ZMART Technical Report

International Aerial Robotics Competition 2019

ZJU's Micro-Aerial Robotics Team(ZMART)[1]

Zhejiang University, Hangzhou, Zhejiang Province, 310027, P.R. China

ABSTRACT

The Zhejiang University's Micro-Aerial Robotics Team (ZMART) has prepared to participate the 2019 International Aerial Robotics Competition (IARC). Our team aims to demonstrate the ability of multiple aerial robots to solve problems together in the same airspace and the ability of man-machine interaction with non-electronic command. The hardware structure, as well as the algorithm structure, will be introduced in this report.

1. INTRODUCTION

Since the intent of the International Aerial Robotics Competition (IARC) is to push new technology of aerial robots, Mission 8 challenges participators with man-machine interaction (non-electronic command and control) and multiple aerial robots solve problems together in the same airspace.

The key to this task is navigation in an indoor environment. Since there is no marker on the ground to calibrate the position, we use the Host robotic to hover over the site to obtain the global image, and process the image to obtain the relative positions of the three Slave robotic in the site, and used this information to correct the position of the Slave robotics.

We set up different task stages, in each stage the Host robotic will issue instructions to three Slave robotics to complete different tasks, including QR code recognition and healing. Gesture recognition is used for workers to switch between different task stag

2. AERIAL ROBOTICS FORMATION

The aerial robotics formation consists of one **Host** robotic and three **Slave** robotics, among which the Host robotic provides global position information of Slave robotics and overall decision-making of formation, the Slave robotics are responsibility for searching, identifying, treating and other tasks in the site.

2.1 Host robotic

2.1.1 Platform

[1] **Team Leaders:** Quan Lun, Zhou Xin

The Matrix 100 is a stable, flexible, and powerful platform developed by DJI, which consists of a flight controller, propulsion system, GPS, DJI Lightbridge, a dedicated remote controller, and a rechargeable battery[1]. This system automatically manages the most complex tasks required for flight.

Figure 1: DJI Matrix 100

2.1.2 Guidance

Guidance is a revolutionary visual sensing system, including five sensor modules and one central processor. Even without GPS, it's possible to achieve hovering accurately within centimeters. Guidance's vision positioning system is effective at an altitude of up to 65 feet (20 meters). Guidance continuously scans the nearby environment and detects obstacles in real time. When used with a DJI flight controller, it can tell our flight system to automatically avoid collision, even at high speeds.

Figure 2: DJI Guidance

2.1.3 Fisheye camera

Fisheye camera has a field of view close to 180 °, which can give a vision of the whole aerial.

2.1.4 Computer

Jetson TX2 is the fastest, most power-efficient embedded AI computing device. There is a latest addition to the industry-leading Jetson embedded platform. This 7.5-watt supercomputer on a module brings true AI computing at the edge. It's built around an NVIDIA Pascal™-family GPU and loaded with 8 GB of memory and 59.7 GB/s of memory bandwidth. It features a variety of standard hardware interfaces that make it easy to integrate it into a wide range of products and form factors[2].

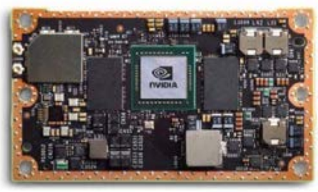

Figure 3: NVIDIA Jetson TX2 Module

2.2 Slave robotic

2.2.1 Platform

DJI F450 is a commonly used flight platform with sufficient holes for mounting various equipment[3]. Quadrotor of this size has a perfect balance of mobility and load capacity. It is used as a slave machine which receives commands from Matrix 100.

Figure 4: DJI F450

2.2.2 Gimbal

There are two purposes for using the gimbal. First, the laser transmitter will be installed on the gimbal so that control the gimbal will precisely help the laser point at the helmet. Second, the gimbal stabilizes the camera to prevent blurring the picture when the plane is moving violently.

