

A Survey on Object Tracking in Aerial Surveillance



Junhao Zhao, Gang Xiao, Xingchen Zhang and Durga Prasad Bavirisetti

Abstract Nowadays the Unmanned Aerial Vehicle (UAV) has been widely used due to its low-cost and unique flexibility. Specifically, the high-altitude operational capability makes it the ideal tool in military and civilian surveillance system, in which object tracking based on computer vision is the core ingredient. In this paper, we presented a survey on object tracking methods in aerial surveillance. After briefly reviewing the development history and current research institutions, we summarized frequently used sensors in aerial platform. Then we focused on UAV-based tracking methods by providing detailed descriptions of its common framework (ego motion compensation, object detection, object tracking) and representative tracking algorithms. Through discussing the requirement of a good tracking system and deficiency of current technologies, future directions for aerial surveillance were proposed.

Keywords Object tracking · UAV · Aerial surveillance · Computer vision

1 Introduction

Nowadays, UAV is widely used and continually expanding its market because of its low-cost, unique flexibility and high-altitude operational capability. Compared to human, UAV can carry out tasks including but not limited to disaster search [1], power lines detection [2], traffic monitoring [3] etc. safely, easily and efficiently. According to Carroll and Rathbone [4], the estimated budget for traffic data collection is about

J. Zhao · G. Xiao (✉) · X. Zhang · D. P. Bavirisetti
School of Aeronautics and Astronautics, Shanghai Jiao Tong University, Shanghai, China
e-mail: xiaogang@sjtu.edu.cn

J. Zhao
e-mail: jhzhao_16@sjtu.edu.cn

X. Zhang
e-mail: xingchen@sjtu.edu.cn

D. P. Bavirisetti
e-mail: bdps1989@sjtu.edu.cn

© Springer Nature Singapore Pte Ltd. 2019

Z. Jing (ed.), *Proceedings of International Conference on Aerospace System Science and Engineering 2018*, Lecture Notes in Electrical Engineering 549,
https://doi.org/10.1007/978-981-13-6061-9_4

Fig. 1 Aerial surveillance over a highway



\$5 million per in an average metropolitan area, while using UAV, we can reduce the total cost by 20% and half of collecting procedures. Therefore, UAV is called the best tool for performing 3D (the Dull, the Dirty and the Dangerous) tasks [5].

Object tracking is one of the hot topics in the field of computer vision, which uses a bounding box locks onto the region of interest (ROI) such as person and vehicle. Give the initial location of the target, then computer can find its location in the next sequences. This technology is one of the important applications used in UAV for ground strike, criminal vehicle etc. as well as plays an important role in other process such as estimating velocity and position of an object [6], UAV landing [7], search and rescue [1]. Particularly, in aerial surveillance, through object tracking technologies, traffic flow over a highway during a period can be estimated and a proactive approach can be adopted for an effective traffic management by identifying and evaluating potential problems before they occur [4], as shown in Fig. 1.

In general, tracking accuracy reflects the tracking performance. Various factors affect this performance such as illumination change, abrupt motion, scale variation and full or partial occlusion [8]. Although the tracking algorithm is more and more robust and efficient, no one can handle all scenarios [8]. In addition, different from static camera, aerial object tracking is also influenced by low-sampling rate, resolution and unstable camera platform, which caused by running vehicle and wind that lead to tracking drifts. When the altitude of the flight is great, the objects on the ground looks so small that it is hard to detect them. Hence, realizing a robust and stable tracking algorithm or system is still an issue to be addressed immediately.

The rest of the paper is organized as follows: Sect. 2 introduces the history and current research institutions of UAV vision, while Sect. 3 summarizes the sensors used in aerial platform. Section 4 mainly discusses the tracking framework and algorithm, collected common datasets and evaluation metrics. Future directions are given in Sect. 5. Finally, Sect. 6 concludes this paper.

2 The Development of UAV Vision

2.1 History of UAV Vision

The first aerial video was captured by Nadal, a famous French photographer in December, 1858 [9]. He used an old-fashioned wet plate camera on the hot air balloon. Then, in World War II, the main belligerent countries used aerial camera to carry out reconnaissance, but this way could not meet the needs of real-time. Then, people concentrated on inventing airborne optoelectronic platform. The famous tactical UAV—"Scout", Israel created, was able to send video, which was obtained through the visible light sensors in the optoelectronic platform, back to display. During the Lebanese war in 1982, Israel became the first country to use the real-time image transferring technology in aerial platform [10]. Since the 1990s, UAVs were used almost exclusively in military applications and they also have been finding commonplace usage in civilian applications. For instance, New Mexico State University used UAV to observe whether fishermen were fishing in legal areas [11].

