

Image-guided Flight Tracking Control System for Multirotor Aerial Vehicles

Ding-Horng Chen,¹ Yen-Ping Chen,¹ and Ming-Yuan Shieh^{2*}

¹Department of Computer Science and Information Engineering,
Southern Taiwan University of Science and Technology,

No. 1, Nan-Tai St., Yung-Kang Dist., Tainan City 71005, Taiwan (R.O.C.)

²Department of Electrical Engineering, Southern Taiwan University of Science and Technology,
No. 1, Nan-Tai St., Yung-Kang Dist., Tainan City 71005, Taiwan (R.O.C.)

(Received June 30, 2018; accepted September 11, 2018)

Keywords: multirotor vehicles, object tracking, aerial images, image-guided flight tracker

Remotely controlled multirotor aerial vehicles have gradually become popular in recent years. There are many applications in society; however, owing to possible damage to surrounding objects during operation, the reduction in operating thresholds is an important issue in their development. In this paper, an automated control system for a multirotor aerial vehicle capable of self-stabilizing during flight and image-guided flight tracking was developed on the basis of an aerial image object tracking algorithm for the real-time transmission of images to a monitor through a wireless network. When combined with the tracking-learning-detector (TLD) object tracking algorithm, aerial images were captured by the smartphone mounted on the aerial vehicle and then transmitted to the ground station operator via the wireless network. Afterward, flight control information was sent back to the aerial vehicle.

1. Introduction

Among the unmanned aerial vehicles (UAVs) developed during the past decades, those with a tuned propulsion system for multirotors have become popular in the field of civil aerial photography because of the reduction in hardware size, increase in flight control processor efficiency, and lower selling price. The scope of the UAV application is quite wide and includes film shooting, geography or ecological research, disaster area rescue, and surveillance activities. Nevertheless, its operation poses great risks. For example, in the field of surveillance, issues such as safety, cost loss, and pretraining not only impose a burden on operators but also overshadow the benefits of surveillance cameras. As a result, the video surveillance system is still mostly based on cameras installed in specific locations.

Traditional video surveillance systems are now quite popular in people's living environments. Various surveillance cameras can be seen in offices, road intersections, or in public areas of housing communities. This shows that modern people attach importance to the safety monitoring of their living environment. However, when the area to be monitored is large

*Corresponding author: e-mail: myshieh@stust.edu.tw
<https://doi.org/10.18494/SAM.2019.2041>

(such as campuses and factories) and the environmental variables are high, the manpower and equipment requirements for monitoring increase significantly. If an automated aerial patrol vehicle capable of path planning is used to carry out anomalous event detection in the air, it can spot moving objects to discriminate abnormal events and send warning messages in various ways. After evaluation, the personnel only needs to select the screen showing the abnormal moving object and command the UAV to follow the target and keep the guards constantly informed of the scene. This type of UAV with semi automated flight capability can greatly reduce monitoring cost and improve the efficiency of personnel use.

The algorithm most commonly used in security monitoring systems mainly examines the separation of the foreground and the background. The foreground is the moving object in the image, and the rest is the background. Such foreground recognition technology is roughly divided into two types: background subtraction⁽¹⁾ and background model training.⁽²⁾ Background subtraction uses a fixed, unaltered background as its baseline condition when comparing similarities across multiple frames and identifying moving or changing perspectives in the foreground. This method is not usually used for dynamic images. Since the background is changing, it is not easy to effectively separate the foreground by the background subtraction technique. The background model training mostly uses neural networks to classify the foreground and the background.

Most of the training methods use multiple frames to perform global operations. Areas with high local similarities are classified and the foreground is identified using time continuity and local movement. This method requires a considerable amount of time for computation. When used in a fixed surveillance camera, the effect of the foreground recognition technology is quite remarkable. However, when used in UAV aerial photography, geometric appearance of the target becomes difficult to distinguish owing to excessive and rapid changes of the background. Moreover, when an aircraft is tracking a target, it may also be accompanied by changes in perspective and distance. In addition, the target may also disappear from view from time to time. Therefore, the ability to redetect the missing target is an important performance indicator.

A good tracking algorithm must be able to solve problem, such as changes in ambient lighting, complex backgrounds, and object occlusion or disappearance, as well as achieve real-time calculations. At present, there are many object tracking algorithms that can be applied to real-time tracking,⁽³⁾ such as Mean Shift,^(4,5) Continuously Adaptive Mean-Shift (CamShift),⁽⁶⁾ Tracking-Learning-Detector (TLD),^(7,8) Tracking-Learning-Matching of keypoints (TLM),^(9,10) Tracking-based Moving Object Detection (TBMOD),⁽¹¹⁾ Template Tracking,⁽¹²⁾ Particle Filters,^(13,14) and Optical Flow.⁽¹⁵⁾ Among these algorithms listed above, Mean Shift is a direct algorithm based on the difference in Gaussian (DOG) mean-shift kernel that is efficient for tracking 2D blobs through scale space.⁽⁵⁾ CamShift is an algorithm modified from Mean Shift and adopts updating searching windows rather than fixed as in Mean Shift. For facing motion tracking, CamShift is more robust against color interference and sudden illumination variation.⁽⁶⁾ Kalal *et al.* proposed a novel algorithm, TLD, for long-term tracking of a human face in unconstrained videos.⁽⁷⁾ TLD compares the results from two tasks: object tracking and detection, and then obtains a better tracking detection on the basis of a learning scheme. The kernel of TLD is the P-N learning algorithm, in which the P expert only identifies

false negatives and the N expert identifies false positives. Both types of expert make errors themselves but their independence enables mutual compensation of their errors.⁽⁷⁾

