UAV Based Target Tracking and Recognition*

Tian Xiang¹, Fan Jiang¹, Gongjin Lan¹, Jiaming Sun¹, Guocheng Liu¹, Qi Hao¹ and Cong Wang²

Abstract—In this paper, we develop a quadrotor UAV based target tracking and recognition system, which includes an intelligent gimbal sub-system for accurate camera positioning and fast image processing. A set of robust consensus-based algorithms are developed for objects tracking, in addition to moving background processing techniques. A neural network learning based database is used to improve target recognition performance. Moreover, a Geographic Information System (GIS) is used to provide geo-location, environmental, and contextual information for the tracked objects. Experimental and simulation results have demonstrated the robustness of the proposed target tracking and recognition framework.

I. INTRODUCTION

In recent years, quadrotor UAV has become increasingly widely used in both military or civil applications due to its small size, low cost, high maneuverability, fast response [1], so that it can replace the pilots and rescue team to implement dangerous missions. Civil applications include aerial crop surveys, inspection of power lines and pipelines, forest fire detection and monitoring, search and rescue, aerial photography, etc. For military applications, UAVs can complete the complex tasks such as transporting materials, battlefield surveillance, border patrol, electronic warfare, etc.

With the recent advances in MEMS, artificial intelligent, digital communications, and sensing technology, UAVs have found more applications in many areas. Besides, computer vision techniques further enable intelligent UAV applications. UAV based autonomous target identification and tracking poses many technical challenges. First, as targets are moving within a changing background, it is important to develop real-time algorithms for varying-background video processing. Second, since the resolution of human facial images is usually low due to the far distances between the UAV and the target, how to identify the targets with high accuracy among the low-resolution facial images becomes a key problem in real-time aerial surveillance for law enforcement agencies. Third, target occlusion by the changing background becomes very difficult to solve since it is hard to determine whether the target leaves the camera Field of View (FoV) or it is occluded by other objects. Besides, algorithms without full utilization of object and GIS databases results in limited identification and tracking performance.

In [2], a set of algorithms are developed to detect and track objects with pre-known shapes for UAVs. The scheme has been proved useful for tracking indoor targets below certain heights. In [3], [4], [5], GIS databases are utilized for UAV navigation systems to generate tracking results in terms of global coordinates. The experimental results illustrate the improved tracking performance with the help of GIS databases. [6] gives an state-of-the-art consensus-based tracker which is capable of tracking deformable objects in real time. A real-time object tracking and identification system with a convolutional neural network running on FPGA is proposed in [7]. The system performance is robust to various environment and target conditions but only suitable for near-distance objects.

In this paper, we propose a framework of real-time UAV based objects tracking and recognition. Compared with most tracking system described above, we develop a more complete and robust system with both GIS environment database and neural network based object database. Besides, a set of consensus-based algorithms are utilized for target tracking, which show better performance than other algorithms in case of occlusion.

The contributions of this paper include
1) Design a UAV based target tracking and recognition system with an intelligent gimbal sub-system;
2) Develop a consensus-based visual tracking algorithm to achieve robust target tracking, as well as image mosaicking techniques to deal with moving background;
3) Utilize a GIS database to provide environmental and contextual information, and improve tracking accuracy;
4) Utilize a neural network based object database for fast feature matching, and high-accuracy target recognition.

This paper is organized as follows. Section II describes the whole system architecture and states problems; Section III presents the target tracking and recognition framework, including image mosaicking, consensus-based tracking algorithm, GIS database, and neural network based database; Section IV provides the results and discussions. Section V concludes the paper and outlines future work.

II. SYSTEM SETUP AND PROBLEM STATEMENT

A. UAV Based Target Tracking and Recognition

Fig. 1 illustrates the UAV based target tracking and recognition system, which contains the UAV system (flight control + gimbal sub-system) and the ground station. The real-time object tracking and recognition scheme is performed within
the UAV system, which implements a set of consensus-based algorithms for target tracking given video images, utilizes a database for feature matching, as well as various sensors for controlling UAV flights and gimbal motions. On the other side, the ground station receives the video stream, the position and orientation information of the quadrotor and gimbal, and fuses them with the GIS database to achieving the accurate target geo-locations.

Fig. 1. Target tracking and recognition system

Pixhawk (an open source autopilot module and software) is selected as UAVs flight control platform [8]. A gimbal system with a 4K HD camera is used to capture the target images. In addition, NVIDIA’s embedded computing board TX1 is selected as the image processing platform. TX1 is OpenCV enhanced and suitable for performing real-time target tracking and recognition tasks, which usually involve high computational complexity.

The development of a gimbal system follows the following steps:

1) **Image acquisition**: under the instruction mode, the UAV points the onboard camera onto the target and continuously acquire the target images, after receiving the instruction on the target initial position, texture, and size from the ground station; under the autonomous mode, the UAV automatically determines the targets of interest, track their movements, continuously acquires their images, and preprocesses the data, including white balance and de-noising.

2) **Video mosaicking**: to deal with the moving backgrounds in the UAV videos, consecutive images within a time window are aligned and stitched together into a panoramic image to get the static background image, such that the targets can be localized more accurately [9].