2.2 Current Research Institution

Medioni from Institute for Robotics and Intelligent Systems, University of Southern California and his research group devote themselves to aerial vision research [12]. They are developing the system of wide area aerial surveillance and aim to build an efficient, scalable framework to provide activity inference from airborne imagery [13]. This system includes image mosaicking, video stabilization, object detection and tracking, and activity inference from wide area aerial videos [14–16].

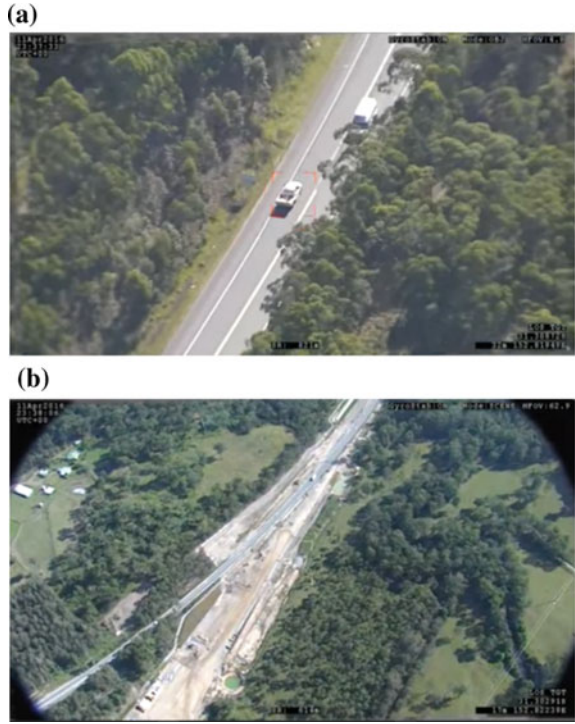
The Air lab of Carnegie Mellon University develops and tests perception and planning algorithms for UAVs [17]. Their research fields include indoor scene understanding, indoor flight in degraded visual environments, micro air vehicle scouts for intelligent semantic mapping etc.

UAV Vision is a company that design and manufacture high performance, lightweight, gyro-stabilized camera payloads for ISR applications [18]. Their sensors can be installed in different aircraft such as fixed wing UAV, multi rotor UAV or rotary wing UAV and carry out various tasks like disaster management, search and rescue. Particularly, when using their CM202U, user can track a moving vehicle from a long distance [19], as shown in Fig. 2.

DJI is a famous company about UAV in China which manufactures and designs UAV, cameras, flight control systems etc. [20]. In civil domain, their products are globally used for music, television and film industries. According to the statistics, DJI is the world's leader in the civilian drone and aerial imaging technology industry, accounting for 85% of the global consumer drone market [21].

In addition, some associated conferences and journals are also offer platforms to UAV researchers and fans. For instance, automated vehicles symposium [22], Inter-

Fig. 2 Tracking a moving vehicle for law enforcement applications



national Conference on Unmanned Aircraft Systems [23] and International Journal of Intelligent Unmanned Systems.

3 Sensors Used in Aerial Platform

Without the airborne optoelectronic platform, the UAV vision cannot be developed. Therefore, the advancement of optoelectronic platform will benefit this technology. This section introduces some common sensors used in aerial platform. Each sensor has its own imaging mechanism and characteristic, which are described in Table 1.

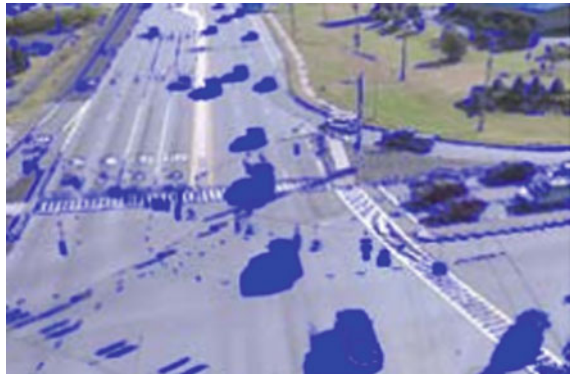
4 Aerial Platform Based Object Tracking

In this section, we first review the object tracking algorithms, which are used in the UAV, followed by common datasets and evaluation metrics.

Table 1 Common sensors and main features

Type of sensor	Main feature
Visible light sensor	Image has high contrast; rich information of color, appearance, shape, geometry and texture; high temporal and spatial resolution
Infrared sensor	It can work day and night; the detection range is from a few thousand meters to more than ten thousand meters; it can be divided into near infrared, mid infrared and long infrared bands
Synthetic aperture radar	It works all day. It has penetrating ability with high resolution
Laser imaging radar	It includes three functions of ranging, velocity measurement and imaging; it has high anti-interference ability, resolution and accuracy
Low light level sensor	Used for night detection, ranging from 800 to 1000 m
Multi-spectral or hyper-spectral sensor	Multiple spectral segments can be used to measure targets precisely at the same time. Its usages are topographic mapping, monitoring and analysis

Fig. 3 False alarms due to moving camera [24]



4.1 Common Framework

Object tracking in aerial surveillance is to estimate the states of the target on the ground through detection algorithm or selecting ROI manually to give the initialized state of it. As shown in Fig. 3, the aerial platform based object tracking consists three main steps. They are (1) Ego Motion Compensation; (2) Object Detection and (3) Object Tracking. Behavior analysis for decision making is the output.