Figure 5: Gimbal Module

2.2.3 RPLIDAR A2

The RPLIDAR A2 is the next generation low cost 360 degree 2D laser scanner (LIDAR) solution developed by SLAMTEC. It can take up to 4000 samples of laser ranging per second with high rotation speed. And equipped with SLAMTEC patented OPTMAG technology, it breakouts the life limitation of traditional LIDAR system so as to work stably for a long time[4].

Figure 6: RPLIDAR A2

2.2.4 Raspberry Pi 3B+

Raspberry PI is a cheap but powerful lightweight computer with a 1.4GHz 4 core CPU and wireless network that can run Linux and perform some low-complexity calculations quickly. It has a large community and plenty of learning materials, which can be very helpful[5].

Figure 7: raspberry Pi 3B+

2.2.5 Wireless communication

According to the experience of the former participating teams, due to the large number of participating teams and wireless electronic equipment on the field, the radio frequency of 2.4G is very unstable. Therefore, we choose the frequency of 433MHz and use the wireless communication module based on SX1278 chip. To avoid failure, we will prepare some modules of other radio frequency within free wireless spectrum as backup plans.

Figure 8: 433MHz Wireless communication module

3. BRIEF DESCRIPTION OF THE TASK STAGES

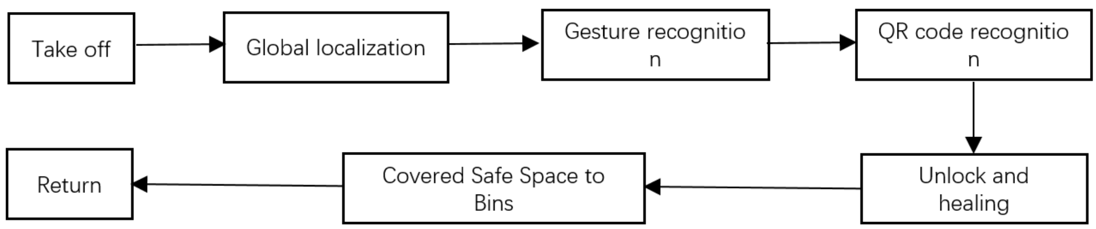

Figure 9: Time line of the task stages

3.1 Take off

After the start signal is sent, the four aerial robotic helpers take off from the starting point and hover. According to the action set in the program, the main aerial robot will fly to a height of 5 meters above the center of the arena, and 3 followers will fly to a height of 3 meters above the ground and into the arena to wait for further instructions.

3.2 Global localization

The main aerial flies to the predetermined observation position, communicates with three followers for confirmation, and sends relative position to followers. The followers consist of one gesture recognition follower and two QR code recognition followers. After receiving the signal from the main aerial robot, they adjust their positions and stand by.

3.3 The human-machine order confirmation

The operator enters the safe space in the center of the arena, and sends gesture signals to the waiting follower. The gesture recognition is successful if the digital tube lights up.

3.4 QR code recognition

The two followers for QR code recognizing fly to the top of bin, take photos of the incomplete QR code, and the photos will be preprocessed and transmitted to the host computer, which will complete the combination, recognition of the QR code and send the password to the operator.

3.5 Unlock and healing

Operators unlock the bins according to the pre-set order of unlocking. Before departure, the operator uses gesture to tell the follower to perform healing once and then hover in the safety zone. Then the operator unlocks all the bins in turn.

When the lock is being unlocked near the bins where there's a follower robot hovering, the follower robot will perform a healing. After the healing is done, the follower robots will follow the operator and do gesture recognition in order to shelter the helmet from the laser.

3.6 Covered Safe Space to Bins

The operator needs to run as fast as possible, so that he won't be hurt by the sentry robot.

3.7 Return

The operator returns. The follower aircrafts go back in turn, and finally the main aerial robot returns after confirming the return from the follower aircrafts.

4. DETAILED TECHNICAL EXPLANATION

4.1 Avoidance

The host M100 uses the guidance visual sensing navigation system to avoid obstacles. The guidance can monitor the environmental information in multiple directions in real time and sense obstacles, which can make the aircraft avoid the possible collisions in high-speed flight.

