Group Member： Jiachun Qi

School： Experimental School Affiliated to Jilin University

Province： JiLin

Country： China

Instructor： Jinhua Li、Liang Hu

Title ： Table tennis recognition and positioning based on monocular vision and Faster-RCNN

Innovative Statement

The participating team stated that this submitted paper was the research work and the research results under the guidance of the instructor. As far as I know, the paper does not contain research results that have been published or written by other people or teams, except as specifically noted and acknowledged in the text. If there is any dishonesty, I am willing to bear all relevant responsibilities.

Team Members：Jiachun Qi **Instructor：Jinhua Li, Liang Hu**

Signature： ________ **Signature：** ________

2018/9 / 13

Abstract

Aiming at the robot vision system for picking up ping-pong balls, the ping-pong recognition algorithm based on Faster-RCNN frame combined with the ground plane constraint localization model of a single frame image was applied to the ping-pong picking robot, so as to achieve the recognition and spatial orientation of ping-pong balls in the image.

For the problem of misrecognition of ping-pong balls based on edge recognition, an improved algorithm based on Faster-RCNN frame for accurate detection and recognition of ping-pong balls is proposed. Firstly, the ping-pong images in various scenes are learned by training the deep network, and then the parameters of model are applied to the actual ping-pong images to obtain the position of the target object in the image and the pixel coordinates of the center of the ping-pong ball. The advantage of the method is proved by experiments, and the issue of accurate recognizing ping pong balls in various scenes is solved.

For the problem of table tennis localization in monocular vision, a ground plane restriction positioning model of a single frame image is proposed. According to the characteristics of ping pong balls and the actual situation of positioning application, this paper establishes a localization model of ground plane constraint based on the three-dimensional geometry and perspective projection principle is established. The center of a ping pong ball's position in the world coordinate system can be calculated by its pixel coordinates in the image, and then the function of table tennis localization was realized.

The results of experiments show that the system can correctly detect and identify ping-pong ball, and carry out spatial localization and ranging on it. Good experimental results of the algorithm improvement and positioning model of the application scenario are obtained and the effectiveness of the proposed algorithm is proved.

Keywords: monocular vision, Faster-RCNN, ping-pong ball, localization model of ground plane constraint of single frame image

Contents

1 Introduction

1.1 Project background and the scene used

After school, especially during the holidays, when I was learning to play table tennis in the gymnasium, every time I was exhausted and had to pick up the balls on the ground after the training. Since then, I started nurturing the idea of designing a ping-pong ball pickup robot. In October 2017, I met Li Haoyuan, a friend of mine who is in college but had returned home for the National Day holiday. The common hobbies and interests made us talk about the “2018 National College Students Digital Product Design Competition”. The theme of this competition is "ball robot", including the ball projection robot, the ball pickup robot and the service robot in stadium, which further strengthens my idea of designing a ping-pong ball pickup robot.

Table tennis is China's national ball, and it is very popular. After training in table tennis venues, there will be hundreds of scattered table tennis balls on the ground, and manual ball-picking usually takes time and effort. Can I design a smart table tennis robot? Can the robot recognize the table tennis balls on the ground and also measure the distance and position to the ball? These were some of the questions I asked myself.

In Fig. 1.1, a standard table tennis court, 14m long and 7m wide. First, the ping-pong pickup robot I intend to design will walk along the dotted line in the figure and blow the ping-pong balls (yellow circles) into the area within 2 meters from the wall. Next, the ping-pong pickup robot will identify the balls in its field of view through the monocular vision system designed in this project and then walk according to the positioning information, and the balls in the field will be sucked into the device one by one.

