A Novel Real-Time Moving Target Tracking and Path Planning System for a Quadrotor UAV in Unknown Unstructured Outdoor Scenes

Yisha Liu [ID], Qunxiang Wang, Huosheng Hu, *Senior Member, IEEE*, and Yuqing He, *Member, IEEE*

Abstract—A quadrotor unmanned aerial vehicle (UAV) should have the ability to perform real-time target tracking and path planning simultaneously even when the target enters unstructured scenes, such as groves or forests. To accomplish this task, a novel system framework is designed and proposed to accomplish simultaneous moving target tracking and path planning by a quadrotor UAV with an onboard embedded computer, vision sensors, and a two-dimensional laser scanner. A support vector machine-based target screening algorithm is deployed to select the correct target from multiple candidates detected by single shot multibox detector. Furthermore, a new tracker named TLD-KCF is presented in this paper, in which a conditional scale adaptive algorithm is adopted to improve the tracking performance for a quadrotor UAV in cluttered outdoor environments. According to distance and position estimation for a moving target, our quadrotor UAV can acquire a control point to guide its fight. To reduce the computational burden, a fast path planning algorithm is proposed based on elliptical tangent model. A series of experiments are conducted on our quadrotor UAV platform DJI M100. Experimental video and comparison results among four kinds of target tracking algorithms are given to show the validity and practicality of the proposed approach.

Index Terms—Path planning, quadrotor unmanned aerial vehicle (UAV), real-time target tracking, unstructured outdoor scenes.

I. INTRODUCTION

TRACKING and path planning are essential tasks for intelligent robot systems working in complex indoor/outdoor environments, from biped walking robots [1], wheeled robots [2] to aerial robots [3], [4]. In the last decades, fast-growing unmanned aerial vehicles (UAVs) have been utilized in wide range of military and nonmilitary tasks to perform searching, patrolling, target tracking, and surveillance in complex outdoor environments [5]–[8]. There have been a variety of studies on UAVs or other mobile robots to carry out these missions, but most of these applications focus on moving target tracking in general outdoor environments without cluttered obstacles. However, when a human target enters in unstructured outdoor scenes, such as groves or forests, UAVs should perform real-time target tracking, obstacle avoiding, and path planning simultaneously. This in turn imposes great challenges for small UAVs with limited computer power.

In recent years, a new method for automatic detection of cars in UAV images acquired over urban contexts was presented in [7], in which only car detection was investigated. Chen *et al.* [8] studied the problem of quadrotor tracking a moving target in cluttered indoor environments. They put Apriltag on the target (a mobile robot) to make detection and tracking easy-to-implement, which is however impossible for real-world applications. In [9], a vision-based quadrotor platform was built and tested flying through an unknown indoor scene with high accuracy. In our previous work [10], a novel object detection system using three-dimensional (3-D) laser scanning data was proposed to deal with cluttered indoor scenes, but 3-D laser scanner is too heavy for our UAV platform. In [11], a small UAV equipped with a gimbaled camera accomplished the task of tracking an unpredictable moving ground vehicle that was running on structured roads without obstruction from trees. This greatly reduced the difficulty of tracking and path planning.

Giusti *et al.* [12] introduced a real-world flying demonstration in which a quadrotor UAV was autonomously working in forest scenes. The problem of perceiving forest or mountain trails from a single monocular image acquired was investigated. A deep neural network for visually perceiving the direction of a forest trail from a single image was trained to guarantee that a quadrotor can perform forest trial tracking robustly. Compared with the work in [8], [9], and [11], the work in [12] is a more challenging task since the forest scenes are much more cluttered. However, apart from trail tracking, autonomous obstacle avoidance was not investigated in [12] since the forest trail is wide enough for a quadrotor to navigate around.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

2   IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS
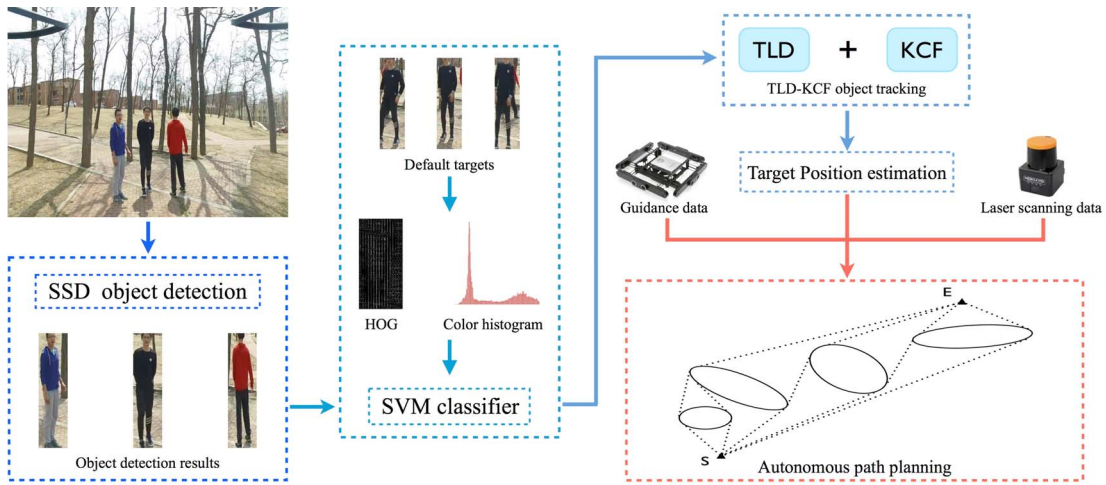

Fig. 1.   System framework for quadrotor to conduct moving target detection, tracking, and path planning simultaneously in unknown cluttered outdoor scenes.

In this paper, a new system framework is proposed to perform moving target tracking and path planning in groves or forests simultaneously. Since 3-D laser scanner is too heavy for our UAV platform, we have to use light vision sensors and two-dimensional (2-D) laser scanner in our research here. It should be noted that the human target in the groves or forests is not walking along the trails but moving randomly in our experiments. Our quadrotor UAV platform should not only avoid the trees in real time but also perform robust tracking when the target may be blocked by the trees frequently. Considering the diversity of targets and their mobility, single shot multibox detector (SSD) algorithm is adopted to provide multiple candidate targets, and then a support vector machine (SVM)-based target screening is used to find the correct one.