F450 uses laser radar combined with artificial potential field method to avoid obstacles. Lidar can detect the distance, bearing and speed information of obstacles. Artificial potential field approach is a real-time robot path planning method, and is widely used for autonomous mobile robot obstacles avoidance due to its elegant mathematical analysis and simplicity. Combined with the detection data from the laser scanner, artificial potential field approach can determine admissible and reachable place for its path planning. In order to avoid dynamic Sentry robots, speed and acceleration factors are introduced in the traditional artificial potential field method.

4.2 Detection and localization

In our UAVs formation system, there are several important components. Firstly, we divide our aerial robotic helpers into two parts, one master drone and three followers (slave drones). These aircrafts need to have a basic function of localization. Then, the system should assure that aircraft performs the task within the arena. It can be achieved by the boundary detection module. Moreover, the mission 8 requires the aircraft to heal some wounds for the person and identify friend or foe, so the object detection module is necessary[6].

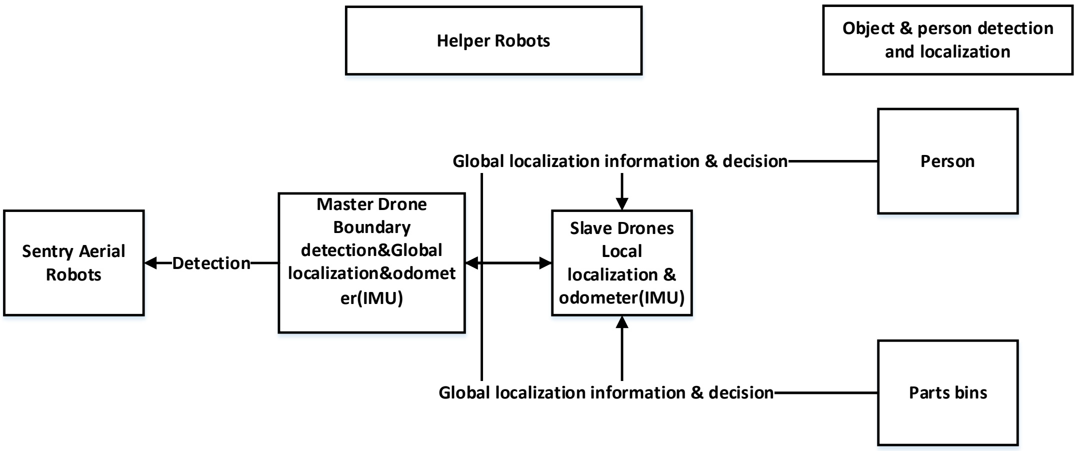

Figure 10: Detection and localization

4.2.1 Localization

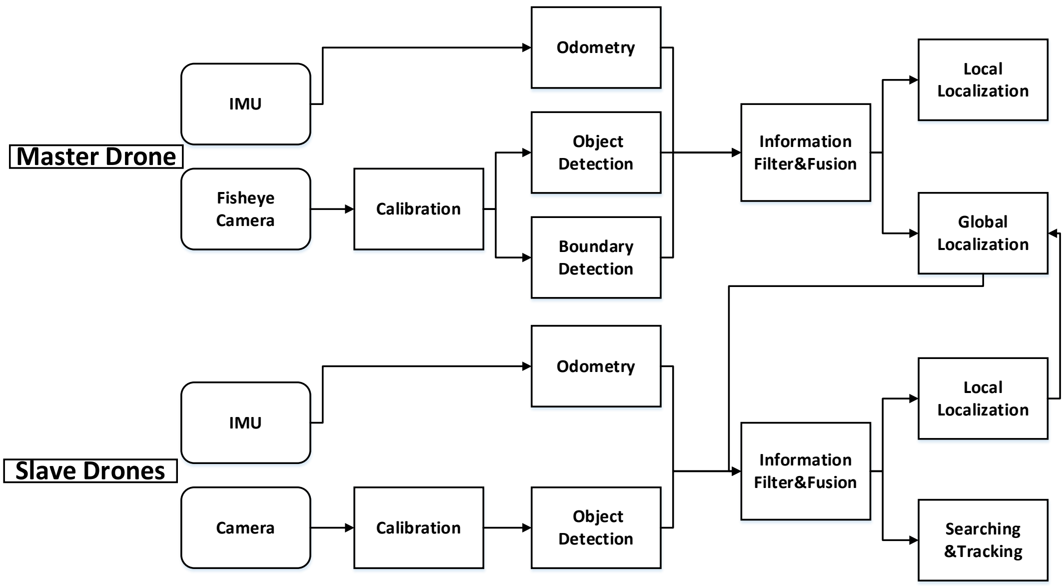

Figure 11: localization

In this mission, a prior map can't be built into micro-processor, so the pressure for localization is main concern. The purposes of localization are keeping the UAV remained in the arena and computing the velocity orientation for decision making. The foundation of localization algorithm is the global localization of the master drone and a local Odometry. The data from IMU and electronic compass are processed by the Odometry[6].

Firstly, the master drone will fly and hover over the field relying on Odometry. And then, we extract the boundaries on the arena as landmark to implement global localization. For high accuracy global localization, the master drone will defish the distorted fisheye image first and then apply detection algorithms to the undistorted image. The question is how to precisely detect those lines and boundaries. We used some traditional computer vision algorithms to detect the lines and used support vector machine to detect the boundaries.

