IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received June 10, 2019, accepted July 9, 2019, date of publication July 18, 2019, date of current version August 9, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2929760

Detection, Tracking, and Geolocation of Moving Vehicle From UAV Using Monocular Camera

XIAOYUE ZHAO[1], FANGLING PU[1], (Member, IEEE), ZHIHANG WANG[1], HONGYU CHEN[1], AND ZHAOZHUO XU[2], (Student Member, IEEE)

[1]School of Electronic Information, Wuhan University, Wuhan 430072, China
[2]Electrical Engineering Department, Stanford University, Stanford, CA 94305, USA

Corresponding author: Fangling Pu (flpu@whu.edu.cn)

This work was supported by the National Key Research and Development Program of China under Grant 2016YFB0502600.

ABSTRACT Unmanned aerial vehicles (UAVs) have been widely used in urban traffic supervision in recent years. However, the detection, tracking, and geolocation of moving vehicle based on the airborne platform suffer from small object sizes, complex scenes, and low-accuracy sensors. To address these problems, this paper develops a framework for moving vehicle detecting, tracking, and geolocating based on a monocular camera, a GPS receiver, and inertial measurement units (IMUs) sensors. First, the method based on YOLOv3 was employed for vehicle detection due to its effectiveness and efficiency for small object detection in complex scenes. Then, a visual tracking method based on correlation filters is introduced, and a passive geolocation method is presented to calculate the GPS coordinates of the moving vehicle. Finally, a flight control method in terms of the previous image processing results is introduced to lead the UAV that is following the interesting moving vehicle. The proposed scheme has been built on a DJI M100 platform on which a monocular camera and a microcomputer Jetson TX1 are added. The experimental results show that this scheme is capable of detecting, tracking, and geolocating the interesting moving vehicle with high precision. The framework demonstrates its capacity in automatic supervision on target vehicles in real-world experiments, which suggests its potential applications in urban traffic, logistics, and security.

INDEX TERMS Unmanned aerial vehicle, YOLOv3, object geolocation, moving vehicle tracking.

I. INTRODUCTION

The prosperity of unmanned aerial vehicles (UAVs) combined with image processing algorithms has led to the expansion of the application fields of UAVs. Using UAVs to detect, track or geolocate moving vehicles has attracted the interests of researchers. These types of drones with tracking and geolocating frameworks attained wide prospects in traffic safety inspection, road surface monitoring, traffic flow monitoring and urban security protection [1] because the drones are unaffected by the ground traffic congestion. Moreover, they can flexibly respond to scene changes in a timely manner.

Most projects focus on the motion detection and visual tracking of moving objects from the quadrotor using a monocular visible light camera. For example, in the MODAT system [2], moving objects are detected by a motion segmentation method. In DATMO [3], moving objects are detected by optical flow and tracked by a Kalman filter. This system was implemented on the AscTec Pelican quadrotor with a firefly camera. However, these projects did not play a role in object monitoring because they are limited to estimating the object in images. A system focused on visual tracking and geolocation from a helicopter using simple sensors was developed in [4]. The Lucas-Kanade optical flow and motion segmentation method were applied to extract the object and track it. The object was geolocated by a pseudo-stereo vision technique. However, the background of aerial images is complex and the noise may be processed as moving objects, such that all these problems affect the speed and accuracy of image processing algorithms.

To improve the performance of the system, a thermal camera or other equipment with high precision supplements the load equipment of UAVs. In paper [5], Sean proposed a system to detect, track and locate wildlife with a multirotor UAV. The moving wildlife was extracted by blob detection and geolocated by a simple trigonometric algorithm.

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

In another tracking and geolocation system [6], the real-time kinematic (RTK) global positioning systems (GPS) and thermal camera were installed on the multirotor UAV. The stationary object was chosen manually, and the object visual tracking was implemented through the mean-shift method. However, the estimation of the moving object was not presented in [6]. Moreover, the ground reference points must be set in advance for geolocation. With the supplement of the thermal camera and RTK equipment, the moving objects could be well detected but could not yet be identified. Furthermore, the UAVs were controlled manually to keep the objects in the field of view (FOV) in the presented schemes.

This paper focuses on using a monocular camera and low-cost sensors, such as a single-frequency GPS receiver and inertial measurement units (IMU), to implement detection, tracking, and geolocation of moving vehicles from a small quadrotor. The main challenges involve three parts. First, the small size of objects, the variation of FOV, and complicated background make it difficult to detect and track moving vehicles. The target vehicles are relatively small in the video frame because of the distance between the UAV and the vehicles. When the UAV moves, the altitude angles change with high frequency, which causes variation in the FOV. To detect vehicles, these challenges must be tackled first. Second, the accuracies of most equipped airborne sensors are low, introducing challenges to vehicle geolocation. Limited by space and cost, most small UAVs are equipped with only one visible light camera. Auxiliary equipment that incorporates the position with high precision, such as achieved with a binocular camera and laser rangefinder, is not included in the standard configurations. Third, the algorithms for detection, tracking and geolocation must be run on board in real time. Therefore, to make the UAV follow the interested vehicle automatically, the introduced algorithms must be installed and run on the airborne microcomputer.

We present a framework for moving vehicle detection, tracking and geolocation using an airborne platform with a microcomputer. Our key contributions in this work are summarized as follows: (1) Instead of using motion detection methods, we applied You Only Look Once version 3 (YOLOv3) to detect small vehicles from the airborne video. YOLOv3 is introduced to determine the position of initial pixels of the target vehicle with high performance. It minimizes the human intervention in the system. In the experiments, we compared YOLOv3 with other state-of-the-art object detection methods, and the result indicates that the YOLOv3 has the best performance in small object detection in terms of speed and accuracy. (2) A geolocation method using a monocular camera and low-cost sensors is introduced. A visual tracking method is used to calculate the vehicle pixels' position in each frame before the geolocation algorithm is implemented. The geolocation method is presented based on the projection model of the camera on the UAV. As the final result, the trajectory of the moving vehicle is displayed on the satellite map of the ground tablet. (3) A flight control method based on the results of visual tracking and geolocation was developed to make the UAV follow the interested moving vehicle. A completed framework of moving vehicle monitoring is introduced.

The experimental results show that in the aerial image with the size of 1920×1080 pixels, the minimum size of the vehicle that can be detected is 30×30 pixels, and the error of geolocation is within 5 m when the UAV flies at a height of no more than one hundred meters. In terms of speed, the detection, tracking and geolocation methods can complete tasks in real time. Limited by the on-board processor and storage, the model obtained through YOLOv3 is pregenerated. The configured UAV has a potential application in disaster relief, city security and investigation for possible target recognition and tracking. For example, in cases wherein relevant personnel are unable to closely approach the target, it could be applied in the monitoring of the transport vehicles.

The rest of this paper is organized as follows. The related work on detection, visual tracking and geolocation is discussed in Section II. The framework and the corresponding image processing methods are described in Section III. The methods of controlling the UAV to follow the vehicle are also presented in this section. The experiments and results are shown in Section IV. The conclusion is given in Section V.