According to our comparison tests between four tracking algorithms tracking learning detection (TLD), kernelized correlation filter (KCF), TLD-KCF, and generic object tracking using regression networks (GOTURNs), the newly proposed TLD-KCF tracker has a superior tracking performance and the reduced computation costs suitable to limited onboard computing power. Moreover, a low-cost and novel path planning algorithm is proposed based on an elliptical tangent model and environmental constraints generated from multisensor data. This novel system framework has been successfully tested in a series of flight experiments, thereby demonstrating its validity and practicality in a real-world implementation. An experimental video can be viewed at the website.[1]

The rest of this paper is organized as follows. Section II briefly introduces our proposed system framework and the small UAV platform. Section III presents visual target detection and tracking with a small UAV in cluttered outdoor environments. In Section IV, a novel path planning algorithm is proposed based on elliptical tangent model, including tracking constraints generation and a path tracker design approach. Experiments are conducted by using a DJI M100 quadrotor, and results are presented in Section V to show the feasibility

---

[1]http://v.youku.com/v_show/id_XMjc5NDY1NjU2NA==.html?spm=
a2hzp.8244740.0.0

Fig. 2.   Quadrotor UAV is simultaneously exploring and tracking in a grove, a typical cluttered unknown scene in this paper.

and effectiveness of the proposed approach. Finally, a brief conclusion and future work are presented in Section VI.

II. SYSTEM FRAMEWORK

Moving target detection and tracking are common tasks for UAVs working in structured environments, and a variety of practical cases have been introduced in real-world outdoor applications. However, most of these cases are only tested in open outdoor space, such as square, field, road, and trail. In this paper, we investigate how a quadrotor UAV is able to simultaneously track a moving target and plan a reliable path in cluttered and unstructured outdoor scenes, such as groves or forests. Fig. 1 shows a system framework that is proposed to carry out simultaneous moving target detection, tracking, and path planning by a quadrotor UAV in unknown grove scenes.

Fig. 2 shows that a small quadrotor UAV is accomplishing autonomous exploring and human target tracking tasks in unknown cluttered outdoor environments simultaneously. Monocular color images are obtained by the onboard camera at a frequency of 25 Hz, and SSD algorithm is utilized to detect the candidate human targets in input images. In order to select the correct target from these candidates, histogram of oriented gradients (HOG) and color histogram features are extracted from the subimage of each candidate target, and then an SVM-based classifier is adopted to determine the target.

After tracking object initialization, a TLD-KCF approach is presented to perform moving object tracking. Moreover, a target position estimation algorithm is used to roughly estimate the relative position relationship between the selected target in the image and the camera. In our proposed framework, the core part is how to achieve both visual tracking and safe flight path planning in an unstructured environment cluttered with obstacles. Richer local environmental information is perceived by a laser range finder and DJI Guidance so that the reliable constraints for the quadrotor's flight can be generated. Based on the flight constraints and target position estimation results, a novel path planning algorithm using elliptical tangent model is proposed and used in the quadrotor UAV path planner and tracker design.

It should be noticed that we aim to provide a practical UAV visual tracking system which is easy-to-carry and can work in unknown and unstructured scenes autonomously. Since DJI Matrice 100 platform includes a commercial flight controller, our research is focused on how to implement and seamlessly integrate the UAV's various tasks, e.g., visual tracking, obstacle avoidance, and path planning, in complex outdoor scenes.

III. AUTONOMOUS MOVING TARGET DETECTION AND TRACKING FOR QUADROTOR UAV WITH MONOCULAR VISION

A. Target Detection and Initialization

SSD is a fast single-shot object detector for multiple categories based on convolutional neural network [13]. This algorithm was proposed by Liu *et al.* [13] in 2016. It is an improved version of YOLO [14] and can ensure both the speed and accuracy of the object detection compared with faster R-CNN [15] and YOLO. A fixed-size collection of bounding boxes is produced when images are input into SSN network, and scores for the presence of object class instances in those boxes are also given. After that, a nonmaximum suppression algorithm is used to produce the final detection results [13].

Although SSD can provide the object position and its category in the image, it cannot tell the difference between the objects belonging to the same category. Therefore, if we want to track a specific human target, an initialization approach for this target has to be applied so that it can select the correct target from multiple candidates in each image. In this paper, we adopted the SVM [16] as the classifier with HOG [17] and color histogram features to perform target screening. The classifier is trained offline with images belonging to the target.

During the quadrotor UAV's initial target searching stage, SSD-based target detection algorithm will show multiple candidate human targets in series of subimages. In order to tell which candidate is the correct one, the features extracted from subimages will be input into SVM classifier. If a correct target is selected from these candidates, its position and size in the image will be stored and used to accomplish the initialization of target tracking algorithm. Otherwise, the quadrotor will continue to search until the target is found. Similarly, when the target is lost during the target tracking process, the same strategy is used to refind the target.
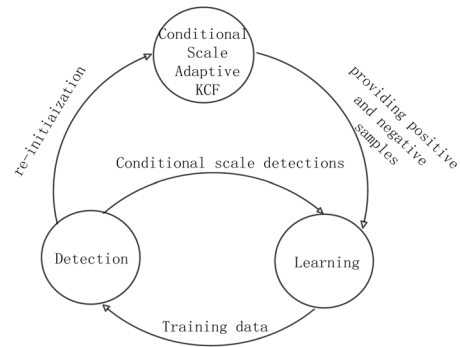

Fig. 3. Block diagram of the TLD-KCF algorithm framework.

B. Target Tracking by TLD-KCF

TLD is a robust framework for target tracking, which was proposed by Kalal *et al.* [18] in 2012 to perform long-term tracking of unknown objects in a video stream. It has three components, namely tracking, learning, and detection, in the TLD framework. Its tracking component is based on median-flow tracker [19], but there may be tracking failure cases especially in complex outdoor environments.

KCF was proposed by Henriques *et al.* [20] in 2014, which is a high-speed tracker and running at hundreds of frames-per-second. KCF can reduce both storage and computational burden significantly by using circulant matrices. In our experiments, we find that KCF may have a poor performance while tracking a high-speed moving target or a target in low-frame-rate video. This means that the target displacement between adjacent frames cannot be too large; otherwise the tracker will fail and cannot be recovered once the target is lost.