After that, followers (slave drones) will start executing the task and the Odometry can provide basic position information. However, the quality of Odometry is not high enough to provide reliable position. The longer the distance the more serious the error. Therefore, validation should be applied[6]. In our plan, UAVs are able to communicate electronically amongst themselves. And based on the global localization, master drone will calculate followers' global location and velocity, orientation, and send the localization information to the followers (slave drones) to fuse with Odometry. This step can correct the accumulated error of IMU Odometry and ensure accurate execution of following tasks.

Another application of localization information is to compute the position and velocity orientation of the person. The position and orientation of person are the key factors for the followers (slave drones)'s task allocation and decision making.

Thus, the whole global localization and local localization process is a fusion of visual, pose estimation, object detection and boundary detection. It's a special kind of visual-inertial Odometry and it's also a combination of GPS and IMU.

4.2.2 Object detection

In this UAVs system, two kinds of cameras will be installed onboard. A downward faced fisheye camera A with 150 degrees of field angle will be installed on the belly of our master quad-rotor drone. A forward-sight camera B will be installed on followers (slave drones) for object detection. Videos taken by fisheye camera will be processed in DJI Manifold which is a TX2 board and Videos taken by camera B will be processed in intel NUC with i5 CPU.

The camera A is used in searching mode. When the master UAV is hovering, down-sight camera is used to search and locate followers (slave drones), person, the hostile sentry aerial robots globally and precisely. Camera A is also used to detect the boundary of competition area for adjusting UVAs location and avoiding UVAs from flying out of the area. After detecting and locating, the master UAV is used to lead the slave UAVs to approach target until the person enter the scope of forward-sight camera B. One function of Camera B for forward-sight is detecting and tracking the person precisely. Another function of Camera B is to recognize gestures and parts bins[6].

4.3 Design of fully enclosed propulsor

In order to ensure our aerial robots are man-safe, which means their propulsors are completely enclosed, we designed a shield for our robots. In order to minimize the weight of the shield and its influence on the propeller lift, the upper and lower sides of the shield are designed as hollow mesh structures. At the same time, in order to facilitate the installation of propellers, a certain width of space is reserved on the upper side. The overall framework of the shield will be made by 3-D printing with resin as the material. It is estimated that the weight of each shield will reach 240g.