II. RELATED WORK

To track and geolocate a moving vehicle automatically from the UAV, three parts of the work are involved. The first one is small vehicle detection in an aerial video. The second one is object visual tracking, and the third one is geolocation.

A. OBJECT DETECTION

Most studies have focused on using traditional motion detection methods on UAVs to detect small objects, such as temporal difference and optical flow. For example, features can be estimated by optical flow, and the segmentation of moving objects is achieved through the background subtraction method [7]. The moving cameras, complicated background, and small objects are all considered in these methods, so they work well in both speed and accuracy. However, these motion detection methods could not recognize the objects of a certain type.

The detection methods based on deep learning have made a great breakthrough in recent years. The two-stage methods and the one-stage methods are two branches of them. Regarding the two-stage algorithms, which are represented by R-CNN [8] and fast R-CNN [9], the detection is divided into two phases, region proposal generation and classification. They focus on improving detection accuracy while sacrificing the detection speed. For the one-stage algorithms, region proposal generation is eliminated, and the probability and position coordinate are directly obtained through a single network. For example, the single-shot multibox detector (SSD) [10] creates bounding box candidates at a given position and scale and then calculates their actual bounding box and score for each class. These algorithms work well in terms

of speed. However, most of them perform worse than the two-stage algorithms in detection accuracy and small object detection.

According to the requirements in this paper, the vehicle detection algorithm needs real-time performance, which means the algorithm processes images at 25 frames per second (FPS) or a higher speed. In this case, we consider one-stage algorithms. As mentioned above, most of them have poor performance in small object detection, such as SSD. YOLO [11] and its enhanced versions, which are denoted as YOLOv2 [12] and YOLOv3 [13], are the typical methods in one-stage algorithms. YOLOv3 has higher performance in both speed and accuracy and works especially well in small object detection. It has almost the same precision as two-stage detectors [14]. Therefore, we choose YOLOv3 as our vehicle detection method.

B. VISUAL TRACKING

There are two major categories of visual tracking algorithms, generative model methods and discriminative model methods. Regarding the generative model methods, such as the particle filter [15] or mean-shift tracking [16], the target region is modeled in the current frame, and then, the most similar region to the model will be predicted as the target pixels' position. For example, Tehran University proposed a visual tracking and geolocation project and chose the mean-shift as the visual tracking solution [6]. These methods are not appropriate for the video captured from a UAV using a visible light camera because they are susceptible to the complex background; they are only suitable for the cases where the pixel size of the object is relatively large and the object moves at a low speed.

Discriminative model methods take the target region as a positive sample and the background region as a negative sample. The classifier is trained based on machine learning methods. Then, the optimal region will be predicted by using the classifier in the next frame. Methods based on the correlation filter (CF) are the state-of-the-art methods of this type, such as kernelized correlation filter (KCF) [17]. The pixel size of the bounding box adapts to the pixel size of the target vehicle. The interested target could be tracked robustly and quickly even if the pixel size of the object is small and the background is complex. In recent years, new methods have emerged for solving the problems that KCF could not address, such as large deformation objects or boundary effects. As an example, the spatially regularized discriminative CF (SRDCF) [18] has been suggested by enforcing the spatial penalty on correlation filter coefficients. However, these methods improve the tracking performance at the price of increasing the computational complexity.

In this paper, the vehicle should be tracked in real time. The target vehicle does not have large deformation and is less affected by the boundary effect. Considering both speed and accuracy, KCF is adopted as visual tracking method for aerial videos.

C. OBJECT GEOLOCATION

According to the working mechanism, geolocation based on UAV can be classified into active methods and passive methods [19]. Active methods are based on the laser rangefinder or other apparatus [20], [21]. However, these devices are not available on small airborne platforms.

For small UAVs, GPS and IMU can provide the location and attitude of the UAV, so the passive methods are widely chosen. The GPS position of the object is calculated by image analysis methods. However, it is difficult to geolocate the moving objects. Therefore, researchers focused on the geolocation of stationary objects. For example, some projects established the projection model and reduced the positioning error by a recursive least square filter [22], [23]. Furthermore, a circular trajectory above the stationary object is necessary for UAV. In [24], they kept the drone flying at a constant altitude, and then, the stationary strawberry crop was geolocated by the Kalman filter. Nevertheless, they are not available for geolocation of the moving objects.

Some scholars tried to geolocate the moving objects through referenced data. In paper [25], the object was geolocated by registering the current image with the images provided by a geographic information system (GIS) database. The availability of the methods depended on the accuracy of the image registration. Generally, when the accuracy increases, the computational complexity follows the rise achieved. Others sought faster methods by introducing the digital terrain model (DTM). The aerial images were corrected, and the object location was calculated by matching the DTM data [26], [27]. The accuracy of the results was significantly affected by the reference data, and the ground reference points were shown to be indispensable. The application range of DTM is limited.

III. METHODOLOGY

A. SYSTEM DESCRIPTION

The functional modules of this vehicle monitoring system are shown in Fig. 1. Manual control at the ground is set to deal with an emergency. The operation of this system consists of three steps. After taking off, the UAV flies to the appropriate height. In step 1, the vehicle detection module obtains the aerial images data from the data acquisition module. The small vehicles are detected automatically, and the pixel position of the object is sent to the next module as the initial pixel position for vehicle visual tracking.

In step 2, the data acquisition module provides image data and flight motion parameters to the visual tracking and geolocation module. Based on the initial pixel position of the vehicle and aerial video, the current vehicle pixel position on each video frame is calculated by the visual tracking method. These pixel positions, along with the flight motion parameters, are applied to calculate the GPS coordinates of the target vehicle. The flight motion parameters here include the UAV location and attitude obtained from GPS and IMU. These computed coordinates are sent to the ground control

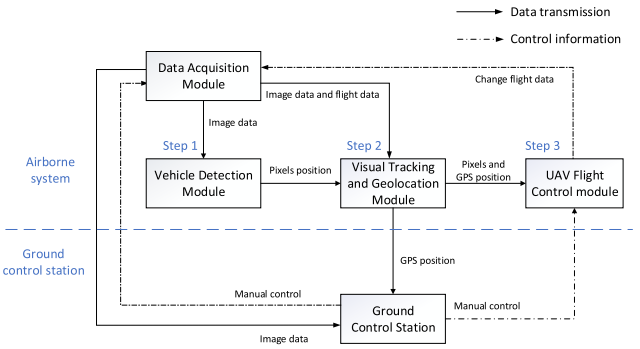

FIGURE 1. System module composition.

