Design of Unfixed-point Receiving Scheme Based on Logistics UAV

Yang Haoran [1, a], She Jiajun [1, b] and Zhang Min[1,c]

University, College of Electrical Engineering

[1]School of Jinan, Jinan University Zhuhai Campus, Zhuhai 519070, China;

[a]3223950053@qq .com, [b]1419838340@qq .com, [c]1308079945@qq .com

Abstract

With the Internet plus business model, people's shopping methods have changed a lot .The logistics industry that accompanies this has reduced people's concern about the convenience of online shopping .However, the logistics industry currently needs a large number of labor force, in order to solve the problem of labor force, many enterprises began to use logistics UAV .The purpose of this scheme is to solve the problem that when UAVs are transporting goods, they must stop at fixed points designed beforehand by the logistics party . The specific process is that UAVs first arrive at the designated location (the specific coordinates and heights selected by the user), then use obstacle avoidance system and hover at a certain height during the descent process, determine whether they are the users through face recognition, and finally control them by the user's gesture .Flight to Accurate Cargo Delivery without Fixed Point

Keywords

UAV; obstacle avoidance system; face recognition; human-computer interaction; human posture; computer vision.

1. Introduction

Unmanned aerial vehicles (UAVs) are gradually being widely used in today's society. The distribution and operation center of UAVs used by Amazon Company in the field of logistics is built on the prototype of honeycomb and equipped with fully automated loading robots .Unmanned aerial vehicles (UAVs) responsible for the delivery of goods can stop automatically in the operation center and load the next single delivery task with the help of the system. However, in the last step of delivery, special operators or pre-set landing points are required .As the last step of logistics delivery, it is also one of the most demanding labor forces in current logistics .In order to realize the automatic transmission of this last step, the design of this scheme proposes one of the solutions

2. Design and purpose

2.1 System design

In order to realize the above scheme, a real-time operating system for UAV-human interaction is designed, which includes two parts: monitoring and processing system and UAV operation system . The processing system includes user identification system, UAV operation authority submission security system, human body attitude recognition and speech recognition operation system .

According to the inherent characteristics of human shape, the human pose recognition system calculates the feature points of human model surface by using the information of section circumference, average radius and curvature, and extracts the position of human joint center .The position of the head and the posture of the human body can be determined .

User identity recognition system uses the user's face information in the user information database set by the transportation company . After the human head is recognized by the human body posture recognition system, the system will provide feedback by recognizing and reading the face information .Through the similarity detection of the returned face data and the face data in the database, the identity can be confirmed only when the matching degree reaches a certain value .

The human posture operating system recognizes the human posture by using the human skeleton, and then operates the UAV by posture, so that it can hover or land in a specific position .

The speech recognition operating system recognizes the user's voice and generates the operation instructions for UAV, so that the user can control the UAV independently .

The use of 5G technology in image data transmission can improve the recognition rate and speed in the process of operation, and can respond quickly when remote operation is needed .

After the UAV is designed by the distributor, it can make the UAV fly to the high altitude of the predetermined location without obstacles through the internal autonomous flight control system, and then complete the transmission through the user's autonomous operation .

The system design diagram is shown in Figure 1 .