Ego Motion Compensation. Ego Motion Compensation is for the image stabilization because of the moving camera by registering video frames onto a reference plane. It is the basic step, otherwise we will get false alarms in the next step for pixel intensity of the background changing, as shown in Fig. 3.

The compensation algorithm can be divided into gray-level [25] based, feature [26] based and transform domain [27] based. In aerial surveillance, feature based methods are often used. Through extracting feature information such as corners, points, lines and edges etc. of two images to carry out the match between them and establish affine model to finish the registration.

Object Detection. The means of detection are various. If we focus on a suspicious object, we can select ROI manually. When the UAV flies high to monitor the traffic condition, there are many vehicles on the ground and detection algorithm is needed, in which false alarms may occur. Usually, optical flow [28], frame differencing [29] and background subtraction [30] are common methods for detection.

Optical flow is defined as the apparent motion of the brightness patterns or the feature points in the image, which can be calculated from the movement of pixels with the same brightness value between two consecutive images [31]. Frame differencing uses the difference of two adjacent frames to detect moving objects. The concept of background subtraction is using the gray difference between the current image and background image to detect object. In addition, parallax, similar appearance, objects merge or split, occlusion etc. will affect detection accuracy [32].

Object Tracking. Object tracking method can be divided as generation methods [33] and discriminant methods [34]. The former methods means in the current frame modeling the object area and in the next frame to find the most similar area as the predicted location. The latter methods means in the current frame extract the features of object and background as positive and negative samples respectively to train classifier. In the next frame, use the classifier to distinguish foreground and use the result to update the classifier. Now, discriminant method is popular because it is more robust.

There are three main modules in object tracking [8]. First, target representation scheme: define an object as anything that is of interest for further analysis [35]. Second, search mechanism: estimate the state of the target objects. Third, model update: update the target representation or model to account for appearance variations.

In aerial surveillance tracking, data associative trackers which belongs to generation methods are often used. It takes input as a number of data points of the form (X, t) where X is a position (usually in 2- or 3-space), and t is the timestamp associated with that position [36]. The tracker then assigns an identifier to each data point indicating its track ID of each object.

Behavior Analysis. Behavior analysis includes the recognition of event, group activity, human roles, and traffic accident prediction etc. It is the output of aerial surveillance tracking for administrators to make decisions. Probabilistic network method is widely used because of its robustness to small changes of motion sequences in time and space scales.

Its function is to define each static posture of a movement as a state or a set of states. Through network to connect these states and using probability to describe the switching between state and state. Hidden Markov Models [37] and Dynamic Bayesian Networks [38] are the representations.

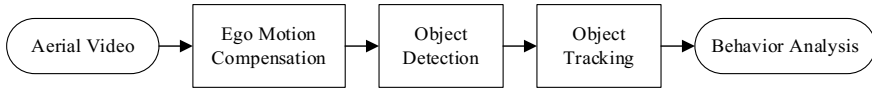


Fig. 4 Common framework of aerial platform based object tracking

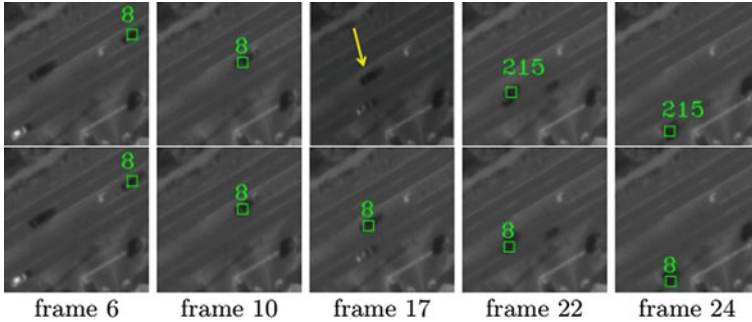


Fig. 5 Results of DBT only (first row) and DBT & LCT (second row)

4.2 Object Tracking Algorithms

In [39], Medioni et al. presented a methodology, in 1997, to perform the analysis of a video stream took from an UAV whose goal was to provide an alert mechanism to a human operator. It was the beginning of their video surveillance and monitoring (VSAM) project. The main procedure follows Fig. 4. In [32, 40], they also followed this and plotted the object trajectory showed in the mosaic image to help to infer their behavior. Nevertheless, these methods cannot deal with the effect of parallax which will lead to false alarm of detection.

