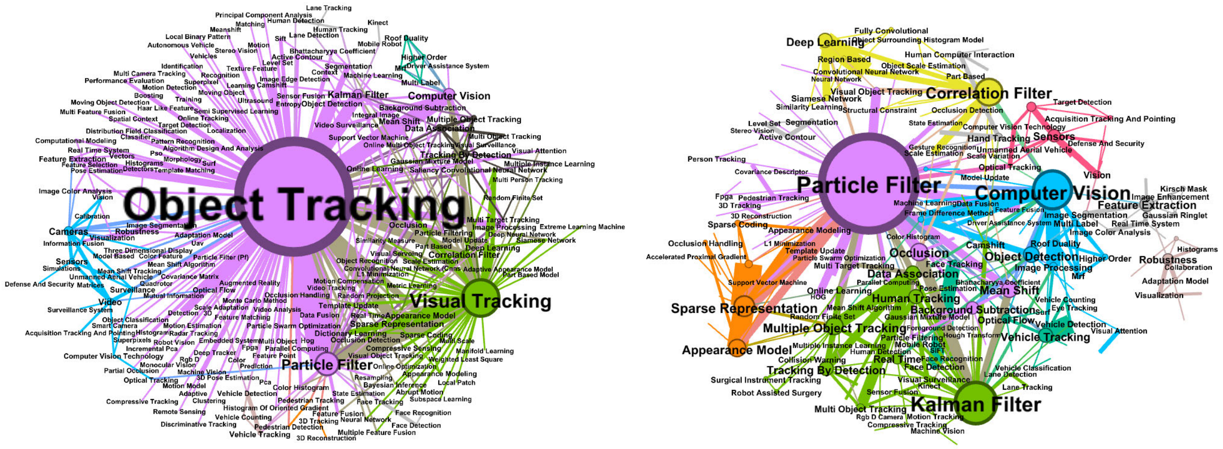
As listed in Table 5, the most cited paper [37] reviews the development in visual object tracking and builds a fully annotated dataset with 50 sequences. Its authors summarize the commonly used tracking methods and make the codes publicly available. Besides, new robustness evaluation methods are proposed in that paper to evaluate tracking performance more accurately.

The second most cited paper [38] proposes a long-term single-target tracking algorithm. The traditional tracking algorithms are combined with traditional detection algorithms to tackle the deformation and partial occlusion of the



(a) (b)

FIGURE 4. The co-occurrence keywords in the 2000s. (a) shows the co-occurrence of keywords in the 2000s and (b) shows the co-occurrence of filtered keywords during that decade.



(a) (b) .

FIGURE 5. The co-occurrence keywords in the 2010s. (a) shows the co-occurrence of keywords in the 2010s and (b) shows the filtered keywords in the 2010s.

tracked target during the tracking process. The detector's errors are estimated, and the learning process is modeled to update the online appearance features to get more accurate tracking results.

In the third most cited paper [39] presents a new approach of target representation and localization in visual object tracking (VOT). The authors use the isotropic kernel to regularize the feature histogram-based target representations, and use the spatial mask to induce spatially-smooth similarity. Then, they formulate the localization problem of the basin of attraction of the local maxima and use the mean-shift procedure to perform the optimization.

G. KEYWORDS

After viewing the information from a macroscopic aspect, we extract the keyword information from every literature, in order to better understand the research trends and difficulties in this field.

First, we unify the extracted keywords, by combining those which share similar meanings but have different forms, such as single or plural form, uppercase or lowercase, and abbreviation or full name. Then, we construct the keywords co-occurrence relationship network based on the WOS Collection covering the past 30 years. The corresponding figure is produced by Gephi, and the more frequently a keyword occurs, the bigger node it possesses. In order to have a more comprehensive understanding about target tracking in recent years, we analyze the keywords data by dividing the past 30 years into three decades, as shown in Fig. 3(a), Fig. 4(a) and Fig. 5(a). Consequently, the most frequent keywords are object tracking and visual tracking, the nodes of which are too big to find other meaningful keywords. To improve visualization, we reshape the co-occurrence of keywords after removing the object tracking and visual tracking keywords. The corresponding figures are shown in Figs. 3(b), 4(b) and 5(b), respectively. In this way, the research trends, popular methods, current research progresses, and the conundrums which need to be addressed in the tracking field can be easily obtained.

Fig. 3(a) shows the keywords co-occurrence in the first decade from 1990 to 1999. However, only 534 keywords and 1443 edges can be collected during this period, due to limited development of the visual tracking field and restricted computer configuration. According to the analysis results in Fig. 3(b), optical flow, dynamic programming and Kalman filter are popular keywords related to visual tracking. During that period, researchers also studied the subject of motion estimation and detection. Since neural network flourished in the 1980s, it was introduced to the visual tracking field as a popular technique.

Fig. 4(a) shows the keywords co-occurrence from 2000 to 2009. The figure contains 3266 nodes and 10262 edges, indicating that visual tracking in the 2000s got a rapid growth compared with the 1990s. However, the major keywords did not change much in this period. As Fig. 4(b) shows, in addition to what has been discussed, several new methods were proposed to deal with the challenges in visual tracking, including particle filter, mean-shift and kernel tracking. Except the real-time property, the robustness of tracking algorithms has also been concerned. In addition to the motion model, the establishment of a reliable appearance model gained attention.

Fig. 5(a) shows the co-occurrence of keywords from 2010 to 2019. There are totally 9442 keywords and 33075 edges in this figure, which indicates that visual tracking has made rapid and enormous progress in the 2010s. The number of keywords becomes nearly 3 times of the number in the 2000s and 20 times of the number in the 1990s. After removing the big but informationless nodes (i.e., object tracking and visual tracking), as shown in Fig. 5(b), Kalman filter, particle filter and mean-shift methods remain popular. Besides, machine learning has gained much attention of researchers. Among them, deep learning and convolutional neural network (CNN) are widely used in the visual tracking field to extract the features of various objects. Correlation filter, sparse representation, and pose tracking have acquired a growing popularity during the past decade. In terms of the feature extraction process, appearance model and object motion are two dominant aspects which have been widely utilized. The affinities of object candidates are calculated in order to determine whether and where these candidates should be connected. Besides, the cooperation among different topics in computer vision has become more and more frequent than the earlier 20 years. For example, image segmentation, recognition and detection are combined with visual tracking, thereby making the tracking methods more diverse and more effective.

In order to have a quantitative analysis of the keywords evolvement to better understand the development of the tracking field, we use the growth statistics of each keyword and list the top 15 ones that grow rapidly in the latest years, calculated as

$$C_i = \sum_{i=1}^{n-1} \frac{(f_{i+1} - f_j)}{f_i} \times 100\%, \tag{2}$$

where f_i is frequency of a particular keyword in the i^{th} time interval. The relevant result diagram is shown in Table 7. The last column shows the keyword growth rate from 1999 to the present. As can be seen, correlation filter, deep learning, particle filter, and sparse representation have the fastest growth rate, and the traditional methods such as color histogram, data association, mean shift, and template matching still maintain a high usage rate.

The above analysis of keywords is based on the occurrence frequency. Although it is natural, only focusing on these nodes is far from an in-depth understanding of the tracking field. Consequently, we analyze the frequently used methods in each decade to get the development trends, as shown in Fig. 6. For the first decade, optical flow tracking was a mainstream. In addition to diverse versions of Kalman filter, motion was primarily modeled. The Hough transform and Gabor wavelet were also used to extract the information of image boundaries and predict the object locations. For the establishment of tracking model and the setting of parameters, dynamic programming, RANSAC and other models were employed.

For the second decade, the newly used methods are marked in red, except the keywords that have appeared in the last decade. Consequently, particle filtering gained much progress. For extracting the features of images, HOG and SIFT were frequently used due to their good performance. Tracking methods such as mean shift and template matching were newly proposed. In addition, the Markov method was introduced to solve tracking problem, such as HMM, CRF and MRF. Non-parametric models were also proposed, including k-means clustering, SVM, and random forest.

Since 2010, massive novel keywords have appeared. The correlation filtering method is widely used in tracking to predict the location of a single object. In addition, the deep learning technology has been progressing rapidly. CNN, siamese network, LSTM and other deep learning methods have been widely used to obtain high-quality appearance and motion models. Using modules such as attention to deep learning also makes the model more accurate. It is worth mentioning that Re-ID has also been introduced into object tracking to obtain a better appearance model for objects. Tracklet-based tracking catches attention and structured models based on MRF and CRF are used to build tracklet association models and obtain final tracking results. In addition, for feature selection, PCA and spectral clustering are widely used. Such methods reduce the dimension of feature vector while preserving its representation ability.