3) **Target recognition**: extract the feature points of the images, and match those features with the target models stored in the database; if the matching is successful, then continue the target tracking, otherwise re-localize the target from the images for subsequent processing.

4) **Target tracking**: implement the consensus-based algorithm to perform the online target detection, localization, tracking and prediction; estimate the motion and the shape of the target continuously.

5) **Gimbal and UAV control**: estimate the position and attitude of the UAV according to the predicted position of the target and data from the sensors, then control the motion and movement of the gimbal and the UAV to keep the target remain in the vicinity of the image center.

6) **Ground command center**: the UAV transfers the target position to the ground station, which is integrated with the GIS database to determine the global coordinates of the target and provide the target trajectory prediction using geographical information such as road or river; besides, the user can interact with the UAV through the ground command center such as select the target from the video monitor or change the UAV flight course.

**B. Gimbal System**

The development of a gimbal system follows the following steps:

1) Choose FPGA/DSP/GPU as the core processor, the flexibility of FPGA can provide great convenience for varieties of algorithms verification, as well as its scalability;

2) Use brushless PMSM motors to meet the requirement of SWaPC, and carbon fiber materials are chosen as the gimbals shell; use low-power devices to reduce the power consumption; use low-cost devices to reduce the cost;

3) Employ the modular design principle to achieve the system scalability and choose standard data interfaces such as Ethernet;

4) Perform the signal integrity design and robust design for the circuit to achieve high system reliability.
The hardware components of the gimbal system include the core processor, various sensors (IMU, barometer, GPS, camera), actuators (brushless motors) and wireless transceiver, as shown in Fig. 3.

The hardware components of the gimbal system include the core processor, various sensors (IMU, barometer, GPS, camera), actuators (brushless motors) and wireless transceiver, as shown in Fig. 3.

The goal of this study is to develop a quadrotor UAV based target tracking and recognition system that can

1) provide an intelligent gimbal sub-system for target tracking and recognition
2) track multiple objects in case of occlusion
3) use a GIS database to provide environmental and contextual information
4) utilize a neural network based database for fast feature matching

III. TARGET TRACKING AND RECOGNITION

A. Image Mosaicking

In order to align the sequential frames, there must be common features within the captured images that can be matched between the neighboring image frames. SIFT algorithm has been used to implement the feature matching [10], which is invariant to scale, orientation and affine distortion so that it can produce a more precise matching between successive video frames.

After finishing the feature matching between the successive frames, the extracted feature points will be used to compute a geometric transformation matrix for warping the images so that they can be aligned with the previous frames. Finally, consecutive aligned images are stitched together into a panoramic image. The process is described in Fig. 5.

B. Object Tracking

A robust and accurate object tracking scheme is central for other image processing stages, such as target recognition and target geo-localization. In an urban environment, targets are frequently occluded by other objects, such as trees, buildings, etc. Targets themselves may have certain 3D structure and are often deformed in 2D images, posing a serious challenge on the tracking algorithm. To address these challenges, we develop a robust visual tracking algorithm using consensus-based temporal learning.

1) The scheme uses the SIFT algorithm to extract the image feature points and their descriptors. After obtaining the target ROI from the operator, the algorithm will use features inside the ROI as initial target model.
2) For subsequent frames, target tracking is performed in an evaluate-and-update manner. The bounding box of the target is determined by taking a consensus vote on inliers, like in CMT [6]. Meanwhile, outliers in subsequent frames are also evaluated using a statistical Bayesian approach to identify outliers that are in fact part of the target, and add back to the model. False inliers rejected by the consensus are also removed from the model. To minimize false matches, feature points with multiple occurrences in the image is also removed from the matching process, as such points degrade the matching quality.
3) Target occlusion poses a challenge on target tracking in urban environments. Although our tracker is robust from partial occlusions, frequent total occlusion will
severely disrupt the functions of subsequent video processing and target-following UAV flight control. To tackle this problem, an Extended Kalman Filter (EKF)-based state observer is employed with the output of the tracking algorithm. This ensures a steady and continuous position and scale update by using predictions, which is essential for a stable control system.

C. Integration of UAV Images and GIS database

Since the position and orientation of UAVs' camera are not accurate enough due to the GPS deviation and measurement errors of IMU sensors, the UAVs usually can’t accurately localize targets in terms of global coordinates merely based on the target position within the images. In order to solve this problem, the method of matching the UAV camera images with a GIS database is proposed to help UAVs localize targets and predict target movements in terms of global coordinates and with a higher precision (Fig. 6).

The key is to extract features of HD images captured on the UAV and matching them with the images stored in the GIS database. Our scheme uses motion detection and estimation, and edge detection techniques to extract images features. Then estimate the zoom ratio and camera angles of the images in terms of the data format of the GIS database. Based on both the GIS data and targets position, we can predict the trajectories of targets, which will improve the range and accuracy of target tracking.