The main purpose of this study is to create a multirotor aerial vehicle with a semiautomated flight tracking capability that can follow any target without putting or installing any sensors on the object before tracking. The UAV can take shots through its own camera and track the target during flight. The tracked object may be a human being, a car, or any moving entity. Before tracking, the target is selected by a human controller in the aerial image displayed by the ground station software; then, follow-up is done automatically during flight. Furthermore, in this work, we designed and verified that the image taken by the aircraft can be instantly sent to the ground station via a wireless network. Hence, aerial images were transmitted to the ground station through digital signals for back-end operations, thus avoiding the traditional first-person view (FPV) problem of noise in radio image transmission. This also reduced the number of equipment attached to the UAV and the complexity of its circuitry.

2. System Architecture

2.1 Multirotor aerial vehicle

The multirotor aerial vehicle used in this article has a four-axis X-shaped configuration and a wheelbase of 770 mm. It has 2 pairs of 15 in. carbon-fiber-tipped propellers, a 4010–310 KV motor, and a PLATINUM PRO 30A electronic transmission by Hobbywing. The flight control board is an ArduPilot PIXHAWK, the GPS accuracy is 8 m, and the core is a Tarot carbon fiber rack of No. 650 equipped with a 6S 5800 mAh lithium polymer battery. When the safe return flight voltage is set to 22.6 V, autohover is about 12 min. The mounted smartphone responsible for taking and transmitting images is a Sony Lt29i. The appearance of the UAV is shown in Fig. 1.



Fig. 1. (Color online) Exterior view of the UAV (excluding paddle).

2.2 Lead zirconate titanate (PTZ) and image quality

The PTZ used in this work was designed by our group, and assembled by printing the framework with a three-dimensional (3D) printer (as shown in Fig. 2). The framework was made of polylactic acid (PLA) plastic and carbon fiber. The pitch axis was controlled through a brushless motor with an outside diameter of about 27 mm. The PTZ control board was the Simple BGC, as shown in Fig. 3.

Twelve shock-absorbing balls were mounted at the connection between the top of the PTZ and the aerial vehicle so as to reduce high-frequency oscillation caused by the impact of the UAV rotors. In addition, an inertial measurement unit (IMU) is attached to the camera platform at the bottom of the PTZ to measure the camera's view angle.

2.3 Flight mode

The multirotor aerial vehicle in this work possesses the following flight modes:

2.3.1 Self-stabilizing mode

In this mode, the UAV measures its altitude of flight with a gyroscope and an accelerometer. It keeps the level of the aerial vehicle perpendicular to the direction of gravity to make the UAV as stable as possible so that when the operator changes the pitch or the roll angles, the aircraft automatically returns to a horizontal position to prevent overturning due to operational errors.

2.3.2 Fixed height mode

Under the self-stabilizing mode, the fixed height mode uses a barometer to maintain the height of the multirotor aerial vehicle. This is mainly used to control the motion of the UAV and prevent it from falling owing to a lack of lift caused by the tilting of the body. With this mode, the operator does not need to use the throttle frequently to maintain altitude.

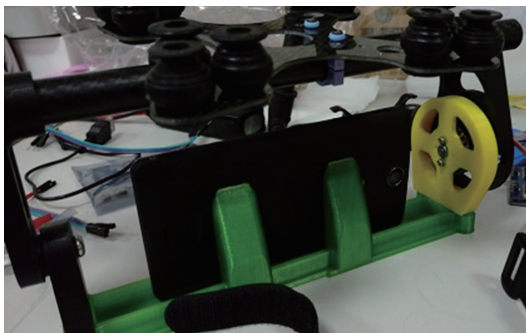


Fig. 2. (Color online) Appearance of the self-made PTZ.

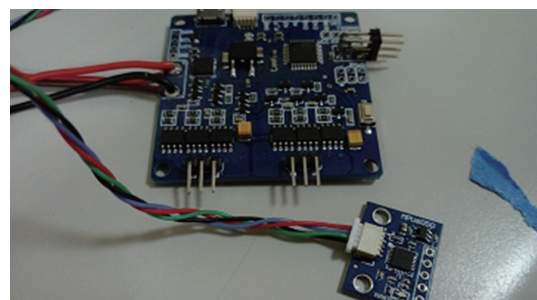


Fig. 3. (Color online) Simple BGC control board for the PTZ and its IMU.