In Fig. 6, we list several newly proposed methods in the tracking field. In order to get an in-depth understanding,

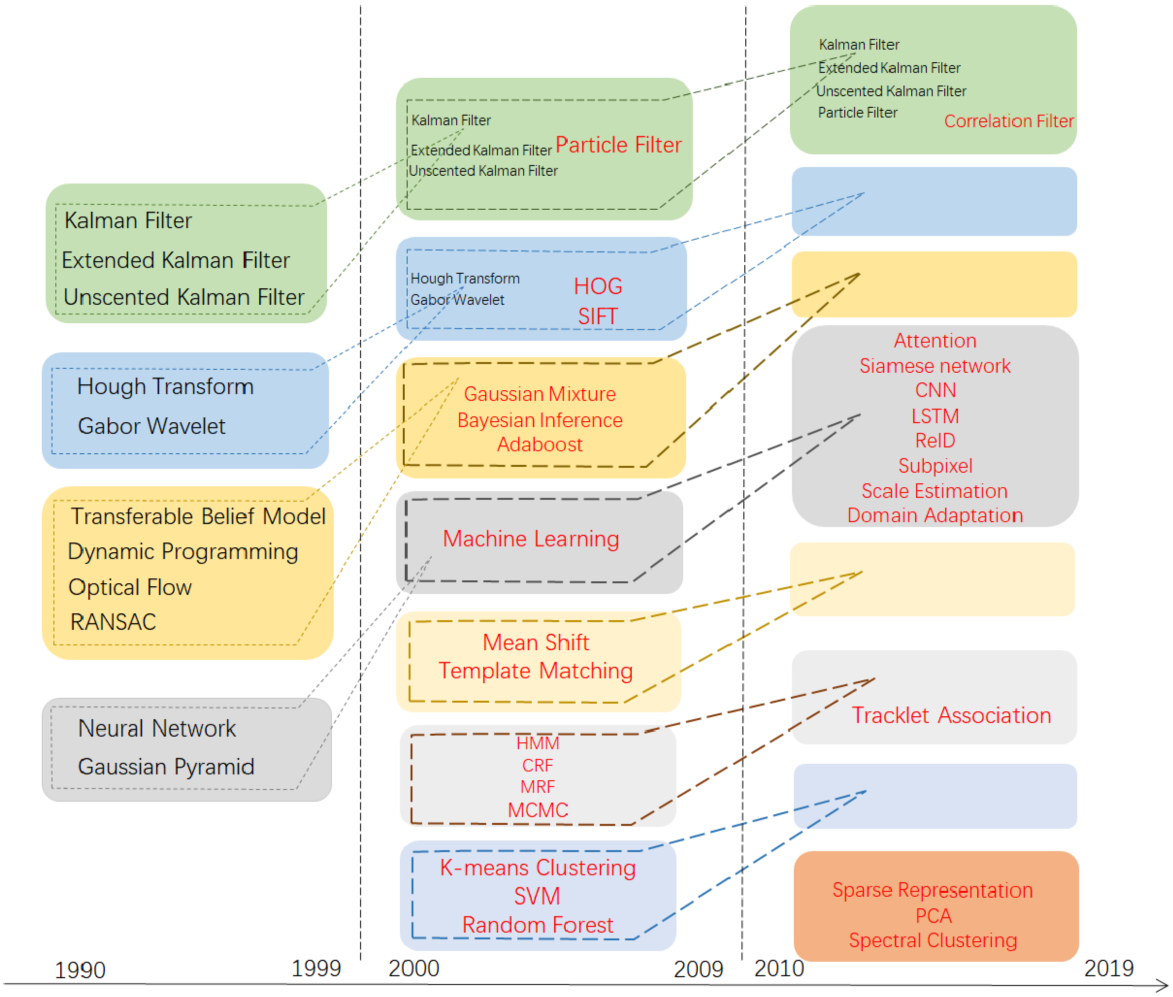


FIGURE 6. The development trend of tracking methods from 1990 to 2019.

we also list the techniques, used datasets and publication years of the methods in Table 6.

H. ANALYSIS OF IMPORTANT METHODS

In order to acquire detailed information of the visual tracking field, we select popular methods according to highly-cited papers during the past 30 years, including mean-shift, optical flow, Kalman filter, particle filter, tracking-by-detection, and correlation filter.

1) MEAN-SHIFT

The concept of mean shift was firstly proposed in 1975 by Fukunaga [57]. However, it was not until 2000 that mean-shift was introduced to the visual tracking field [58]. The method generally refers to an iterative step. In other words, when the offset of current point is calculated, it is moved to a new starting point. The process repeats until the terminal condition is met. This method has been widely used due to its robustness and computational efficiency. However, when applying the method to visual tracking, searching for object location in the next frame causes high computational complexity, especially when the number of objects is large. Worse still, the mean shift method requires a proper initialization.

Collins *et al.* [59] propose a method for selecting locally discriminative tracking features during the mean-shift process. They present an online automatic method to set candidate features and convert them into adjustable features by calculating log likelihood ratio between target and background. A two-class variance ratio is used to value those features and the most discriminative features are selected. Finally, weighted images and features are applied in the mean shift algorithm to determine the best positions of target between two frames.

2) OPTICAL FLOW

Optical flow is a description of motion information. The optical flow method was first proposed in 1981 by Horn and Schunck [60], who linked the two-dimensional velocity field with the gray scale, introduced the optical flow constraint equation, and proposed the basic algorithm of optical flow calculation. When applying optical flow to the visual

TABLE 6. The latest papers with novel methods.

Index	Method	Used dataset	Technique	Year
1	SiamRPN++: Evolution of Siamese Visual Tracking with Very Deep Networks [45]	OTB-2015, VOT2018, UAV123, LaSOT, and TrackingNet	Siamese networks	2019
2	Visual Tracking via Adaptive Spatially-Regularized Correlation Filters [46]	OTB-2015, TC128, VOT2016, VOT2017, and LaSOT	Correlation filtering	2019
3	Robust Visual Tracking via Nonlocal Regularized Multi-View Sparse Representation [47]	Datasets proposed in [37]	Sparse representation; Multi-view learning; Dual group structure	2019
4	Unsupervised Deep Tracking [48]	OTB-2015, VOT2016, and Temple-Color	Siamese network; Unsupervised learning	2019
5	Online Multi-Object Tracking with Dual Matching Attention Networks [49]	MOT16 and MOT17	Dual matching attention networks	2018
6	Data Association for Multi-Object Tracking via Deep Neural Networks [50]	MOT15	Bidirectional LSTM	2019
7	Multi-object Tracking with Neural Gating Using Bilinear LSTM [51]	MOT17, MOT15, ETH, KIT-TI, PETS09, TUD-Crossing, and AVG-TownCentre	Bilinear LSTM	2018
8	A Directed Sparse Graphical Model for Multi-Target Tracking [52]	Sprint, Football, Baseball, AFL and Basketball	Hidden Markov model; Domain Adaptation	2018
9	A Hierarchical Feature Model for Multi-Target Tracking [53]	MOT15, MOT16, and MOT17	Unsupervised dimensionality reduction; Bayesian filtering	2017
10	Tracking without Bells and Whistles [54]	MOT15, MOT16, and MOT17	Person re-identification; Siamese network	2019
11	Multi-Camera Multi-object Tracking [55]	EPFL Terrace Sequence and Duke MTMC	Generalized Maximum Multi Clique optimization; Iterative Hankel total least squares	2019
12	Online Inter-Camera Trajectory Association Exploiting Person Re-Identification and Camera Topology [56]	DukeMTMC	Orientation-driven person reidentification (ODPR); Camera topology estimation	2018

TABLE 7. Keywords trends.

Keyword	1995-1999(%)	2000-2004(%)	2005-2009(%)	2010-2014(%)	2015-2019(%)	Growth(%)
Correlation Filter	0.00	0.00	0.00	0.02	1.78	7026.46
Deep Learning	0.00	0.00	0.00	0.02	1.43	5638.19
Particle Filter	0.00	0.23	4.83	4.90	3.35	1933.80
Sparse Representation	0.00	0.00	0.05	0.75	2.13	1526.46
Convolutional Neural Network	0.00	0.00	0.00	0.05	0.58	1056.89
Color Histogram	0.00	0.12	1.04	0.40	0.25	689.83
Data Association	0.00	0.12	0.73	0.82	0.79	530.12
Pedestrian Tracking	0.00	0.00	0.05	0.30	0.44	523.60
Mean Shift	0.00	0.23	1.66	1.22	0.46	521.60
Dictionary Learning	0.00	0.00	0.00	0.10	0.58	478.45
Classification	0.00	0.00	0.05	0.27	0.25	421.52
UAV	0.00	0.00	0.05	0.25	0.28	391.94
Pose Estimation	0.00	0.12	0.36	0.17	0.56	376.11
Template Update	0.00	0.00	0.05	0.17	0.37	348.16
Background Subtraction	0.00	0.12	0.47	0.75	0.51	327.64

tracking issue, researchers should first process the sequence of consecutive video frames and use certain target detection method to detect the foreground objects. If a frame has a foreground target, one should find its representative feature points, and seek for the best position estimation where these feature points are located in the next frame. This aims to get the association of feature points of foreground target between two frames. After iterations, the long trajectory can be obtained.