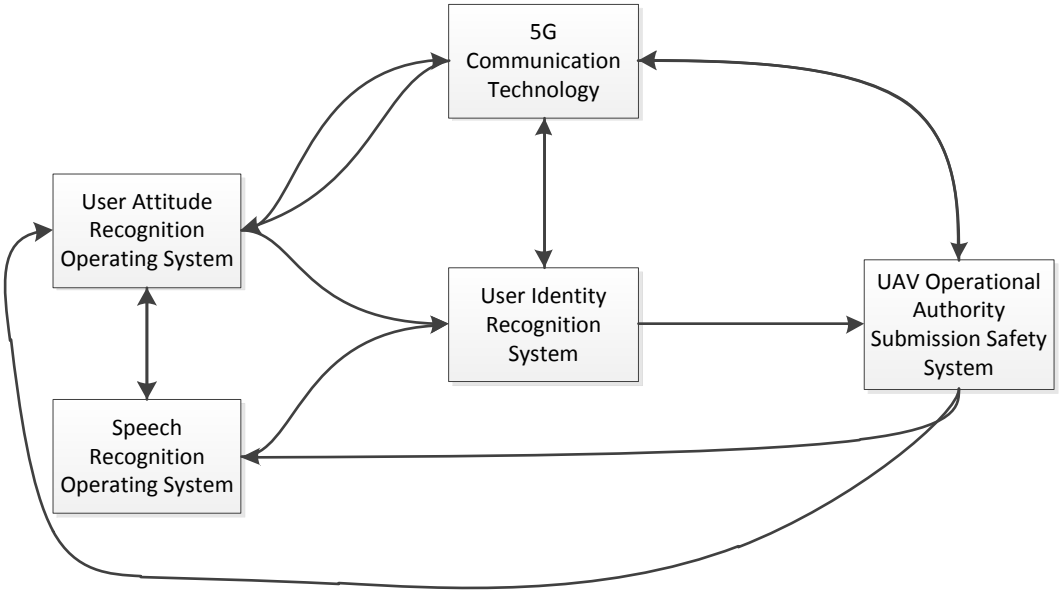

Fig . 1 Systematic Design Overview

2.2 Background and purpose

As an automatic flight platform, UAV has high efficiency and extensibility in many aspects . It has been widely used in various fields of digitalization and automation .For example, aerial photography, automated unmanned patrol, fast logistics, efficient farmland spraying of drugs or water resources, accurate laser mapping, mobile face recognition and security applications of UAVs .New requirements are put forward for over-the-horizon remote control, image real-time cloud processing, and high-definition video stream return capabilities, which promote the deep integration of UAV and mobile network .Targets in images are extracted by cameras or sensors, and multi-layer convolution neural network is used to classify the types of targets .The attitude information of human body is recognized or the operation instructions are obtained by speech recognition . The UAV reacts accordingly according to the recognition results to achieve the effect of controlling the UAV .The 5G network will be extended to three-dimensional coverage of the ground and air from "human" to "object" to realize low-altitude digitization, thus promoting the improvement of UAV supervision .

There are many problems in the process of delivering goods on time and on demand .As the "artery system" to ensure the normal operation of the city, the predicament of urban distribution has been exposed under the strong purchasing flow of e-commerce in recent years .At present, "the last kilometer" still depends on the electric tricycle or two feet of the courier. There is no logistics vehicle

channel to ensure smooth logistics. Most of the logistics freight vehicles are restricted by traffic in the city, and there is still a problem that logistics vehicles have no place to stop .Full use of human foot transport will lead to extreme inefficiency and a sharp rise in human costs .Taking express delivery as an example, after finishing the express delivery, terminal distribution center needs to send special courier to carry express, and then deliver it to each relevant logistics container or hand over directly to the recipient .In order to solve the problem of labor force, many enterprises begin to use logistics UAVs .The purpose of this scheme is to solve the problem that when UAVs are transporting goods, they must stop at fixed points designed beforehand by the logistics party . The specific process is that UAVs first arrive at the designated location (the specific coordinates and heights selected by the user), then use obstacle avoidance system and hover at a certain height during the descent process, determine whether they are the users through face recognition, and finally control them by the user's gesture .Flight The use of 5G high-speed communication channel in real-time communication can help improve the accuracy of operation and identification and achieve the accurate delivery of goods without fixed-point .

3. Specific principles

3.1 Recognition and Attitude Recognition

Firstly, the human body posture recognition needs to determine the region of human body and distinguish the parts of human body. [Recognition content can be roughly divided into head, trunk, hand,] training with the method of human body image depth learning, which can accurately identify the approximate position of human body within a certain range of distance]

After recognizing the approximate position of the human body, the skeletal nodes of the human body will be recognized to recognize the posture information of the human body.

Human pose recognition operation system can first recognize the position of human body, use primary neural network to calculate the coordinates of human joint pixels in color image, map the coordinates of human joint in color image to the coordinates of depth map, calculate joint thermal map, input depth image and hot spot image into secondary neural network to estimate the position of human joint in 3D .The human joint is defined as 15 joint points (Fig. 2). The human 3D joint position prediction is divided into two levels of network. The human body state of color image is predicted in the primary neural network. By calculating the likelihood value of human joint of each pixel, the maximum coordinate of human joint is taken as the joint coordinate.