In this paper, we investigate the problem of autonomous visual tracking of a moving target in cluttered outdoor scenes with a quadrotor UAV. Since KCF can be implemented in a few lines of code, it is very suitable for our quadrotor platform with limited computing resources. However, trees and other obstacles in the cluttered testing environments will cause the occurrences of target shielding and KCF will have poor tracking performance in these cases. Moreover, if the target in the field of view is too close or too far from the camera, it will result in significant changes of the target's size in the image. Then the bounding box will easily drift and eventually lead to target tracking failure. In order to improve the tracking performance, a new tracker named TLD-KCF is proposed which can perform fast and robust target tracking due to the conditional scale adaptive algorithm. The block diagram of the TLD-KCF algorithm framework is shown in Fig. 3.

In Fig. 3, the conditional scale adaptive KCF component is a KCF-based tracker that estimates the motion of the target between consecutive frames, and the tracker can adaptively vary with changes of the target's scale. The Detection component is a detector that scans images at a regular interval, which can relocalize the position of the target and reinitialize the target of the tracker. The learning component supervises the performance of tracker and detector, estimates the errors of tracker, and provides training samples for the detector to

reduce tracking errors and improve tracking accuracy in the future.

As introduced in [20], the circulant matrix in KCF can be expressed as follows:

$$X = C(x) = F \bullet \text{diag}\ (\widehat{x}) \bullet F^H \tag{1}$$

where F is a constant matrix that does not depend on x, and $\widehat{x}$ denotes the discrete Fourier transform (DFT) of the generating vector. For sample set (x_i, y_i), we need to find a linear regression equation $f(z) = \omega^T z$ to minimize the squared error over samples x_i and their regression targets y_i

$$\arg \min_{\omega} \sum (f(x_i) - y_i)^2 + \lambda \|\omega\|^2 \tag{2}$$

where $\omega = (X^T X + \lambda I)^{-1} X^T y$, X has one sample per row x_i, and each element of y is a regression target y_i.

For frequently used kernel functions, we can solve ω by solving the dual space coefficient α. α can be solved in Fourier domain as

$$\widehat{\alpha} = \frac{\widehat{y}}{\widehat{k}^{xx} + \lambda} \tag{3}$$

where k^{xx} is the first row of the kernel matrix $K = C(k^{xx})$ and a hat ˆ denotes the DFT of a vector. More generally,

$$k^{xx'} = \exp\left(-\frac{1}{\sigma^2}\left(\|x\|^2) + \|x'\|^2\right) - 2F^{-1}\left(\widehat{x}^* \odot \widehat{x}'\right)\right) \tag{4}$$

where $\odot$ is the element-wise product, and the kernel correlation of two arbitrary vectors, x and x', is the vector $k^{xx'}$ with elements

$$k_i^{xx'} = k\left(x', P^{i-1}x\right). \tag{5}$$

Then in the next frame, we can calculate the response in Fourier domain by

$$\widehat{f}(z) = \left(\widehat{k}^{xz}\right)^* \odot \widehat{\alpha} \tag{6}$$

where $\hat{x}$ can be learned in the model. According to the response peak, the position of the target can be estimated precisely.

In real-world applications, there exists a problem of the target's size change in the image under some conditions. For example, the acceleration and distance between the target and the quadrotor is bigger than a given threshold. In order to solve this problem, a conditional scale adaptive algorithm is proposed in this paper to improve the performance of KCF. We employ the bilinear interpolation to enlarge the image space to enhance the robustness and the tracking accuracy. Here, we define a fixed image template size as $S_0 = (s_x, s_y)$ and a scale adjustment vector $K_s = \{k_1, k_2 \ldots k_n\}$, which are named scale pooling in this paper. Then the response peak can be calculated as follows:

$$\arg \max F^{-1}\widehat{f}\left(z^{k_i}\right) \tag{7}$$

where z^{k_i} is the scale sample patch, which is resized to S_0. On the basis of the response peak, the bounding box can be adjusted and the target position can also be confirmed.
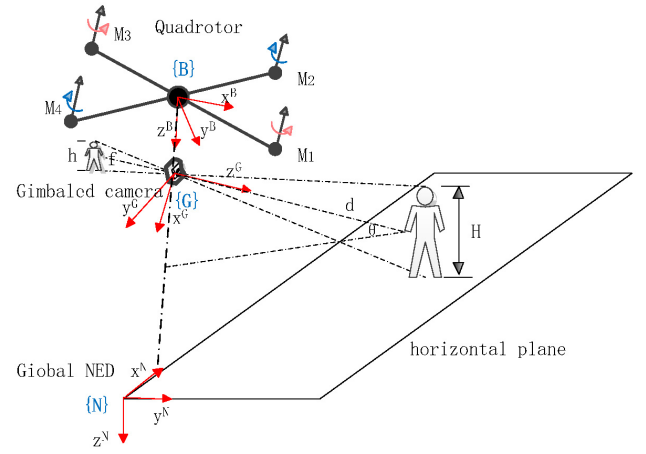

Fig. 4. Three coordinate systems and object imaging relationship.

C. Distance and Position Estimation for Moving Target

The relative distance between the target and the quadrotor can be estimated approximately. As shown in Fig. 4, there are three coordinate systems which are North-East-Down frame $\{N\}$, UAV body frame $\{B\}$, and the gimbaled camera frame $\{G\}$. The monocular camera is fixed in the center bottom of the quadrotor and facing forward in order to keep the tracking target in the center of the camera view.

In Fig. 4, f is the focal length of the camera and c is the optical center of the lens. The light emitted by the object passes through the camera's optical center and is then imaged on the image plane. Suppose that the distance between a target with a height of H and the optical center of the lens is d, and the length of the object in the sensor is h. Then there is a proportional relationship among these parameters: $f/d = h/H$. If H can be given or obtained in advance, the distance between the target and the camera can be estimated with $d = Hf/h$. Otherwise, the UAV can use the data in the target initialization stage and the data in current observation to estimate the distance d as follows:

$$d = \frac{\text{len}_0}{\text{len}}d_0 \tag{8}$$

where d_0 is the laser range finder data between the target and the camera and len_0 is the size of the bounding box of the target which are both obtained during the target initialization stage, and len is the size of the bounding box of the target detected currently. The units of len_0 and len are pixels.