Paragios and Deriche [61] apply edge-based models in motion estimation and tracking. Specifically, they use the optical flow method to derive local motion estimation. In the condition that global illumination stays unchanged during short time period, the regions of moving objects are successfully discriminated from the background areas.

3) KALMAN FILTER

Kalman filter is an algorithm that produces estimates of unknown variables using a series of measurements observed over time. Here we choose three representative papers to illustrate recent progress of this method.

Many tracking algorithms use Kalman filter to predict object position in next frames. The methods are primarily distinguished according to application scenarios. For example, in [62], which could be one of the most influential works in this field, Doyle *et al.* combine the methods of optical flow and Kalman filter.

Cox and Hingorani [63] introduce multiple hypothesis tracking algorithm into the visual tracking field, which provides a Bayesian framework to model the initial and terminal states of a trajectory. Besides, the position of diagonal points is predicted by Kalman filter and Mahalanobis matching is used to calculate the distances. Experiments show that their algorithm is robust.

4) PARTICLE FILTER

Particle filter is an important method in the field of visual tracking. It is mainly applied to predict object position, similar to the role of Kalman filter. However, they differ from many aspects. For example, particle filter can be employed in nonlinear and non-Gaussian noise situations, while Kalman filter is constrained to Gaussian hypothesis. This means that particle filter has a better versatility than Kalman filter. Nevertheless, as the versatility grows, the computational cost goes up. To reduce this cost, lots of modifications for particle filter have been proposed. Some typical variants like Gaussian Particle Filter (GPF) and Gaussian Sum Particle Filter (GSPF) [64] show better performance than the original one. Here, we introduce two influential works using particle filter for visual tracking.

Ross *et al.* [41] use particle filter to estimate motion parameters. By using the method, their tracker is able to recover from the drift caused by short-term large pose changes. First, the object to be tracked is determined in the first frame, and the appearance information of the object is initialized. Then, particles are drawn from particle filter by dynamical model. For each particle, calculate the distance between the candidate target and the center of the feature subspace. Then the tracking result is the candidate point with the shortest distance.

Arulampalam. *et al.* [65] review both optimal and suboptimal Bayesian algorithms for non-Gaussian tracking problems. For the latter, extended Kalman filter approximates grid-based methods and particle filters are introduced in detail. Especially for the particle filter, Sequential Importance Sampling (SIS) Algorithm and other related Particle Filters are introduced and compared in tracking.

5) TRACKING-BY-DETECTION

The tracking-by-detection methods have gradually caught attention in recent years. With the development of detection methods, visual tracking can be conducted based on the results of detection. The merits of automatic initialization and strong robustness bring this kind of methods into wide use.

Huang *et al.* [66] propose a hierarchical approach to matching different detections into the final tracks. The author propose a three-level measurement to analyze the tracklet affinity and refine the consequences based on position, size and appearance. Besides, network flow model is used as another solution to the tracking problem.

Wang *et al.* [67] propose a tracklet affinity estimation model based on online target-specific appearance and coherent dynamics. They construct stable tracklet nodes in network flow by learning the target-specific metric and correcting tracklets. Furthermore, motion, spatio-temporal and exit constraints are imposed to build the affinity formulas which constitute the edge weights of network flows.

Bae and Yoon [68] put forward a tracklet-confidence based association method. Detectability and continuity are used to calculate the tracklet confidence. The tracklets with high confidence are locally associated while the low-confidence tracklets are globally associated with other candidates. Besides, incremental linear discriminant analysis is used to project the characteristics of the trajectory sheet into another feature space to learn the high-area feature, where the trajectory of the same target can be more robustly clustered.

6) CORRELATION FILTER

In 2010, Bolme *et al.* [69] introduced correlation filter to the tracking field for the first time. Correlation filter is based on the idea that the more similar two signals, the higher the correlation value. Then in the tracking field, this method has been used to find the maximum response of the tracked target.

Henriques *et al.* [70] use the innovative circulant matrix to generate training samples. The advantage is that the resulting data matrix is circulant, so DFT (Discrete Fourier Transform) diagonalization can be used to reduce the amount of computation. They find that using this sample to train the linear regression model is equivalent to the correlation filter, which greatly speeds up the tracking. At the same time, the authors consider the case of kernel regression, and propose a kernelized correlation filter (KCF) and a dual correlation filter using a linear kernel.

In [71], Danelljan *et al.* combine CNN and correlation filter for visual tracking. They utilize the CNN method to extract features as the input for discriminative correlation filter based tracking frameworks. In this way, the model trimming can be easily realized and the activations are of low dimension. Their experiments show that the obtained tracking results are competitive to the best ones in OTB, ALOV300++ and VOT2015.

7) MULTI-OBJECT DATA ASSOCIATION

Compared with single-object tracking, multi-object tracking involves optimal matching between multiple targets. The problem is often solved with the graph theory by constructing association models. The commonly used graph theory methods include probabilistic graphical models, like Bayesian Network, Condition Random Field (CRF), Markov Random Field (MRF), and graph models, such as Network Flow (NF) and Bipartite Graph Match. Usually, the progress focuses on tracklets, which are scattered pieces of tracks. By constructing the nodes and affinity edges between different tracklets, the final tracks can be calculated through a matching process. The difference among methods mainly lies in the types of graphical models. Except [66], [67] using Network Flow introduced previously, we list several papers using different graphical models to obtain the tracking results.

Yang and Nevatia [72] propose an online learning-based tracking method using conditional random field. First, the nodes of the model are tracklet pairs whose head and tail intervals meet certain threshold conditions. Then, based on the motion model and appearance model, the energy functions of unary and binary terms are defined to distinguish the affinity of tracklets. Worth mentioning, the online learned discriminative appearance model is adopted to obtain the positive and negative training samples. The final trajectories are obtained by minimizing the total energy function.

Ullah and Cheikh [73] use the min-cost flow to solve the tracking problem and the Google Inception model to get the appearance similarity. The appearance and motion similarities are used to calculate the similarity of two nodes. Their experimental results show that their proposed method performs well on long sequences.

Wu et al. [74] combine face clustering label and face trajectory tracking through hidden Markov random field to transform into a Bayesian inference problem, and propose an effective coordinate degradation method. The method reduces tracking failures caused by linking tracklets with different clustering labels, and clustering the same target detections in the long term makes the clustering more reliable.

Wojke *et al.* [75] establish bipartite graph for the historical trajectories and current detections, and adopt the Hungarian algorithm to obtain the final online tracks. Based on appearance model trained on the person Re-ID dataset, and the existing trajectories, the authors use Kalman filter to predict the next track position. The Mahalanobis distance is used to calculate the candidate detection and the predicted position at the next moment, and the cascade matching is used to take account of objects that appear more frequently.

Ullah *et al.* [76] use the Bayesian filtering to model the MOT tracking problem. The HOG descriptor and the constant velocity model is used to get the appearance and motion features respectively. The bipartite graph matching is used to get the final tracking results. Furthermore, Ullah *et al.* [77] propose a synergetic novel appearance model for multi-object tracking. The Siamese neural network trained with contrastive loss function is used to increase the discrimination ability of similar targets.

8) MULTI-VIEW MULTI-OBJECT TRACKING

Compared with single-camera multi-object tracking, the multi-view multi-object (or multi-target multi-camera, MTMC) tracking needs to realize object matching between different views in different cameras besides the tracklet association in a single camera. We list several papers to explain the MTMC work.

Chen *et al.* [78] propose a global graphical model of similarity measure that integrates single-camera tracking and multi-camera tracking. The minimum cost flow model is used to calculate the final correlation tracklets. Innovatively, the authors disassemble a tracklet into input node, output node and associated edge. The confidence of the tracklet is considered as the affinity of the edges to make more comprehensive use of information. The tracklets involved in the model include all the tracklets inside and among the cameras, in order to obtain the global optimal results.

Ristani and Tomasi [79] use an adaptive weighted triplet loss for training the appearance model, and use linear motion model to predict motion affinity. For the association, there are three steps: one-second long tracklets computation, single-camera trajectories formation and multi-camera identities formation. The correlation clustering method is used in data association by considering both cost and accuracy. The standard hierarchical reasoning and sliding time window techniques are incorporated to reduce the computational complexity.