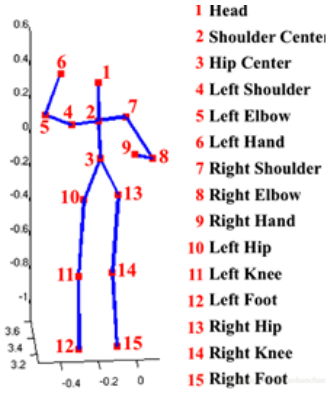

Fig. 2 Hot spots of human joints

The primary neural network usually uses top-down algorithm: mainly divided into two stages, pedestrian detection and single-person attitude estimation. Pedestrian detection has a great impact on the latter one. Usually, a better performance detector is used, and then the detected pedestrian frame is used as the input of single-person attitude estimation.

Secondary neural network binds depth image and multi-channel joint thermo gram as input, and further optimizes the results of 2D joint detection by convolution neural network to obtain 3D human joint posture .Using the spatial pyramid pooling method, a convolution neural network is designed,

which is not limited by the size of the input image .The flow chart of the algorithm is roughly as follows:

Firstly, through selective search, 2000 candidate windows are searched for the detected images.

Feature extraction stage .Convolutional neural network is used for feature extraction, and pyramid pooling is used .The specific operation of this step is as follows: input the whole picture to be detected into CNN, extract one-time features, get the feature map, then find the regions of each candidate box in the feature map, and then use the pyramid space pooling for each candidate box to extract the feature vector of fixed length .SVM algorithm is used to classify and recognize feature vectors .Then the distance between the joint points is calculated . Firstly, the actual distance between the person and the camera is obtained by using the depth information of the scene .The actual distance d from the target to the attitude sensor is calculated by using the obtained depth value.

$$d = K\tan(Hd_{raw} + L) - O \quad . \tag{1}$$

The transformation formulas from the pixel coordinates of the depth image to the actual coordinates are as follows:

$$\begin{cases} x_{world} = (x_{image} - \dfrac{w}{2})(z_{world} + D')F\dfrac{w}{h} \\ y_{world} = (y_{image} - \dfrac{h}{2})(z_{world} + D')F \\ z_{world} = d \end{cases} . \tag{2}$$

The actual coordinates of the joint points can be obtained by combining formula (1) and formula (2). Finally, the distance between the two joint points can be obtained by using Euclidean distance.

$$D(X,Y) = \sqrt{(x_1 - y_1)^2 + (x_2 - y_2)^2 + (x_3 - y_3)^2} \tag{3}$$

After getting the distance between the nodes, we can judge the instructions expressed by the human body's posture behavior by comparing the motion patterns of the joint points in the existing data in the database .By identifying and returning instructions, the UAV can operate at a relatively long distance through the user's attitude, thus achieving good human-computer interaction .

3.2 User Identity Recognition

At present, face recognition technology is becoming more and more mature. It can identify the consignee's identity and transfer the control rights of UAV by recognizing face .User identity recognition system can confirm user's identity through user's face information and voice information. Head recognition can be trained by deep learning method of human head image, and searched in the range of human body position that has been found. After finding the head position, the user can read the face information and control UAV to adjust its angle, then the face image and the human body position can be trained .The voice information is transmitted back to the background, and the facial feature points and voiceprint feature points are compared through the network connection to the identity database of the logistics company to confirm the user's identity. By this way, the UAV's safe operation authority can be transferred to the designated personnel.

The algorithm used to recognize facial feature points is that the problem of facial feature point location can be regarded as learning a regression function F, using image I as input and output theta as feature points (face shape): theta = F (I) .That is to say, learning multiple regression functions {f1, . . .Fn-1,fn} to approximate functions

F: Theta = F (I) = FN (fn-1 (. . .F1 (theta 0, I, I), I)

Theta I = fi (theta i-1, I), I = 1, . . .N

The input of cascaded fi for the current function depends on the output of the previous function fi-1. The learning objective of each FI is to approximate the real position of the feature point, and theta-0 is the initial shape .In general, fi does not directly return to the real position theta, but to the difference between the current shape theta I-1 and the real position theta: theta I = theta-theta i-1 .The basic idea of this algorithm is to find the common points on the faces of 68 points.

