IOP Conf. Series: Materials Science and Engineering **719** (2020) 012077 doi:10.1088/1757-899X/719/1/012077

Research on UAV Point Landing based on Visual Navigation

Boyong He, Chen Meng and Liaoni Wu

School of Aerospace Engineering, Xiamen University, Xiamen City, Fujian Province, 361005, China

Email: heboyong@qq.com

Abstract. In this paper, the point landing technology based on visual navigation is studied on a rotor UAV flight platform. By contrasting different flight datum, several common image recognition and tracking algorithms are compared such as blob analysis, CamShift and KCF. Thus, automatic and manual methods are respectively designed for identifying ground targets, with rapidity, accuracy and robustness. After testing on UAV onboard equipment, the point landing technology based on visual navigation is finally achieved.

1. Introduction

With the rapid development of UAV industry, intelligent UVA has been put on the agenda. An important aspect of UAV intelligence is the development of visual navigation, so researches on UAV visual navigation have been gradually emphasized. UAV landing technology is one of the key technologies of UAV whose visual system can be used as a complement to the existing sensor system. Therefore, the research of vision based on UAV landing technology has a positive significance[1-3]. This paper mainly introduces the principle of image recognition and target tracking and the implement in opencv(Open Source Computer Vision Library)[4]. Rapidity, accuracy and robustness of visual programs are compared among four groups flight datum of different light and background. After testing on UAV onboard equipment, the point landing technology is eventually realized.

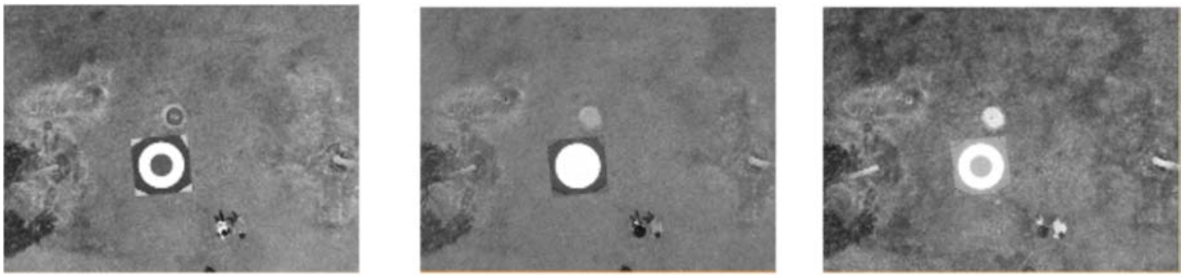
2. Navigation Approach

2.1. The Design of Image Pre-Processing

The main purpose of image pre-processing is to eliminate irrelevant information in the image, recover useful information, enhance the detectability of related information and simplify the data to the utmost, thereby improving the reliability of feature extraction, image segmentation, matching and recognition. Image pre-processing covers the following approaches: color-space conversion, gray-scale image transformation, smoothing, thresholding, region features and so on. Different image pre-processing is usually required in different computer vision fields. The following will introduce the image pre-processing algorithms involved in this paper.

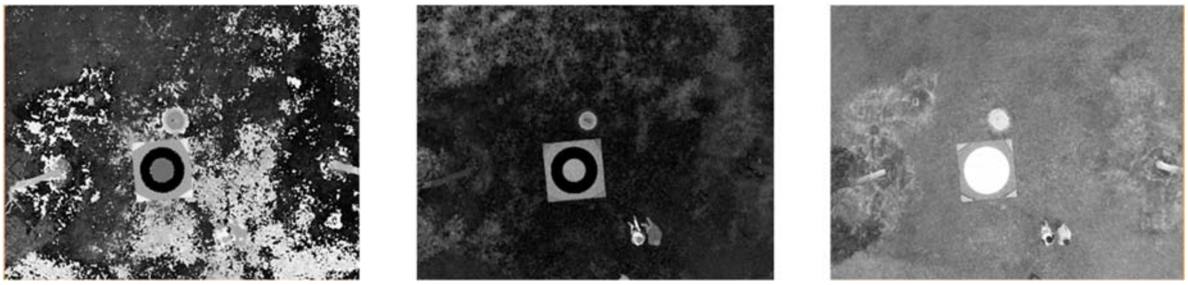
In image processing, the most common color space is the RGB model, as shown in Fig.1. However, The HSV color space is insensitive to light conditions and can reflect color changes better, as shown in Fig.2. Therefore, the HSV color space can be employed to perform color segmentation of a specified color.