Suppose that the position of the bounding box of a target in the image is shown in Fig. 5, in which Δw and Δh are offset values of the center of the bounding box to the image center in the horizontal and vertical directions. The relative position between the camera and the target can be estimated with the following equations:

$$\alpha = \frac{\Delta w}{w}\theta_w, \quad \beta = \frac{\Delta h}{h}\theta_h$$
$$\Delta x = d\ \tan\alpha, \quad \Delta y = d, \quad \Delta z = d\ \tan\beta \tag{9}$$

where θ_w and θ_h represent the horizontal viewing angle and vertical viewing angle of the camera, α and β represent the relative angle between the target and the camera in horizontal
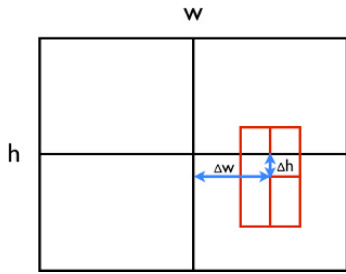
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LIU *et al.*: NOVEL REAL-TIME MOVING TARGET TRACKING AND PATH PLANNING SYSTEM FOR QUADROTOR UAV

5

Fig. 5.   Position of a target's bounding box in the image.

Fig. 6.   DJI Guidance system.

Fig. 7.   Laser scanning data used for obstacle distance estimation.

Fig. 8.   In the cost map, the size of obstacle should be expanded accordingly.

and vertical directions, Δx, Δy, and Δz represent the relative distance between the target and the camera in three axes of x, y, and z. The right-handed coordinate system is used in this paper.

Based on the results of distance and position estimation for a moving target, a control point can be given to guide the autonomous fight of a UAV. In addition, the estimation values provided in (9) can also be used to control the rotation angle of the 3-axis gimbal of Zenmuse X3 camera, which can make the target in the center of the image as much as possible.

IV. PATH PLANNING FOR UAV IN CLUTTERED SCENES

A. Local Environment Perception and Constraints Generation

In the process of target tracking, a UAV has to fly at a fixed height and perform obstacle avoidance in cluttered outdoor scenes simultaneously. In this paper, our Matrice 100 platform uses ranging data acquired by DJI Guidance system and HOKUYO laser scanner to estimate both its current flying height and distances to surrounding obstacles.

DJI Guidance system includes a central processor with vision algorithms on the chip and five sensor modules, each of which is integrated with a visual camera and an ultrasonic sensor (see Fig. 6). By reading the Guidance data, UAV can estimate the distances of the obstacles in five directions (front, rear, left, right, and bottom) in real time, and the data of downward distance is used as the flight height of the UAV.

Since the ranging data acquired from DJI Guidance system is not accurate enough, the obstacle distances in the front, left, and right directions are estimated by combining ranging data from Guidance and HOKUYO laser. HOKUYO UTM-30LX is a 2-D laser scanner and the measuring distance ranges from 0.06 m to 10.0 m. Its frequency is 40 Hz, the detection angle is 270°, and the angular resolution is 0.25°.

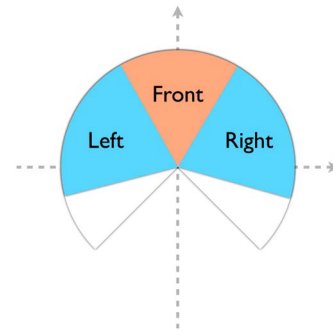Matrice 100 is about 1 m long, 1 m wide, and 0.2 m high. In order to prevent the laser data from being affected by the
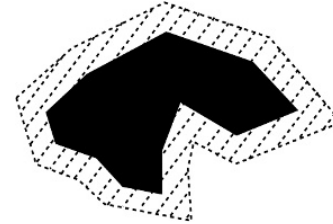
moving part of the quadrotor, the ranging data less than 0.71 m will be eliminated. So the possible obstacle distances obtained by the UAV's laser scanner are from 0.71 m to 10 m. Based on these laser data, a local map is built with a fixed-size grid division algorithm (scale is 10 cm). As shown in Fig. 7, 60° laser data in the front direction are used to estimate the front obstacles, and 75° laser data on both sides are used to estimate the obstacles on the left and right.

B. Path Planning Algorithm Based on Elliptical Tangent Model

In the field of mobile robot path planning, configuration space is a very popular approach. In this approach, the geometric center of a mobile robot is used to replace the whole robot to perform path planning. Considering the size of the robot itself, the size of obstacles in the map should be expanded accordingly. Fig. 8 shows such an example, where the black part represents the actual size of the obstacle, while the region labeled with dotted lines represents the expansion area. The UAV's path planning task is completed based on the cost map.

1) Basic Algorithm I (Elliptical Fitting for Obstacle Region): In the cost map, a minimum external ellipse is generated for the obstacle region. Considering the efficiency and precision, the algorithm of least squares fitting of ellipses [21] is adopted to perform elliptical fitting of boundary points near the obstacle's contour edge. Sometimes there are still a few points outside the ellipse [see Fig. 9(a)]. In order to enable the ellipse to enclose all corresponding obstacle region [see Fig. 9(b)], the following algorithm is designed.

1) Judge whether the ellipse contains all boundary points of the current obstacle region. If all boundary points are included, the algorithm ends; otherwise, go to step 2).
2) Increase a unit length of the long axis of the ellipse, and then judge whether the current ellipse contains all

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

6

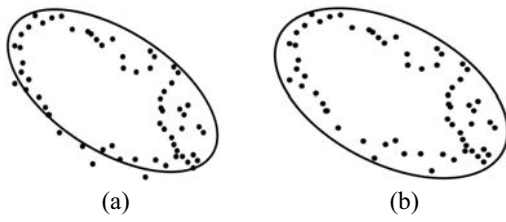IEEE TRANSACTIONS ON SYSTEMS, MAN, AND CYBERNETICS: SYSTEMS

Fig. 9. (a) Example of a few points outside the ellipse. (b) Example of the ellipse enclosing all points belonging to an obstacle.

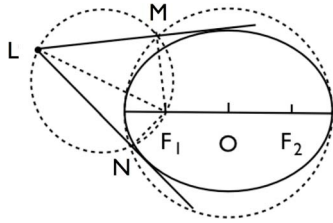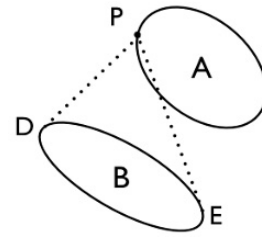Fig. 10. Line LM and LN are two tangents of the ellipse O.

Fig. 11. Line PD and PE are the elliptical tangents, but these two tangents pass through the inside region of ellipses A.