Compared with the principle of face, we need to increase the number of search feature points. Similarly, we use the above principle to make and train a deep convolution neural network, train it to generate 128 measurements for face, and then get the Euclidean distance of 128D values through the following formula .The system will give it a threshold of Euclidean distance that is considered to be the same person, that is, beyond this threshold we will determine that they are the same person.

Compared with the principle of voiceprint, for voiceprint recognition system, if we start from the point of view of the user's voice content, it can be divided into two categories: content-related and content-independent voice information .Content-related means that the system assumes that the user only talks about the system prompt content or the content allowed in a small range, while content-independent does not limit the content that the user says to include various voice information .The former only needs recognition system to deal with the difference of voice characteristics between different users in a small range . Because the voice content is similar, it only needs to consider the difference of the voice itself. The latter needs to consider not only the specific difference between users' voices, but also the different content because of the unrestricted content .It is difficult to make a difference in speech .At present, there is a technology between them, which can be called limited content correlation. The system will randomly collocate some numbers or symbols. Users need to correctly read out the corresponding content to recognize voiceprint. This randomness makes every voiceprint collected in text correlation recognition have a difference in content timing, which is just in line with the widespread existence on the Internet .Computer digital string (such as digital verification code) can be used to verify identity, or combine with other biological features such as face to form a multi-factor authentication method .

It uses the I Vector framework proposed by N.Dehak in 2009 at the level of machine learning model and the complete training and recognition framework of voiceprint recognition algorithm .(Fig . 3)

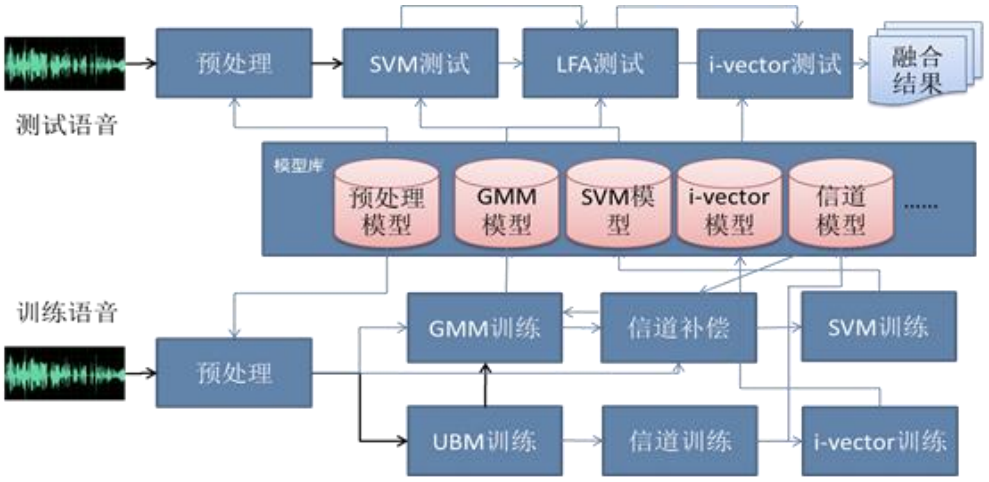

Figure 3 Speech Recognition Flow Chart

The framework consists of three model algorithms:

GMM is a clustering algorithm, and each component is a clustering center .In other words, the parameters of the model can be calculated without knowing the classification of samples (including implicit variables) .Then the trained model is used to differentiate the classification of samples .SVM support vector machine method is based on statistical learning theory and structural risk minimization principle . According to limited sample information, it seeks the best compromise between model complexity and learning ability in order to obtain the best generalization ability .The Bottleneck feature based on DNN replaces the LP acoustic feature in DNN/i-vector model by calculating

sufficient statistics to realize speaker recognition . The essential reason why DNN has powerful classification ability is that it learns more features from data that are more conducive to specific classification tasks .Therefore, DNN is often used as a feature extraction tool, especially for open-set recognition tasks, such as face recognition and speaker recognition. Bottleneck feature is a typical application of DNN as a feature extraction tool. The DNN model contains a layer of hidden layer with fewer nodes, called Bottleneck layer, whose excitation value can be regarded as input signal .A low-dimensional representation, namely Bottleneck feature, is proposed in this paper. While using phoneme-based DNN to achieve more accurate frame alignment, we use speaker-based DNN model to extract Bottleneck features that contain more speaker information to calculate sufficient statistics together to achieve better classification results.