Recently, in 2017, they used detection-based tracker (DBT) and local context tracker (LCT) simultaneously to track the vehicles on the ground [41]. Because the object in the airborne images is small, gray and the displacement of a moving target is large, rely merely on DBT is unreliable. By introducing LCT, which explores spatial relations for a target to avoid unreasonable model deformation in the next frame, to relax the dependency on frame differencing motion detection and appearance information. Here, DBT explicitly handles merged detections in detection association. The results showed that this method has a high detection rate, except for its high computation time and inability to long-term occlusion (Fig. 5).

Ali et al. proposed COCOA system for tracking in aerial imagery [42]. The whole framework likes [32, 40]. The system works well but the scenario is simple and no vehicles merging occurs. In [43], they used motion and appearance context for tracking and re-acquiring. It is the first time to use context knowledge in aerial imagery processing. Briefly, the appearance context is used to discriminate whether the objects are occluded or not and similar motion context of the unoccluded objects is used to predict the location of occluded ones as shown in Fig. 6. Obviously, it

Fig. 6 Using motion context to handle occlusion



can handle occlusion while it needs reference knowledge and does not take slow or stopped vehicles into account.

Perera et al. proposed a tracking method under the conditions of long occlusion and split-merge simultaneously [44]. The object detection is performed by background modeling which flags a pixel whether it belongs to foreground or background. A simple nearest-neighbor data association tracker is used, in which Kalman filter updates the position and velocity of objects. Long occlusion is solved by tracklets linking according one-to-one correspondence. In terms of merges and splits, suppose two objects A and B merge for a while, so that

$$\begin{aligned} A &= \{T_{a,1}, \dots, T_{a,m}, T_{c,1}, \dots, T_{c,o}, T_{d,1}, \dots, T_{d,p}\} \\ B &= \{T_{b,1}, \dots, T_{b,n}, T_{c,1}, \dots, T_{c,o}, T_{e,1}, \dots, T_{e,q}\} \end{aligned} \quad (1)$$

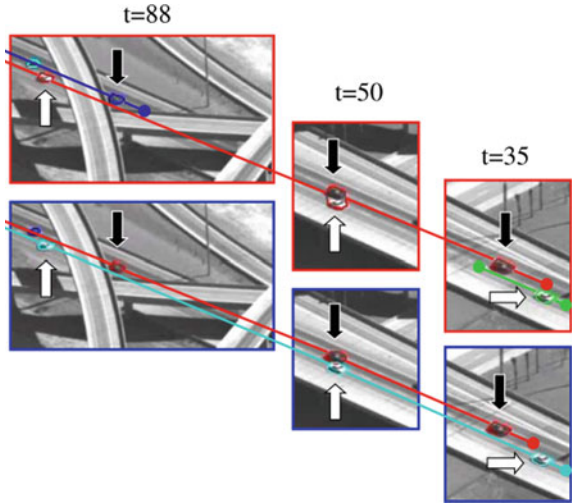
where T_c represents the merging period, the rest are splitting one. Using the pairwise assumption:

$$\begin{aligned} P(A, B) &= P(\{T_{a,1}, \dots, T_{a,m}\}) \times P(\{T_{b,1}, \dots, T_{b,n}\}) \\ &\quad \times P_m(\{T_{a,m}, T_{b,n} \rightarrow T_{c,1}\}) \times P(\{T_{c,1}, \dots, T_{c,o}\}) \\ &\quad \times P_s(\{T_{c,o} \rightarrow T_{d,1}, T_{e,1}\}) \times P(\{T_{d,1}, \dots, T_{d,p}\}) \\ &\quad \times P(\{T_{e,1}, \dots, T_{e,q}\}) \end{aligned} \quad (2)$$

with P_m and P_s denote the probability of a merge and split, respectively. The results showed that tracker will not be confused after two vehicles merging and continue tracking on the same vehicle when they split, as shown in Fig. 7. However, this method needs 30 frames to initialize the background model and does not take slow or stopped vehicles into account.

Xiao et al. proposed a joint probabilistic relation graph approach to detect and track vehicles [45], in which background subtraction is used because it can make up the drawback of three-frame subtraction that it is hard to detect slow or stopped vehicles. Vehicle behavior model is exploited to estimate potential travel direction and speed for each individual vehicle. In line with expectations, more stopped and

Fig. 7 The images (red border) confused while the images (blue border) show linking after merge processing



slow vehicles are detected, while due to the overlap when two vehicles merge, the detection accuracy will be affected. In the results, track identifications were missed that all detected vehicles were marked as the same color.

Keck et al. realized the real-time tracking of low-resolution vehicles for aerial surveillance [36]. The airborne images are characterized that they have 100 megapixels, which will increase the burden of computation. To solve this problem, they divided the large images into tiles and set TileProcessors to process tiles in parallel. FAST-9 algorithm (feature based), three-frame difference and Kalman filter are used to perform registration, detection and tracking respectively. Through quantitative results, the detection and tracking accuracy are high. Because of parallelism, the efficiency of computation can meet the need of real-time. However, occlusion, merge and split, same appearance of objects will affect these accuracy.