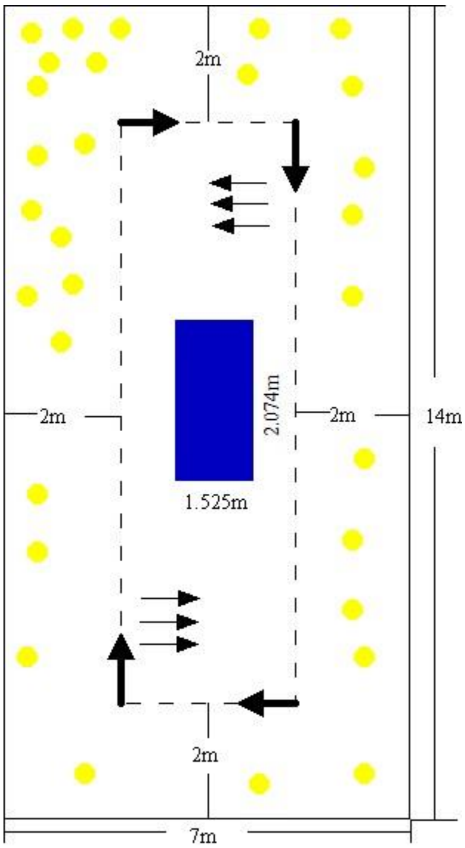


Fig 1.1 Schematic diagram of the usage scenario

The visual system is an important part of the vision-based ping-pong ball pickup robot. The robot uses the built-in camera to identify and locate the balls scattered on the ground. The main difficulty lies in how to accurately identify and position the balls in complicated settings, and whether the monocular vision can meet specific requirements. These are the main issues to be solved in this paper. Machine vision, as the entrance to artificial intelligence and the visual information of the Internet, is a very important foundation for the machine to perceive the world. The solution of these problems has important research significance for the detection of targets, the positioning of mobile robots, visual navigation, and target tracking.

1.2 Research Background

In the study of Ping-Pong ball recognition，Ling Bao developed a table tennis ball system of the embedded architecture [1]. This automatic ping-pong pickup system integrated embedded technology, measurement control, and image processing. He used the Hoff round transform algorithm and centroid calculation of table tennis and their location to identify balls. But all the circles contained in the image are considered, that is, the circle contained in the image is considered to be a table tennis ball, and the actual scene will include factors such as the field baffle and the illumination change, which may affect the detection result, and there may be a case of misidentification.

Ji Yunfeng et al. designed an image recognition algorithm to identify table tennis ball in game pictures, based on a large number of functions in OpenCV vision library [2]. The recognition of table tennis in game pictures was achieved through binarization, setting ROI area, corrosion and expansion, and finding pixel blocks. The algorithm achieves a good recognition result, but has a large time complexity. For an application system with high real-time requirements, such as ping-pong ball pickup robot, it did not meet the time requirement.

Wang Xiaolong et al. designed an automatic ping-pong system which integrated embedded image processing and measurement control technology, based on the third-party library OpenCV [3]. Table tennis balls in the video were identified by Hough circle transform and SVM classification algorithm. The training of the SVM classifier requires a large amount of sample data, and the composition of the sample data has a great influence on the performance of the classifier. In addition, the training of the classifier also takes a certain amount of time. This is not suitable for this particular application scenario for different table tennis venues.

For positioning, in the “automatic picking up table tennis embedded system based on OpenCV”, its precise positioning module is mainly done by color sensor [3]. In the “design and realization of table tennis picking robot”, the vision module displays the captured image on an LCD. The position of the ball on the LCD corresponds its actual position and while the distance is detected with an ultrasonic module [8]. In the utility model patent "a kind of intelligent table tennis picking robot", during the running of the robot, the ultrasonic sensor always detects whether there is any table tennis ball scattered in front of the car, if not, it keeps driving, and continuously checks whether there is table tennis ball. If there is table tennis ball, it will control the manipulator action and throw the table tennis ball into the basket.

In all, the main research focus of this paper includes the monocular vision technology and the Faster-RCNN framework of the table tennis ball recognition algorithm. The improved Hough transform algorithm will be used to accurately identify the center of mass and radius of the table tennis ball as well as the ranging and the localization model based on single image by horizon constraint.

