HITCSC Technical Paper for IARC Mission 8 in 2019

Bohan Cui, Qianjin Wei, Yinbao Niu, Zibo Lv
Zhaochen Lin, Shengbo Jiang, Xiyang Liu
Harbin Institute of Technology

Abstract

The HITCSC Team is preparing for the IARC 8th mission. We are currently working on a safe and reliable system. Based on our understanding of the 8th mission and last year's experience, we first introduced our platform and safety design. Then we presented the current efforts and related work of system's key parts, including voice command recognition, target detection, positioning, recognition, control and planning.

1 INTRODUCTION

IARC mission 8 takes the Human-machine collaboration as the key challenge, requiring system to have non-electronic man-machine interaction capabilities, and at the same time, some new abilities are proposed and emphasized compared to the previous tasks such as the information fusion, swarm interaction, air target identification and so on. Due to the participation of human beings, there is also high requirement for the safety of aircraft. Of course, this is also the necessary condition for UAVs to be deployed indoors. According to our understanding of tasks, we believe that how to recognize short voice instructions quickly and accurately is the key point in the process of human-machine interaction. This also requires that UAVs are smart enough and have good strategies so that they can perform tasks independently and automaticly when they hear "mission" level voice commands. For UAV assistants, besides avoiding enemy UAVs, how to allocate tasks reasonably and avoid collision with each other in the indoor environment without global information and sufficient feature points is also challenging .

2 SYSTEM OVERVIEW

2.1 Flight Platform

According to the requirements of the mission, we temporarily choose DJI MAVIC2 Zoom as the flight platform. Mavic 2 Zoom, with a 1/2.3 inch double optical zoom camera, has 31 minutes' flight time. The forward-looking obstacle detection system can detect object in 0.5~20 meters with the 40 degrees' horizontal and 70 degrees' vertical view angle.

Figure 1: DJI Mavic 2 Zoom

2.2 Security Consideration

In order to ensure the safety of personnel and other aircrafts in the field during mission execution, in addition to the obstacle avoidance system of flight platform, we add electrical safety switch between battery and power interface of platform so that power can be cut off when it is necessary. Referring to DJI mavic Pro fully enclosed shield, we designed a shield for mavic 2 to prevent high-speed rotating propellers from hitting people. Although safety design reduces flight time and motion performance of aircraft, it is still a necessary design. Relevant accessories are still being tested and improved.

2.3 Function Model

In the current design, our system consists of the following modules: human-machine interaction, perception, planning and control. The relationship between each module in our simulation system is shown in the following figure.

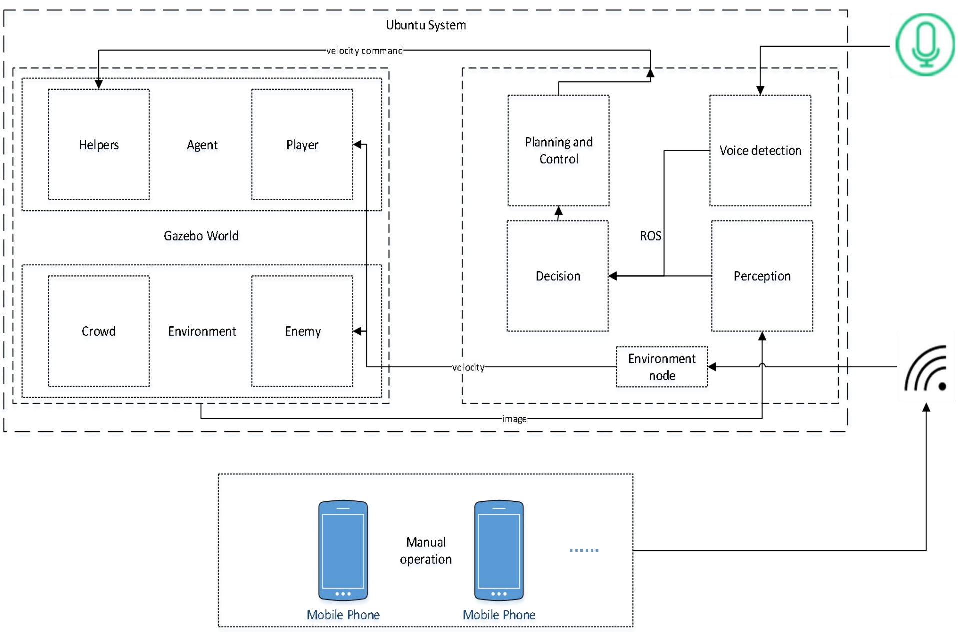

Figure 2: Simulation System Overview

3 HUMAN-VEHICLE INTERACTION

According to the IARC mission 8, between operators and aircrafts, there must not be any electric communication except for voice, gesture and etc. Considering the operator needs to take the risk of being exposed to the enemies in order to give the command with gesture, we decide to use voice command. Based on the voice command, we have two methods.