Second, it is necessary to transform the gray-scale. Performing a gray-scale transformation on a color image can significantly reduce the amount of image processing calculation and greatly increase the speed of image processing. According to theoretical calculations, it can increase the calculation speed by two times. Binarization is to convert the gray value of the point on the image into 0 or 255. The collection property of the image is only related to the position of the pixel value of 0 or 255, and no longer involves the multi-level value of the pixel, so that the processing becomes simple, and the data processing and compression amount small. Finally, we use the method of erosion and dilation to eliminate noise and divide the individual elements[5]. Fig.3 depicts the results of image pre-processing.



(a) (b) (c)

Figure 1. RGB single channel image



(a) (b) (c)

Figure 2. HSV single channel image

(a) (b) (c)

Figure 3. Image pre-processing. (a): gray scale transformation, (b): binarization, (c): erosion and dilation

2.2. The Design of Image Recognition

There are many algorithms for image recognition, which are generally divided into two categories. One of them bases on gray matching, and the other bases on feature matching. Gray matching is mainly used in space of one-dimensional or two-dimensional sliding template image matching, and the differences among them mainly lie in the template and the selection of related guidelines. Generally, this kind of method has a high efficiency, but the large amount of calculation and slow speed; Feature-based matching extracts features such as points, lines and regions from the original image as matching resources, and then it is performed. Generally, matching speed is faster, but matching accuracy is not high. Comparing common image recognition algorithms, this paper choose the recognition based on color and blob analysis to test[6].

After color-space conversion, there are six color channels, R(red), G(green), B(blue), H(hue), S(saturation), V(value), respectively. The main of recognition based on color is selecting one color

IOP Conf. Series: Materials Science and Engineering **719** (2020) 012077 doi:10.1088/1757-899X/719/1/012077

channel with enough contrast. The result of image pre-processing is a binary image, and then use find-Contours, a function in opencv, to extracting outlines. These outlines are screened by shapes and areas. At last the target is identified by the outline. Fig.4 depicts the results of image recognition based on color .

The key of blob detection is the analysis of blob features[7]. In computer vision, Blob refers to a connected area with similar colors, textures, and other features in an image. In practice, change the threshold value at first, and run binarization continuously with threshold step. Next, use find-Contours to Extract connected domains and calculate the center of each connected domain. Some close points are classified as a group corresponding to a blob feature. Then, estimate the blob feature and radius of the target. Eventually, run Simple-Blob-Detector, a function in opencv, to detect the target, as shown in Fig.5.

Figure 4. Image recognition based on colour

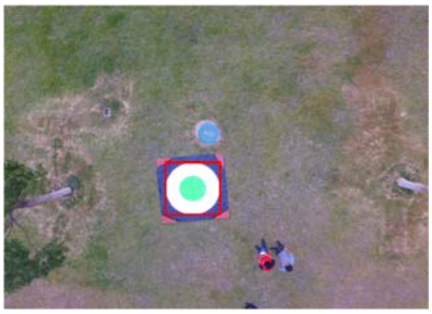


Figure 5. Image recognition with blob detect

2.3. The Design of Target Tracking

The target recognition algorithm needs to process and compare the whole image, while the target tracking algorithm only needs to recognize and calculate the initialized target, which is relatively faster and more accurate. According to the target tracking method, the tracking algorithm can be divided into two categories: generation class method and discriminant method. The generation method refers to the modeling of the target area in the current frame, and the next frame is looking for the most similar area of the model, that is, the prediction position, like Kalman-filter, particle filter, mean-shift method; The discriminant method, also called tracking detection (tracking-by-detection), refers to training classifier based on the current frame target area as the positive sample and the background area as the negative sample. The next frame uses a trained classifier to find the best region, and the typical method is TLD (tracking-learn-detect) algorithm. In recent years, the correlation filtering, such as the KCF/DCF algorithm, is also widely used. After the comparison of real time and accuracy, finally, three kinds of tracking methods, CamShift and KCF are selected to test in this paper[8-9].