1.3 Algorithm flowchart

This project will use the Faster-RCNN deep learning framework for table tennis ball recognition, combined with the ground plane constrained positioning model based on single frame image. This will be applied to static ping-pong ball detection to realize the table tennis recognition and positioning function in the image. The working principle of the proposed algorithm is shown in Fig. 1.2.

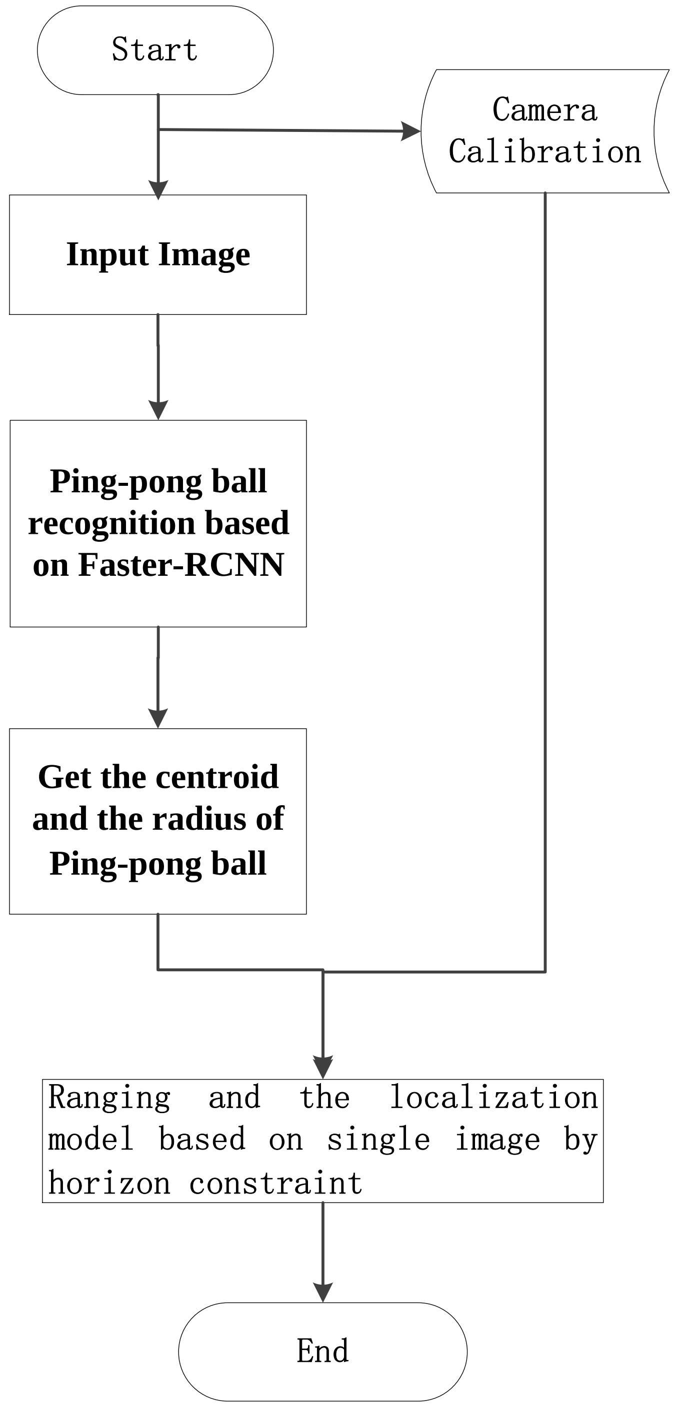


Fig 1.2 Flow chart of table tennis recognition and localization algorithm

2 Algorithm of Detection and Identification of ping-pong Ball

2.1 Principle of Faster-RCNN

Faster-RCNN is a framework for region detection and recognition of images. By feeding the target image into CNNs network for training, feature extraction and region generation, network algorithm are used to detect the target object.