2.3.3 Hover mode

A GPS is used to position the aircraft. When the UAV is in the hover mode, the heading, position (via GPS), and altitude of the aircraft are fixed by its sensors. When the UAV moves owing to external forces, it will automatically return to the original hover position after the flight control system detects that the GPS position has been changed.

2.3.4 Automatic mode

The controller connects the ground station with the multirotor aerial vehicle and then uses the ground station software to draw several waypoints. After the multirotor aircraft has taken off, it can be flown automatically to these waypoints for patrol missions without manual manipulation through a GPS, an electronic compass, or a barometer.

2.3.5 Automatic return mode

The automatic return mode is also called the Return to Launch (RTL) mode. When switched to this mode, the multirotor aircraft first climbs to the preset return altitude and then travels at this altitude to the starting point during its take-off (also called Home Point). Upon returning to the starting point, it slowly descends to the ground to land.

2.3.6 Image-guided follow mode

The image-guided follow mode is the flight mode developed in this study. Its principle is to send images captured by the camera mounted on the multirotor aircraft to the ground station so that the controller can select the object of interest for tracking. Afterward, the object tracking algorithm is used to calculate the position of the object in each image frame continuously sent by the UAV. The positions of the aerial vehicle and the tracked object calculated from data such as the position of the tracked object, the camera shooting angle, and the height of the aerial vehicle give the ground station the necessary information in order to decide whether to let the aerial vehicle advance, retreat, or turn accordingly so as to maintain the distance between the aircraft and the tracked object and achieve the purpose of following the target.

This function requires the UVA and its PTZ camera to be self-stabilizing. The PTZ camera must be able to correct or minimize blurred vision during motion of the aircraft. The camera's shooting angle requires the PTZ to send the gyroscope parameters. The height of the image capture point is detected by the altimeter on the aerial vehicle. The main image tracking operation is performed by the ground station. This mode does not need to track the target or send a GPS signal. It can perform tracking operations through the images generated by the UAV.

3. Object Tracking Algorithm

The object tracking algorithm in this study uses the TLD framework, and its architecture is shown in Fig. 4. This decomposes the single-target, long-term tracking into three subtasks, namely tracking, learning, and detection. These three subtasks help one another during the task. When the target is detected or is being tracked, the learning mechanism instantly adds new target samples to the classifier to improve the entire online model, making tracking and detection more accurate.

3.1 Tracking algorithm

The tracking algorithm used by the TLD is based on the Lucas–Kanade optical flow method, also known as the median optical flow method, and its flowchart is shown in Fig. 5. This method calculates the optical flow with two frames of difference and estimates the motion of all pixels within the image in the time difference between t and $t + \Delta t$. During the tracking of the object, the 10×10 feature points uniformly distributed in the frame are first set in the bounding box to indicate the location of the previous target.

Then, the target is tracked by the Lucas–Kanade optical flow method in order to give a forecast of the direction of the feature point motion in the previous and current frames. The Forward–Backward Error proposed by the TLD author was used to check the mechanism and distinguish the feature points that have tracking success and tracking failure. Feature points with 50% tracking effectiveness were utilized, and the other 50% were used to predict the target position of the current frame.

Figure 6 illustrates the calculation of the Forward–Backward Error, which is effective in tracking the result of the Lucas–Kanade optical flow method and the calculation of each feature point during tracking. To calculate the effectiveness of tracking X_t of image I_t to X_{t+k} of image I_{t+k} , tracking using the Lucas–Kanade optical flow method must first be done on X_t of frame I_t . At this moment, X_{t+k} of I_{t+k} is then backtracked to the previous frame by the Lucas–Kanade

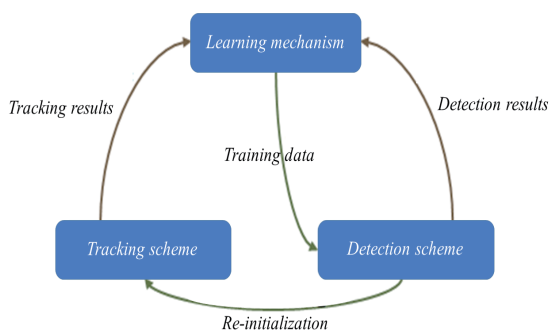


Fig. 4. (Color online) TLD framework.

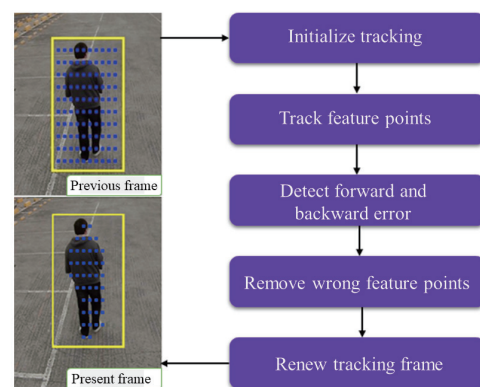


Fig. 5. (Color online) Flowchart of the tracking algorithm.