Ristani *et al.* [80] introduce a new dataset for MTMC tracking tasks and define new indexes of MTMC tracking performance by emphasizing the correct identities. Besides, they establish a baseline MTMC tracker for future comparisons.

V. COLLABORATION PATTERN ANALYSIS

In this section, collaboration patterns on the author and the author-keyword levels are analyzed to show the partnership and important research directions in the tracking field.

A. CO-AUTHORS ANALYSIS

In the field of scientific research, the value of cooperation is huge, therefore we analyze the authors's collaboration relationship to discover active groups. This is done through constructing the co-authorship network using network analysis methods, where the nodes represent the authors and the edges are drawn when collaboration exists. Here we use Gephi [26] to visualized the graphs and obtain basic statistics for further analysis. We first introduce several important indicators in bibliographic analysis, including the average degree, the diameter, the average shortest path and clustering coefficient.

Given a graph $G = (V, E)$, where V and E are a set of vertices and edges respectively. An edge $e_{i,j}$ connects vertice v_i with vertice v_j. The average degree for an undirected graph G can be calculated as

$$\overline{deg}(G) = 2 * \frac{\Sigma_{v_i} deg(v_i)}{|V|}, \tag{3}$$

where $deg(v_i)$ is the degree of a vertice $v_i \in V$. The average degree of a co-authors graph denotes the number of collaboration relationships.

The average shortest path is defined as the average distance between any two nodes in a graph G, which is calculated as:

$$\bar{d}(G) = \frac{\Sigma_{j,k} d(v_j, v_k)}{n(n-1)}, \tag{4}$$

where $d(v_j, v_k)$ denotes the distance between vertices v_j and v_k. The average shortest path indicates the closeness of collaboration within the group.

The diameter D of an undirected graph G is defined as the ‘‘longest shortest path’’ between any two vertices v_j and v_k,

$$D = \max_{j,k} d(v_j, v_k), \tag{5}$$

where $d(v_j, v_k)$ denotes the distance between vertices v_j and v_k. It shows the distance between the two farthest nodes in the graph.

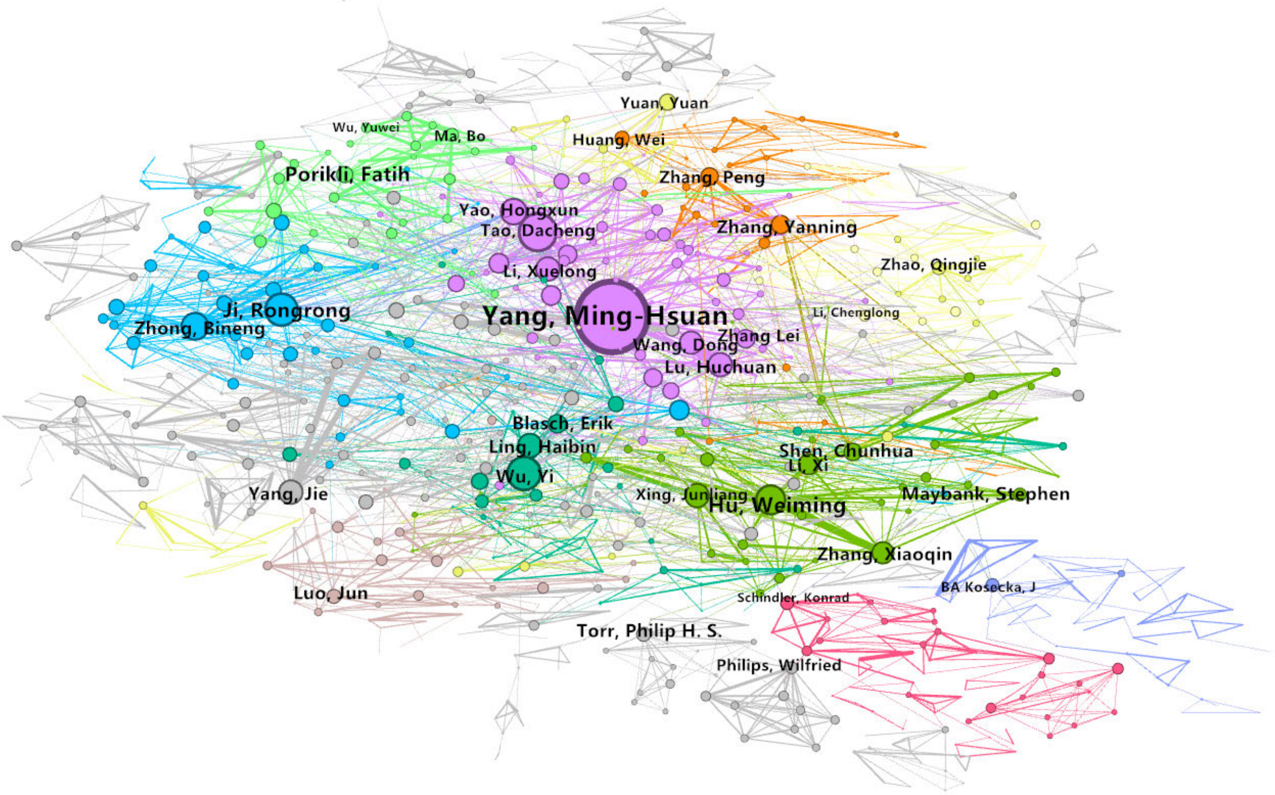


FIGURE 7. **The co-occurrence of authors from 1990 to 2019.**

The clustering coefficient C_i for an undirected graph for a vertice v_i is defined by the proportion of links between the vertices within its neighborhood and the the mumbler of links,

$$C_i = \frac{\{e_{jk} : v_j, v_k \in N_i, e_{j,k} \in E\}}{k_i(k_i - 1)}, \tag{6}$$

where k_i is the number of neighbors for vertice v_i, and $|N_i|$ is the compacity of the neighborhood for vertice v_i. The average clustering coefficient indicates how actively researchers in this group cooperate with others.

B. AUTHOR-KEYWORD ANALYSIS

In order to make a specific and concise elaboration of the collaboration patterns on the author level, we analyze the four largest communities. The corresponding co-author figure is showed in Fig. 7.

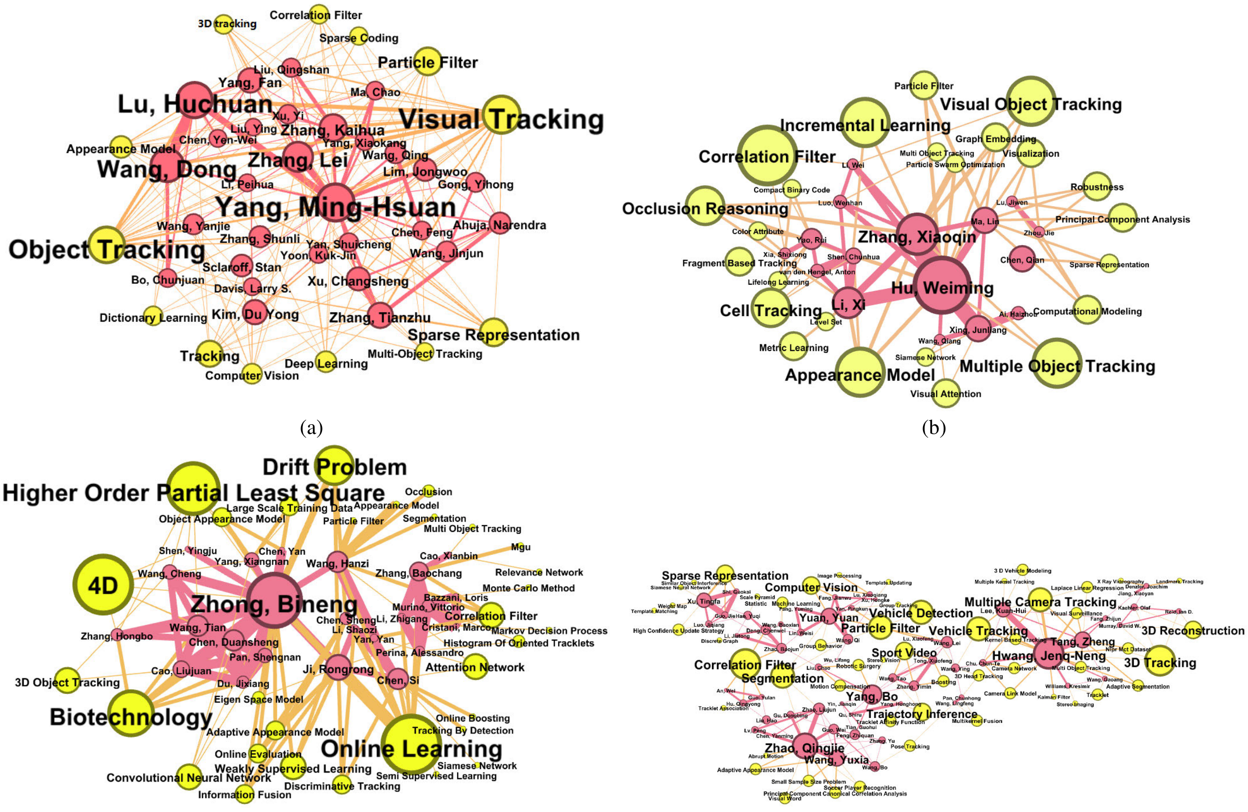
The largest community has 135 nodes and 362 edges. In this community, the average degree is 5.363, the average shortest path is 3.331, the diameter is 8, and the clustering coefficient is 0.663. The critical researchers Ming-Hsuan Yang is with the School of Engineering, University of California, Merced, and Huchuan Lu is with Dalian University of Technology, China.