	Type	Filters	Size	Output
	Convolutional	32	3 × 3	256 × 256
	Convolutional	64	3 × 3 / 2	128 × 128
1×	Convolutional	32	1 × 1	
	Convolutional	64	3 × 3	
	Residual			128 × 128
	Convolutional	128	3 × 3 / 2	64 × 64
2×	Convolutional	64	1 × 1	
	Convolutional	128	3 × 3	
	Residual			64 × 64
	Convolutional	256	3 × 3 / 2	32 × 32
8×	Convolutional	128	1 × 1	
	Convolutional	256	3 × 3	
	Residual			32 × 32
	Convolutional	512	3 × 3 / 2	16 × 16
8×	Convolutional	256	1 × 1	
	Convolutional	512	3 × 3	
	Residual			16 × 16
	Convolutional	1024	3 × 3 / 2	8 × 8
4×	Convolutional	512	1 × 1	
	Convolutional	1024	3 × 3	
	Residual			8 × 8
	Avgpool		Global	
	Connected		1000	
	Softmax			

FIGURE 2. YOLOv3 architecture.

station and displayed on the satellite map of the tablet in real time.

In step 3, after calculating the current pixel position and GPS position of the target vehicle at the exact moment in step 2, the motion of the vehicle is analyzed. The attitude angles and GPS position of the UAV at the next moment are calculated in the flight control module to change the flight control parameters of the UAV and make the UAV follow the vehicle automatically. This keeps the target vehicle always in the FOV of the UAV, which makes the vehicle visual tracking and geolocation more accurate in step 2.

The methods included in these modules will be described in the remainder of this section.

B. VEHICLE DETECTION

The vehicles in aerial images are detected through the YOLOv3 algorithm. The network structure of YOLOv3 is shown in Fig. 2 [13]. All the input images of various image sizes in our experiment are resized to 416×416.

Three training strategies are employed in YOLOV3 and contribute to its speed and accuracy, including the Darknet-53 network, independent logistic classifiers, and multiscale prediction. The Darknet-53 network is introduced as the feature extractor. It is a feature extracting deep neural network

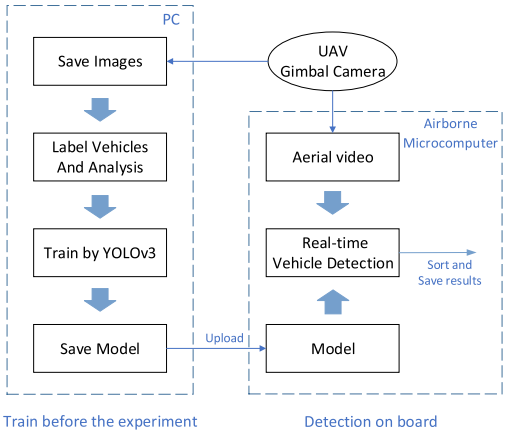

FIGURE 3. Technology road map of vehicle detection.

with 53 convolutional layers [14]. The input image is divided into grid cells. Regarding the 416×416 images, there are 13×13 grid cells. Each grid cell predicts a certain number of bounding boxes, which contain widths and heights and their coordinates and the objectness scores. The classes that the box may contain are predicted using multilabel classification. Independent logistic classifiers are applied instead of softmax for better performance.

YOLOv3 outperforms other methods in small object detection for multiscale prediction. As shown in Fig. 2, YOLOv3 predicts boxes at 3 different scales. K-means is utilized to cluster the bounding box priors. Nine clusters are chosen and divided up across scales. For the first scale, several convolutional layers are added from the base feature extractor. Then, the feature map from 2 layers is taken previously and upsampled. An earlier feature map in the network is merged with the upsampled features using concatenation. This combined feature map is processed to predict boxes for the second scale after adding a few convolutional layers. The same process is performed for the final scale. This method helps to obtain more meaningful semantic information from the upsampled features and finer-grained information from the earlier feature map [13]. In case each cell predicts 3 boxes at each scale, the output tensor is $13 \times 13 \times [(5 + C) \times 3]$, where C is the number of object classes.

The technology road map of vehicle detection is shown in Fig. 3. Aerial images are collected and preprocessed before the experiments. All the vehicles in images are labeled. These labeled images will be divided into a training set, validation set, and testing set. Car is the category we care about in our experiment. YOLOv3 is applied to train the model before the experiment. The model composed of the weight of the network is saved and installed in the microcomputer on the small UAV for the vehicle detection on board.

When several vehicles are detected in the video frame, they will be automatically numbered in order of confidence from small to large. This helps to quickly determine the interested vehicle to track. The vehicle with the highest confidence is usually treated as the interested target, or we can select the number of the vehicle we interested in. Additionally,

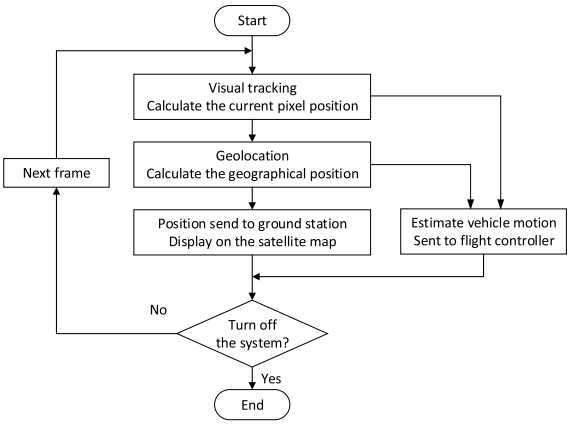

FIGURE 4. The flowchart of the visual tracking and geolocation algorithm.

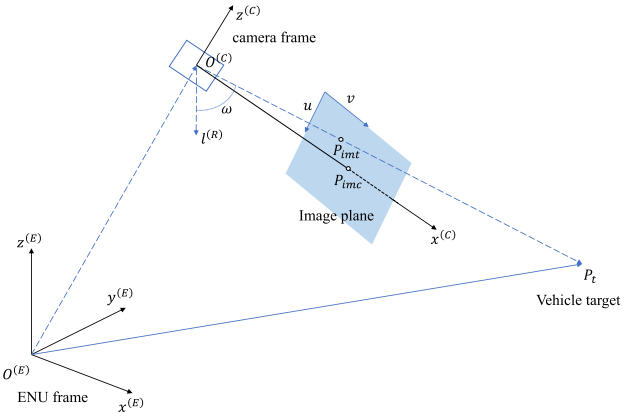

FIGURE 5. The projection model of camera on UAV.

the minimum pixel size of the selected type of vehicles that the system can detect will be estimated when the vehicles are labeled.

C. VISUAL TRACKING AND GEOLOCATION

The flowchart of the visual tracking and geolocation part is shown in Fig. 4. The initial pixel position of the target vehicle obtained by the vehicle detection algorithm, along with the aerial video and flight motion parameters, is input into both the visual tracking and geolocation algorithm. The pixel position and GPS coordinates of the vehicle will be output as reference data for the flight control algorithm. The position coordinates of the vehicle are sent to the ground control station, and the vehicle is displayed on the satellite map of the tablet.

1) VISUAL TRACKING BY KCF

Before the GPS coordinates are calculated, the vehicle is tracked visually by KCF to obtain the pixel position of the vehicle on each video frame.

KCF uses the circulant matrices to collect positive and negative samples. The models to be tracked are trained by kernel ridge regression [17], and the operations of matrices are transformed into the dot-product of vectors with very low computational complexity according to the diagonalization of circulant matrices by the discrete Fourier transform. In addition, KCF allows for the use of the histogram of oriented gradient (HOG) features with multiple channels instead of raw pixels, which gives an important boost to the performance of visual tracking.