Some state-of-the-art methods are summarized in Table 2.

4.3 Common Datasets

Some common datasets of airborne imagery that can be used for object tracking are collected and listed below:

VIVID datasets [48]. This datasets is created for tracking ground vehicles from airborne sensor platforms. Its functions includes ground-truthed data set, some baseline tracking algorithm and a mechanism for compare yours results with ground-truth.

UAV123 dataset [49]. All videos in this dataset are captured from low-altitude UAVs. It contains a total of 123 video sequences and more than 110 K frames.

Table 2 Object tracker in aerial surveillance, their components and performance

References	Registration	Detection	Tracking	Slow or stopped	Occlusion	Merge and split	Track identifications	Computational efficiency
[46]	Feature points	Motion pattern	Motion pattern	✗	✓	✗	✓	N/A
[14]	Gray-level	Background subtraction	Motion pattern	✗	✗	✗	✓	High
[41]	✗	Optical flow	Local context	✓	✗	✓	✓	High
[43]	✗	Manually	Motion context and appearance context	✗	✓	✗	✓	N/A
[47]	✗	✗	Meanshift	✗	✗	✗	✗	N/A
[44]	KLT features	Background model	Nearst-neighbor data association	✗	✓	✓	✓	Low
[45]	Geography features	Three-frame subtraction & background subtraction	Graph match & vehicle behavior model	✓	✗	✗	✗	High
[36]	Feature points	Three-frame difference	Kalman filter	✗	✗	✗	✓	High
[13]	Feature points	Background model	Graph-based model	✗	✓	✗	✗	N/A

CLIF 2006 dataset [50]. This datasets is established by Air Force Research Labs of America. It is used for the research of aerial surveillance. Its features are high altitude, large field of view and small objects.

SEAGULL dataset [51]. A multi-camera multi-spectrum (visible, infrared, near infrared and hyper-spectral) image sequences dataset for research on sea monitoring and surveillance. The image sequences are recorded from a fixed wing UAV flying above the Atlantic Ocean.

In addition, Image Sequence Server dataset [52], WPAFB 2009 dataset [53], UCF Aerial Action Data Set [54] and UCLA Aerial Event Dataset [55] are also common aerial image datasets.

4.4 Evaluation Metrics

When tracking algorithm is performed, the results should be evaluated both qualitatively and quantitatively to illustrate whether the algorithm is robust or not.

Qualitative evaluation. Generally, we use a bounding box or more to contain the object(s) in pixels we want to track. Then, evaluation is carried out by our eyes.

If the tracking algorithm is robust and accurate, the bounding box will lock on the appearance of the object as much as possible, whenever occurs illumination change, occlusion, abrupt motion etc. Otherwise, when it drifts, it is weak and inaccurate.

Quantitative evaluation. Only qualitative evaluation is not persuasive. Quantitative evaluation always couples with qualitative one.

In [8], author introduces four metrics for evaluation of single object. They are Center Location Error (CLE), Distance Precision (DP), Overlap Precision (OP) and Frames Per Second (FPS).

CLE refers to the Euclidean distance between the estimated location and ground-truth location of the object. The smaller the value is, the better the performance is.

$$\text{CLE} = \sqrt{(x - x_0)^2 + (y - y_0)^2} \quad (3)$$

DP refers to the percentage of the frames whose CLE is smaller than a threshold among the whole sequences. The higher the value is, the better the performance is.

$$\text{DP} = \frac{N_{\text{CLE} \leq \text{th}}}{N} \times 100\% \quad (4)$$

OP refers to the percentage of the frames in which the overlap rate between bounding box and area of ground-truth is higher than a threshold. The higher the value is, the better the performance is.

$$\text{OP} = \frac{N_{\phi \geq \text{th}}}{N} \times 100\%, \phi = \frac{A_{\text{output}} \cap A_{\text{ground truth}}}{A_{\text{output}} \cup A_{\text{ground truth}}} \quad (5)$$

FPS refers to the how many frames the algorithm can process in one second. The higher the value is, the better the performance is:

$$\text{FPS} = N/t \quad (6)$$

5 Future Directions

Although those state-of-the-art methods have lower false alarms and higher tracking accuracy, some issues are still “bottlenecks” that constrain the further development of UAV-based tracking.

(1) Appearance change

When object moving, the pose and shape may change. Additionally, illumination variation will also affect tracker.

(2) Occlusion

Because object lost in view during occlusion, trackers may not resume tracking when occlusion ends.

(3) Complex background

Due to the high-altitude angle of surveillance, objects may drown in background that brings difficulty to detection.

(4) Merge and split

When objects merge, some trackers consider them as one object that lose identities even switch track ID.