The second largest community has 80 nodes and 204 edges. The average degree is 5.1, the average shortest path length is 4.477, the diameter is 11, and the clustering coefficient is 0.673. Weiming Hu from Institute of Automation, Chinese Academy of Sciences acts as the key author in this community.

The third largest community has 67 nodes and 218 edges, and the core author is Bineng Zhong who is with Huaqiao University, China. The average degree is 6.507, the average shortest path length is 3.525, the diameter is 10, and the clustering coefficient is 0.698. The members of these subgroups come mainly from Chinese and Italian institutions.

There are 61 nodes and 112 edges in the fourth largest community. The average degree in this community is 3.672, the average shortest path length is 5.924, the diameter is 15, and the clustering coefficient is 0.616. Qingjie Zhang from Beijing Institute of Technology, plays a central role.

Fig. 8 shows the author-keyword co-occurrence networks of the top 4 clusters. As can be seen, researchers centered on Ming-Hsuan Yang and Huchuan Lu mainly study the topics of multi-object tracking, sparse representation, and deep learning methods. Except 2D tracking, they are also interested in 3D tracking. Besides, both online and offline tracking are studied by the researchers in this cluster. The second community focuses on multiple object tracking. Additionally, they also engage in cell tracking and occlusion reasoning. For the neural network model, they focus on attention and siamese networks. The robustness of tracking methods and correlation filter are paid much attention. The third cluster pays attention to drift and online learning in tracking. They are also interested in Markov decision process to solve the



(a)

(b)

(c)

(d)

FIGURE 8. The author-keyword co-occurrence in the top 4 largest communities. (a) shows the author-keyword co-occurrence in the largest community while (b), (c) and (d) show the second, third and fourth clusters.

multi-object tracking problem. The fourth community studies the sparse representation, trajectory inference, multiple camera tracking, and correlation filter. They pay much attention to sports video, in which the tracked persons have similar appearance. According to the investigation, occlusion is one of the major conundrums that need to be tackled, and various deep learning methods coupled with traditional ways are frequently utilized to get better appearance and motion models.

Fig. 9 shows the keywords that frequently occur in recent publications and their relations to major authors. Conclusively, neural network, correlation filter, and attention are three major branches in dealing with the tracking issue. Specifically, in the neural network branch, CNN and recurrent neural network (RNN) are built to achieve better performance. Ming-Hsuan Yang, Huchuan Lu, Fatih Porikli and other productive authors are interested in both the correlation filter and neural network. For the attention model, Weiming Hu and Peng Wang conduct in-depth research. Besides, optical flow and tracklet association methods are studied by Jenq-Neng Hwang and Dacheng Tao. Worth mentioning, pose tracking has gradually attracted the attention of researchers due to the improvement of image understanding. Unlike traditional tracking methods, pose tracking can obtain pedestrian attitudes and produce more detailed representation. Nassir Navab and Luc Van Gool study this issue a lot. Besides, reinforcement learning [81], GAN [82], and interaction model [83] are introduced to object tracking by Li Fei-Fei, who enriches the research content of the visual tracking field and stimulates new research hotspots.

VI. THOUGHTS AND PROSPECTS

The field of visual tracking has experienced rapid development in the past 30 years, particularly in the past decade. However, there still exist challenges leading to accuracy drop, ID switch and trace loss. In this section, we suggest some thoughts and prospects.

A. DIRECT TRACKING

By saying direct tracking we mean that end-to-end tracking could be the leading hotspot in the future.

The tracking-by-detection method is a general idea in the past decades, whereas the tracking accuracy is tremendously influenced and restricted by detection methods. Although various tracking methods have been proposed, such as Kalman filter, mean-shift and other traditional methods, challenges (e.g., occlusion, objects with similar appearance, route crossing, and objects re-appearance) still seriously bother the researchers.

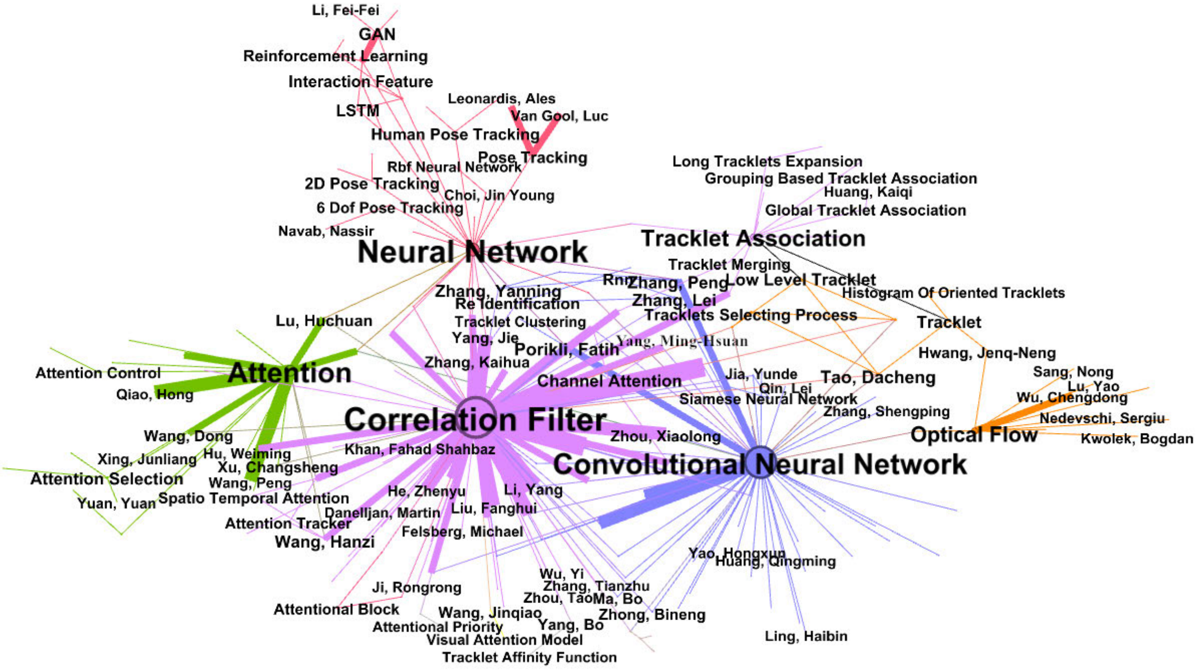


FIGURE 9. The co-occurrence of frequent keywords and popular authors in recent years.

Machine learning especially deep learning has regained its popularity in computer vision. Deep learning models, such as CNN and RNN, bring great benefits to visual detection, image classification, and image-to-image translation. The deep leaning models, which are end-to-end method, surpass traditional methods that utilize hand-crafted features. The end-to-end tracking method takes raw data as input and directly outputs the results, so there is no need for human intervention in the feature extraction process. This can give the model more flexibility to adjust automatically according to data and increase the overall fitness between the model and data. Consequently, current tracking features that heavily rely on human experience could be replaced by automatically learned features, which should be more powerful. In this way, researchers can study the direct tracking approach without explicitly dealing with the detection problem.

B. EXACT TRACKING

Tracking with detailed information makes the process of tracking more exact.

Currently, the bounding-box-based tracking are mainly used. When the targets are occluded with each other heavily, such methods contain pixels of other targets which cause interference. As a result, the drift and ID switch frequently occur. For now, the tracking representation is gradually changing from simple bounding boxes to more exact modes, such as pose [84] and mask [85]. Compared with bounding box tracking, such methods could remove background information and more precisely understand the appearance, movement and other characteristics of a certain target. Therefore, it could improve the tracking accuracy and reliability.