![Fig. 6. Target global coordinates and static background](image)

The ground station first receives a rough estimate of target geo-location information from the UAV based on the GPS and IMU readings, which can be used to estimate the camera height and angles, and then search it in the GIS database to determine the GIS region where the tracked object is present. After the determination of the target GIS region, the images transformation and registration are performed between UAV video streams and the GIS database images. After the successful image registration, the local coordinates of target positions can be converted into global coordinates. Furthermore, the trajectories of targets can be better predicted by utilizing the geographic structures of the environment, such as road, bridge, and river. The process is illustrated in Fig. 7 and Fig. 8. Given the huge size of the GIS database, it should be installed on the ground station. The resolution of UAV video streams is also optimized to achieve a trade-off between image matching performance and transmission costs.

The angle $\theta$, altitude $h$, and 3D position $(x_1, y_1, z_1)$ of the UAV can be obtained by the onboard IMU and GPS module.

\[
\begin{align*}
x &= x_1 \\
y &= y_1 - h \tan \theta \\
z &= z_1 - h
\end{align*}
\] (1)

Then the target location can be estimated in terms of those parameters by

![Fig. 7. Process of video scenario and GIS integration](image)

![Fig. 8. GIS as a service](image)

as shown in Fig. 9.

![Fig. 9. The principle of geopointing](image)

However, such an estimated position is not accurate due to the errors caused by sensors and GPS readings. As a result, it requires the GIS database to provide more geographic information of the target surroundings, which can improve target tracking accuracy, and provide target global coordinates, and predict the target trajectory with more geographical limits.

D. Integration of Neural Network Based Database

To improve the target recognition accuracy, which is degraded by target occlusion, target obscuration with long measurement distance, we propose a scheme by integrating both the GIS database and a neural network based target
database to achieve long-distance and high-precision targets recognition performance for UAVs. The system setup is illustrated in Fig. 10.

*First*, edge detection and target segmentation are used to extract target features before feature normalization. *Then*, a neural network is trained to establish the target database, where each target has multiple images in different scales and perspectives. During the training, the weights are adjusted to construct the feature models for targets correspondingly, which can maximize the feature distances among different targets, and reduce the feature distances for the same target, and update the target database, as shown in Fig. 11.

The target classifier is first trained on the ImageNet[11] dataset, then fine-tuned on a pedestrian/vehicle dataset captured by ourselves. The proprietary dataset not only includes various kinds of targets, but also takes into account a variety of target poses and surroundings. The structure of the recognition neural network includes five convolution layers and two fully connected layers. Meanwhile, a similar neural network with three convolution layers is trained in parallel, which trades for speed at the expense of accuracy.

The ground station and the UAV implements two sets of neural networks, which are synchronized at a certain rate. First, the ground station implements a target database based on artificial neural networks and trains them offline. Then, the trained weights are updated to the neural network based classifier in the UAV, which performs the target recognition in real time. Meanwhile, new target images acquired by the UAV will also be sent to the ground station to update the target database. The implementation of the neural network based target database is shown in Fig. 12.

Fig. 13 shows the target recognition probability of different vehicle images using convolution neural networks. One challenge for UAV based target recognition is that target images are often acquired from the top perspective, the number of feature points will be reduced, and the feature differences among targets will be reduced accordingly.

The tracking algorithm without EKF estimation is then tested on a real-world scenario car chase dataset. The dataset features many challenges including similar objects, partial/total occlusion and change of target pose, illustrated

IV. Experiment Results
in Fig. 14. We compared the performance of our algorithm against CMT using two criteria, feature points tracked and total frames of tracking loss. The result is shown in Fig. 15 and Fig. 16. Total occlusion by land obstacles (trees, bridges, etc.) is marked as red.

From Fig. 15, we can see that the number of feature points tracked of our algorithm is consistently far larger than that of CMT. The speed of recovery from tracking loss due to occlusions is also faster. The unstable number of features is due to the rapid update of the model, which still needs better tweaking in terms of learning rate.

Fig. 16 shows the total number of frames that do not have a good target lock. Lower cumulative tracking loss indicates that our algorithm has significantly lower probability of losing track in partial occlusions by nearby vehicles and road signs in urban environments, proving its robustness.

At the current stage, we are performing more experiments for human and vehicle tracking and recognition with the proposed UAV-based system. More results will be obtained in near future. With the help of GIS and neural network databases, the system performance has been much improved.

V. CONCLUSIONS

This paper presents a framework for UAV based targets tracking and recognition. The intelligent gimbal system is designed to provide a platform for objects tracking and recognition. Within our scheme, SIFT is used to extract the image feature points and descriptors, and a robust visual tracking algorithm using consensus-based temporal learning is applied to realize the tracking process. Besides, an Extended Kalman Filter (EKF)-based state observer is employed to solve the target occlusion problem. The GIS database has been integrated within the scheme to improve the target localization and prediction accuracy, and provide target global coordinates. The neural network target database has been integrated to improve target recognition performance. Experiment results demonstrate that the proposed algorithm can achieve better performances than CMT in various challenging situations.

ACKNOWLEDGMENT

This paper is partly supported by Southern University of Science and Technology Research Committee (No. FRG-SUSTC1501A-29 and FRG-SUSTC1501A-44). We are also grateful to the precious help from our lab engineers Yunbo Yang and Miaolin Hou.

REFERENCES