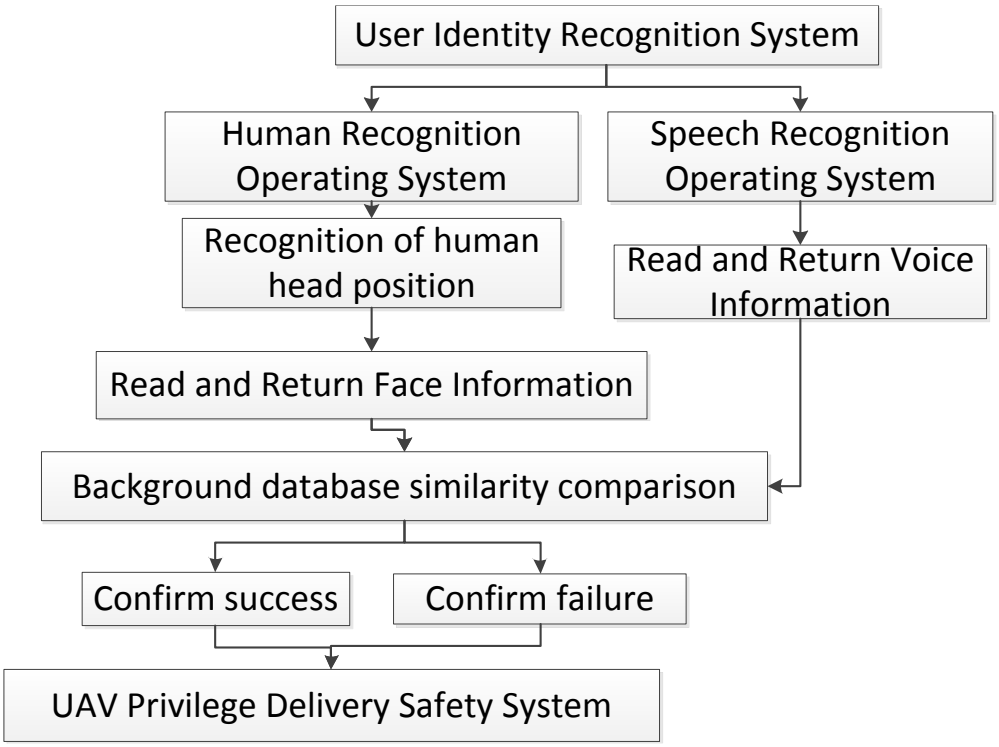

Fig. 4 Flow chart of user identification system

3.3 Control and transmission part

In order to control UAV's relative operation through the result of user identification, when the user's identity is confirmed, the UAV's operation authority will be delivered to the user, and then the user will operate the UAV .When the user's identity is wrong, the UAV will wait for a period of time with the same operation authority. When the UAV waits for a period of time, it will automatically return to the pick-up site.

The human posture operating system recognizes the human posture by using the human skeleton, and then operates the UAV by posture, so that it can hover or land in a specific position .Speech recognition operating system can recognize the user's instruction content through the user's voice acquisition and return to the background using the interface of the existing speech recognition platform .If there are corresponding instructions in UAV flight control system through platform semantics, UAV broadcasting operation and implementation of instructions content can be carried out . In the middle of the operation, users can carry out additional instructions to meet the human-computer voice interaction in a relatively close range.