CamShift(Continuously Adaptive Mean-Shift)is an improved algorithm for meanshift. Meanshift is a non-parametric method for density function gradient estimation, which locates the target by iterative optimization to find the extremum of the probability distribution[10-11]. Based on opencv, we calculate the H-component histogram of target in HSV color space at first, and use the target's color histogram model to convert the image into a color probability distribution, and initialize the size and position of a search window. Next, call the command of meanshift in opencv, and it can automatically track and adjust the center and size of the target window eventually. CamShift can effectively solve the problem of target deformation and occlusion. Furthermore, it has low requirements for system resource, fast processing speed and is insensitive to the changes in the shape and size of the target. Especially, it can automatically adjust the size of the search box and apply under a simple background. While the background is extremely complicated or there are more interferences in similar target colors, the tracking will fail[12]. Fig.6 depicts the results of CamShift.

As mentioned, KCF (Kernelized Correlation Filter) is a tracking method belonging to tracking-by-detection[13]. In opencv, it trains a target detector during the tracking process. The target detector is employed to judge whether the next frame predicted position is the target or no. Then, the latest target detector is employed to update the training set and update the target detector. When train the target detector, the target region is generally selected as a positive sample, and the target surrounding region is a negative sample. Especially, it collects samples by the circulant matrix in the region around the target, and uses the method of ridge regression to train the target detector. As consequence, it reduces the amount of calculation, increase the speed of operation and meets the real-time requirement, as shown in Fig.7.

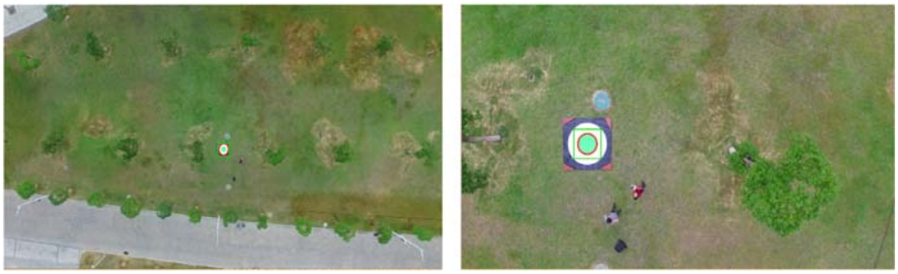


Figure 6. Target tracking with CamShift in different sclaes

Figure 7. Target tracking with KCF in different sclaes

2.4. The Design of Position Estimation

The position estimation needs to understand the transformation relationship between the pixel plane coordinate system, the image plane coordinate system, the camera coordinate system and the world coordinate system. Besides, several significant parameters contain the followings: internal reference of camera (focal length and pixel size), a rotation matrix between the camera coordinate system and the world coordinate system(angles of pitch, roll and yaw), and the height of the UAV relative to the ground. More specifically, internal reference of camera is supplied by the manufacture, and the rotation matrix is measured by the attitude measurement of the UAV, and GPS offers the relative height at the same time[14-15].

According to these known parameters, the relative position between the UAV and the ground target can be obtained. No more explanations here as this section is not the focus of our work.

2.5. The Process of Visual Navigation

According to the above recognition algorithms, we design a visual navigation strategy that automatically recognizes ground targets. Besides, through the above tracking algorithms, we can identify manually by select the target on the ground station in real time. Due to the powerful computing power of our onboard equipment, these two image recognition algorithms or these two target tracking algorithms are able to run simultaneously. Furthermore, there is a filter to combine the results between these two image recognition algorithms or these two target tracking algorithms. After the process by filter, the image processing equipment is able to output the center of target's coordinate in the image coordinates in 10 hz, as shown in Fig.8.

IOP Conf. Series: Materials Science and Engineering **719** (2020) 012077 doi:10.1088/1757-899X/719/1/012077