3.1 Online Voice Progressing

The programming of the speech dictation function, we have made a program based on Java and IFLYTEK online voice progress service. This program can upload the voice we record to the service and translate it to the text. Then we detect the command from the text we get. Besides, we have completed an available APP in Android environment. This APP can carry out the dictation of short text.

3.2 Off-line Voice Progressing

Considering the Network interference, the online progressing method could have significant delay, so we also made an off-line program by ourselves.

3.2.1 Voice Detection Based On Spectrogram

We use spectrogram which contain both Timing information and frequency information, to recognize voice command. In order to do this, we need to make a STFT (Short Time Fourier Transfer) to the voice we record. The basic idea of this method is to do a STFT of a short time period (10ms~30ms), to obtain the frequency information of this period. After doing STFT on all the periods of the voice, we can get an image which shows the texture of the voice. Then use 2-D CNN to learn the texture of each command.

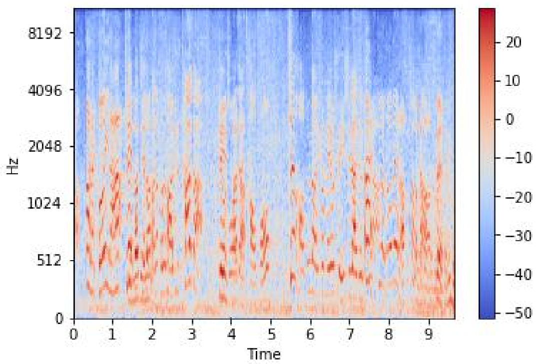

Figure 3: Spectrogram Example

3.2.2 1-D CNN Voice Progressing

Considering the voice is saved as a 1-D array, which means the amplitude of the sound, by applying the 1-D CNN on the array, we can let the program to learn the feature of voice. Then put the output to the RNN or dense layers to complete the classifying mission.

4 PERCEPTION

4.1 Target Detection

The UAV platform selected in our competition plan is the Dji Mavic2 whose PTZ camera has a resolution of 12 million effective pixels and supports 2x zoom. The camera's lens focal length ranges from 24mm to 48mm and the viewing angle ranges from 83° (24mm) to 48° (48mm). The PTZ camera's pitching motion range is -90° to +30° and yow direction motion range is -75° to +75°.

According to the experience of last year's competition, the recognition algorithm based on color features is not reliable when the light of the venue changes a lot. In addition, the flying height and attitude of the drone frequently change and the direction and speed of the participating players also have great uncertainty which cause the problem that the target characteristics change greatly in the view of the drone. Considering those questions, traditional methods based on features and optical flow methods also have significant limitations and it is difficult to design the recognition algorithm when the details of the venue are known very little to the teams.

CNN (Convolutional neural network) have become ubiquitous in comuputer vision ever since AlexNet won the ImageNet Challenge in 2012. Compared with traditional methods, convolutional neural networks have the advantages of high accuracy and adaptability. We use tiny-YOLOv3 as the detection network, which has a good performance in trade-off of speed and accuracy. Due to platform ' s limited computational resources, in order to ensure the timeliness of information processing, and to take into account the accuracy of target detection, it is necessary to optimize the network specifically. Learning from Mobilenets, we using blocks suitable for mobile devices, such as depth separable convolution and linear bottleneck module, to redesign the backbone features network of tiny-YOLOv3. The redesigned neural network has a better performance on mobile devices. Figure 4 show the detection result of the network.

Figure 4: Detection Result

4.2 Bin Localization

This part is based on PNP algorithm for camera pose estimation. The problem of PNP refers to the problem of calculating the projection relationship between N feature points in the world and N image points in image imaging, so as to obtain the camera or object pose.

The specific solution principle is as follows:

$$\begin{pmatrix} x \\ y \\ 1 \end{pmatrix} \sim \begin{bmatrix} f_x & 0 & c_x \\ 0 & f_y & c_y \\ 0 & 0 & 1 \end{bmatrix} \cdot \begin{bmatrix} r_{11} & r_{12} & r_{13} & t_1 \\ r_{21} & r_{22} & r_{23} & t_2 \\ r_{31} & r_{32} & r_{33} & t_3 \end{bmatrix} \cdot \begin{bmatrix} X \\ Y \\ Z \\ 1 \end{bmatrix} \tag{1}$$