In addition, due to the improvement of camera resolution, part-based tracking such as face tracking and gait tracking becomes popular. When the scene is crowded, each object occupies few pixels, so that very limited information could be captured. Therefore, part-based tracking can be effective as long as the concerned part is obtained.

3D tracking [86] will become another trend which makes use of depth information. In this way, the occlusion problem in 2D tracking could be better solved by applying top view information. It can be used from autonomous driving to visual SLAM, showing a wide range of applications.

C. VIRTUAL-REAL INTERACTION

The tracking field has less data available for learning tracking features under diverse scenarios. The parallel vision approach [7], [8] could augment the data and improve the existing tracking effect via virtual-real interaction.

With the rapid development of the computer graphics and virtual/augmented reality technologies, the virtual environments can be built more and more realistically, making virtual-real interaction practicable. OpenStreetMap, CityEngine, and Unity3D can generate virtual environment. The virtual-real interaction idea brings benefits to the training and testing of a visual tracking algorithm. It can produce sufficient video/image data to overcome the difficulty in collecting and annotating large-scale diversified training data. In this way, those situations that own few or no training samples can be obtained, such as extremely adverse weather, terrible traffic accidents, and some military applications. Furthermore, the most important merit of virtual-real interaction is that the ground truth of synthetic video/image data can be generated automatically, without the need of manual annotation which

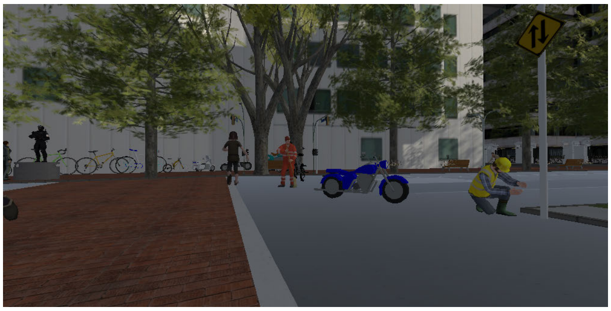


FIGURE 10. A virtual image from ParallelEye.

is always time-consuming and expensive, not to mention the accuracy loss brought up by manual annotation in complex situations such as low illumination.

Moreover, in the virtual space, one is able to produce different kinds of virtual data by controlling the environmental condition and virtual objects, thereby enlarging the size of dataset. The diversity introduced by this process can also help to train a more robust visual tracking model. As shown in Fig. 10, Li *et al.* [87] release a virtual dataset ParallelEye that includes different scenes and diverse weathers, benefitting the virtual training and testing. They build virtual artificial scenes using simulation tools to simulate the scenarios whatever may occur in real scenes, and automatically generate accurate annotations and obtain large-scale diversified virtual dataset. Gaidon *et al.* [88] generate "Virtual KITTI" dataset by cloning the KITTI dataset and automatically obtain annotation information. According to their experimental results, the pre-training on Virtual KITTI dataset improves the performance of tracking in real world.

VII. CONCLUSION

In this paper, the visual tracking literatures from 1990 to 2019 on the WOS are retrieved and analyzed, and the co-occurrence patterns of authors and keywords are obtained by using statistical analysis and network analysis methods.

According to the analyses, Chinese institutions have published the most papers on visual tracking, but the most highly cited papers come mainly from USA and UK. Since each country/region has its own research strengths, the cooperation between different countries/regions becomes a compelling need to achieve mutual communication and information sharing among researchers. Therefore, we use the network analysis method to build the co-authors, co-keywords and author-keyword co-occurrence networks to visualize the collaboration patterns and research topics. The authors' cooperation has gradually strengthened, more knowledge in relevant fields has been introduced, and the research content between groups has become greatly intersected. Generally speaking, the research hotspots in visual tracking gradually change to more accurate tracking. We also propose some thoughts and prospects on future visual tracking research. In our opinion, direct tracking, exact tracking, and virtual-real interaction will receive much attention from the visual tracking community.

REFERENCES

[1] S. S. Rautaray and A. Agrawal, "Vision based hand gesture recognition for human computer interaction: A survey," *Artif. Intell. Rev.*, vol. 43, no. 1, pp. 1–54, 2015.

[2] Y. Liu, K. Wang, and D. Shen, "Visual tracking based on dynamic coupled conditional random field model," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 3, pp. 822–833, Mar. 2016.

[3] K. Wang, W. Huang, B. Tian, and D. Wen, "Measuring driving behaviors from live video," *IEEE Intell. Syst.*, vol. 27, no. 5, pp. 75–80, Sep./Oct. 2012.

[4] H. Sun, "The study of military affairs target recognition and tracking method based on wavelet analysis," Ph.D. dissertation, School Mech. Manuf. Automat., Changchun Univ. Sci. Technol., Jilin, China, 2008.

[5] K. Q. Huang, X. T. Chen, Y. F. Kang, and T. N. Tan, "Intelligent visual surveillance: A review," *Chin. J. Comput.*, vol. 38, no. 6, pp. 1093–1118, 2015.

[6] B. D. Anderson and S. Vongpanitlerd, *Network Analysis and Synthesis: A Modern Systems Theory Approach*. Chelmsford, MA, USA: Courier Corporation, 2013.

[7] K. Wang, C. Gou, and F.-Y. Wang, "Parallel vision: An ACP-based approach to intelligent vision computing," *Acta Autom. Sinica*, vol. 42, no. 10, pp. 1490–1500, 2016.

[8] K. Wang, C. Gou, N. Zheng, J. M. Rehg, and F.-Y. Wang, "Parallel vision for perception and understanding of complex scenes: Methods, framework, and perspectives," *Artif. Intell. Rev.*, vol. 48, no. 3, pp. 299–329, 2017.

[9] A. Yilmaz, O. Javed, and M. Shah, "Object tracking: A survey," *ACM Comput. Surv.*, vol. 38, no. 4, 2006, Art. no. 13.

[10] A. W. M. Smeulders, D. M. Chu, R. Cucchiara, S. Calderara, A. Dehghan, and M. Shah, "Visual tracking: An experimental survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 36, no. 7, pp. 1442–1468, Jul. 2014.

[11] X. Li, W. Hu, C. Shen, Z. Zhang, A. Dick, and A. Van Den Hengel, "A survey of appearance models in visual object tracking," *ACM Trans. Intell. Syst. Technol.*, vol. 4, no. 4, p. 58, Sep. 2013.

[12] H. S. Parekh, D. G. Thakore, and U. K. Jaliya, "A survey on object detection and tracking methods," *Int. J. Innov. Res. Comput. Commun. Eng.*, vol. 2, no. 2, pp. 2970–2979, 2014.

[13] Z. Hou and Z. Han, "A survey of visual tracking," *Acta Autom. Sinica*, vol. 32, no. 4, pp. 603–617, 2006.

[14] E. Trucco and K. Plakas, "Video tracking: A concise survey," *IEEE J. Ocean. Eng.*, vol. 31, no. 2, pp. 520–529, Apr. 2006.

[15] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 32, no. 3, pp. 478–500, Mar. 2010.

[16] N. Buch, S. A. Velastin, and J. Orwell, "A review of computer vision techniques for the analysis of urban traffic," *IEEE Trans. Intell. Transp. Syst.*, vol. 12, no. 3, pp. 920–939, Sep. 2011.

[17] X. Gu, Y. Mao, and Q. Li, "Survey on visual tracking algorithms based on mean-shift," *Comput. Sci.*, vol. 39, no. 12, pp. 16–24, 2012.

[18] S. Y. Chen, "Kalman filter for robot vision: A survey," *IEEE Trans. Ind. Electron.*, vol. 59, no. 11, pp. 4409–4420, Nov. 2012.

[19] L. Wan-Yi, W. Peng, and Q. Hong, "A survey of visual attention based methods for object tracking," *Acta Autom. Sinica*, vol. 40, no. 4, pp. 561–576, 2014.

[20] H. Zhang, S. Hu, and G. Yang, "Video object tracking based on appearance models learning," *J. Comput. Res. Develop.*, vol. 52, no. 1, pp. 177–190, 2015.

[21] T. Wang, X. Wang, S. Tang, Y. Lin, W. Liu, Z. Liu, B. Xiu, D. Shen, X. Zhao, and Y. Gao, "Collaborations patterns and productivity analysis for IEEE T-ITS between 2010 and 2013," *IEEE Trans. Intell. Transp. Syst.*, vol. 15, no. 6, pp. 2360–2367, Dec. 2014.

