

Find my mates

Intelligent science and technology
Jiachen Liang
1711523

Abstract

Our task name is "find my matters". The task of the robot is to get more details of the designated guest for the operator who only knows the name of the guest participating in the banquet. In the aspect of vision, the main tasks are to detect human body, record guest information and provide photo taking tasks. In terms of detecting human body and recording guest information, Baidu API is used to realize high-precision recognition of human body characteristics. At the same time, in order to eliminate the possible interference of off-site personnel, wave detection is added to the detection of human body information to improve the accuracy of the detection object. After the determination of the guest, the guest position information is released to the navigation, and the feature information and relative position information of the designated guest are further detected, and finally the guest is photographed.

Key words: find my mates , body detect , feature detect , relative position , taking photos

1 Introduction

1.1 Overall introduction:

The background of the task is that many guests attend the banquet, but the host only knows the names of the guests, so the robot is sent to collect the characteristic information of the designated guests. After the robot detects the banquet guest, it goes to the guest, finds out the designated guest by asking the guest's name, records the relevant information of the guest and returns to the host after taking photos, and introduces the recorded guest information by voice.

This paper mainly introduces the visual part of "find my matters". It mainly involves human detection, wave recognition, feature recognition and photo recording.

2 Technical realization

2.1 Human detection

The first task is to detect the human body and move closer to the human body. That is to find the human target. In this regard, there are two main implementation methods: Baidu API human body recognition interface and human body detection using Yolo.

2.1.1 Yolo detection:

Yolo is a target detection system, which is widely used for its fast and accurate recognition. Yolo applies a single neural network to the full image. This network divides the image into regions and predicts bounding boxes and probabilities for each region. These bounding boxes are weighted by the predicted probabilities.

Yolov3.weights, the training model of Yolo, has been used to realize the accurate detection of human body. In Figure 1, the recognition rate of human body detection with Yolo is relatively high. The detection speed of CPU version of Yolo is very slow, GPU version of Yolo is significantly faster, and video stream can be processed continuously without delay. After using Yolo to detect the object, it will return to identify the type of object and its corresponding position. By judging the object type and moving to the corresponding position, we can find the guest. Yolo detects the pictures in the field of vision, selects the detection results with the category of "person", extracts the coordinates of the upper left corner of the detection box and the corresponding length and width from the detection results, and publishes the regionofinterest type message. However, Yolo has a relatively large demand for computing power. It can't be opened for a long time on ordinary laptops. A long-term opening may lead to a computing crash (CUDA error or memory overflow, etc.). In the stage of looking for customers, we decided not to use Yolo, but to use Baidu API instead.



Figure 1. Yolo detection

2.1.2 Baidu API human detection:

API is the abbreviation of application programming interface. Internet service providers (such as Baidu, Weibo, etc.) have a large amount of data. We can query the data, but we need to follow a certain format / protocol, otherwise the service providers do not know our needs and cannot process them, and we can not understand the data when we get them. Through the API interface specification, both sides of the communication can understand the information and data transmitted by the other side, and also simplify the operation. Baidu API is a series of application program interfaces opened to the public by Baidu promotion platform. Developers can directly interact with search promotion platform by calling API, develop their own application based on API, and realize the management of promotion account and advertisement.

Baidu API has a high recognition accuracy, and runs in the cloud, so it does not need its own notebook with high performance and convenient use. However, compared with Yolo, the recognition delay is large and it is difficult to achieve continuous recognition. Therefore, in order to ensure the consistency and fluency of task realization in the experiment, we perform frame skipping processing on the image collected by the camera. In addition, the identification of Baidu API depends on the smoothness of the network, so it is necessary to ensure that the network is in good condition and highly dependent during the use process.

In the process of using Baidu API for human body detection, because the returned results of Baidu API detection can include the location results of parts of the human body detected, the combination of human body detection and wave detection can be realized through Baidu API. These contents are introduced in detail in the wave detection section.

2.2 Wave detection

At present, there are two methods of wave detection. One is to identify and judge the direction of human fingers through openpose, and detect the posture of human body. Another way is to call the human detection interface of Baidu API to get the position description of human posture. Both of them are determined by the relative relationship among elbow, shoulder and wrist positions, and they have good and stable recognition accuracy in most cases.

2.2.1 Baidu API:

When Baidu API detects human body, the image data collected by the directly subscribed camera cannot be detected. Therefore, the format of the incoming image needs to be stored in the local area first, and then the result of reading the local image into the API function interface.

Recognition rules of waving posture: for the left hand and right hand of the human body, when the position of the wrist joint is higher than the elbow joint and the wrist joint is outside the shoulder joint, the human body is considered to be waving. Record all detected position information of waving human body.

Considering the principle of women first, we call Baidu API to further detect the gender of the waving guests screened from the top. If it is a female guest, we will release the location information of the guest and advance in the direction of the guest. Otherwise, a guest's location information will be released from the detection object.

2.2.2 Openpose

Openpose[1] is an open source human pose estimation model based on Caffe framework. It can detect key points of human, such as neck, elbow, wrist, shoulder, etc. This is a fast matching method, which first detects the key points, and then matches the key points belonging to the same person. As shown in Figure 2, an RGB image is input, which will extract features through convolution network. Then, it is divided into two branches, one is to extract partial confidence mapping, the other is to calculate partial associated fields (PAFS). Then it uses two part matching to get the part Association, and establish an undirected graph to describe the possibility that different key points belong to the same person. Finally, based on PAFS, ungarian algorithm is used for multi-person parsing.