Firstly, four feature points are used to obtain the world coordinates and image pixel values. Then the camera distortion parameters and built-in parameters are obtained by camera calibration. The rotation matrix R and translation matrix T are solved by the function of OPENCV. Finally, the camera attitude is expressed by rotation matrix. The specific formulas are as follows:

(1)Solving Rotation Angles of Three Axis Based on Rotation Matrix

$$R = \begin{pmatrix} r_{11} & r_{12} & r_{13} \\ r_{21} & r_{22} & r_{23} \\ r_{31} & r_{32} & r_{33} \end{pmatrix} \tag{2}$$

$$\theta_z = a\tan 2(r_{21}, r_{11}) \tag{3}$$

$$\theta_y = a\tan 2(-r_{31}, \sqrt{r_{31}^2 + r_{33}^2}) \tag{4}$$

$$\theta_x = a\tan 2(r_{32}, r_{33}) \tag{5}$$

(2)According to the rotation and translation matrix of the camera coordinate system in the world coordinate system, the coordinates of the camera in the world coordinate system are solved, and the world coordinates of the camera relative to the reference point are determined. As shown in the figure below, the yellow dot is the reference point.

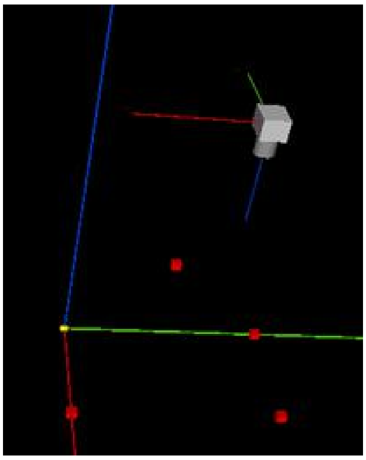

Figure 5: Camera Position Example

(3)The space coordinates and attitude angles of the camera are shown in the following figure.

TABLE 1. POSITION AND ATTITUTE OF CAMERA

X(mm)	527.296
Y(mm)	-4.99289
Z(mm)	356.245
Pitch（angle）	-177.973
Yaw（angle）	-56.1131
Roll（angle）	-88.4027

4.3 QR Recognition

Correct scanning of QR code requires the correct stitching of four parts of QR code first. In order to achieve this goal, we first use deep learning to extract the general location of the IPAD screen. After the image is preprocessed, we can get a rough area of screen which will be processed by our program.

Figure 6: Original 4th Part of QR Code

We use contour detection, rectangular detection, convex Hull points detection and other methods for the first step of image processing. Then, we use boundary detection, corner detection, linear intersection detection and other methods to obtain the data which is needed for image correction. After filtering the edges and intersections, we get the necessary points for screen correction, and then we use mathematical operations to correct the screen contour to a rectangle.

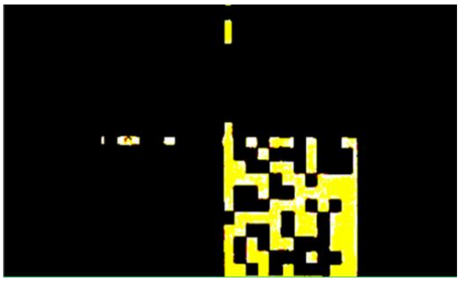

Figure 7: Corrected 4th Part of QR Code

Finally, we intercept the four screens for the transformation of image size and image mixing to stitch them together and get the complete QR code.

Figure 8: Stitched QR code

Lastly we scan the stitched QR code to get the information it contains. The results we get finally are shown as followed.

TABLE 2. Scan Result

Type	QR-Code
Number	2468

5 PLANNING AND CONTROL

5.1 Path Planning

In the mission 8 of IARC, there are four quadrotors involved in search, tracking, identification and treatment tasks. In these tasks, the major task we need to achieve is to obtain a safe path that make each quadrotor can reach their target points quickly.

5.1.1 Target Searching

For the sake of efficiency, we give up the scheme of unified path planning by global mapping and adopt the scheme of distributed path planning of each quadrotor. Since the competition field is small, and the success rate of target search algorithm is relatively stable and efficient, the quadrotors can find the targets which they need to track relatively quickly, it is not necessary to build the global map to assist in searching the targets. In the preparation step after the start of the task, our quadrotors will fly to a certain height to search and identify objects in a certain range in front of them. They will find the targets quickly and wait for the operator's command to assign the tracking target.

5.1.2 Path Planning Based On Field Of View Constraint

Since the global position information of the construction map is lost, the relative position information of the target points is the essential input information in the process of path planning. In the scheme of this paper, the relative position information to the target points is acquired by the front-facing camera of the quadrotors, so it is necessary to keep the tracking target points always within the camera field of view of the quadrotors in the process of path planning.