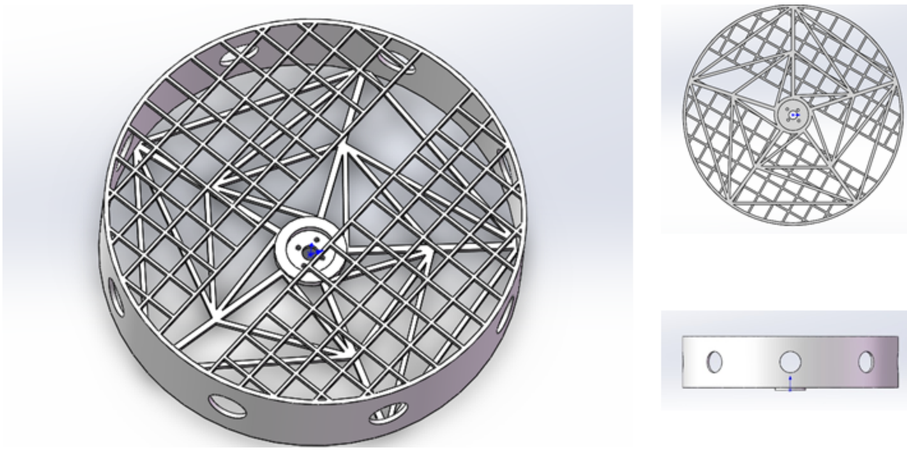

Figure 12: fully enclosed propulsor

In the follow-up test, in order to further reduce the influence of protective shield on aerial robots, nylon rope may be used to replace the mesh structure in the figure above, and the side is also designed as hollow mesh structure to reduce the quality of protective shield. At the same time, the use of nylon material instead of resin material for 3D printing will be considered to increase the strength of protective shield.

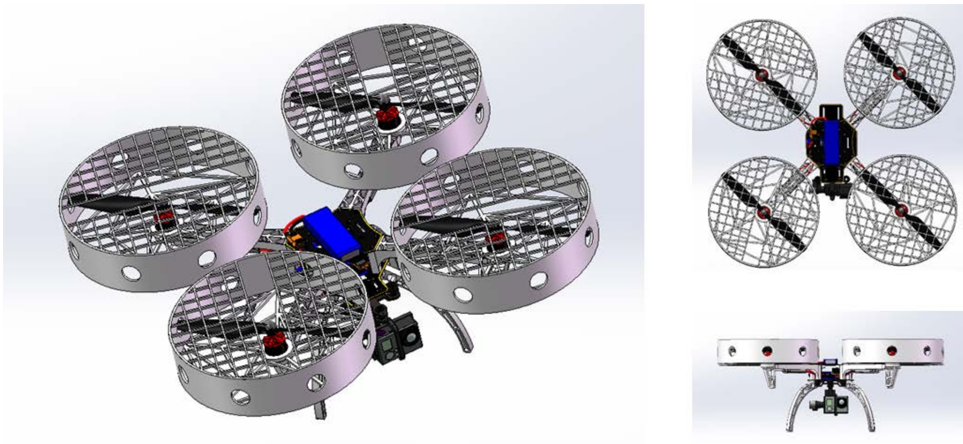

Figure 13: The sketch of an aerial robot equipped with a protective shield

4.4 Hand-Command Recognition

4.4.1 General Framework

Each sub-drone has the ability to recognize the hand gestures of Drone-operator.

The sub-drone first collects the image by down-looking camera and waits for the available signals. Once receiving available signals, it begins a process of translating hand signal into command. The hand signal processing is carried out on the raspberry Pi. Then the sub-drone uses LED digital tube as a feedback, and the sub-drone executes the task according to the command after the confirmation of the drone operator.

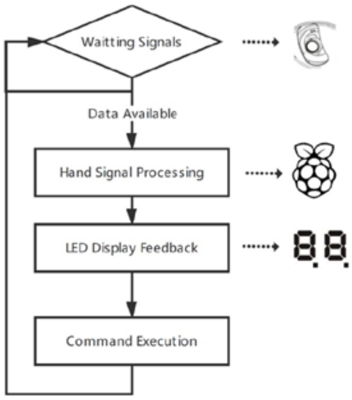

Figure 14: Hand-Command Recognition Framework

And the key points of software processing are presented as follow:

4.4.2 Software Process1: Bilateral Filtering

Bilateral filtering is a method that combines the spatial proximity of image with the similarity of pixel value. During filtering, the spatial proximity information and color similarity information are taken into account. While filtering noise and smoothing image, edge preservation is also achieved.