2) VEHICLE GEOLOCATION

The object geolocation method works by employing three components. The first is the transformation of the pixel coordinate frame to the East-North-Up (ENU) frame. The second is the depth estimation of the object. The third is conversion of the ENU frame to the GPS coordinates.

The projection model of the camera on UAV is shown in Fig. 5. The shaded area is the imaging plane of the camera. $\boldsymbol{P}_t$ is the target vehicle. $\boldsymbol{P}_{imt}$ is the corresponding position of $\boldsymbol{P}_t$ on the imaging plane. $\boldsymbol{P}_{imc}$ is the center of the image. The distance from the optical center of the camera to $\boldsymbol{P}_{imc}$ is the focal length f of the camera.

Step 1: Establish the transformation formula of the pixel coordinate frame (uv) to the ENU frame (E) through the camera frame (C).

Assume that all coordinate frames are right-handed orientations. The ENU frame is the world coordinate frame with its origin at the defined home location, as shown in Fig. 5. The origin of the camera frame is the optical center of the camera on the UAV. The x-direction points to the front of the camera, and y-direction points to the left.

The ENU frame can be coincident with the camera frame through rigid motion. The rigid motion includes rotating and translational motion. The rotating motion from the ENU frame to the camera frame can be decomposed into three rotations around the three axes. The rotating matrix from the ENU frame to the camera frame is expressed as

$$\boldsymbol{R}_{(E)}^{(C)} = \boldsymbol{R}_x\boldsymbol{R}_y\boldsymbol{R}_z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & cos\theta & sin\theta \\ 0 & -sin\theta & cos\theta \end{bmatrix} \times \begin{bmatrix} cos\gamma & 0 & sin\gamma \\ 0 & 1 & 0 \\ -sin\gamma & 0 & cos\gamma \end{bmatrix} \times \begin{bmatrix} cos\varphi & sin\varphi & 0 \\ -sin\varphi & cos\varphi & 0 \\ 0 & 0 & 1 \end{bmatrix}. \tag{1}$$

$\boldsymbol{R}_{(E)}^{(C)}$ is the rotation matrix from the ENU frame to the camera frame. $\boldsymbol{R}_x$, $\boldsymbol{R}_y$ and $\boldsymbol{R}_z$ represent the matrix of rotation around the x, y and z axis, respectively. γ, θ, and φ are the pitch, roll and yaw angles of the camera, respectively.

The translational motion is described as a vector

$$t = \boldsymbol{P}_{uav}^{(E)}. \tag{2}$$

$\boldsymbol{P}_{uav}^{(E)}$ is the position of the UAV in the ENU frame.

The final transformation of the camera frame to the ENU frame is presented as

$$\boldsymbol{P}_t^{(E)} = \boldsymbol{R}_{(C)}^{(E)}\boldsymbol{P}_t^{(C)} + \boldsymbol{P}_{uav}^{(E)}. \tag{3}$$

$\boldsymbol{P}_t^{(E)}$ denotes the position of the target vehicle in the ENU frame. $\boldsymbol{R}_{(C)}^{(E)}$ is the reciprocal of $\boldsymbol{R}_{(E)}^{(C)}$. $\boldsymbol{R}_{(C)}^{(E)}$ means the rotation

matrix from the camera frame to the ENU frame. $\boldsymbol{P}_t^{(C)}$ represents the position of the target in the camera frame.

Assume that $\boldsymbol{P}_t^{(uv)} = \boldsymbol{P}_{imt} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T$ is the pixel position of target in the homogeneous form. Then, to establish the transformation formula of the pixel coordinate frame to the ENU frame, $\boldsymbol{P}_t^{(C)}$ should be converted to $\boldsymbol{P}_t^{(uv)}$ in Equation 3.

The position of point $\boldsymbol{P}_{imt}$ in the camera frame can be calculated by

$$\begin{aligned} \boldsymbol{P}_{imt}^{(C)} &= \begin{bmatrix} f & \frac{c_x - u}{\alpha} & \frac{c_y - v}{\beta} \end{bmatrix}^T \\ &= \begin{bmatrix} 0 & 0 & f \\ -f & 0 & 0 \\ 0 & -f & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{f_x} & 0 & \frac{-c_x}{f_x} \\ 0 & \frac{1}{f_y} & \frac{-c_y}{f_y} \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} u \\ v \\ 1 \end{bmatrix} \\ &= \begin{bmatrix} 0 & 0 & f \\ -f & 0 & 0 \\ 0 & -f & 0 \end{bmatrix} \boldsymbol{K}^{-1} \boldsymbol{P}_t^{(uv)} = \boldsymbol{R}_f \boldsymbol{K}^{-1} \boldsymbol{P}_t^{(uv)}. \end{aligned} \tag{4}$$

Here, $\boldsymbol{K}$ represents the camera intrinsics. f_x, f_y, c_x and c_y are the parameters of $\boldsymbol{K}$. c_x and c_y represent half of the image width and half of the image height, respectively, and $f_x = \alpha f$ and $f_y = \beta f$ describe the relationship between the pixels' distance and the real distance (meters), respectively. f is the focal length of the camera.

The depth of the target vehicle is denoted as d. d is the distance between the vehicle P_t and the optical center of the camera in the ENU frame. Combining Equations 3-4 and Fig. 5, the transformation formula from the pixel position to the ENU frame can be written as

$$\begin{aligned} \boldsymbol{P}_t^{(E)} &= \boldsymbol{R}_{(C)}^{(E)} \boldsymbol{P}_t^{(C)} + \boldsymbol{P}_{uav}^{(E)} \\ &= \boldsymbol{R}_{(C)}^{(E)} d \frac{\boldsymbol{P}_{imt}^{(C)}}{|\boldsymbol{P}_{imt}^{(C)}|} + \boldsymbol{P}_{uav}^{(E)} \\ &= \boldsymbol{R}_{(C)}^{(E)} \frac{d \boldsymbol{R}_f \boldsymbol{K}^{-1} \boldsymbol{P}_t^{(uv)}}{|\boldsymbol{R}_f \boldsymbol{K}^{-1} \boldsymbol{P}_t^{(uv)}|} + \boldsymbol{P}_{uav}^{(E)}. \end{aligned} \tag{5}$$

After calculating d, the target position in the ENU frame ($\boldsymbol{P}_t^{(E)}$) has been calculated by the pixel position of the target in the aerial image ($\boldsymbol{P}_t^{(uv)}$).

Step 2: Calculate the depth of the target vehicle.