According to the characteristics of obstacle avoidance sensor carried by our quadrotors, we choose the Dynamic Window Approach to overcome the shortcoming of narrow perception range of obstacle detection of our quadrotors.

In this path planning task, the position of the target points and the rotation range of the airborne cradle head determine the range of the quadrotor course angle, while the range of the course angle and the performance of the quadrotors themselves determine the range of the safe flight velocity of the quadrotors. In the dynamic window approach, each control cycle will firstly generate a set of physically reachable velocity and heading angle based on the performance limitation of quadrotor, which is called the dynamic window. Then, based on the position limit of the target point, the current quadrotor's course angle limit and the obstacle ahead limit, each group of velocity and course angle in the dynamic window is comprehensively examined through an evaluation function to select the current optimal velocity and course angle. By iterating through this process, the target points can be tracked safely and quickly.

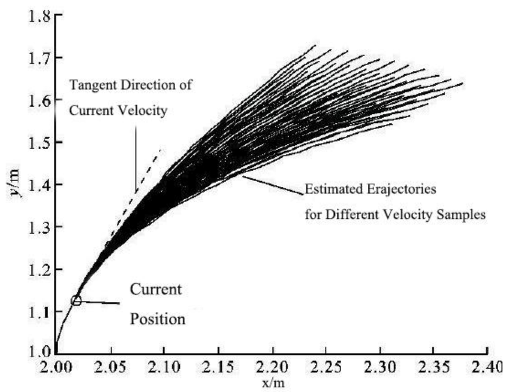

Figure 9: The result of Velocity Sampling

5.2 Motion Control

The movement of the UAV in three directions is controlled respectively to track the target in a given position. Speed command is adopted for control. Taking horizontal X direction as an example, the system block diagram is shown in figure 10. The control method adopted is Active Disturbance Rejection Control(ADRC). Input the signal of desired relative position into the Tracking Differentiator(TD), then TD will output the tracking value of input signal as well as its differential. The system can get the relative position of the target through measurement and calculation with the help of camera which is fixed on the gimbal. Then input the relative position into the Extended State Observer(ESO), which can output the estimate of relative position as well as its differential, and at the same time it can estimate the disturbance, the error signals are obtained by subtracting the output of TD from the output of ESO, then design an appropriate control law to get the control signal. By using the observation value of disturbance to compensate the control signal, the UAV control system is turned into a second order integral system.

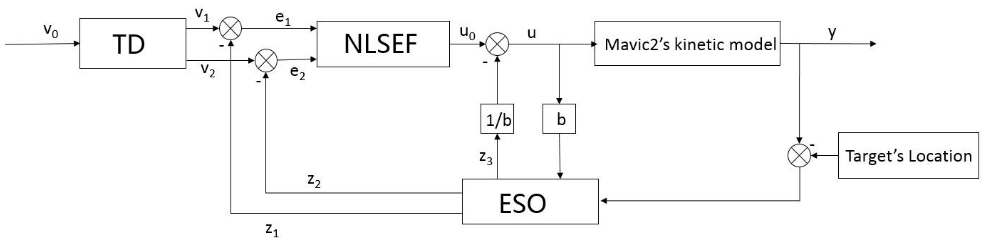

Figure 10: The Control System Block Diagram

Since there are disturbances such as wind in the environment, there are many uncertainties in the modeling of UAV. The reason for using Active Disturbance Rejection Control is that it can estimate the disturbance synchronously and compensate it without the exact model of UAV. At the same time, there is an uncertain delay in obtain the relative position, so the system needs to predict the output.

6 CONCLUSION

This article introduces the main parts of the team's current system for 8th mission with related design and functions by now. The modules are still being tested and integrated. We believe that in the next challenges, performance of each functional module will be greatly improved.

7 REFERENCES

[1]Redmon, J., & Farhadi, A. (2018). Yolov3: An incremental improvement. arXiv preprint arXiv:1804.02767.

[2]Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., & Chen, L. C. (2018). Mobilenetv2: Inverted residuals and linear bottlenecks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (pp. 4510-4520).

[3]Zhang, C. G., & Xi, Y. G. (2002). Mobile robot path planning based on rolling windows. Syst Eng Electron, 24(6), 63-9.

[4]Han, J. (2009). From PID to active disturbance rejection control. IEEE transactions on Industrial Electronics, 56(3), 900-906.

[5]Gao, Z. (2006, June). Scaling and bandwidth-parameterization based controller tuning. In Proceedings of the American control conference (Vol. 6, pp. 4989-4996).

[6]Sithara, A., Thomas, A., & Mathew, D. (2018). Study of MFCC and IHC Feature Extraction Methods With Probabilistic Acoustic Models for Speaker Biometric Applications. Procedia computer science, 143, 267-276.