Faster-RCNN (R corresponds to "Region") is one of the best algorithms for target detection based on deep convolution network. This algorithm combines RPN network and CNNs network, connects the target areas obtained by RPN directly to ROI pooling layers. It is an end-to-end target detection CNNs network.

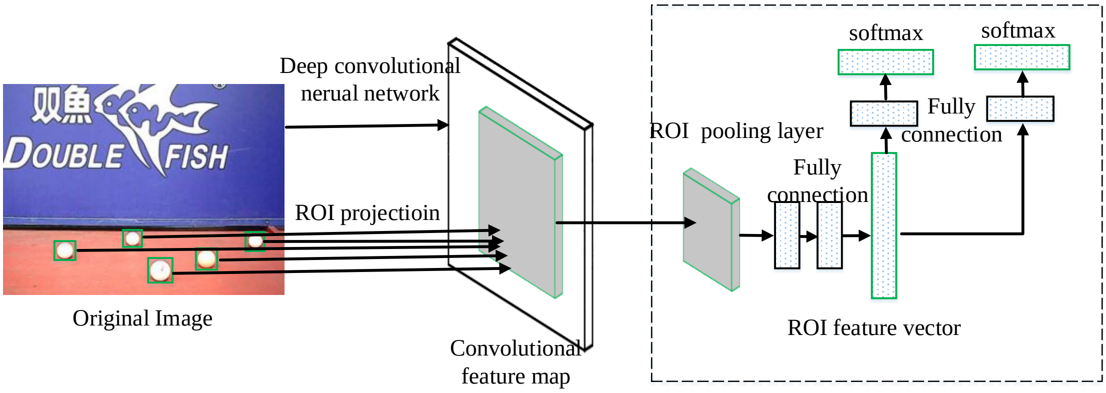


Fig. 2.1 Faster-RCNN system framework for detection of ping-pong ball

2.2 Data preprocessing

In order to detect and recognize the ping-pong ball in the picture captured by the monocular camera, it is necessary for the computer to conduct supervised learning. The learning process is a training process based on the deep convolution neural network. This process is accomplished by a PC machine with better performance. Faster-RCNN algorithm is trained and verified on the dataset with VOC2007 format, and achieved good learning results. Based on the VOC2007 dataset format, the images are replaced by the ping-pong ball images, and trained after labeling the areas with ping-pong balls. All data needed are organized with the following formats:

(1) All images with ping-pong ball are moved into the folder with "JPEGImages" name. A total of 425 ping-pong ball images in two different scenes were captured, with the name: 000001.jpg, 000002.jpg …… and so on, some of them are shown in Fig. 2.2.

Fig 2.2 Part of ping-pong ball training dataset

(2) The positions and types of ping-pong ball need to be labeled in the training dataset, in order to detect the area and identify the type of ping-pong ball accurately by the neural network. This process is finished by the XML label file. Each image corresponds to an XML file, and these XML files are placed uniformly in the Annotations folder. The specific example of the XML file is shown in Fig. 2.3.

```xml
<annotation>
    <folder>JPEGImages</folder>
    <filename>000001.jpg</filename>
    <path>.\VOC2007\JPEGImages\000001.jpg</path>
    <source>
        <database>Unknown</database>
    </source>
    <size>
        <width>500</width>
        <height>375</height>
        <depth>3</depth>
    </size>
    <segmented>0</segmented>
    <object>
        <name>ball</name>
        <pose>Unspecified</pose>
        <truncated>0</truncated>
        <difficult>0</difficult>
        <bndbox>
            <xmin>295</xmin>
            <ymin>326</ymin>
            <xmax>333</xmax>
            <ymax>366</ymax>
        </bndbox>
    </object>
</annotation>
```

Fig 2.3 the format of XML file

As shown in Fig. 2.3, the size of the image is specified by the "size" tag, and all ping-pong ball positions are specified by the "object" tag. The "bndbox" tag, which may have more than one, indicates the location of the Ping-pong ball. In order to label images accurately, an open source software called "label image"[13] was used to label images. The specific annotation process is shown in Fig. 2.4.