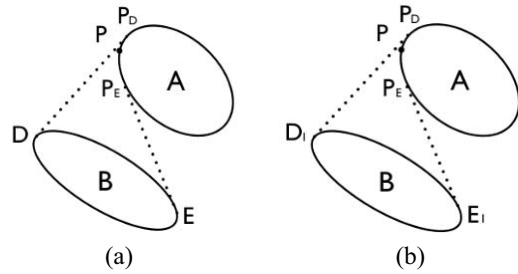

Fig. 12. (a) Generate tangent DP_D and EP_E from points D and E to ellipse A. (b) Generate tangent $P_D D_1$ and $P_E E_1$ from points P_D and P_E to ellipse B.

boundary points of the obstacle region. If all boundary points are included, the algorithm ends; otherwise, increase a unit length of the short axis of the ellipse and go to step 1).

2) Basic Algorithm II (Elliptical Tangent Generation From Outside Point): To reduce the computational burden, we propose an algorithm using auxiliary circle to calculate the tangent of the ellipse. Suppose that there is an ellipse with the focus points F_1 and F_2, and L is a point outside the ellipse. Two tangents from point L to ellipse O can be obtained by the following steps.

1) Generate an auxiliary circle for the ellipse O; the circle center is O and the diameter is the long axis of the ellipse.
2) Generate a circle with diameter of LF_1, which intersects with the auxiliary circle of the ellipse O at points M and N.
3) Line LM and LN are the elliptical tangents (see Fig. 10) and they are the candidate paths in the UAV's path planner.

3) Basic Algorithm III (Approximate Common Tangent Generation): When a path is planned between two adjacent ellipses, it is not feasible to generate elliptical tangents from a point of an ellipse and use them as the candidate path. As shown in Fig. 11, A and B are two adjacent ellipses and P is a point of A. Line PD and PE are two elliptical tangents, but these tangents pass through the inside region of ellipses A, which means that the UAV may collide with the obstacles in region A and it is absolutely unacceptable.

To meet the requirement of the UAV's real-time path planning, we use an approximation algorithm to obtain the approximate common tangents between two ellipses as follows.

1) Generate tangent DP_D from point D to ellipse A (two tangents are generated and the one near the point P is selected). Similarly, tangent EP_E will be generated from point E to ellipse A.

2) Generate tangent $P_D D_1$ from point P_D to ellipse B (select the one near point *D*). Similarly, tangent $P_E E_1$ will be generated from point P_E to ellipse B.

Fig. 12(a) and (b) shows two examples to illustrate step 1) and 2), respectively. According to the steps introduced above, approximate common tangents can be generated which can be used as safe paths connecting two adjacent ellipses.

4) Main Algorithm (Path Planning Based on Elliptical Tangent Model): Set S and E as the start point and endpoint in the UAV's path planning. The detailed steps of path planning are as follows.

1) Backtrack from E and get the line segment ES. If this segment does not collide with any obstacle, it will be the optimal path and the path planning task ends; otherwise, go to step 2).
2) When backtracking from E and the line segment ES collides with an obstacle, perform elliptical fitting for this obstacle (using Basic Algorithm I) and obtain minimum external ellipse O_0 [see Fig. 13(a)].
3) Generate tangents ES_1, ES_2 from point E to ellipse O_0 and SE_1, SE_2 from point S to ellipse O_0, respectively (using Basic Algorithm II) [see Fig. 13(b)].
4) There are four new subpaths generated, which are S_1E, S_2E, SE_1, and SE_2. For each subpath, go back to step 1) and perform path planning recursively. In the process of path planning, Basic Algorithm III will be used to generate the path connecting two ellipses. The recursive process ends until all subpaths have no collisions with any obstacle [see Fig. 13(c)].
5) Store all possible paths between points S and E. For any two points of each path, if there is no obstacle between two points, connect them. Then according to the rule that the line segment is the shortest between two points, optimize each path [see Fig. 13(d)].
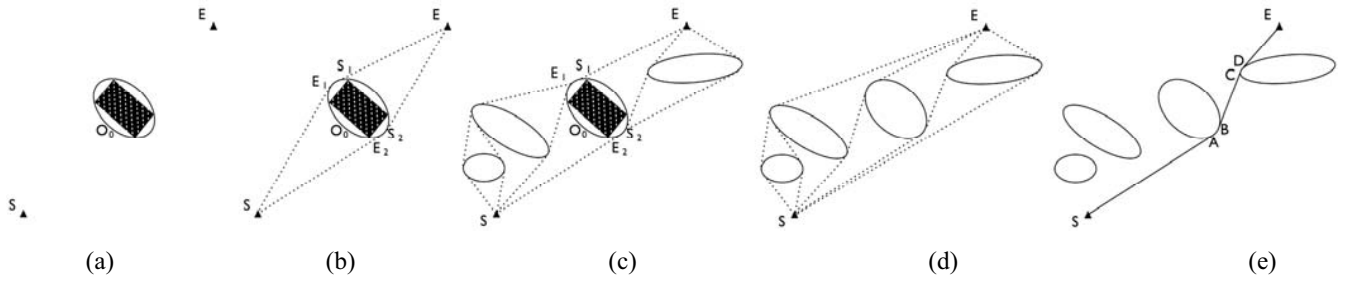
This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LIU *et al.*: NOVEL REAL-TIME MOVING TARGET TRACKING AND PATH PLANNING SYSTEM FOR QUADROTOR UAV 7

Fig. 13. Some detailed steps of path planning algorithm (a) step 2), (b) step 3), (c) step 4), (d) step 5), and (e) step 7).

6) Score every path from S to E with the following equation which includes distance cost and rotating angle cost:

$$P(S, E) = D(S, E) + Y(S, E) \tag{10}$$

where $P(S, E)$ is the total cost, $D(S, E)$ is distance cost in this path, and $Y(S, E)$ is rotating angle cost in this path.

7) Calculate the cost values of all possible paths and select the one with least cost as the output of UAV's path planner. As an example shown in Fig. 13(e), the path SABCDE is the final path planning result for this case, where SA, BC, and DE are line segments while AB and CD are the arcs of the corresponding ellipses.

Until now, the algorithm of UAV's path planning based on elliptical tangent model is completed. This algorithm can select an optimal path with a least cost and also guarantee the smoothness of local paths, which makes it easy for our small UAV to implement autonomous flight in unknown cluttered outdoor scenes.