(5) Computation efficiency

With the improvement of sensors, the rising of megapixels and more objects being tracked, the amount of calculation will be greater. The requirements of efficient algorithm and high performance hardware facilities need to be meet.

In the future, some improvements and innovations maybe realized:

(1) Rarely use traditional generation methods

Detection-based tracking method will be the mainstream for aerial surveillance, in which background information, local models and dynamic model are critical components [8]. Fully using background information can separate object and background well. Local model can fight against appearance change. Dynamic model is used for prediction that search region can be minimized.

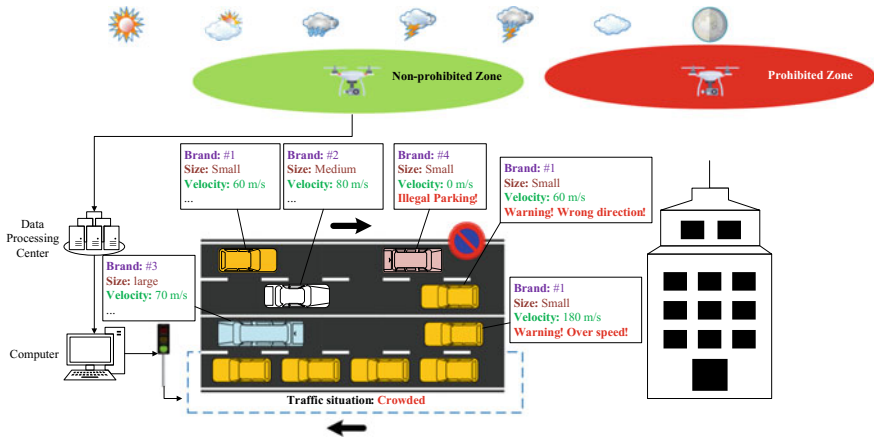


Fig. 8 The UAV-based traffic surveillance system in the future

(2) **Surveillance with AI technology**

Now the Artificial Intelligence (AI) has been widely and deeply studied. The technologies of machine learning (ML) and deep learning (DL) have shown their great power in the field of computer vision, automation and even Go [56]. Through the feature extraction and training of massive data, computers can even compete with humans and they can do some work instead of us. As shown in Fig. 8, if supported by the department of transportation and automobile manufacturer, we can train models by the data (prior knowledge) they offered and use detection and tracking algorithm to realize UAV traffic monitoring. After UAV taking and transferring videos to the data processing center (DPC), the brand, size, velocity etc. can be recognized online. At the same time, situation estimation and congestion judgment will be performed by object tracking. These information will be all received by the operators who are in front of computer. Then they can make decision that whether send out a warning signal or give priority of driving.

(3) **No ego motion compensation**

Advanced technology of UAV stabilization such as flight control and wind estimation may decrease the needs of ego motion compensation that this step can be optional.

(4) **Lower computation burden**

Advanced hardware, processor and efficient algorithms in aerial platform can relax the computation burden that the situation of scene and the processed data can be reflected to the observers in real-time.

(5) **Persistent working ability**

UAVs should meet the needs of working all-time and all-weather if they are used in engineering. Working merely in good weather is far from enough. The technology of waterproof and battery etc. should be progressed as soon as possible.

(6) More open airspace

The limited flying space will make the UAV useless. In the future, more airspace will be available for operators. While UAVs are flying, operators should also obey the flight rules in the non-prohibited zone and keep in mind that prohibited zone is inviolable at all times.

6 Conclusion

This paper presents a survey of object tracking in aerial surveillance. First, the development history and current research institutions are reviewed. Then, frequently used sensors are summarized which are followed by the detailed descriptions of the common frame work and representative tracking algorithms of aerial surveillance. Some suggestions and future directions are proposed for the deficiency of the current technologies in which we conclude that by combining advanced algorithm with AI technology, the UAV can play a greater role in the field of aerial surveillance.

Acknowledgements This paper is sponsored by National Program on Key Basic Research Project (2014CB744903), National Natural Science Foundation of China (61673270), Shanghai Pujiang Program (16PJD028), Shanghai Industrial Strengthening Project (GYQJ-2017-5-08), Shanghai Science and Technology Committee Research Project (17DZ1204304) and Shanghai Engineering Research Center of Civil Aircraft Flight Testing.