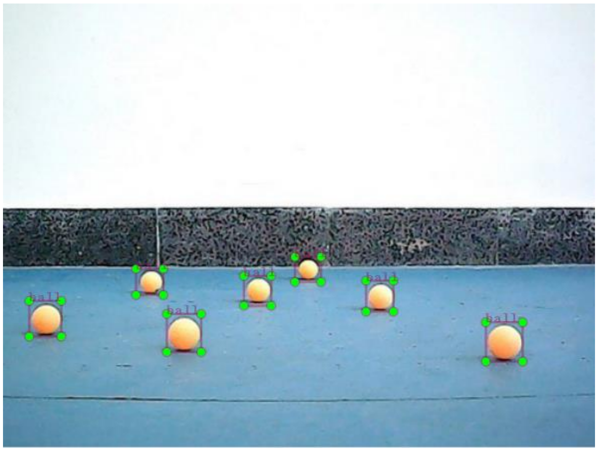


Fig 2.4 Marking ping-pong positions and types by using "labelImage" tool

The training process is initiated after the prepared data. In order to reduce the training time and keep the accuracy rate, the number of training iterations is set to 4000 times. Finally, the training model is saved into CKPT format, which saves the weights and bias value of the model. The training process is based on the Tensorflow framework, which organized the training process with Graph, open sourced by Google.

The training process of the model is finished under the PC machine with the following configurations:

CPU: Intel I5 7300HQ 2.5GHz, 4 kernels.

Memory: 8GB

Video card: NVIDIA GeForce GTX 1050Ti, with 4GB memory.

OS: Windows 10 Home Edition

Python: Version 3.5

Tensorflow: Version 1.2

After the training, the model file with the "vgg16_faster_rcnn_iter_4000.ckpt" header is copied into the development board for detecting the ping-pong ball in the working environments.

In the actual scenes, the model is used to identify ping-pong balls. The system was able to achieve the speed of 5 frames per second, which meets the real-time requirements of the robot. The specific result is shown in Fig. 2.5.

Fig. 2.5 Recognizing results of ping-pong ball in the actual scenes

3 Ranging and the localization based on Monocular vision

3. 1 Camera Calibration

In monocular vision ranging and location, camera calibration is needed to obtain the parameters in the ranging and location model. The process of camera calibration is to convert the pixel plane captured by the camera to the reference imaging plane in spatial space. The parameter calibration from the pixel plane to the camera imaging plane is internal parameter calibration, and the parameter calibration from the imaging plane coordinate system to the reference coordinate system is external parameter calibration.

According to the linear model of the camera, for any point P in the spatial space, its world coordinate is $P_w=[X_w, Y_w, Z_w]^T$, and in the camera coordinate system, it is $P_c=[X_c, Y_c, Z_c]^T$, and projected into the pixel coordinate system, it is $p_{pix}=[x_{pix}, y_{pix}]^T$. The geometric model of the camera can be expressed as follows:

$$Z_c\begin{bmatrix} x_{pix} \\ y_{pix} \\ 1 \end{bmatrix}=\begin{bmatrix} f_x & 0 & x_0 & 0 \\ 0 & f_y & y_0 & 0 \\ 0 & 0 & 1 & 0 \end{bmatrix}\begin{bmatrix} R & T \\ 0 & 1 \end{bmatrix}\begin{bmatrix} X_w \\ Y_w \\ Z_w \\ 1 \end{bmatrix} \tag{1}$$

Where, R is orthogonal rotation matrix, T is translation matrix, three parameters are needed to determine R, and three parameters are needed to determine T, these six parameters are called external parameters of the camera. As long as the relative position of the world coordinate system and the camera coordinate system changes, R and T will change.