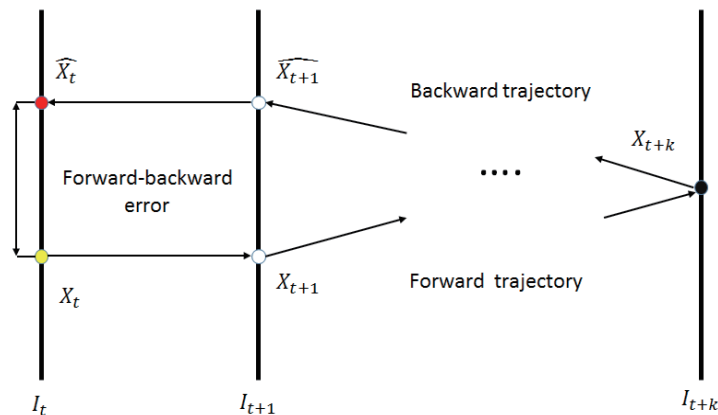


Fig. 6. (Color online) Forward–Backward Error.

optical flow method until \hat{X}_t of I_t is tracked. Then, the Euclidean distance between X_t and \hat{X}_t is calculated. This distance is the tracking error, and the larger this value is, the lower is the validity of the trace. The TLD trace algorithm only performs the Forward–Backward Error check mechanism between the current and previous frames to verify tracking validity. This means that only the target positions in the two images I_{t+k} and I_{t+k-1} are used for decision making.

After the Lucas–Kanade optical flow method, the Median-Flow algorithm sorts the X and Y components of the remaining pixel's motion vectors for tracking after the Forward–Backward Error filtered out the pixels that have failed to track. The position of the current image's new tracking frame can be identified by taking the median value of each motion vector component when the previous tracking frame moved to the current image.

3.2 Detection algorithm

The detection algorithm used in this study contributes to the accuracy of the tracking algorithm by providing error correction information. When the system tracking algorithm is inaccurate, it offers a chance for correction and stops the error from increasing, thus preventing the tracking of the wrong object. In TLD detection, the first step is the scanning of the image in the sliding window. Specifically, the image is scanned according to the specific ratio set in the bounding box after the initialization of the target. The displacement is 10% of the image size during scanning. All scanned patches are sent to the cascaded classifier for mathematical operations. Only after passing through all classifiers will it be treated as foreground.

Figure 7 shows the scanning process. The image block is sent to the cascade classifier for detection. The detection mechanism consists of three parts in series: patch variance classifier, ensemble classifier, and nearest neighbor classifier. Each part of the input is imported from the output of the previous part.

The following sections describe the three parts of the cascade classifier.

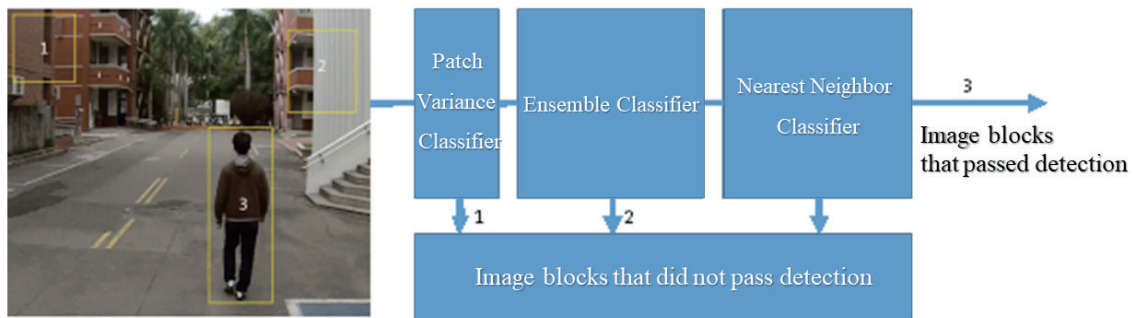


Fig. 7. (Color online) Flowchart of the detection algorithm.

3.2.1 Patch variance classifier

The first part of the classifier is the patch variance classifier. It calculates the variation of the image block in the current and previous frames. After calculating the selected image block of the current frame, the 50% of the number with less variation is used as the input for the ensemble classifier. The other 50% of the value with higher variance was discarded. At this stage, most of the background can be filtered.

3.2.2 Ensemble classifier

The second part of the classifier is the ensemble classifier, similar to a simplified random forest classifier, consisting of a number of basic classifiers, each of which represents a set of pixel comparisons. This instrument uses 13 sets of pixel comparisons to describe the characteristics of an image block for each basic classifier. After passing through the ensemble classifier, the image block obtains a set of posterior probabilities. These probabilities are used to distinguish the image blocks into positive and negative samples.

Its difference from the random forest lies in the fact that each decision tree in the random forests has a different formula for each node in each layer. However, in the TLD ensemble classifier, the formula for the nodes of the decision tree in the same layer is the same. To obtain the feature of its image block, the BRIEF feature is used. Each node of each decision tree randomly selects two pixels with the same X , or the same Y coordinates in the image block to compare the brightness values. When pixel A is larger than pixel B, the feature is 1; otherwise, it is 0. This is the composition of the 13-bit length integer code.