References

1. Chikwanha, A., Motepe, S., & Stopforth, R. (2012, November). Survey and requirements for search and rescue ground and air vehicles for mining applications. In *Proceeding of the 19th Conference on Mechatronics and Machine Vision in Practice* (pp. 105–109).
2. Bian, J., Hui, X., Yu, Y., Zhao, X., & Tan, M. (2017, December). A robust vanishing point detection method for UAV autonomous power line inspection. In *Proceeding of the Conference on Robotics and Biomimetics on IEEE* (pp. 646–651).
3. Ke, R., Li, Z., Tang, J., Pan, Z., & Wang, Y. (2018). Real-time traffic flow parameter estimation from UAV video based on ensemble classifier and optical flow. In *IEEE Transaction on Intelligent Transportation Systems*.
4. Carroll, E. A., & Rathbone, D. B. (2002, January). Using an unmanned airborne data acquisition system (ADAS) for traffic surveillance, monitoring and management. In *Proceedings of the Conference on ASME 2002 International Mechanical Engineering Congress and Exposition* (pp. 145–157).
5. Clapper, J. R., Young, J. J., Cartwright, J. E., & Grimes, J. G. (2009). Office of the secretary of defense unmanned systems roadmap (2009–2034). *United States Department of Defense, Tech. Rep.*
6. Mercado, D., Colunga, G. R. F., Castillo, P., Escareño, J. A., & Lozano, R. (2013, May). Gps/ins/optic flow data fusion for position and velocity estimation. In *Proceeding of the Conference of Unmanned Aircraft Systems on IEEE* (pp. 486–491).

7. Yang, S., Scherer, S. A., Schauwecker, K., & Zell, A. (2013, May). Onboard monocular vision for landing of an MAV on a landing site specified by a single reference image. In *Proceeding of the Conference on Unmanned Aircraft Systems on IEEE* (pp. 318–325).
8. Wu, Y., Lim, J., & Yang, M. H. (2013). Online object tracking: A benchmark. In *Proceeding of the Conference on CVPR* (pp. 2411–2418).
9. Taylor, J. W. R., & Munson, K. (1977). Jane's pocket book of remotely piloted vehicles: Robot aircraft today. *Collier Books*.
10. Huang, M., Zhang, B., & Ding, Y. L. (2018). Development of airborne photoelectric platform at abroad. *Aeronautical Manufacturing Technology*, 9, 70–71.
11. <http://www.borderlandnews.com/stories/borderland/20040319-95173.shtml>.
12. http://iris.usc.edu/people/medioni/current_research.html.
13. Reilly, V., Idrees, H., & Shah, M. (2010, September). Detection and tracking of large number of targets in wide area surveillance. In *Proceeding of the Conference on ECCV* (pp. 186–199).
14. Prokaj, J., Zhao, X., & Medioni, G. (2012, June). Tracking many vehicles in wide area aerial surveillance. In *Proceeding of the Conference on CVPR Workshops* (pp. 37–43).
15. Prokaj, J., & Medioni, G. (2011, June). Using 3d scene structure to improve tracking. In *Proceeding of the Conference on CVPR* (pp. 1337–1344).
16. Prokaj, J., Duchaineau, M., & Medioni, G. (2011, June). Inferring tracklets for multi-object tracking. In *Proceeding of the Conference on Computer Vision and Pattern Recognition Workshops* (pp. 37–44).
17. <http://theairlab.org/>.
18. <https://uavvision.com/about/>.
19. <https://www.youtube.com/watch?v=pgYDoU8BiiE>.
20. <https://www.dji.com/cn>.
21. <http://thedronegirl.com/2017/02/26/dji-yuneeec-autel-mota/>.
22. <http://www.automatedvehiclessymposium.org/avs2018/proceedings>.
23. <http://www.uasconferences.com/>.
24. Yuan, C., Medioni, G., Kang, J., & Coehn, I. (2011, December). Detection and tracking of moving objects from a moving platform in presence of strong parallax. *U.S. Patent 8,073,196*.
25. Barnea, D. I., & Silverman, H. F. (1972). A class of algorithms for fast digital image registration. *IEEE Transaction on Computers*, 100(2), 179–186.
26. Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91–110.
27. Fan, X., Rhody, H., & Saber, E. (2005, December). Automatic registration of multisensor airborne imagery. In *Proceedings of the 34th Conference on Applied Imagery and Pattern Recognition Workshop* (p. 6).
28. Bouchahma, M., Barhoumi, W., Yan, W., & Al Wardi, H. (2017, October). Optical-flow-based approach for the detection of shoreline changes using remote sensing data. In *Proceeding of the 14th Conference on Computer Systems and Applications* (pp. 184–189).
29. Srivastav, N., Agrwal, S. L., Gupta, S. K., Srivastava, S. R., Chacko, B., & Sharma, H. (2017, January). Hybrid object detection using improved three frame differencing and background subtraction. In *Proceedings of the 7th Conference on Cloud Computing, Data Science and Engineering-Confluence* (pp. 613–617).
30. Ahmed, A. H., Kpalma, K., & Guedi, A. O. (2017, December). Human detection using HOG-SVM, mixture of Gaussian and background contours subtraction. In *Proceedings of 13th Conference on Signal-Image Technology and Internet-Based Systems* (pp. 334–338).
31. Chao, H., Gu, Y., & Napolitano, M. (2013, May). A survey of optical flow techniques for uav navigation applications. In *Proceeding of the Conference on Unmanned Aircraft Systems* (pp. 710–716).
32. Cohen, I., & Medioni, G. (1998, November). Detecting and tracking moving objects in video from an airborne observer. In *Proceeding of the Conference on IEEE Image Understanding Workshop* (Vol. 1, pp. 217–222).
33. Cheng, Y. (1995). Mean shift, mode seeking and clustering. *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 17(8), 790–799.