The principle of bilateral filtering is shown as follows[7]:

$$\mathrm{g(i,j)} = \frac{\sum_{(k,l)\in S(i,j)} f(k,l)w(i,j,k,l)}{\sum_{(k,l)\in S(i,j)} w(i,j,k,l)}$$

$\mathrm{g(i,j)}$ is the output point, $\mathrm{S(i,j)}$ refers to the rectangular range of $(2\mathrm{N}+1)*(2\mathrm{N}+1)$ size centered on $(\mathrm{i,j})$, $f(k,l)$ represents multiple input points, $w(i,j,k,l)$ is the value calculated by two Gaussian functions. Assuming that there are n input points, the nth input point is f_n, and the corresponding $w(i,j,k,l)$ is m_n, $\sum_{i=1}^{n} m_i = M$，Thus

$$\mathrm{g(i,j)} = \frac{\sum_{i=1}^{n} f_i\, m_i}{M}$$

In the above formula, $\frac{m_i}{M}$ represents the weight of the first point. The image matrix and the kernel are weighted by convolution operator, and the output value is finally obtained.

As for w，it can be calculated by

$$w = ws * wr$$

ws is a spatial proximity Gaussian function, wr is a pixel value similarity Gaussian function

$$ws = e^{\left(-\frac{(i-k)^2+(j-l)^2}{2{\sigma_s}^2}\right)}$$

$$wr = e^{\left(-\frac{||f(i,j)-f(k,l)||^2}{2{\sigma_r}^2}\right)}$$

σ_s is a parameter of space near range, σ_r is similarity factor, which are input parameters of the program.

4.4.3 Software Process2: Capturing and removing the background

Capturing background involves skin color extraction in HSV space. According to the value of skin color on the three components of HSV, the skin color part of an image can be simply detected.

Background removal involves KNN (K-nearest neighbor method), which classifies by measuring the distance between different eigenvalues.

After the classification is successful, the non-target set can be removed.

Figure 15: Remove background

4.4.4 Software Process3: Image Preprocessing

After removing the background, the image is processed with gray scale, Gaussian blur and binarization. Through the above steps, the image is transferred from RGB image to grayscale image, and the edge is smoothed at the same time.

Figure 16: Image Preprocessing

4.4.5 Software Process4: Finding Contours and Generating Convex Hulls

Then using the opencv functions to realize the binary image topology structural analysis, to find the image contour. According to the Sklansky's algorithm[8], we generate convex hull from the set of edge points.

4.4.6 Software Process5:Find the fingertips

Calculate the distance among all points in the edge point set and the center, remove the points below the center and the points closer to the center from the point set. Finally, we find the points that corresponds to the fingertips. Hand gesture is judged by the number of fingertips, and the result is output to LED digital tube to display as feedback.

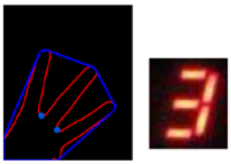

Figure 17: Find Outline and LED Dispaly

4.5 Identification and splicing of QR codes

4.5.1 *Whole process*

The whole process is shown in the figure below. Two QR code recognition robots respectively take code puzzle photos, each robot needs to take two pictures. After pre-processing the photos, the position information of the code and the code of the part will be sent to the host aerial robot in the same size. According to the four parts of the code sent by follower aerial robot, the host completes the splicing and identification of the code, and then obtains the password, and sends the password to the operator.