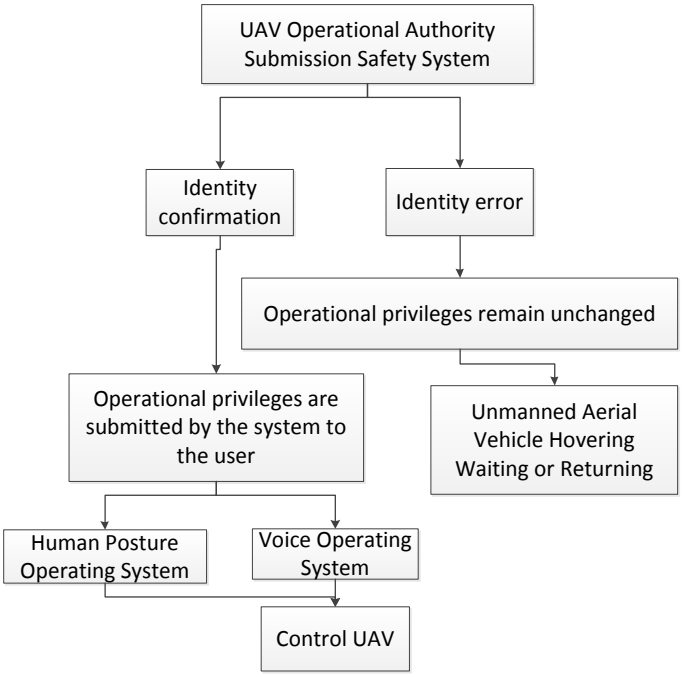

Figure 5 Flow chart of UAV Operational Authority Submission Safety System

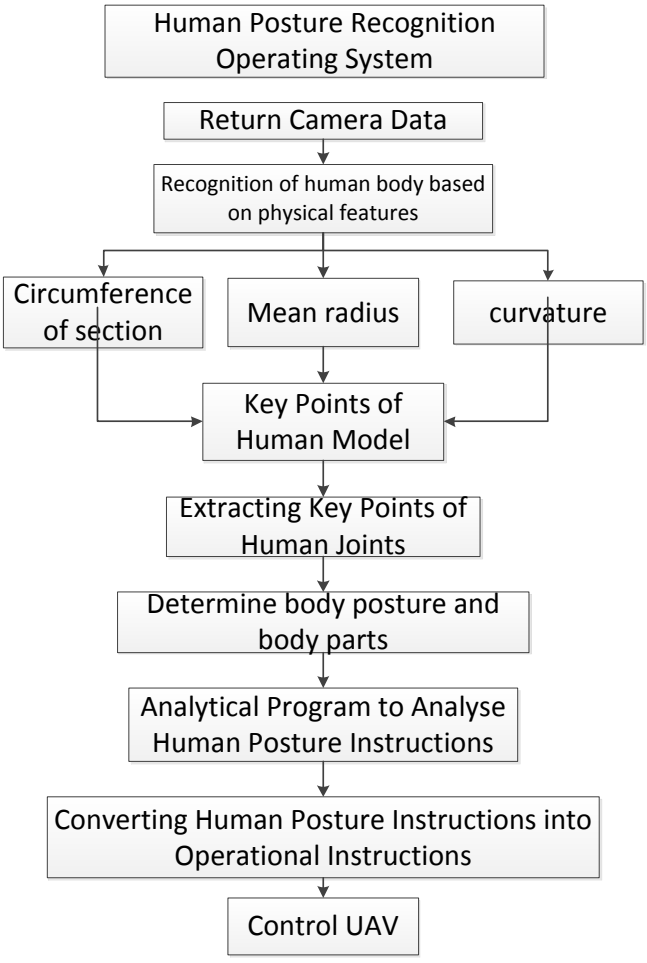

Figure 6 Flow chart of human body attitude recognition operation system

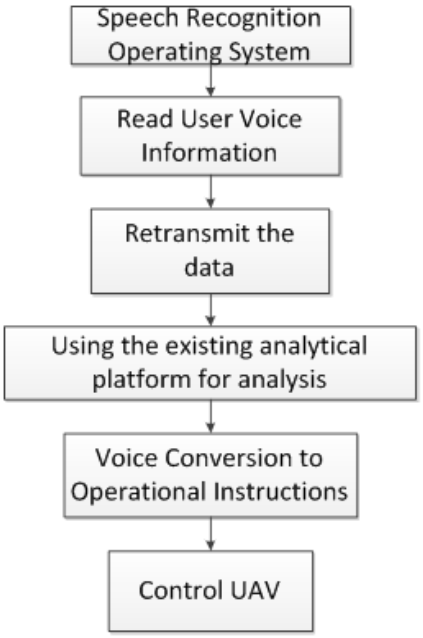

Figure 7 Flow chart of speech recognition operation system

5G network can transmit image data in real time, so that real-time high-speed human-computer interaction can be realized. The autonomous flight control system of UAV can control the UAV through the identified commands, achieve automatic obstacle avoidance through the ultrasonic and laser scanning devices, and can fly to the destination without obstacles at the initial stage of dispatch. It can improve the safety of operation by specific self-control when the user misoperates or makes the UAV possible accidents.