Openpost recognition speed is fast, but it needs a higher computing power support, higher requirements on the computer, and its installation and configuration process is also more cumbersome. Therefore, in the final experiment, we decided to use openpost as an alternative, and use Baidu API to judge posture in the experiment.

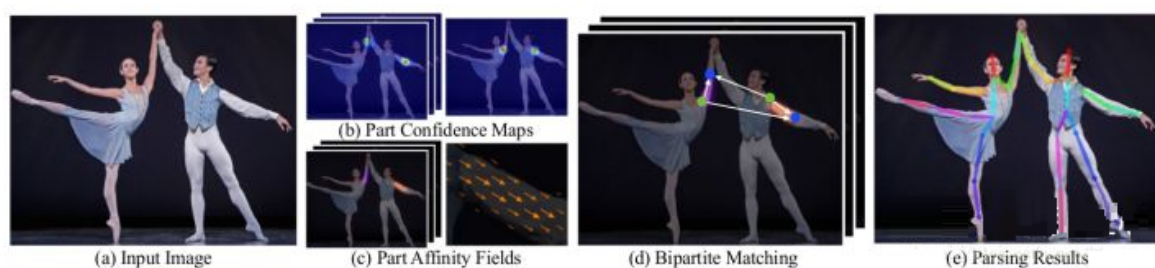


Figure 2. Openpose recognition

Feature detection

Feature detection uses the human feature and face feature detection module of Baidu API, which provides a variety of attribute detection for human body. In the experiment, we mainly select six features: gender, glasses, skin color, age, coat type and coat color. After identifying the corresponding features, we release feature messages to the voice node for later broadcast of guest information description. We directly use the human body position detection information released in the previous human body detection to detect the characteristics of the guest image in the corresponding human body detection frame, and limit the number of detected faces in the image to one face, so as to ensure that the detection information can be matched one by one without mismatch.

Due to the detection of top color, baidu API returns Chinese information, which is not conducive to the English broadcast of information after the robot, so the corresponding color will be converted into English message after the detection of top color.

Figure 3 and Figure 4 show the result of feature detection using Baidu API.



Figure 3. body detection

```
"person_info": [
  {
    "attributes": {
      "upper_wear": {
        "score": 0.77748852968216,
        "name": "短袖"
      },
      "gender": {
        "score": 0.71456223726273,
        "name": "女性"
      },
      "upper_color": {
        "score": 0.78069370985031,
        "name": "棕"
      }
    },
    "location": {
      "height": 644,
      "width": 340,
      "top": 0,
      "score": 0.98054593801498,
      "left": 109
    }
  }
],
```

Figure 4. detection result

Spatial relative location

We also need to describe the spatial location of the guest. To finish it, we use Yolo detection combining with point cloud data to judge.

We use technology of human body detection to detect the position of the person's field of vision, and use Yolo detection to find the position of the furniture's field of vision. Through the point cloud data, we calculate the space coordinates of people and the space coordinates of objects. And then we need to determine the most suitable description of the set of items. To do this, we calculate the distance between each item and the person, once the threshold is exceeded, such as more than 3m, we do not use this item for description. Go further, we choose the most appropriate from the collection of items. We calculate the variance of the point cloud at the position of the item. Since the larger the value, the less accurate the position determination of the item. As a result, we choose the one with the smallest variance as the description.

Taking photos

When the robot finds the target guest, it will have a simple communication with him and ask if the robot is allowed to take photos for the guest. After the guest's consent, the robot will take photos.

In order to ensure that the guests are in the center of the picture and the size of the human image is appropriate, the haarcascades face recognition feature detector of OpenCV is used. By setting the cascade table parameters, the face recognition effect of haarcascades feature detector is optimized. For the detected face information, by comparing the location coordinates of the center point of the detection box, remind the guest to move up, down, left and right; by judging the length and width of the detection box, remind the guest to move forward or backward away from the camera. When the guest's face position is appropriate, the robot informs the guest to take photos and store them.

Future work

In the actual scene, the robot needs to recognize the human body in the designated area, and the wave detection will appear clumsy. In the experiment process, it is found that because everyone's wave habits are not the same, so it depends on the relative position of wrist joint, shoulder joint and elbow joint proposed in the previous article to judge, it is likely that the robot will not be able to detect the customer's wave, and the recognition process It's rigid. After that, it may be considered to add the judgment of position information in the recognition process, so as to limit the detection range to the specified area.

Improvement: in the process of waving detection, when the guests are dense or the robot is not rotating in one direction, it may lead to secondary detection and inquiry of the detected guests. Here, it may be considered to establish a face database, that is, for the detected guests, record their information into the face database, and after each detection of the guests, first determine whether the guest is in the existing face database, if it is a strange guest, move it to the front for further communication; if it is a recorded guest, do not deal with it. Here, we mainly use the face database building technology of Baidu API to write the collected face information into the face database, and support the recognition of whether a given face is a member of the face database.

Although the accuracy of using Baidu API is very high, it has potential instability and uncertainty due to the dependence on the network and the delay of processing speed. At present, because the robot and notebook cannot be shared in the school, it is not convenient to use Yolo detection and openpose for object detection and pose recognition. If there is any condition in the future, the effect of task execution will be tested by combining Yolo and openpose.

References

[1] Zhe Cao and Gines Hidalgo and Tomas Simon and Shih-En Wei and Yaser Sheikh. OpenPose: realtime multi-person 2D pose estimation using Part Affinity Fields[C]. 2018 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). IEEE Computer Society, 2018.