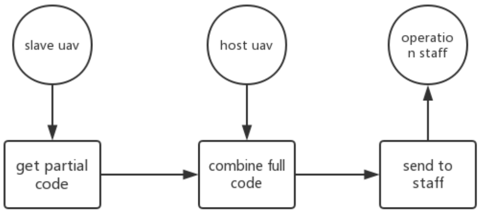

Figure 18: gaining partial code process

4.5.2 *Get partial code by follower UAV:*

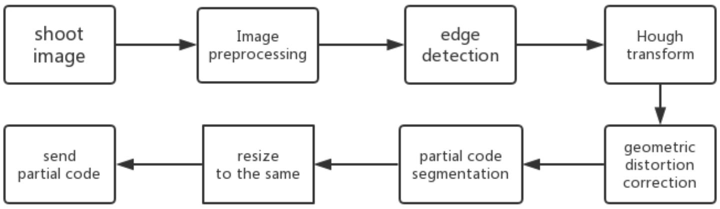

Figure 19: gaining partial code process

The process is shown in the figure below. The follower UAV will preprocess the image first. It includes four steps. (1) Grayscale: reducing storage and speeding up calculation. (2) Median filtering: removing image recognition noise, protecting edge information. (3) Apply adaptive brightness equalization algorithm. (4) Image binarization.

Then it will use Sobel operator edge detection to detect the partial code in the picture, and the four corner points of the partial code are searched. The Hough transform algorithm then is used to detect the partial code rotation to the horizontal state. Combining the four corner points of the partial code, and the geometric distortion correction transform is used to obtain a square partial code standard image without geometric distortion [9]. After that, each of the follower will resize the partial code images to the same before transmits them to the host robot.

4.5.3 *Combine all partial codes by host UAV*

Since it's a two-dimensional code, every two adjacent parts of the code has the same line of information in the interception, the host UAV can get the relative arrangement relation of the four codes by comparing these information of the four partial codes, so as to restore the whole two-dimensional code. The detail operation is to compare three of the four partial codes that contain positioning labels and then combine them. Then the host will combine the last part into the last gap part. After getting the whole QR code, the host decodes it and then sends the password to the operator.

4.6 Treatment

4.6.1 Helmet Detection

We use the KCF algorithm to get the position of the target. In the training phase, by using cyclic matrix theory, take dense sample of the target area in the video, obtain positive and negative samples, extract the characteristics of Histogram of Oriented Gradients (HOG) and utilize all samples to train Ridge Regression Classifier. In the detecting phase, the target position of the previous video frame is taken as the center for dense sampling. All the samples are input into the Ridge Regression Classifier. After the classification of the Ridge Regression Classifier, the position of the target is obtained [10]. Finally, we combine the KCF with HOG & SVM, the detection & tracking can reach a high speed. The procedure is shown in Fig 20.

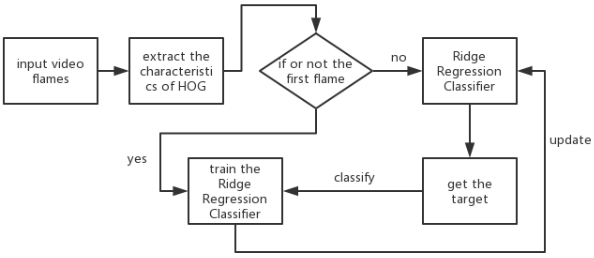

Figure 20: the procedure of helmet detection

In order to solve the problem of the occlusion and that once the target being lost, it can't be found again. We can add the filtering module to improve the KCF algorithm, and fuse the filtering module and the tracking target location information to obtain the accurate target position information and improve the accuracy. The procedure of improved KCF algorithm is shown in Fig 21. Using KCF to detect pedestrian is shown in Fig 22.

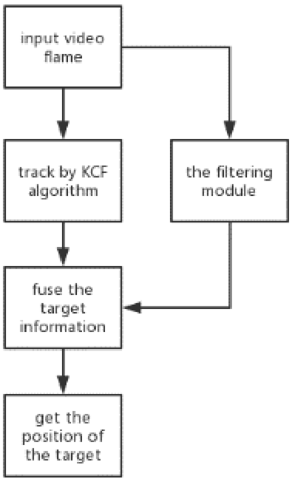

Figure 21: the procedure of improved KCF algorithm(left)
Using KCF to detect pedestrian(right)

4.6.2 Gimbal Control

The procedure of controlling gimbal is shown in Fig 20. Firstly, we use PnP which can find the coordinates of feature points from 2D to 3D. Then the position of the feature points in the global coordinate system is obtained, and then transform the position from global to the UAV coordinate for gimbal control.