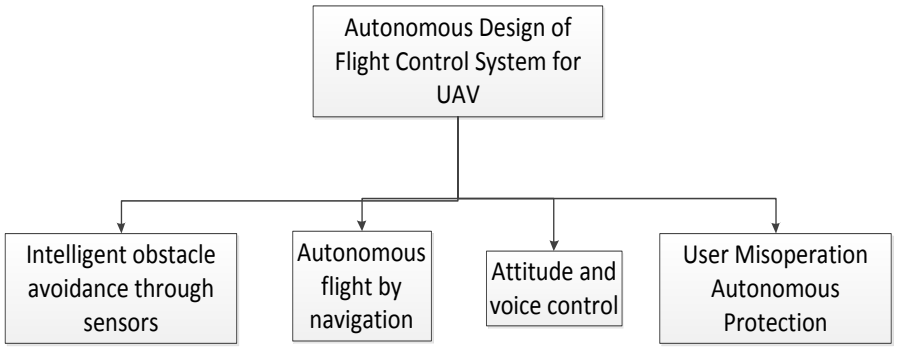

Figure 8 Composition Diagram of Autonomous Design Flight Control System for UAV

4. Conclusion

A logistics delivery method for UAV human-computer interaction using human body attitude information and voice information is designed .The design principle of the whole system can greatly improve the autonomous intelligence of UAV. The optimization of the UAV's use in the logistics industry will reduce the manual operation and realize the full automation of logistics delivery.

Acknowledgements

Natural Science Foundation.

References

[1] Fraundorfer F, Heng L, Honegger D, et al. Vision-based autono-mous mapping and exploration using a quadrotor MAV [C] IEEE/RSJ International Conference on Intelligent Robots and Systems, IEEE, 2012:4557-4564.

[2] Shen S, Michael N, Kumar V. Autonomous indoor 3D exploration with a micro-aerial vehicle [C] IEEE International Conferences Robotics and Automation, IEEE, 2012:9-15.
[3] Dalal N, Triggs B. Histograms of oriented gradients for human detection [C] IEEE Computer Society Conference on ComputerVision and PatternRecognition, CVPR2005, IEEE, 2005:886-893.
[4] GirshickR, Donahue J, Darrell T, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C] Computer Vision and PatternRecognition, IEEE, 2014:580-587.
[5] Uijlings JRR, Sande K E, Gevers T, et al. Selective search for object recognition [J]. International Journal of Computer Vision, 2013, 104 (2): 154-171.
[6] Chen Guiliang, Liu Yuxin, Guo Shijie, etc. A kind of transfer and care robot: 201720534050.7 [P].2017-05-15.Chen G L, Liu Y X, Guo S J, et al.A transfer-care robot: 201720534050.7 [P]. 2017-05-15.
[7] Pishchulin L, Insafutdinov E, Tang S, et al. DeepCut: Joint subset partition and labeling for multi-person pose evaluation [C]//E E Conference on Computer Vision and Pattern Recog-nition. Piscataway, USA: IEEE, 2016: 4929-4937.
[8] Ma Zhao, Li Yibin. Two-dimensional human posture estimation based on multi-level dy namic model [J]. Robot, 2016, 38 (5): 578-587.Ma M, Li Y B. 2D human pose estimation using multi-level dy-namic model [J]. Robot, 2016, 38 (5): 578-587.
[9] Li Bin. Single depth map human joint location [D]. Wuhan: Huazhong University of Science and Technol ogy, 2012.Li B. Key points of human body location based on single Depthmap [D]. Wuhan: Huazhong University of Science and Technol-ogy, 2012.
[10] Shotton J, Sharp T, Kipman A, et al. Real-time human pose recognition in parts from single depth images [C]//IEEE Confer ence on Computer Vision and Pattern Recognition. Piscataway, USA: IEEE, 2011: 1297-1304.