[22] X. Xu, W. Wang, Y. Liu, X. Zhao, Z. Xu, and H. Zhou, "A bibliographic analysis and collaboration patterns of IEEE transactions on intelligent transportation systems between 2000 and 2015," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 8, pp. 2238–2247, Aug. 2016.

[23] J. A. Moral-Muñoz, M. J. Cobo, F. Chiclana, A. Collop, and E. Herrera-Viedma, "Analyzing highly cited papers in intelligent transportation systems," *IEEE Trans. Intell. Transp. Syst.*, vol. 17, no. 4, pp. 993–1001, Apr. 2016.

[24] A. Emrouznejad and G.-L. Yang, "A survey and analysis of the first 40 years of scholarly literature in DEA: 1978–2016," *Socio-Econ. Planning Sci.*, vol. 61, pp. 4–8, Mar. 2018.

[25] A. Zhang, *Protein Interaction Networks: Computational Analysis*. Cambridge, U.K.: Cambridge Univ. Press, 2009.

[26] M. Bastian, S. Heymann, and M. Jacomy, "Gephi: An open source software for exploring and manipulating networks," in *Proc. 3rd Int. AAAI Conf. Weblogs Social Media*, 2009.

[27] M.-H. Yang, D. Kriegman, and N. Ahuja, "Detecting faces in images: A survey," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 1, pp. 34–58, Jan. 2002.

[28] W. Zhong, H. Lu, and M.-H. Yang, "Robust object tracking via sparsity-based collaborative model," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2012, pp. 1838–1845.

[29] C. Ma, J. Huang, X. Yang, and M. Yang, "Hierarchical convolutional features for visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 3074–3082.

[30] L. Wang, L. Wang, H. Lu, P. Zhang, and X. Ruan, "Salient object detection with recurrent fully convolutional networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 41, no. 7, pp. 1734–1746, Jul. 2019.

[31] L. Zheng, Y. Huang, H. Lu, and Y. Yang, "Pose invariant embedding for deep person re-identification," *IEEE Trans. Image Process.*, vol. 28, no. 9, pp. 4500–4509, Sep. 2019.

[32] P. Zhang, W. Liu, H. Wang, Y. Lei, and H. Lu, "Deep gated attention networks for large-scale street-level scene segmentation," *Pattern Recognit.*, vol. 88, pp. 702–714, Apr. 2019.

[33] W. Hu, G. Tian, Y. Kang, C. Yuan, and S. Maybank, "Dual sticky hierarchical Dirichlet process hidden Markov model and its application to natural language description of motions," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 10, pp. 2355–2373, Oct. 2018.

[34] W. Hu, J. Gao, J. Xing, C. Zhang, and S. Maybank, "Semi-supervised tensor-based graph embedding learning and its application to visual discriminant tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 1, pp. 172–188, Jan. 2017.

[35] J. Xing, Z. Niu, J. Huang, W. Hu, X. Zhou, and S. Yan, "Towards robust and accurate multi-view and partially-occluded face alignment," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 987–1001, Apr. 2018.

[36] S. Uddin and A. Khan, "The impact of author-selected keywords on citation counts," *J. Inform.*, vol. 10, no. 4, pp. 1166–1177, 2016. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S1751157716301146

[37] Y. Wu, J. Lim, and M.-H. Yang, "Online object tracking: A benchmark," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2013, pp. 2411–2418.

[38] Z. Kalal, K. Mikolajczyk, and J. Matas, "Tracking-learning-detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 34, no. 7, pp. 1409–1422, Jul. 2012.

[39] D. Comaniciu, V. Ramesh, and P. Meer, "Kernel-based object tracking," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 25, no. 5, pp. 564–577, May 2003.

[40] R. A. Newcombe, S. Izadi, O. Hilliges, D. Molyneaux, D. Kim, A. J. Davison, P. Kohli, J. Shotton, S. Hodges, and A. W. Fitzgibbon, "Kinectfusion: Real-time dense surface mapping and tracking," in *Proc. ISMAR*, vol. 11, 2011, pp. 127–136.

[41] D. A. Ross, J. Lim, R.-S. Lin, and M.-H. Yang, "Incremental learning for robust visual tracking," *Int. J. Comput. Vis.*, vol. 77, nos. 1–3, pp. 125–141, 2008.

[42] B. Babenko, M.-H. Yang, and S. Belongie, "Robust object tracking with online multiple instance learning," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 33, no. 8, pp. 1619–1632, Aug. 2011.

[43] Y. Wu, J. Lim, and M. H. Yang, "Object tracking benchmark," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 9, pp. 1834–1848, Sep. 2015.

[44] M. Isard and A. Blake, "CONDENSATION–conditional density propagation for visual tracking," *Int. J. Comput. Vis.*, vol. 29, no. 1, pp. 5–28, 1998.

[45] B. Li, W. Wu, Q. Wang, F. Zhang, J. Xing, and J. Yan, "Siamrpn++: Evolution of siamese visual tracking with very deep networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4282–4291.

[46] K. Dai, D. Wang, H. Lu, C. Sun, and J. Li, "Visual tracking via adaptive spatially-regularized correlation filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 4670–4679.

[47] B. Kang, W. P. Zhu, D. Liang, and M. Chen, "Robust visual tracking via nonlocal regularized multi-view sparse representation," *Pattern Recognit.*, vol. 88, pp. 75–89, Apr. 2019.

[48] N. Wang, Y. Song, C. Ma, W. Zhou, W. Liu, and H. Li, "Unsupervised deep tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, 2019, pp. 1308–1317.

[49] J. Zhu, H. Yang, N. Liu, M. Kim, W. Zhang, and M.-H. Yang, "Online multi-object tracking with dual matching attention networks," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 366–382.

[50] K. Yoon, D. Y. Kim, M. Jeon, and Y. C. Yoon, "Data association for multi-object tracking via deep neural networks," *Sensors*, vol. 19, no. 3, p. 559, 2019.

[51] C. Kim, F. Li, and J. M. Rehg, "Multi-object tracking with neural gating using bilinear LSTM," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 200–215.

[52] M. Ullah and F. A. Cheikh, "A directed sparse graphical model for multi-target tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. Workshops*, Jun. 2018, pp. 1816–1823.

[53] M. Ullah, A. K. Mohammed, F. A. Cheikh, and Z. Wang, "A hierarchical feature model for multi-target tracking," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 2612–2616.

[54] P. Bergmann, T. Meinhardt, and L. Leal-Taixe, "Tracking without bells and whistles," 2019, *arXiv:1903.05625*. [Online]. Available: https://arxiv.org/abs/1903.05625

[55] W. Liu, O. I. Camps, and M. Sznaier, "Multi-camera multi-object tracking," Sep. 2017, *arXiv:1709.07065*. [Online]. Available: https://arxiv.org/abs/1709.07065

[56] N. Jiang, S. Bai, Y. Xu, C. Xing, Z. Zhou, and W. Wu, "Online inter-camera trajectory association exploiting person re-identification and camera topology," in *Proc. 26th ACM Int. Conf. Multimedia (MM)*, New York, NY, USA, 2018, pp. 1457–1465. [Online]. Available: http://doi.acm.org/10.1145/3240508.3240663

[57] K. Fukunaga and L. Hostetler, "The estimation of the gradient of a density function, with applications in pattern recognition," *IEEE Trans. Inf. Theory*, vol. 21, no. 1, pp. 32–40, Jan. 1975.

[58] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, vol. 2, Jun. 2000, pp. 142–149.

[59] R. T. Collins, Y. Liu, and M. Leordeanu, "Online selection of discriminative tracking features," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 10, pp. 1631–1643, Oct. 2005.

[60] B. K. P. Horn and B. G. Schunck, "Determining optical flow," *Artif. Intell.*, vol. 17, nos. 1–3, pp. 185–203, Aug. 1981.

[61] N. Paragios and R. Deriche, "Geodesic active regions and level set methods for motion estimation and tracking," *Comput. Vis. Image Understand.*, vol. 97, no. 3, pp. 259–282, 2005.

[62] D. D. Doyle, A. L. Jennings, and J. T. Black, "Optical flow background estimation for real-time pan/tilt camera object tracking," *Measurement*, vol. 48, pp. 195–207, Feb. 2014.

[63] I. J. Cox and S. L. Hingorani, "An efficient implementation and evaluation of Reid's multiple hypothesis tracking algorithm for visual tracking," in *Proc. 12th Int. Conf. Pattern Recognit.*, vol. 1, 1994, pp. 437–442.

[64] J. H. Kotecha and P. M. Djuric, "Gaussian sum particle filtering," *IEEE Trans. Signal Process.*, vol. 51, no. 10, pp. 2602–2612, Oct. 2003.