The two-axis gimbal is installed under the quadrotor UAV. The gimbal needs to adjust the angle pitch and angle yaw to drive the camera connected to the gimbal ensuring that the camera's optical axis always points to the moving target and the target of detection. The center point of the box is always at the center of the camera coordinate system.

The tracking of the target helmet by the gimbal means that, according to the Pitch and Yaw detected by the sensor of gimbal, calculating the error between them and the desired angle. By designing the gimbal tracking controller, the driving motor of the gimbal is controlled, and then control the pitch and yaw of the gimbal and realize the tracking of the target by the gimbal.

Due to the limitation of detection, there is a delay between sending a shooting command and performing a shooting action, and the gimbal may be shaken. Therefore, EKF is used to smooth the output yaw and pitch angles, and the uniform linear motion model is for prediction. Finally, the predicted gimbal yaw angle and pitch angle are used as inputs to the cascaded PID controller to control the gimbal.

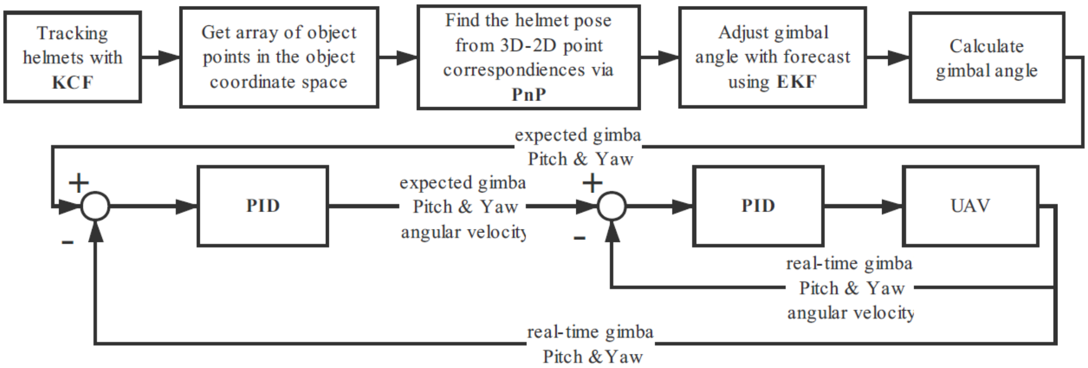

Figure 23: Control flow chart

Reference

[1] DJI. MATRICE 100[DB/OL]. https://www.dji.com/matrice100?site=brandsite&from=insite_search, 2019-5-28.

[2] NVIDIA. Jetson TX2 Module. https://developer.nvidia.com/embedded/buy/jetson-tx2[DB/OL], 2019-5-28.

[3] DJI. Flame wheel arf. https://www.dji.com/flame-wheel-arf/feature[DB/OL], 2019-5-28.

[4] SLAMTEC. RPLIDAR A2. http://www.slamtec.com/en/Lidar/A2[DB/OL], 2019-5-28.

[5] Raspberry Pi Foundation. Raspberry Pi 3 Model B+. https://www.raspberrypi.org/products/raspberry-pi-3-model-b-plus/[DB/OL], 2019-5-28

[6] ZMART Technical Report International Aerial Robotics Competition 2018

[7] Di Liu, Yanru Wu, Yong Xu, Jing Li. Stochastic response of bistable vibration energy harvesting system subject to filtered Gaussian white noise[J]. Mechanical Systems and Signal Processing,2019,130.

[8] Sklansky, Jack. "Finding the convex hull of a simple polygon." Pattern Recognition Letters 1.2 (1982): 79-83.

[9] Wang Junfeng, Tang Peng, Chen Wei, et al. A method for fast detection and extraction of two-dimensional codes based on computer vision detection technology

[10] Liang Shuo, Chen Jinyong, Wu Jinliang, Wang Changli.A Long Time Video Target Tracking Algorithm Based on KCF Framework[J]. Radio Communications Technology,2017,43(02):55-58+82.