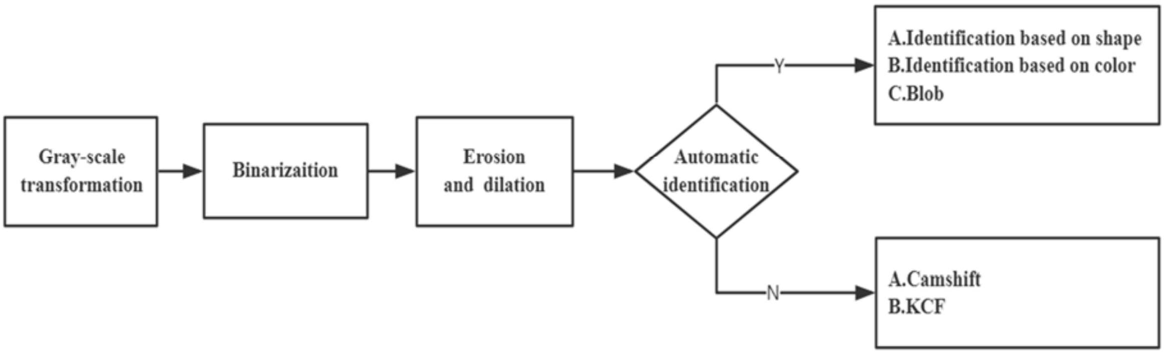


Figure 8. Flow chart of visual navigation

3. Hardware Platform and Software Platform

DJI Matrice 100 flight platform, as shown in Fig.9, is employed as a carrier which allows developers to customize and tailor the system by using the DJI SDK(Software Development Kit)and has flexible access to external devices for control and monitor applications. A NUC computer is equipped to operate GCS(Ground Control Station)and visual programs. An USB camera is added to capture videos. Fig.10 shows the structure of hardware platform[16].

In visual aspect, opencv 3.0 version is utilized to implement multiple universal algorithms on image processing and computer vision in Visual Studio 2015 environment on windows 7 system.

Figure 9. DJI Matrice 100

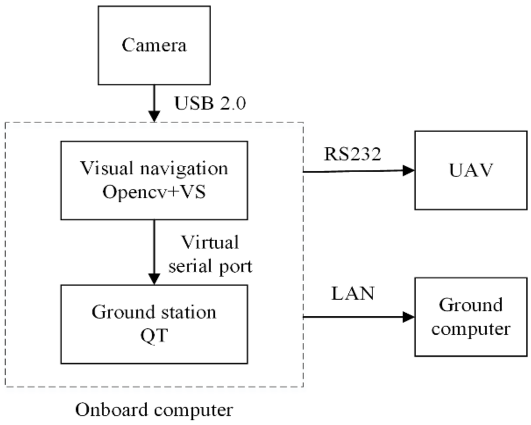


Figure 10. Structure of hardware platform

In flight control aspect, the programs of ground control station is operated in QT development environment on the on-board computer, communicating with the aircraft through a RS232 serial port. Virtual serial ports enable the communication between visual system and ground station, through which the ground station receives relative position information sent by vision program at a frequency of 10 hz, converted desired position information to speed information through PID(proportion-integral-derivative) control and then sent the speed to flight control system after softening speeds.

The ground computer connects with the on-board one via LAN(local area network).It remotely controls the programs of the on-board computer to monitor the health status of the aircraft and the progress of visual navigation in real-time on the ground.

IOP Conf. Series: Materials Science and Engineering **719** (2020) 012077 doi:10.1088/1757-899X/719/1/012077

4. Result and Discussion

4.1. The Comparison between Visual Algorithms

We designed six sets of flight scenes, and run these six sets of visual algorithms, namely the recognition based on color, blob analysis , combined recognition algorithm , CamShift and KCF, combined tracking algorithm. Fig.11.12.13.14 show that the comparison of pixel coordinate in image plane coordinate system among combined recognition algorithm combined tracking algorithm and the real pixel coordinate. From these line charts, three line almost coincide, showing that the error of these two combined algorithms are close to zero. Especially, some incorrect datum equivalent to zero means that the error between these two recognition algorithms or tracking algorithms is larger than the desired value of error, and through the filter, the output of visual system is zero.