We define a reference coordinate frame, which is denoted as (R) in the following equations. Its origin is the optical center of the camera. The axes of it are aligned with the axes of the ENU frame. We define a unit vector $\vec{l}^{(R)} = \begin{bmatrix} 0 & 0 & -1 \end{bmatrix}^T$ there. The unit vector of $\overrightarrow{O^{(C)} P_{imt}}$ in the ENU frame is denoted as

$$\vec{l}_s = \boldsymbol{R}_{(C)}^{(E)} \frac{d \boldsymbol{R}_f \boldsymbol{K}^{-1} \boldsymbol{P}_t^{(uv)}}{|\boldsymbol{R}_f \boldsymbol{K}^{-1} \boldsymbol{P}_t^{(uv)}|}. \tag{6}$$

The angle between $\vec{l}_s$ and $\overrightarrow{O^{(C)} P_t}$ is denoted as ω. ω can be calculated by the inner product of $\vec{l}_s$ and $\vec{l}^{(R)}$. Suppose that the ground on which the vehicles drive is relatively flat; then,

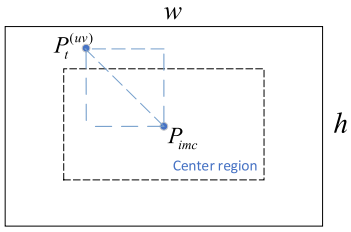

FIGURE 6. Attitude adjustment. The dotted line shows the center region of the aerial image.

the depth of the target vehicle can be expressed as

$$d = \frac{H}{cos\omega} = \frac{H}{\vec{l}^{(R)} \cdot \vec{l}_s}. \tag{7}$$

H is the altitude of the UAV.

Step 3: Calculate the latitude and longitude of the target in terms of the target position in the ENU frame.

To calculate the latitude and longitude of the vehicle, the target position in the ENU frame $\boldsymbol{P}_t^{(E)}$ should be converted to the World Geodetic System-1984 (WGS84) Coordinate System through the Earth-centered inertial (ECI) frame:

$$\boldsymbol{P}_t^{(I)} = \begin{bmatrix} (H + r_E) cosB cosL \\ (H + r_E) cosB sinL \\ (H + r_E) sinB \end{bmatrix} = \boldsymbol{R}_{(E)}^{(I)} \boldsymbol{P}_t^{(E)} + \boldsymbol{P}_{home}^{(I)}. \tag{8}$$

(I) represents the ECI frame in the equations, and $\boldsymbol{P}_{home}^{(I)}$ is the home position in the ECI frame. L, B, and H are the longitude, latitude, and altitude, respectively. r_E is the radius of the Earth. $\boldsymbol{R}_{(E)}^{(I)}$ represents the transformation of the ENU frame and ECI frame [24]. $\boldsymbol{R}_{(I)}^{(E)}$ can be expressed as follows:

$$\boldsymbol{R}_{(I)}^{(E)} = \begin{bmatrix} -sinL & cosL & 0 \\ -sinB cosL & -sinB sinL & cosB \\ cosB cosL & cosB sinL & sinB \end{bmatrix}. \tag{9}$$

Finally, the longitude and latitude of target vehicle P_t can be calculated combining Equations 5, 7 and 8.

D. FLIGHT CONTROL

To keep the vehicle target in the FOV, when the distance between $\boldsymbol{P}_t^{(uv)}$ and $\boldsymbol{P}_{imc}$ along the direction of u and v axis is greater than a certain threshold, the gimbal angles will be adjusted to ensure the $\boldsymbol{P}_t^{(uv)}$ is in the center region of the image. The position of the vehicle in pixel coordinates is $\boldsymbol{P}_t^{(uv)} = \begin{bmatrix} u & v & 1 \end{bmatrix}^T$, as shown in Fig. 6.

When $\boldsymbol{P}_t^{(uv)}$ is outside the center region, the change in the pitch $\Delta\gamma$ and yaw $\Delta\varphi$ can be expressed as

$$\Delta\gamma = arctan(\frac{h/2 - v}{\beta f}) = arctan(\frac{h/2 - v}{f_y}), \tag{10}$$

$$\Delta\varphi = arctan(\frac{w/2 - u}{\alpha f}) = arctan(\frac{w/2 - u}{f_x}). \tag{11}$$

w and h are the width and height of the image, respectively.

To keep the UAV following the target vehicle, the motion of the vehicle should be estimated first. The position of the vehicle in the ENU frame is calculated on each video frame,

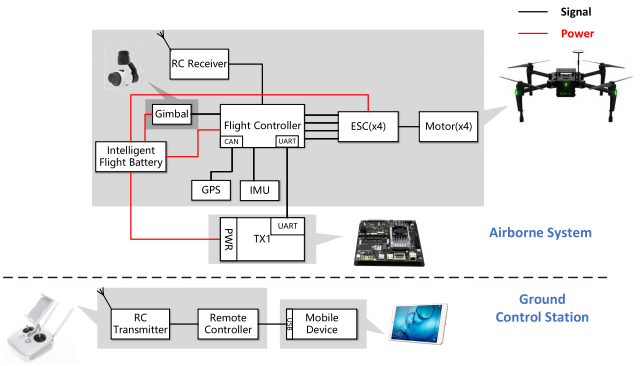

FIGURE 7. System hardware module.

that is, its position is updated every 20 ms. The vehicle position of the current video frame along with the previous one is taken out every second to calculate the speed and direction vectors of the target vehicle. Then, the flight control parameters will be sent to the flight controller, leading the UAV to follow the vehicle.

IV. EXPERIMENTS AND RESULTS

A. HARDWARE COMPOSITION

The framework presented in this paper is implemented on a microcomputer equipped on a small UAV. The UAV configured in the experiments are shown in Fig. 7. DJI Matrice 100 is used as the airborne platform, which has been integrated with the flight controller, gimbal camera, DJI Lightbridge HD video transmission, GPS, IMU, and other sensors. Table 1 shows the main sensors and hardware we used in this system. The camera uses the Sony Exmor R CMOS as the image sensor. The video in our experiment is recorded at 50 frames per second, and the image size of each frame is 1920×1080 pixels. There is also a microcomputer NVIDIA Jetson TX1 attached to the airborne system. The aerial video captured from the camera will be processed on the Jetson TX1 to detect, track and geolocate the target vehicle in real time.

The ground control station communicates with the airborne system through the wireless communication link, Lightbridge. Lightbridge is a dedicated wireless communication link technology in DJI. It works in the 2.4 GHz band, with a maximum communication distance of 5 kilometers. The ground control station consists of a HUAWEI BTV-DL09 tablet PC and DJI C1 remote controller. The remote controller integrates the ground terminal of the Lightbridge transmission system. It can output aerial images and data to the mobile terminal via USB directly. In our system, the tablet PC receives the aerial videos and the GPS coordinates of the target vehicle and displays the trajectory on the satellite map of the tablet in real time.

B. EXPERIMENTAL DATA AND PREPROCESSING

More than 1000 aerial images were collected by DJ Matrice 100 at WUHAN University for the vehicle detection dataset.

TABLE 1. Sensors and hardware composition of the UAV.