A large probability for error arises when only a few sets of pixel comparisons were used as image features. However, combining random forests allows a large number of decision trees to classify and vote on new samples. The decision tree with the largest number of votes is used for the classification of new samples. This makes the device capable of learning the new appearance of a target during classification and greatly improves the accuracy of identification.

3.2.3 Nearest neighbor classifier

The third part of the classifier is the nearest neighbor classifier. After the image block selected by the sliding frame passes through the patch variance classifier and the ensemble classifier, dozens of image blocks remain in the nearest neighbor classifier. The nearest neighbor classifier compares these remaining image blocks with online models for similarity. If the similarity is higher than the preset threshold (usually preset to 0.6), it is considered as a positive sample. In contrast, if the similarity is lower than the threshold, this means that the classification is wrong, and it is regarded as a negative sample.

3.2.4 P–N learning mechanism

The P–N learning mechanism as shown in Fig. 8 is a method of learning that uses structured unmarked data. The structural characteristics of the data are utilized to assign positive and negative labels to unmarked data sets. The P expert is capable of finding the missing information in the detector. The N expert identifies the part of the data that has been detected by mistake.

For example, when the image marker near the object trajectory is positive and the marker far away from the trajectory is negative, the structure is guaranteed to be temporal. When only one positive sample is selected from all positive ones, the structure is guaranteed to be spatial. The positive and negative constraints can be used simultaneously, and the combination of these two can be used to correct classifier error. These constraints are used to process an entire set of unlabeled datasets, and at the same time, learn from them; therefore, it is possible to obtain effects that are different from when the classifier is used on only a single sample.

P–N learning consists of four parts: (1) a classifier awaiting learning; (2) a training sample set: some classified samples; (3) supervised learning: method of training a classifier using a training sample set; (4) P–N experts: function expressions for generating positive and negative samples during the learning process. The relationship among these four parts is shown in Fig. 8.

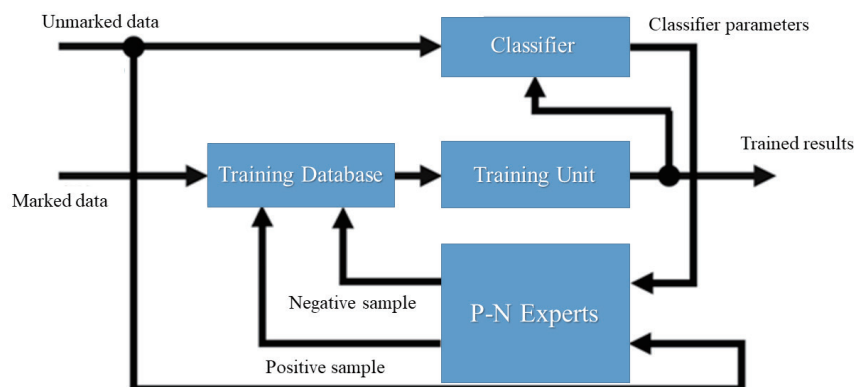


Fig. 8. (Color online) P–N learning mechanism.

Some of the marked samples and a large number of the structured unmarked samples comply with the following learning strategies:

1. Use marked samples to train the classifier and adjust the corresponding predetermined constraints accordingly.
2. Use the resulting classifier to classify the unmarked data and seek out samples that have classifier results inconsistent with structural constraints.
3. Correct their markers and add obtained data to the training set in order to retrain the classifier.

In this paper, the P–N learning mode was applied to the TLD algorithm. By using the constitutive property of the data, target positioning and target model were updated frame by frame. Here, the tracked target was treated as a single marked sample and the video as unmarked data. The tracker was used to predict the target position. Furthermore, the positive sample near the current position and the negative sample far from the target position were used to update the model. This strategy enables the tracker to adapt to the appearance and background of the new target. However, once the tracker makes an error, the learning mode fails. This problem can be solved by simultaneously training a classifier capable of generating positive samples and distinguishing negative samples during tracking.

4. Flight Tracking Algorithm

According to the operation result of the object tracking algorithm, it is possible to know the pixel coordinates of the tracked object in the aerial image. By correcting the shooting angle of the camera, we were able to display the tracked object at the center of the aerial image, that is, the yaw and pitch angles of the aerial vehicle or the PTZ were altered to determine the distance between the UAV and the tracked object. Afterward, the distance between the aircraft and the tracked object was controlled automatically.

4.1 Vision tracking and control

To make the multirotor UAV follow the target successfully during flight, the flight tracking algorithm used in this study kept the position of the tracked object close to the center in the aerial view (image); otherwise, the calculation error for the positions of the multirotor aerial vehicle and the tracked object would be too large owing to lens deformation error. When this happens, subsequent flight control would be affected.

When the multirotor UAV is located at a certain distance from the tracked object (for example, more than 20 m apart), according to the structure characteristics of the continuous image, it is assumed that under normal conditions, the object will not move out of the viewable range within a very short image frame. Therefore, when the multirotor aircraft constantly modifies its field of vision and makes sure that the target is in the center of the field of vision, continuous tracking of the object is possible.