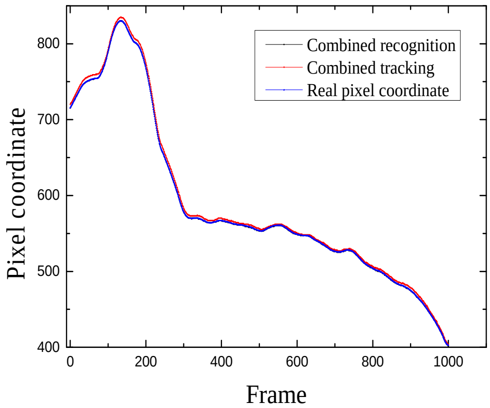


Figure 11. Flight under strong light

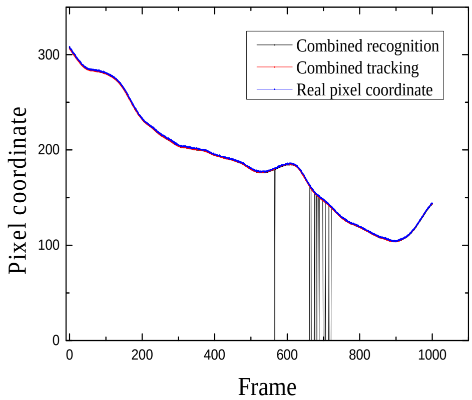


Figure 12. Flight under weak light

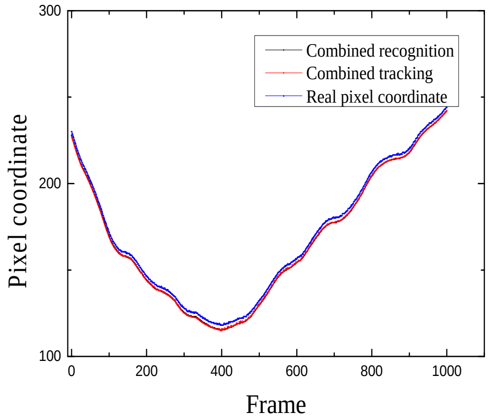


Figure 13. Flight under complex underground

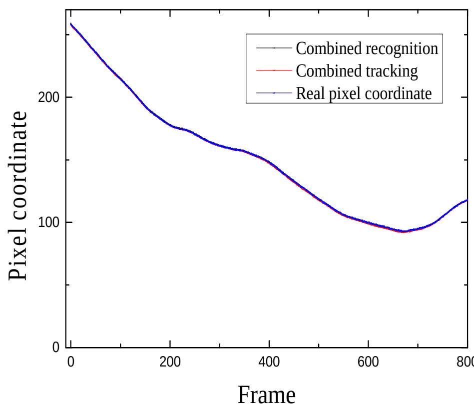


Figure 14. Flight under single underground

Table 1 shows the characteristics among these algorithms, and the computing time is what time it costs to calculate a picture (960*560 dpi) of a frame. In brief, combined recognition algorithm and combined tracking algorithm have higher accuracy and processing speed, which can meet the accuracy and real-time requirements of autonomous landing. Through selecting the color space and changing value of the thresholding visual system is able to accomplish recognition and tracking in most of the scenes. To sum up, this visual system meets the requirement of rapidity, accuracy and robustness, with simple configuration.

IOP Conf. Series: Materials Science and Engineering **719** (2020) 012077 doi:10.1088/1757-899X/719/1/012077

Table 1. Comparison between visual algorithms

Algorithm	Time	Characteristic
Image recognition based on color	2.33 ms	Simple, high processing speed and accuracy, especially on target in similar color.
Blob detect	14.56 ms	Robust, high processing speed and accuracy, but affected by complex background.
KCF	9.66ms	High processing speed and accuracy, but without self-adaptability aimed at varied target, without loss message.
CamShift	12.87 ms	High processing speed and accuracy, with self-adaptability aimed at target in brightly color, able to change target at any times.

4.2. Flight Experiment
The main process of point landing contains the following steps:

- Manipulate manually the UAV to the height of 10 meters. Make sure that the camera has been collected with the ground target and communicate normally with the ground station.
- Select ground targets manually or recognize them automatically, then visual procedure executes image pre-processing, target tracking or image recognition and the solution of relative position orderly at a frequency of 10 Hz. At the same time, the real distance will be sent to the ground station program at a frequency of 10Hz. The relative position received by the station will be converted into the desired speed with PID control. After softening, the forward desired speed and the lateral one will be sent to flight control system at the vertical speed of 0. These mentioned above are the first phase of visual navigation.
- Whether the relative position reaches the setting value will be determined by visual program loop. If so, the aircraft will hover for one second and begin to decline. Otherwise repeat the second step.
- The second phase of visual navigation will start. The ground station will send the forward and lateral desired speed to flight control system, and keep the vertical speed at 0.5m/s.
- When the height decline to 2 meters, the visual navigation will terminate. At the same time, the forward and lateral desired speed will reduce to 0, and keep the vertical speed at 0.5m/s.
- The experimental result shows that the visual system run well under different light or backgrounds. Due to the altitude error measured by GPS, and the simple position estimation, the error between the landing point and the center of target exist within 3 meters.