Hardware	Type specification	Description
Flight controller	DJI N1	Adjust motor speed to control drone flight.
IMU (magnetometer, accelerometer and gyro)	-	Calculate the attitude information of UAV, and improve location accuracy.
GPS	NEO-M8N-0-01	Measure latitude, longitude and altitude. Accuracy: $2.5m$
Video transmission	DJI LightBridge	Complete the image transmission task, transmit flight information and receive remote control signals.
Gimbal camera	Zenmuse X3	Provide aerial video and attitude angles.
Barometer	MS5607	Measure air pressure and help drone to hover. Accuracy: $1.5m$
Brushless Motor	DJI 3510	Provide the power of drone flight.
Electronic speed control	E SERIES 620D	Input DC power, output three-phase AC power, and drive motor rotation.
Propeller	1345s	Provide lift.
Battery	TB47D, TB48D	-

FIGURE 8. Aerial images captured at WUHAN university. The vehicles are marked with red boxes.

The altitude of UAV is approximately one hundred meters, and the shooting angles are oblique or vertical downward. All the vehicles in images were labeled in advance. This dataset included 700 frames with the image size of 1920×1080 pixels and more than 3000 small target samples. The types of vehicles include sedans, Jeeps, taxis, sport utility vehicles, and other compact cars. All these samples were marked as car. Some images labeled are shown in Fig. 8. The dataset was divided into three subsets, where 60% was employed as the training set, 20% as the validation set, and 20% as the test set.

The detection model is trained on an Intel CPU i7-7700 k machine with GTX1080tI and 32 GB RAM memory in advance. The aerial images are applied to fine-tune the YOLOv3 model based on convolutional weights pretrained on ImageNet. During training, the batch size was set to 32. The learning rate was initialized to 0.001 with the decay rate set to 0.1 for every 10 thousand iterations. When the number of iterations was 20000, the training result had converged, and the model was saved for the following vehicle detection.

All the image process algorithms and models, including YOLOv3, the pretrained model, KCF and object geolocation,

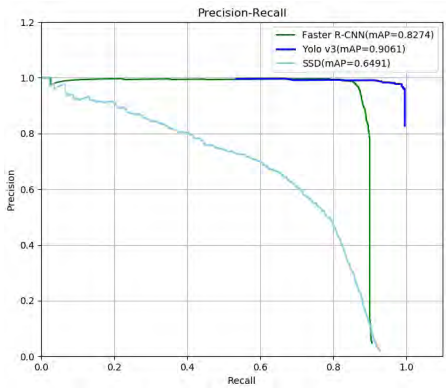

FIGURE 9. Precision-recall curves of the state-of-the-art detection methods.

TABLE 2. The accuracy and speed of YOLOv3, SSD and faster R-CNN.

Methods	mAP	Time
Faster R-CNN	82.74%	0.318 s
YOLOv3	90.61%	0.017 s
SSD	64.91%	0.032 s

are installed on Jetson TX1. The camera is calibrated in advance to obtain the intrinsics [28].

C. RESULTS AND ANALYSIS

1) EFFECTIVENESS OF THE METHODS

To prove the validity of YOLOv3 in this small vehicle monitoring system, several state-of-the-art detection methods were compared with it, including Faster R-CNN [31] and SSD. They were trained with the same parameters and the dataset described above. The models are saved after training. Fig. 9 shows their precision-recall (PR) curves over the validation set. YOLOv3 has the largest area between the curve and the coordinate axis. That is, YOLOv3 has the best performance in detecting the small target, followed by faster R-CNN.

Table 2 shows their mean average precision (mAP) and their average time cost on processing each image on Jetson TX1. YOLOv3 and faster R-CNN achieve a similar mAP, which are both larger than SSD. It means they have better performance than SSD for small object detection. However, in terms of speed, SSD and YOLOv3 meet the real-time requirements of our system. It is demonstrated that YOLOv3 is the best choice for the system developed in this paper.

To verify the effectiveness of the detection methods, we applied object detection in the aerial images. Some small vehicle detection results using YOLOv3 are shown in Fig. 10. The threshold of confidence is set to 0.5. As seen from Fig. 10, the accuracy and recall rate of YOLOv3 is very high even when the target size is relatively small.

To demonstrate the robust of KCF, Fig. 11 shows the tracking result in the presence of interference. And we zoomed into the vehicles of Fig. 11. The screenshots of the tracked vehicle are shown in Fig. 12.

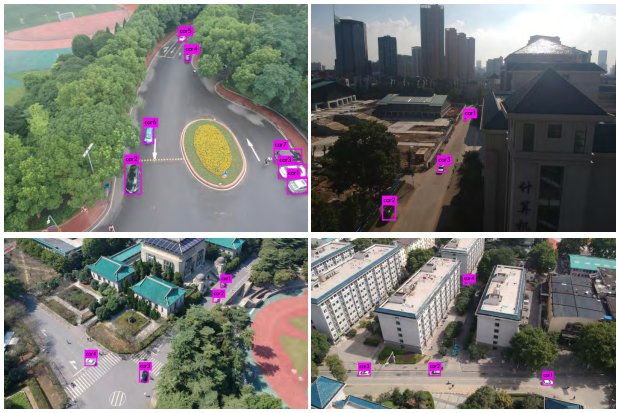

FIGURE 10. Some small vehicle detection results using YOLOv3.

FIGURE 11. Tracking results. When some interfering vehicles appear, the tracking algorithm could still guarantee its high performance.

FIGURE 12. Tracking results. We zoom into the vehicles in Fig. 11.

As shown in Fig. 11 and Fig. 12, when some interfering vehicles of a similar color and similar type appear, the tracking method is still able to track target vehicles steadily. The accuracy of tracking also ensures the accuracy of vehicle geolocation.

FIGURE 13. Geolocation result on an aerial image.

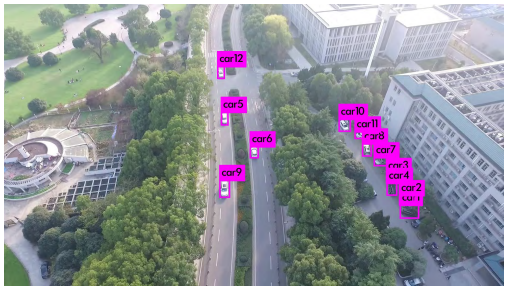

FIGURE 14. The vehicle detection result. The targets detected are numbered in order of confidence.

To test the object geolocation method, we experimented on a single aerial image. The left figure in Fig. 13 is an aerial image captured by DJI M100, and the red box represents the target position, which is the center of the playground. After calculating the longitude and latitude of the target, we marked the geolocation result, ground truth, and the UAV position on the Google Earth, which is shown in the right figure in Fig. 13. The error between the result and ground truth is within five meters.

2) EXPERIMENT RESULTS OF THE PROPOSED FRAMEWORK

A complete experiment is performed at WUHAN University to demonstrate this vehicle monitoring system. Once the system is working, vehicles in the aerial video will be detected automatically and numbered in order of confidence from small to large by Jetson TX1, as shown in Fig. 14. The threshold of confidence is set to 0.5. Based on this result, the vehicle with the highest confidence is treated as the interested target, or we can choose the interested target by selecting the number of the vehicle we are interested in. The pixel position of this bounding box is sent to the visual tracking and geolocation algorithm as the initial pixel position of the target vehicle in the aerial video.