As shown in Fig. 9, when the tracked object was located on the left or right side of the center of the field of vision, to maintain the object's position at the center, the multirotor aircraft

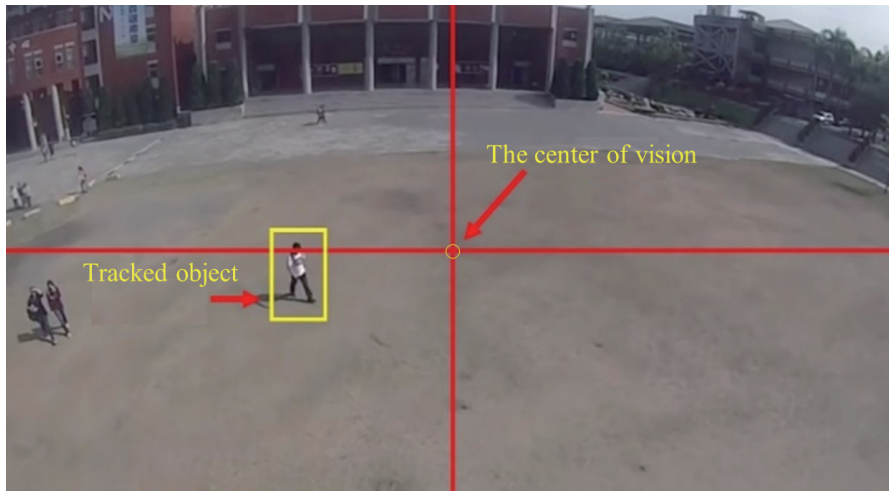


Fig. 9. (Color online) Actual situation showing the deviation of the tracked object from the center of vision.

performed yaw correction according to the difference between the current target and the center of the field of vision. When the tracked object returned to the effective tracking area in the center of the field of vision, the relative distance to the tracking object could be calculated, and the information regarding the position of the tracked object could be obtained again.

When there is a large deviation in the vertical distance of the tracked object and the center of the field of vision, control is not made by the multi rotor UAV but by the PTZ mounted on it. By adjusting the pitch of the PTZ, the image is set to be located near the center of the field of vision.

The effective tracking range is determined by the degree of image distortion caused by the camera lens. To maintain the characteristics of the tracked object and prevent image distortion from affecting the tracking results, the effective tracking area must not be too large. If the effective tracking area is set to be too small, when the UAV corrects the field of vision, there will be a tendency to make more than necessary corrections. This causes the aircraft to unceasingly adjust the field vision, thus preventing it from tracking the object smoothly. Therefore, the effective tracking range must be set according to the specifications of different multirotor aerial vehicles and their PTZs.

4.2 Flight tracking

After tracking the object in the image and aligning it in the center of the field of vision, the ground station sends back flight control instructions to the aerial vehicle to allow it to follow the target. In order to monitor the movement of the object, the relative distance between the UAV and the tracked object must be obtained. When the tracked object moves away from the multirotor UAV, the size shown in the image gradually decreases. However, because of the possible deformation of the tracked object and the different shooting surface, determining the change in distance by tracking the target's size information in the image is unreliable.

Therefore, together with the camera, gyroscopes, accelerometers, and various sensors on the flight control board of the PTZ are used to measure the distance.

Its difference from the random forest lies in the fact that each decision tree in the random forests has a different formula for each node in each layer. However, in the TLD ensemble classifier, the formula for the nodes of the decision tree in the same layer is the same. To obtain the feature of its image block, the BRIEF feature is used. Each node of each decision tree randomly selects two pixels with the same X or the same Y coordinates in the image block to compare the brightness values. When pixel A is larger than pixel B, the feature is 1; otherwise, it is 0. This is the composition of the 13-bit length integer code.

The position of the UAV can be obtained through GPS. The calculation method for the relative distance of the tracked object is shown in Fig. 10. H , the height of the UAV from the ground, can be measured by the air-pressure sensor of the UAV. After the tracking system aligns the center of vision with the bottom of the tracking object bounding box, the gyroscope acquires the vehicle's top view angle. After the angle of the camera PTZ, θ , is obtained, the straight line distance S of the UAV from the tracked object, as well as its horizontal distance D , can be calculated using Eqs. (1) and (2).

$$S = H \times \sec \theta \quad (1)$$

$$D = H \times \tan \theta \quad (2)$$

After obtaining the distance between the multirotor UAV and the tracked object, the ground station transmits the flight control command so as to adjust the position of the aircraft and recalculate the relative distance in each frame. When the distance changes again, the target position of the UAV is updated in order to maintain its distance from the tracked object.