34. Kalal, Z., Matas, J., & Mikolajczyk, K. (2010, June). PN learning: Bootstrapping binary classifiers by structural constraints. In *Proceeding of the Conference on CVPR* (pp. 49–56).
35. Yilmaz, A., Javed, O., & Shah, M. (2006). Object tracking: A survey. *ACM Computing Surveys (CSUR)*, 38(4), 13.
36. Keck, M., Galup, L., Stauffer, C. (2013, January). Real-time tracking of low-resolution vehicles for wide-area persistent surveillance. In *Proceeding of the Conference on Applications of Computer Vision* (pp. 441–448).
37. Gong, S., & Xiang, T. (2003, October). Recognition of group activities using dynamic probabilistic networks. In *Proceeding of the Conference on Computer Vision* (pp. 742–749).
38. Pavlovic, V., Frey, B. J., & Huang, T. S. (1999). Time-series classification using mixed-state dynamic Bayesian networks. In *Proceeding of the Conference on CVPR* (Vol. 2, pp. 609–615).
39. Medioni, G., & Nevatia, R. (1997). Surveillance and monitoring using video images from a UAV. In *Proceeding of the Conference on IUW*.
40. Cohen, I., & Medioni, G. (1998). Detection and tracking of objects in airborne video imagery. In *Proceeding of the Conference on CVPR Workshop on Interpretation of Visual Motion*.
41. Chen, B. J., & Medioni, G. (2017, March). Exploring local context for multi-target tracking in wide area aerial surveillance. In *Proceeding of the Conference on Applications of Computer Vision* (pp. 787–796).
42. Ali, S., & Shah, M. (2006, May). COCOA: Tracking in aerial imagery. In *Airborne Intelligence, Surveillance, Reconnaissance (ISR) Systems and Applications III* (Vol. 6209, p. 62090D).
43. Ali, S., Reilly, V., & Shah, M. (2007, June). Motion and appearance contexts for tracking and re-acquiring targets in aerial videos. In *Proceeding of the Conference on CVPR* (pp. 1–6).
44. Perera, A. A., Srinivas, C., Hoogs, A., Brooksby, G., & Hu, W. (2006, June). Multi-object tracking through simultaneous long occlusions and split-merge conditions. In *Proceeding of the Conference on CVPR* (Vol. 1, pp. 666–673).
45. Xiao, J., Cheng, H., Sawhney, H., & Han, F. (2010). Vehicle detection and tracking in wide field-of-view aerial video. In *Proceeding of the Conference on CVPR* (pp. 679–684).
46. Yu, Q., & Medioni, G. (2009, June). Motion pattern interpretation and detection for tracking moving vehicles in airborne video. In *Proceeding of the Conference on CVPR* (pp. 2671–2678).
47. Fang, P., Lu, J., Tian, Y., & Miao, Z. (2011). An improved object tracking method in UAV videos. *Procedia Engineering*, 15, 634–638.
48. Collins, R., Zhou, X., & Teh, S. K. (2005). An open source tracking testbed and evaluation web site. *IEEE International Workshop on Performance Evaluation of Tracking and Surveillance*, 2, 35.
49. Mueller, M., Smith, N., & Ghanem, B. (2016, October). A benchmark and simulator for uav tracking. In *Proceeding of the Conference on ECCV* (pp. 445–461).
50. <https://www.sdms.afrl.af.mil>.
51. Jha, M. N., Levy, J., & Gao, Y. (2008). Advances in remote sensing for oil spill disaster management: state-of-the-art sensors technology for oil spill surveillance. *Sensors*, 8(1), 236–255.
52. http://i21www.ira.uka.de/image_sequences/.
53. <https://www.sdms.afrl.af.mil/index.php?collection=clif2006>.
54. http://crev.ucf.edu/data/UCF_Aerial_Action.php.
55. Shu, T., Xie, D., Rothrock, B., Todorovic, S., & Chun Zhu, S. (2015). Joint inference of groups, events and human roles in aerial videos. In *Proceedings of the Conference on CVPR* (pp. 4576–4584).
56. Wang, F. Y., Zhang, J. J., Zheng, X., Wang, X., Yuan, Y., Dai, X., ... Yang, L. (2016). Where does AlphaGo go: From church-turing thesis to AlphaGo thesis and beyond. *IEEE/CAA Journal of Automatica Sinica*, 3(2), 113–120.