5. Conclusions
Image recognition and target tracking, the most essential technologies of visual navigation on UAVs were studied in this paper. The flight datum on various scenes demonstrated contrasts in rapidity, accuracy and robustness among six visual navigation algorithms. Thus, automatic and manual methods were respectively designed for identifying ground targets, with rapidity, accuracy and robustness. Onboard experiments suggested the feasibility of point landing technology based on visual navigation. in practice, the ground targets were effectively identified while the landing point had large data errors. Therefore, some improvement opinions were put forward as below for further work.

- GPS has relatively large errors when approaching the ground. Thus, ultrasonic or laser radar is more applicable for measuring the height above ground.
- The relative position estimation algorithm should be improved to accurately solve the true relative position between UAVs and the ground target.
- The cooperative target lacks the indication of forward direction towards the head of UAVs while this direction after landing need to be clarified in the actual landing scenario. Therefore, a ground target with strong directivity should be designed.

IOP Conf. Series: Materials Science and Engineering **719** (2020) 012077 doi:10.1088/1757-899X/719/1/012077

6. References

[1] Yang S, Scherer S A, Schauwecker K, et al. Autonomous Landing of MAVs on an Arbitrarily Textured Landing Site Using Onboard Monocular Vision[J]. Journal of Intelligent & Robotic Systems, 2014, 74(1-2):27-43.

[2] Maravall D, Lope J D, Fuentes J P. Vision-based anticipatory controller for the autonomous navigation of an UAV using artificial neural networks[J]. Neurocomputing, 2015, 151:101-107.

[3] Saripalli S, Montgomery J F, Sukhatme G S. Visually guided landing of an unmanned aerial vehicle[J]. IEEE Transactions on Robotics & Automation, 2003, 19(3):371-380.

[4] Kalal Z, Mikolajczyk K, Matas J. Forward-Backward Error: Automatic Detection of Tracking Failures[C]// International Conference on Pattern Recognition. IEEE, 2010:2756-2759.

[5] Kong W, Zhang D, Wang X, et al. Autonomous landing of an UAV with a ground-based actuated infrared stereo vision system[C]// Ieee/rsj International Conference on Intelligent Robots and Systems. IEEE, 2013:2963-2970.

[6] Bay H, Ess A, Tuytelaars T, et al. Speeded-Up Robust Features (SURF)[J]. Computer Vision & Image Understanding, 2008, 110(3):346-359.

[7] Wang L, Ju H. A Robust Blob Detection and Delineation Method[C]// International Workshop on Education Technology and Training & 2008 International Workshop on Geoscience and Remote Sensing. IEEE Computer Society, 2008:827-830.

[8] Henriques J F, Rui C, Martins P, et al. High-Speed Tracking with Kernelized Correlation Filters[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2015, 37(3):583-596.

[9] Weiss S, Siegwart R. Real-time metric state estimation for modular vision-inertial systems[C]// IEEE International Conference on Robotics and Automation. IEEE, 2011:4531-4537.

[10] Ranganatha S, Gowramma Y P. An integrated robust approach for fast face tracking in noisy real-world videos with visual constraints[C]// International Conference on Intelligent Computing and Control. IEEE, 2018:772-776.

[11] Yen S H, Wang C H, Chien J C. Accurate and robust ROI localization in a CamShift tracking application[J]. Multimedia Tools & Applications, 2015, 74(23):10291-10312.

[12] Wang C, Zhan S U. CamShift Target Tracking Based on Kalman and Surf[J]. Software Guide, 2018.

[13] Lehtola V, Huttunen H, Christophe F, et al. Evaluation of Visual Tracking Algorithms for Embedded Devices[M]// Image Analysis. 2017:88-97.

[14] Li L, Liu Y H, Jiang T, et al. Adaptive Trajectory Tracking of Nonholonomic Mobile Robots Using Vision-Based Position and Velocity Estimation[J]. IEEE Trans Cybern, 2018, 48(2):571-582.

[15] Chenavier F, Crowley J L. Position estimation for a mobile robot using vision and odometry[C]// IEEE International Conference on Robotics and Automation, 1992. Proceedings. IEEE, 2002:2588-2593 vol.3.

[16] Dey V, Pudi V, Chattopadhyay A, et al. Security Vulnerabilities of Unmanned Aerial Vehicles and Countermeasures: An Experimental Study[C]// International Conference on Vlsi Design and 2018, International Conference on Embedded Systems. IEEE Computer Society, 2018:398-403.