In addition, when the relative distance is obtained, since the position of the multirotor UVA is known through the GPS, the GPS position of the tracked object can be calculated by adding

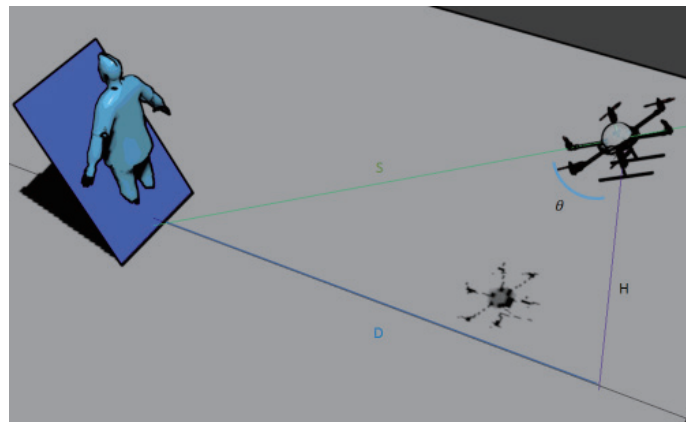


Fig. 10. (Color online) Schematic diagram of distance calculation.

the current vector coordinates of the UAV and the GPS coordinates. It is not necessary to attach a sensor to the target in advance. The object can be tracked by image guidance.

From the perspective of the image tracking algorithm, maintaining the distance in the tracking method to control the distance between the multirotor UAV and the tracked object makes the size of the tracked object in the image measurable to a certain extent. This is beneficial for the calculation of size reduction and the tracking success of the tracking algorithm. If only the tracking route is planned and the distance between the multirotor UAV and the tracked object is not maintained, for example, when the object being tracked is too far away from the multirotor UAV during activity, the amount of tracking calculations can be greatly increased to make sure that all possible dimensions of the target for good tracking are considered.

This, in turn, affects the tracking response time. Another case is when the field of view of the multirotor UAV is increased by increasing the height to reduce the motion time of the UAV. Although this ensures that the tracked object will not escape from the field of view, it changes the viewing angle and the size of the object in the image. This situation may also lead to tracking failures. Therefore, in this work, we maintained the distance between the UAV and the target when tracking flights, rather than focusing on planning a more efficient or more power-saving flight path.

5. Experimental Results

A self-assembled four-axis aerial vehicle was used in this study for flight experiments that monitor different targets to test the effectiveness and reliability of the system. This system was integrated into the open source ground station software Mission Planner. A smartphone and its PTZ were mounted on the UAV. Aerial images were transmitted to the ground station via the wireless network on the mobile phone. After receiving the image, the ground station analyzed it and calculated the position of the tracked object. These data were sent to the flight control command and the multirotor UAV to enable the aircraft to track the target during its flight.

In the first experiment, the tracking time was about one minute. The location was in the campus open space. The fps of image transmission was about 7, and one image frame was tested every 400 ms as a sample, as shown in Fig. 11. The flight process was successful in tracking the object; therefore, no nonrelevant image was recorded. The tracked object was always displayed on the screen and not blocked. Furthermore, speed was similar throughout the process. The experimental results showed that among the 150 entries, 144 were correct and 6 were incorrect; thus, the accuracy was 96%.

In the second experiment, we aimed to track the person wearing red trousers among other people. The tracking time was about one minute and thirty seconds. The location was also in the campus open space. There were 90 frames sampled per second. Figure 12 illustrates the flight processes, in which most of the tracking results were successful except in the case of more than one person. In the experiment, the tracked person was close to other people many times. Although the detection was disturbed during the tracking process, it could still frame the tracked person when people separated.

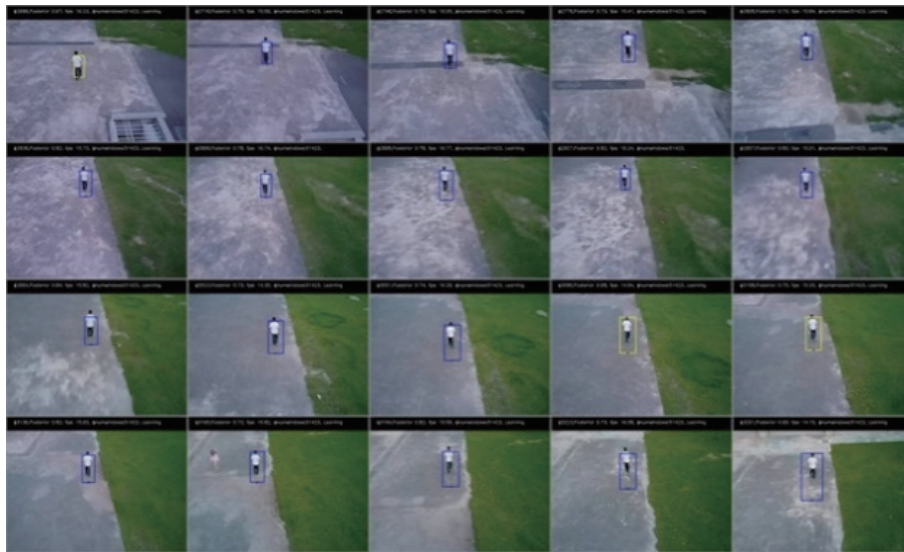


Fig. 11. (Color online) Experimental results of flight tracking using 4G LTE (1st experiment).



Fig. 12. (Color online) Experimental results of flight tracking with disturbance using 4G LTE (2nd experiment).