The vehicle denoted as car8 is chosen as the interested target. The tracking result is shown in Fig. 15. The bounding box adapts to the pixel size of the target vehicle. In terms of speed, KCF processes each frame within 10 ms, and the vehicle can be visually tracked in real time.

During this experiment, the home position is taken as the GPS coordinates of the UAV at the start of the log. The geolocation algorithm was run at the same sampling rate of the aerial video. To show the final result of geolocation, the GPS coordinates are sent to the ground station. The trajectory of the interested vehicle is displayed on the satellite map of Amap, as shown in Fig. 16. The GPS position of UAV is also displayed on the map. To estimate the error of geolocation, the calculated positions of the interested vehicle are compared with ground truth every 20 ms. The absolute error along the x- and y-axis in the ENU frame is shown in Fig. 17. As seen from the figure, the error of geolocation is within 5 m.

FIGURE 15. Several video frames of the visual tracking result. The target vehicle is in the red bounding box. The center region of the image is indicated by the green box.

FIGURE 16. The result of geolocation at several moments. The GPS position of the target vehicle and UAV is shown on Amap. The trajectory of the target vehicle is drawn with the red line.

In addition, the vehicle target is always in the FOV of UAV, which is the result of the flight control algorithm. The center region of the image is defined as a rectangular region with a width of 900 pixels and a height of 600 pixels, which is drawn with the green line in Fig. 15. The time for changing attitude angles is set to one second.

D. DISCUSSION

1) PRE-TRAINING FOR OBJECT DETECTION

In this system, subject to the computing power and memory limitations of the microcomputer, the model used in the object detection is pretrained before the experiments. This limits the emergency response capability of the system. When the type of interested vehicle is not available in the training data, the model can be updated to include the type of vehicle through increasing the diversity of the sample and additional experiments.

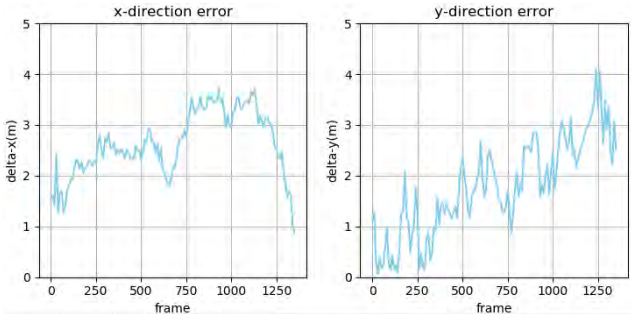

FIGURE 17. Error of vehicle geolocation. X-direction points east. Y-direction points north.

2) OBJECT TRACKING WITH INTERFERENCE

Tracking is the most important intermediate step of the whole framework. The performance of tracking is related to the quality of the object geolocation and flight control. In most of the experiments, the tracking algorithm was effective except for long-time occlusions above the vehicles. For example, the target vehicle drives into the tunnel. As can been seen in Fig. 11 and Fig. 12, without considering the long-time occlusion, our tracking method is robust even in the presence of interference.

3) OBJECT GEOLOCATION ERROR ANALYSIS

The geolocation of the target vehicle is calculated based on the pixel position of the target and the flight data of the UAV. According to this calculation process, the error of target geolocation comes from the following aspects. First, it originates with the system errors, mainly including the measurement errors of the GPS and IMU. These components introduce errors to the position information and attitude information of the UAV.

Second, the visual tracking errors also contribute. Visual tracking ensures the accuracy of target geolocation when the vehicle target is tracked successfully. However, if there is a long-time occlusion, the target vehicle will be lost, and the target geolocation is unable to work.

The third source is the error caused by calculations or experimental assumptions. Because the height of the target vehicle in the experiment is much lower than the height of the UAV, the depth is directly calculated by ignoring the height of the target vehicle. This experimental assumption introduces a slight error to the result of the target geolocation. In addition, due to the large number of sine, cosine, and matrix inversion calculations in the geolocation process, the accuracy of the intermediate process will be affected, resulting in a slight error in the final geolocation result.

To reduce these errors as much as possible, an initialization process is required. Before the system starts, a reference target with known GPS coordinates is geolocated with different attitude angles several times. The deviations will be input into the system as feedback data to reduce these errors.

4) MORE DETAILS

In the experiments, we chose 100 meters as the flight height for two reasons. First, when the UAV performs a vehicle monitoring mission in the city, the flight altitude should be as high as possible compared to the buildings. Second, to perform target detection and tracking tasks, the vehicles in the picture should be visible to the naked eye. Because the camera of the UAV may be obliquely downward, the pixel size of the vehicle farther away may be smaller. Therefore, we choose 100 meters as the safe flying height.

In addition, to keep the UAV following the moving target, the speed of the vehicle should not exceed the maximum speed of the UAV. The maximum speed of the UAV is 17 m/s (61 km/h). This speed accords with the actual situation of vehicles driving in urban traffic.

V. CONCLUSION

In this paper, we presented a complete framework for moving vehicle detection, tracking and geolocation using an airborne platform with a microcomputer. The moving vehicle with a small pixel size in the video is detected based on the model trained by YOLOv3. After the interested target is chosen, a visual tracking algorithm based on the correlation filter is applied to the task of target tracking. A passive geolocation method is presented to calculate the GPS coordinates of the moving vehicle. To keep the target in the FOV of the UAV, a flight control method based on the results of visual tracking and geolocation is also introduced.

We established an airborne experimental system, which is composed of DJI M100 UAV and the Jetson Tx1. The detection, tracking, and geolocation algorithms are all running on board. When the small UAV follows the moving target vehicle, the location of the vehicle is sent to the ground control station and displayed on the tablet. The experimental results show that YOLOv3 is the best choice for a framework that implements real-time vehicle detection. The interested vehicle can be tracked and kept in the FOV of the UAV. The GPS coordinates of the moving vehicle computed by our method are within an error of 5 meters compared to the ground truth points set in advance.

Our moving vehicle detection, tracking and geolocation framework can be implemented on small UAV with no ground reference. Moreover, all the methods mainly depend on the low-accuracy sensors on board. The results show that the presented vehicle monitoring system is a cost-effective and high-performance solution. Further, a navigation algorithm will be added to improve this system and make it more intelligent.

REFERENCES

[1] C. Hu, Z. Zhang, Y. Tao, and N. Wang, "Decentralized real-time estimation and tracking for unknown ground moving target using UAVs," *IEEE Access*, vol. 7, pp. 1808–1817, 2018.

[2] A. W. N. Ibrahim, P. W. Ching, G. L. G. Seet, W. S. M. Lau, and W. Czajewski, "Moving objects detection and tracking framework for UAV-based surveillance," in *Proc. PSIVT*, Nov. 2010, pp. 456–461.