V. EXPERIMENTAL RESULTS AND ANALYSIS

A. Platform

A quadrotor UAV platform, DJI Matrice 100 (see Fig. 14), is used in our experiments, which is equipped with a monocular vision sensor DJI ZENMUSE X3 [see Fig. 15 (a)], an embedded computer Manifold, a visual sensing system DJI Guidance and GPS. Moreover, a 2-D laser scanner HOKUYO UTM-30LX [see Fig. 15 (b)] is installed in the UAV platform by us in order to improve its real-time obstacle detection performance.

In our experiments, the onboard embedded computer named DJI Manifold is installed on our quadrotor UAV platform and the DJI SDK provided by DJI is used to support our programming. The onboard HOKUYO laser, Zenmuse X3, Guidance, GPS, and flight controller are connected to Manifold. In this paper, robot operating system is running in Manifold.

B. Experimental Results

The test environments are the groves on the campus of Dalian University of Technology, which are typical complex outdoor scenes and have various unstructured environment characteristics, such as cluttered trees, pedestrians, grasslands, and brick roads. In a series of experiments, different people wearing different clothes are selected as tracking targets who walk through the woods quickly. Many pedestrians are also walking in the woods randomly during our
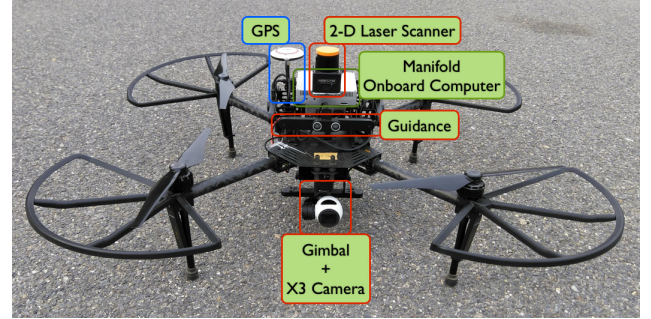

Fig. 14. Matrice 100 UAV platform used in our experiments.

Fig. 15. (a) Monocular vision sensor. (b) 2-D laser scanner.

experiments and they bring unpredictable interference to our experiments.

Fig. 16 shows three groups of experimental results of moving target tracking. The tracking results for a human target in the UAV's vision images are shown in Fig. 17. A group of experimental results of the UAV's autonomous flight through the cluttered woods are shown in Fig. 18. It should be noted that all these experiments are implemented on our quadrotor UAV's onboard computer and perform in real time. Although the computational resources of the embedded computer are limited, our quadrotor UAV can accomplish real-time moving target tracking and path planning in cluttered outdoor environment simultaneously.

More experimental videos can be viewed at the website.[2] As shown in the video, the human targets are wearing different clothes in different experiments, and there will be other interference targets in the tracking process (some pedestrians were walking through the experimental sites randomly). All these can prove the validity of our approach.

[2]http://v.youku.com/v_show/id_XMjc5NDY1NjU2NA==.html?spm=a2hzp.8244740.0.0

TABLE I
MEAN ERROR AND PERCENTAGE OF ERROR OF TARGET DISTANCE ESTIMATION AT DIFFERENT DISTANCE BETWEEN TARGET AND CAMERA

Distance between target and camera (m)	3	4	5	6	7	8	9	10	11	12	13	14
Mean error (cm)	8.6	16.5	17.8	23.8	32.3	34.8	38.2	45.9	55.3	68.6	85.7	93.3
Percentage of error	2.9%	4.1%	3.6%	2.3%	4.6%	4.4%	4.2%	4.6%	5.0%	5.7%	6.6%	6.7%

Fig. 16.   Three groups of experiment results: a quadrotor is tracking three different human targets walking through the woods in first, second, and third row.

Since the human target does not move at a fixed speed, it is important for the UAV to estimate the relative distance between the moving target and itself accurately. A testing experiment is designed and carried out to verify the accuracy of the relative distance estimation algorithm in this paper. Let the UAV work in hovering flight mode at a fixed height of 1.7 m. A fixed-height target stops at different locations ranging from 3 m to 14 m with an interval of 1 m. Based on the algorithm proposed in Section III-C, ten groups of distance estimation values are obtained at each location, which means total 120 groups of data are obtained at 12 locations. The mean error and percentage of error of target distance estimation at different distance between target and camera are given in Table I. Taking into account the system errors existing in the experiment, the distance estimation algorithm proposed in this paper is a valid one and can be used in real-world applications.

C. Comparison of Four Target Tracking Algorithms

In this section, four kinds of target tracking algorithms, i.e., TLD, KCF, TLD-KCF, and GOTURN, are compared in time-cost and lost track ratio. TLD, KCF, and TLD-KCF (our tracking algorithm) have been introduced in Section III-B.

GOTURN was proposed by Held *et al.* [22] in 2016, and this tracker uses a regression-based approach and is trained offline to learn a generic relationship between appearance and motion. Compared with previous trackers using networks, GOTURN is a much faster tracker and can be used in real-time applications.

Considering the limited computational resources provided by our UAV's onboard computer, time-cost is a crucial evaluation criterion. Table II presents the comparison result of time-costs for the four tracking algorithms running on our UAV's onboard computer. It can be seen that the average time-cost of GOTURN is 122.8 ms/frame, which is about 3.5 times that of TLD-KCF. In our UAV platform, the monocular camera gets a video stream at a frame rate of 25 Hz. Since the tracking frequency of GOTURN is only 8 Hz, it cannot meet the requirements of the UAV's real-time visual tracking for a moving target in cluttered scenes. The tracking frequency of TLD-KCF is about 29 Hz, and the tracking frequency of TLD and KCF are much faster, so all of these algorithms are fast enough in this paper.

The lost track ratio is an effective evaluation criterion to measure the tracking results with reference to a ground truth.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LIU *et al.*: NOVEL REAL-TIME MOVING TARGET TRACKING AND PATH PLANNING SYSTEM FOR QUADROTOR UAV

9

Fig. 17. Tracking results for a human target in the quadrotor's vision images.

Fig. 18. Quadrotor's autonomous flight through the cluttered woods.