[65] M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp, "A tutorial on particle filters for online nonlinear/non-Gaussian Bayesian tracking," *IEEE Trans. Signal Process.*, vol. 50, no. 2, pp. 174–188, Feb. 2002.

[66] C. Huang, B. Wu, and R. Nevatia, "Robust object tracking by hierarchical association of detection responses," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2008, pp. 788–801.

[67] B. Wang, G. Wang, K. L. Chan, and L. Wang, "Tracklet association by online target-specific metric learning and coherent dynamics estimation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 3, pp. 589–602, Mar. 2016.

[68] S.-H. Bae and K.-J. Yoon, "Robust online multi-object tracking based on tracklet confidence and online discriminative appearance learning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2014, pp. 1218–1225.

[69] D. S. Bolme, J. R. Beveridge, B. A. Draper, and Y. M. Lui, "Visual object tracking using adaptive correlation filters," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2010, pp. 2544–2550.

[70] J. F. Henriques, R. Caseiro, P. Martins, and J. Batista, "High-speed tracking with kernelized correlation filters," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 37, no. 3, pp. 583–596, Mar. 2015.

[71] M. Danelljan, G. Hager, F. Shahbaz Khan, and M. Felsberg, "Convolutional features for correlation filter based visual tracking," in *Proc. IEEE Int. Conf. Comput. Vis. Workshops*, Dec. 2015, pp. 58–66.

[72] B. Yang and R. Nevatia, "Multi-target tracking by online learning a CRF model of appearance and motion patterns," *Int. J. Comput. Vis.*, vol. 107, no. 2, pp. 203–217, 2014.

[73] M. Ullah and F. A. Cheikh, "Deep feature based end-to-end transportation network for multi-target tracking," in *Proc. 25th IEEE Int. Conf. Image Process. (ICIP)*, Oct. 2018, pp. 3738–3742.

[74] B. Wu, S. Lyu, B.-G. Hu, and Q. Ji, "Simultaneous clustering and tracklet linking for multi-face tracking in videos," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2013, pp. 2856–2863.

[75] N. Wojke, A. Bewley, and D. Paulus, "Simple online and realtime tracking with a deep association metric," in *Proc. IEEE Int. Conf. Image Process. (ICIP)*, Sep. 2017, pp. 3645–3649.

[76] M. Ullah, F. A. Cheikh, and A. S. Imran, "HoG based real-time multi-target tracking in Bayesian framework," in *Proc. 13th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS)*, Aug. 2016, pp. 416–422.

[77] M. Ullah, H. Ullah, and F. A. Cheikh, "Single shot appearance model (SSAM) for multi-target tracking," *Electron. Imag.*, vol. 2019, no. 7, pp. 1–466, 2019.

[78] W. Chen, L. Cao, X. Chen, and K. Huang, "An equalized global graph model-based approach for multicamera object tracking," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 27, no. 11, pp. 2367–2381, Nov. 2017.

[79] E. Ristani and C. Tomasi, "Features for multi-target multi-camera tracking and re-identification," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 6036–6046.

[80] E. Ristani, F. Solera, R. Zou, R. Cucchiara, and C. Tomasi, "Performance measures and a data set for multi-target, multi-camera tracking," in *Proc. Eur. Conf. Comput. Vis.* Springer, 2016, pp. 17–35.

[81] Y. Xiang, A. Alahi, and S. Savarese, "Learning to track: Online multi-object tracking by decision making," in *Proc. IEEE Int. Conf. Comput. Vis.*, Dec. 2015, pp. 4705–4713.

[82] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi, "Social-GAN: Socially acceptable trajectories with generative adversarial networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 2255–2264.

[83] A. Sadeghian, A. Alahi, and S. Savarese, "Tracking the untrackable: Learning to track multiple cues with long-term dependencies," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Oct. 2017, pp. 300–311.

[84] M. Andriluka, U. Iqbal, E. Insafutdinov, L. Pishchulin, A. Milan, J. Gall, and B. Schiele, "Posetrack: A benchmark for human pose estimation and tracking," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 5167–5176.

[85] P. Voigtlaender, M. Krause, A. Osep, J. Luiten, B. B. G. Sekar, A. Geiger, and B. Leibe, "MOTS: Multi-object tracking and segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2019, pp. 7942–7951.

[86] F. Mueller, F. Bernard, O. Sotnychenko, D. Mehta, S. Sridhar, D. Casas, and C. Theobalt, "Ganerated hands for real-time 3D hand tracking from monocular RGB," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 49–59.

[87] X. Li, K. Wang, Y. Tian, L. Yan, F. Deng, and F. Wang, "The ParallelEye dataset: A large collection of virtual images for traffic vision research," *IEEE Trans. Intell. Transp. Syst.*, vol. 20, no. 6, pp. 2072–2084, Jun. 2019.

[88] A. Gaidon, Q. Wang, Y. Cabon, and E. Vig, "Virtual worlds as proxy for multi-object tracking analysis," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2016, pp. 4340–4349.

YATING LIU received the B.S. degree from the Civil Aviation University of China, in 2014. She is currently pursuing the Ph.D. degree with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences and the University of Chinese Academy of Sciences. Her research interests include visual object detection and tracking, machine learning, and intelligent transportation systems.

KUNFENG WANG (M'11–SM'18) received the Ph.D. degree in control theory and control engineering from the Graduate University of Chinese Academy of Sciences, Beijing, China, in 2008. After that, he joined the Institute of Automation, Chinese Academy of Sciences, where he became an Associate Professor with the State Key Laboratory for Management and Control of Complex Systems. From December 2015 to January 2017, he was a Visiting Scholar with the School of Interactive Computing, Georgia Institute of Technology, Atlanta, GA, USA. In August 2019, he moved to the Beijing University of Chemical Technology, as a Professor with the College of Information Science and Technology. His research interests include intelligent transportation systems, intelligent vision computing, and machine learning. He currently serves as an Associate Editor for the IEEE Transactions on Intelligent Transportation Systems.

XUESONG LI received the B.S. degree in automation from the University of Electronic Science and Technology of China, Chengdu, China, in 2017. He is currently pursuing the M.S. degree with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences, and the University of Chinese Academy of Sciences. His research interests include visual tracking, image processing, and machine learning.

TIANXIANG BAI received the B.S. degree from Zhejiang University, in 2013. He is currently pursuing the Ph.D. degree with the State Key Laboratory for Management and Control of Complex Systems, Institute of Automation, Chinese Academy of Sciences and the University of Chinese Academy of Sciences. His research interests include robotics, reinforcement learning, and unmanned aerial vehicles.

FEI-YUE WANG (S'87–M'89–SM'94–F'03) received the Ph.D. degree in computer and systems engineering from the Rensselaer Polytechnic Institute, Troy, NY, USA, in 1990. He joined the University of Arizona, in 1990 and became a Professor and the Director of the Robotics and Automation Lab (RAL) and Program in advanced research for complex systems (PARCS). In 1999, he founded the Intelligent Control and Systems Engineering Center with the Institute of Automation, Chinese Academy of Sciences (CAS), Beijing, China, under the support of the Outstanding Oversea Chinese Talents Program from the State Planning Council and 100 Talent Program from CAS, where, he was appointed as the Director of the Key Lab of Complex Systems and Intelligence Science, in 2002. In 2011, he became the State Specially Appointed Expert and the Director of The State Key Laboratory for Management and Control of Complex Systems. His current researches focus on methods and applications for parallel systems, social computing, and knowledge automation. He is also an elected Fellow of INCOSE, IFAC, ASME, and AAAS. In 2007, he received the Second Class National Prize in Natural Sciences of China and awarded the Outstanding Scientist by ACM for his work in intelligent control and social computing, the IEEE ITS Outstanding Application and Research Awards, in 2009 and 2011, and the IEEE SMC Norbert Wiener Award, in 2014. He was the Founding Editor-in-Chief of the *International Journal of Intelligent Control and Systems*, from 1995 to 2000, the Founding EiC of the *IEEE ITS Magazine*, from 2006 to 2007, and the EiC of the IEEE Intelligent Systems, from 2009 to 2012, and the IEEE Transactions on ITS, from 2009 to 2016. He is currently the EiC of *China's Journal of Command and Control*. Since 1997, he has served as the General or Program Chair of more than 20 IEEE, INFORMS, ACM, ASME conferences. He was the President of the IEEE ITS Society, from 2005 to 2007, the Chinese Association for Science and Technology (CAST), USA, in 2005, the American Zhu Kezhen Education Foundation, from 2007 to 2008, and the Vice President of the *ACM China Council*, from 2010 to 2011. Since 2008, he has been the Vice President and the Secretary General of Chinese Association of Automation.