[3] G. R. Rodríguez-Canosa, S. Thomas, J. D. Cerro, A. Barrientos, and B. MacDonald, "A real-time method to detect and track moving objects (DATMO) from unmanned aerial vehicles (UAVS) using a single camera," *Remote Sens.*, vol. 4, no. 4, pp. 1090–1111, 2012.
[4] K. M. Han and G. N. DeSouza, "Geolocation of multiple targets from airborne video without terrain data," *J. Intell. Robot. Syst.*, vol. 62, no. 1, pp. 159–183, Apr. 2011. doi: 10.1007/s10846-010-9442-7.
[5] S. Ward, J. Hensler, B. Alsalam, and L. F. Gonzalez, "Autonomous UAVs wildlife detection using thermal imaging, predictive navigation and computer vision," in *Proc. IEEE Aerosp. Conf.*, Big Sky, MT, USA, Mar. 2016, pp. 1–8.
[6] H. R. Hosseinpoor, F. Samadzadegan, and F. DadrasJavan, "Pricise target Geolocation and tracking based on UAV video imagery," in *Proc. ISPRS-Archives*, vols. XLI–B6, Jul. 2016, pp. 243–249. doi: 10.5194/isprsarchives-XLI-B6-243-2016.
[7] S. Minaeian, J. Liu, and Y.-J. Son, "Effective and efficient detection of moving targets from a UAV's camera," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 2, pp. 497–506, Feb. 2018.
[8] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proc. CVPR*, Columbus, OH, USA, Jun. 2014, pp. 580–587.
[9] R. Girshick, "Fast R-CNN," in *Proc. ICCV*, Santiago, Chile, Dec. 2015, pp. 1440–1448.
[10] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, and A. C. Berg, "SSD: Single shot multibox detector," in *Proc. ECCV*, Amsterdam, The Netherlands, Oct. 2016, pp. 21–37.
[11] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. CVPR*, Las Vegas, NV, USA, Jun. 2016, pp. 779–788.
[12] J. Redmon and A. Farhadi, "YOLO9000: Better, faster, stronger," in *Proc. CVPR*, Honolulu, HI, USA, Jul. 2017, pp. 7263–7271.
[13] J. Redmon and A. Farhadi, "YOLOv3: An incremental improvement," 2018, *arXiv:1804.02767*. [Online]. Available: https://arxiv.org/abs/1804.02767
[14] V. Kharchenko and I. Chyrka, "Detection of airplanes on the ground using YOLO neural network," in *Proc. MMET*, Kiev, Ukraine, Jul. 2018, pp. 294–297.
[15] 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.
[16] D. Comaniciu, V. Ramesh, and P. Meer, "Real-time tracking of non-rigid objects using mean shift," in *Proc. CVPR*, Hilton Head, SC, USA, Jun. 2000, pp. 142–149.
[17] 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.
[18] M. Danelljan, G. Häger, F. S. Khan, and M. Felsberg, "Learning spatially regularized correlation filters for visual tracking," in *Proc. ICCV*, Santiago, Chile, Dec. 2015, pp. 4310–4318.
[19] C. Xu, D. Huang, and F. Kong, "Small UAV passive target localization approach and accuracy analysis," *Chin. J. Sci. Instrum.*, vol. 36, no. 5, pp. 1115–1122, May 2015.
[20] Y. Yang, "Research and implementation of geo-locating technology of moving object based on vision tracking by UAV," (in Chinese), *J. Sichuan Ordnance*, vol. 35, no. 3, pp. 137–140, Mar. 2014.
[21] R. He, X. Tian, H. Liu, and Q. Xi, "A UAV target localization approach based on Monte-Carlo Kalman filter," *J. Northwestern Polytech. Univ.*, vol. 35, no. 3, pp. 435–441, Jun. 2017.
[22] P. Pai and V. P. S. Naidu, "Target geo-localization based on camera vision simulation of UAV," *J. Opt.*, vol. 46, no. 4, pp. 425–435, Dec. 2017.
[23] S. Xu, L. Yafei, H. Zhongxi, S. Shangqiu, and G. Zheng, "Passive geo-location for ground target with multiple measurements using fixed wing UAV," in *Proc. CGNCC*, Nanjing, China, Aug. 2016, pp. 2477–2482.
[24] C. A. Garcia and Y. Xu, "Target Geolocation for agricultural applications via an Octorotor," *IFAC-PapersOnLine*, vol. 49, no. 16, pp. 27–32, 2016.
[25] A. Arun, M. M. Yadav, K. Latha, and K. S. Kumar, "Self directed unmanned aerial vehicle for target geo-localization system," in *Proc. IEEE ICCEET*, Mar. 2012, pp. 984–990.
[26] J. Zhuo, L. Sun, Y. Yang, and X. Zhao, "A target localization method for UAV image sequences based on DEM matching," in *Proc. ISCID*, Zhejiang, China, Dec. 2016, pp. 215–218.
[27] M. Hamidi and F. Samadzadegan, "Precise 3D GEO-location of UAV images using GEO-referenced data," in *Proc. ISPRS-Archives*, vol. XL-1/W5, pp. 269–275, Nov. 2015. doi: 10.5194/isprsarchives-XL-1-W5-269-2015.
[28] Z. Zhang, "A flexible new technique for camera calibration," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 22, no. 11, pp. 1330–1334, Nov. 2000. [Online]. Available: http://10.5194/isprsarchives-XL-1-W5-269-2015
[29] M. Campbell and M. Wheeler, "A vision based Geolocation tracking system for UAV's," in *Proc. AIAA Guid. Navigat. Control Conf. Exhibit*, Keystone, CO, USA, Aug. 2006, p. 6246.
[30] P. Meir, N. Ceccarelli, and P. R. Chandler, "Vision-based target geolocation using micro air vehicles," *J. Guid. Control Dyn.*, vol. 31, no. 3, pp. 597–615, Jun. 2008.
[31] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Proc. Adv. Neural Inf. Process. Syst.*, 2015, pp. 91–99.

XIAOYUE ZHAO received the B.E. degree in electronic and information engineering from Wuhan University, Wuhan, China, in 2017, where she is currently pursuing the M.S. degree in information and communication engineering. Her research interests include image processing and applications in unmanned aerial vehicles.

FANGLING PU received the Ph.D. degree in photogrammetry and remote sensing from Wuhan University, in 2005, where she has been an Associate Professor with the School of Electronic Information, since 2000. She was a Research Assistant with the Institute for Theoretical Information Technology, RWTH Aachen University, in 2007. Her research interests include systems and networking, signal and information processing, and data mining.

ZHIHANG WANG received the B.E. degree in electronic and information engineering from Wuhan University, Wuhan, China, in 2017, where he is currently pursuing the M.S. degree in electronic and communication engineering. His research interest includes unmanned aerial vehicle (UAV) application development.

HONGYU CHEN received the B.E. degree in electronic and information engineering from Wuhan University, Wuhan, China, in 2017, where he is currently pursuing the M.S. degree in information and communication engineering. His research interests include unmanned aerial vehicle (UAV) position and navigation.

ZHAOZHUO XU (S'13) received the B.E. degree in electronic engineering from Wuhan University, Wuhan, China, in 2017.

He is currently a Graduate Researcher with Stanford University. His research interests include machine learning and computer vision, especially applications on object detection and time series modeling.