The parameters needed for calibration are f_x, f_y, x_0 and y_0, which are also called the internal parameters of the camera. Here, f_x and f_y denote the equivalent focal length of f in the direction of X_{pix} and Y_{pix}, respectively, while (x_0, y_0) denotes the coordinates of the center of the projection plane in the pixel coordinate system. Monocular camera calibration is to compute internal parameters by equations when

p_{pix} in the pixel coordinate system and P_w in the world coordinate system are known. This topic will use the Zhang Zheng You calibration method to calibrate the camera.

3. 2 Ranging and the localization model based on single image by horizon constraint

According to the actual application scene of table tennis location, combined with the principle of stereo geometry and perspective projection, this paper intends to establish a constrained location model. That is, when the camera and the target object are on the same horizon, and the camera has a certain degree of inclination in the vertical direction, the constrained location model can be established, as shown in Fig.3. 1. The point C is the center of the camera; its vertical projection O_w on the horizon can be treated as the origin; the projection direction of the optical axis on the horizon is X_w; the camera direction on the vertical ground is Z_w; and the world coordinate system is established with the right direction of the vertical X_w camera as Y_w.

As illustrated in Fig. 31, Cc is the effective focal length f of the camera, and CO_w represents the height h of the camera, (x_c, y_c) is the intersection point of the optical axis of the camera and the image plane, and it is also the coordinate origin of the image plane coordinate system, usually set (0, 0). The coordinates of the centroid P of table tennis on the ground in the image plane coordinate system are expressed as (x_p, y_p), θ is the cameras' angle, and the radius of table tennis ball is r.

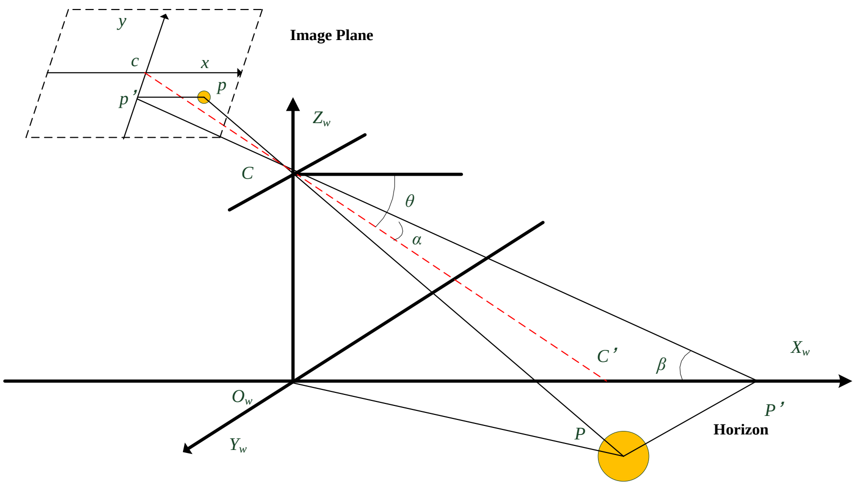


Fig 3.1 Ranging and the localization model based on single image by horizon constraint（a）

As illustrated in Fig. 3.1, $\beta=\theta-\alpha$, $\alpha=\arctan(\frac{y_p}{f})$, according to the above illustrations, the spatial coordinates of the centroid P of table tennis can be treated as (X_w, Y_w , Z_w).

(1) Compute the ordinate value $X_w(P)$ of table tennis centroid P in the world coordinate system.

In $\Delta CO_wP'$:

$$X_w = O_wP' = \frac{h-r}{\tan\beta} = \frac{h-r}{\tan(\theta - \arctan(\frac{y_{pix}-y_0}{f_y}))} \tag{2}$$

$X_w(P)$ is the ordinate value of table tennis centroid P in the world coordinate system.