This demonstrates that the TLD algorithm can still track when the object is partially blocked by other people, in which the disturbed object may be regarded as a temporary deformation of the tracking target. When coupled people separate, the TLD algorithm can still successfully distinguish between the target and the disturber. The experimental results showed that among the 90 entries, 73 were correct and 17 were incorrect; thus, the accuracy was 81.11%.

6. Conclusions

In this study, a multirotor UAV with self-stabilizing flight control capability was modified and a smartphone was installed on the self-assembled PTZ. An image transmission APP was also developed for the smart phone, and object tracking algorithms were studied on the basis of aerial images. Then, a multirotor UAV automatic tracking system capable of image-guided tracking during flight was constructed. This facilitated the real-time transmission of images from the multirotor UAV to the monitor via a wireless network. The monitor was able to select targets through this image, and the UAV was able to automatically follow the object when it moved. From the experimental results, the object tracking algorithm TLD used in the experiments retained a good tracking ability for aerial image perspectives. Furthermore, it possessed a definite tracking ability in environments with large changes in light.

Acknowledgments

The authors would like to express their sincere thanks to the reviewers for their invaluable comments and suggestions. This study was supported by the Ministry of Science and Technology, Taiwan, under Grant numbers MOST102-2218-E-281-004, MOST103-2221-E-281-024, and MOST107-2221-E-218-026-MY2.

References

- 1 D. Gutchess, M. Trajkovics, E. Cohen-Solal, D. Lyons, and A. K. Jain: Proc. 8th IEEE Int. Conf. Computer Vision. ICCV 2001 (IEEE, 2001) 733–740.
- 2 S. H. Oh, S. Javed, and S. K. Jung: Proc. 11th Int. Conf. Frontiers of Information Technology (IEEE, 2013) 13–18.
- 3 J. Shi and C. Tomasi: Proc. IEEE Computer Society Conf. Computer Vision and Pattern Recognition (IEEE, 1994) 593–600.
- 4 Y. Cheng: IEEE Trans. Pattern Anal. Mach. Intell. **17** (1995) 790.
- 5 D. Comaniciu and P. Meer: IEEE Trans. Pattern Anal. Mach. Intell. **24** (2002) 603.
- 6 T. Zou, X. Tang, and B. Song: Proc. 2011 Int. Conf. Multimedia Information Networking and Security (IEEE, 2011) 11–15.
- 7 Z. Kalal, K. Mikolajczyk, and J. Matas: IEEE Trans. Pattern Anal. Mach. Intell. **34** (2012) 1409.
- 8 A. J. Sinisterra, M. R. Dhanak, and K. Von Ellenrieder: Ocean Eng. **133** (2017) 197.
- 9 G. Nebehay and R. Pflugfelder: Proc. 2013 Int. Conf. Distributed Smart Cameras (IEEE, 2013) 1–6.
- 10 T. Wang and H. Ling: IEEE Trans. Pattern Anal. Mach. Intell. **40** (2018) 1494.
- 11 H. Shen, S. Li, J. Zhang, and H. Chang: Proc. 2013 IEEE Int. Conf. Image Processing (IEEE, 2013) 3093–3097.
- 12 I. Matthews, T. Ishikawa, and S. Baker: IEEE Trans. Pattern Anal. Mach. Intell. **26** (2004) 810.
- 13 M. S. Arulampalam, S. Maskell, N. Gordon, and T. Clapp: IEEE Trans. Signal Process. **50** (2002) 174.
- 14 F. Gustafsson, F. Gunnarsson, N. Bergman, U. Forssell, J. Jansson, R. Karlsson, and P.-J. Nordlund: IEEE Trans. Signal Process. **50** (2002) 425.
- 15 J. L. Barron, D. J. Fleet, and S. S. Beauchemin: Int. J. Comput. Vision **12** (1994) 43.

About the Authors



Ding-Horng Chen received his B.S. degree from the Department of Mathematics, and M.S. and Ph.D. degrees from the Department of Computer Science and Information Engineering in 1989, 1993, and 2001 respectively, all in National Cheng Kung University, Tainan, Taiwan. In 1996, he joined the Department of Computer Science and Information Engineering, Southern Taiwan University of Science and Technology, Tainan, Taiwan, where he is currently an associate professor. He is an IEEE member and served as the vice chairman of the Signal Processing Society, Tainan Chapter in 2007 and 2008.



Yen-Ping Chen received his B.S. degree in 2012 and his M.S. degree in 2015 from Southern Taiwan University of Science and Technology, Taiwan. Since 2015, he has been an image-processing engineer at Innolux Corporation, Taiwan. His research interests are in digital image processing.



Ming-Yuan Shieh received his B.S. degree in 1992 and his Ph.D. degree in 1998 from National Cheng-Kung University, Taiwan. From 2000 to 2008, he was an assistant professor at Southern Taiwan University of Science and Technology, Taiwan. Since 2008, he has been an associate professor at Southern Taiwan University of Science and Technology, Taiwan. His research interests are in intelligent control, digital image processing, and humanoid robotics.