TABLE II
TIME-COST FOR FOUR TRACKING ALGORITHMS

Algorithms	TLD	KCF	TLD-KCF	GOTURN
Average time cost (*ms/f*)	25.3	12.6	34.2	122.8

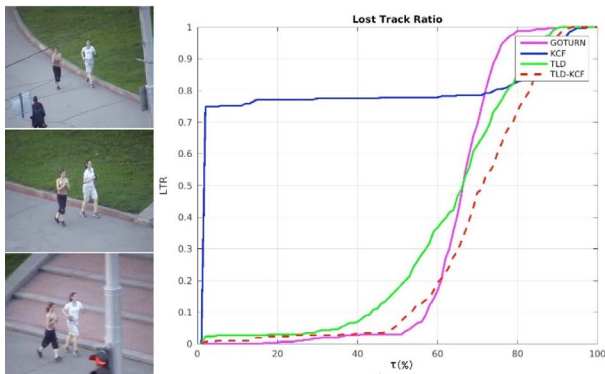Fig. 19. Comparison results of lost track ratio using a group of human targets in VOT 2014.

Fig. 20. Comparison results of lost track ratio using another group of human targets in VOT 2014.

As introduced in [23], the lost track ratio defines a comprehensive measure of tracking performance. The smaller the area under the lost-track ratio curve is, the better the tracking result is. Here, we use VOT 2014 as the benchmark dataset to compare the tracking performance of TLD, KCF, TLD-KCF, and GOTURN. There are 25 sequences in VOT 2014 dataset showing various target objects in challenging backgrounds. 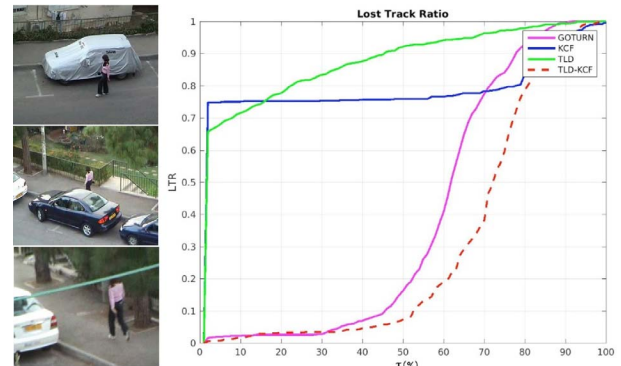Two groups of human targets and two groups of cars are selected to test the performance of these four tracking algorithms. Comparison results of lost track ratio using different human targets and cars in VOT 2014 dataset are given in Figs. 19–22, respectively.

According to the lost track ratio curves shown in above figures, TLD-KCF shows a superior tracking performance than that of TLD and KCF. As compared with GOTURN, TLD-KCF shows a better performance than that of GOTURN when the target does not have salient features. As shown in Figs. 19 and 20, the performance of TLD-KCF is better than that of GOTURN when the tracking target is human. When a car is tracked in complex outdoor scenes, the performance of TLD-KCF is still better than that of GOTURN (see Fig. 21). When the target is changed to a black car with the relatively clean background, the performance of GOTURN is better than that
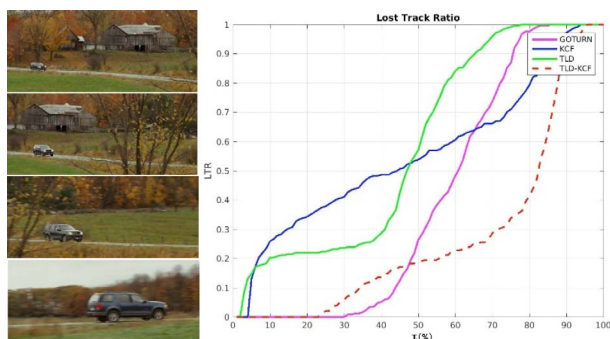
Fig. 21.   Comparison results of lost track ratio using a group of cars in VOT 2014.
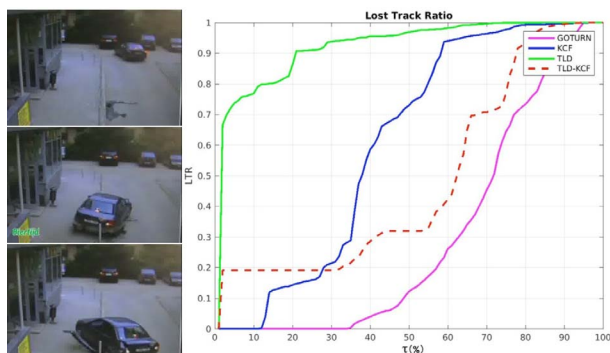

Fig. 22.   Comparison results of lost track ratio using another group of cars in VOT 2014.

of TLD-KCF, but the performance of TLD-KCF is still better than that of TLD and KCF significantly (see Fig. 22).

In this paper, we want to find a practical solution for a quadrotor UAV to accomplish autonomous visual tracking of moving targets in cluttered outdoor scenes. Considering both the tracking accuracy and time-cost, TLD-KCF is our first choice to accomplish the tracking task robustly with our quadrotor platform.

VI. CONCLUSION

This paper has been focused on how to accomplish vision-based moving target detection and tracking, as well as real-time path planning with a small UAV flying in unstructured and cluttered outdoor scenes. To accomplish real-time moving target tracking tasks, SSD algorithm has been adopted to detect multiple candidate targets from an input image, and then an SVM-based target screening algorithm has also been used to eliminate the false targets and find the correct one. A new tracking algorithm, TLD-KCF, has been proposed to improve the tracking performance significantly, and its low computation cost is suitable for the real-time tracking task. Then, a novel path planning algorithm has been proposed based on an elliptical tangent model, which can perform feasible path planning without map building. Experimental results and videos have shown that the proposed approach is a practical solution for a UAV to accomplish autonomous moving target tracking in cluttered outdoor environments.

In the future research, we plan to further improve the performance of our moving target tracking algorithms when the target is moving at a high speed (e.g., a human is running instead of walking). Moreover, other object detection and tracking algorithms will be studied to further improve our UAV's robustness in real-world applications.

REFERENCES

[1] L. Wang *et al.*, "A UKF-based predictable SVR learning controller for biped walking," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 43, no. 6, pp. 1440–1450, Nov. 2013.

[2] M. Gupta, S. Kumar, L. Behera, and V. K. Subramanian, "A novel vision-based tracking algorithm for a human-following mobile robot," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 47, no. 7, pp. 1415–1427, Jul. 2017.

[3] Z. Cao *et al.*, "Image dynamics-based visual servoing for quadrotors tracking a target with a nonlinear trajectory observer," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2720173.

[4] G. Lai, Z. Liu, Y. Zhang, and C. L. P. Chen, "Adaptive position/attitude tracking control of aerial robot with unknown inertial matrix based on a new robust neural identifier," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 27, no. 1, pp. 18–31, Jan. 2016.

[5] S. Minaeian, J. Liu, and Y.-J. Son, "Vision-based target detection and localization via a team of cooperative UAV and UGVS," *IEEE Trans. Syst., Man, Cybern., Syst.*, vol. 46, no. 7, pp. 1005–1016, Jul. 2016.

[6] D. Cavaliere, V. Loia, A. Saggese, S. Senatore, and M. Vento, "Semantically enhanced UAVs to increase the aerial scene understanding," *IEEE Trans. Syst., Man, Cybern., Syst.*, to be published, doi: 10.1109/TSMC.2017.2757462.

[7] T. Moranduzzo and F. Melgani, "Detecting cars in UAV images with a catalog-based approach," *IEEE Trans. Geosci. Remote Sens.*, vol. 52, no. 10, pp. 6356–6367, Oct. 2014.

[8] J. Chen, T. Liu, and S. Shen, "Tracking a moving target in cluttered environments using a quadrotor," in *Proc. IEEE/RSJ Int. Conf. Intell. Robots Syst. (IROS)*, Daejeon, South Korea, 2016, pp. 446–453.

[9] M. Blösch, S. Weiss, D. Scaramuzza, and R. Siegwart, "Vision based MAV navigation in unknown and unstructured environments," in *Proc. IEEE Int. Conf. Robot. Autom.*, Anchorage, AK, USA, May 2010, pp. 21–28.

[10] X. Zhang, Y. Zhuang, H. Hu, and W. Wang, "3-D laser-based multiclass and multiview object detection in cluttered indoor scenes," *IEEE Trans. Neural Netw. Learn. Syst.*, vol. 28, no. 1, pp. 177–190, Jan. 2017.

[11] S. A. P. Quintero and J. P. Hespanha, "Vision-based target tracking with a small UAV: Optimization-based control strategies," *Control Eng. Pract.*, vol. 32, pp. 28–42, Nov. 2014.

[12] A. Giusti *et al.*, "A machine learning approach to visual perception of forest trails for mobile robots," *IEEE Robot. Autom. Lett.*, vol. 1, no. 2, pp. 661–667, Jul. 2016.

[13] W. Liu *et al.*, "SSD: Single shot multibox detector," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, 2016, pp. 21–37.

[14] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Las Vegas, NV, USA, 2016, pp. 779–788.

[15] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017.

[16] T. Joachims, *Making Large Scale SVM Learning Practical*. Dortmund, Germany: Universität Dortmund, 1999.

[17] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit.*, San Diego, CA, USA, 2005, pp. 886–893.

[18] 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.

[19] Z. Kalal, K. Mikolajczyk, and J. Matas, "Forward-backward error: Automatic detection of tracking failures," in *Proc. Int. Conf. Pattern Recognit.*, Istanbul, Turkey, 2010, pp. 2756–2759.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

LIU *et al.*: NOVEL REAL-TIME MOVING TARGET TRACKING AND PATH PLANNING SYSTEM FOR QUADROTOR UAV   11

[20] 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.

[21] R. Halir and J. Flusser, "Numerically stable direct least squares fitting of ellipses," in *Proc. Int. Conf. Central Europe Comput. Graph. Visual.*, Pilsen, Czech Republic, 1999, pp. 125–132.

[22] D. Held, S. Thrun, and S. Savarese, "Learning to track at 100 FPS with deep regression networks," in *Proc. Eur. Conf. Comput. Vis.*, Amsterdam, The Netherlands, 2016, pp. 749–765.

[23] T. Nawaz and A. Cavallaro, "PFT: A protocol for evaluating video trackers," in *Proc. 18th IEEE Int. Conf. Image Process. (ICIP)*, Brussels, Belgium, 2011, pp. 2325–2328.

Yisha Liu received the B.S. and Ph.D. degrees in control theory and engineering from the Dalian University of Technology, Dalian, China, in 2005 and 2011, respectively.

She is an Associate Professor with Information Science and Technology College, Dalian Maritime University, Dalian. Her current research interests include mobile robot and UAV's road detection, path planning, object detection and tracking, and outdoor scene understanding.

Qunxiang Wang received the bachelor's degree in measurement and control technology and instrument from Dalian Jiaotong University, Dalian, China, in 2014 and the master's degree in control theory and engineering from the Dalian University of Technology, Dalian, in 2017.

His current research interests include visual object detection, tracking, and UAV's autonomous navigation in complex outdoor environments.

Huosheng Hu (M'94–SM'01) received the M.Sc. degree in industrial automation from Central South University, Changsha, China, in 1982 and the Ph.D. degree in robotics from the University of Oxford, Oxford, U.K., in 1993.

He is a Professor with the School of Computer Science and Electronic Engineering, University of Essex, Colchester, U.K., leading the Robotics Research Group. His current research interests include behavior-based robotics, human–robot interaction, embedded systems, multisensor data fusion, machine learning algorithms, mechatronics, pervasive computing, and service robots. He has published over 500 papers in journals, books, and conferences in the above areas.

Prof. Hu was a recipient of number of Best Paper Awards. He has been a Program Chair or a member of Advisory Committee of many IEEE international conferences, such as the IEEE International Conference on Robotics and Automation, International Conference on Intelligent Robots and Systems, the International Conference on Mechatronics and Automation, the International Conference on Robotics and Biomimetics, International Conference on Information and Automation, and International Conference on Automation and Logistics. He currently serves as the Editor-in-Chief of the *International Journal of Automation and Computing* and Online Robotics Journal, and the Executive Editor of the *International Journal of Mechatronics and Automation*. He is a Founding Member of the IEEE Robotics and Automation Society Technical Committee on Networked Robots. He is a Fellow of the Institution of Engineering and Technology and the Institute of Measurement and Control.

Yuqing He (M'12) was born in Weihui, China, in 1980. He received the B.S. degree in engineering and automation from Northeastern University at Qinhuangdao, Qinhuangdao, China, in 2002 and the Ph.D. degree in pattern recognition and intelligent system from the Shenyang Institute of Automation, Chinese Academy of Sciences, Shenyang, China, in 2008.

He is currently a Full Professor with the State Key Laboratory of Robotics, Shenyang Institute of Automation, Chinese Academy of Sciences. In 2012, he was a Visiting Researcher with the Institute for Automatic Control Theory, Technical University of Dresden, Dresden, Germany. His current research interests include nonlinear estimation and control and autonomy of mobile robot systems.