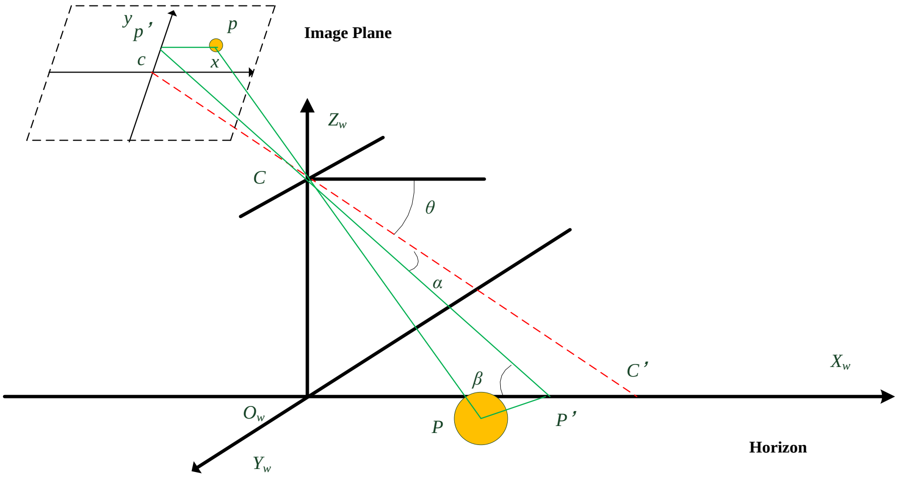


Fig 3.2 Ranging and the localization model based on single image by horizon constraint (b)

As illustrated in Fig 3.2, when the centroid P projection coordinates P' coordinate value is less than the projected C' coordinate value of the optical center C, $\beta=\theta+\alpha$,then:

$$X_w = O_wP' = \frac{h-r}{\tan\beta} = \frac{h-r}{\tan(\theta + \arctan(\frac{y_{pix}-y_0}{f_y}))} \tag{3}$$

(2) Compute the abscissa value $Y_w(P)$ of table tennis centroid P in the world coordinate system.

As illustrated in Fig. 3.1, $\Delta pCp' \cong \Delta PCP'$, and then $\frac{pp'}{PP'} = \frac{Cp'}{CP'}$,

$$PP' = \frac{CP'}{Cp'} \times pp'.$$

In $\Delta CO_wP'$, $CP' = \frac{h-r}{\sin\beta} = \dfrac{h-r}{\sin(\theta - \arctan(\frac{y_{pix}-y_0}{f_y}))}$.

$$Cp' = \sqrt{f^2 + (y_p - y_c)^2} \,.\, pp' = (x_p - x_c)$$

And then,

$$Y_w(P) = PP' = \frac{CP'}{Cp'} \times pp' = \frac{(h-r)\times(x_p - x_c)}{\sin(\theta - \arctan(\frac{y_{pix}-y_0}{f_y}))\times\sqrt{f^2 + (y_p - y_c)^2}} =$$

$$\frac{(h-r)\times\dfrac{f(x_{pix}-x_0)}{f_x}}{\sin(\theta - \arctan(\frac{y_{pix}-y_0}{f_y}))\times\sqrt{f^2 + \left(\dfrac{f(y_{pix}-y_0)}{f_y}\right)^2}} = \frac{(h-r)\times\dfrac{(x_{pix}-x_0)}{f_x}}{\sin(\theta - \arctan(\frac{y_{pix}-y_0}{f_y}))\times\sqrt{1+\dfrac{(y_{pix}-y_0)^2}{f_y{}^2}}}$$

$$= \frac{(h-r)\times(x_{pix}-x_0)\times f_y}{\sin(\theta - \arctan(\frac{y_{pix}-y_0}{f_y}))\times\sqrt{f_y{}^2 + (y_{pix}-y_0)^2}\times f_x}.$$

$Y_w(P)$ is the abscissa value of table tennis centroid P in the world coordinate system.

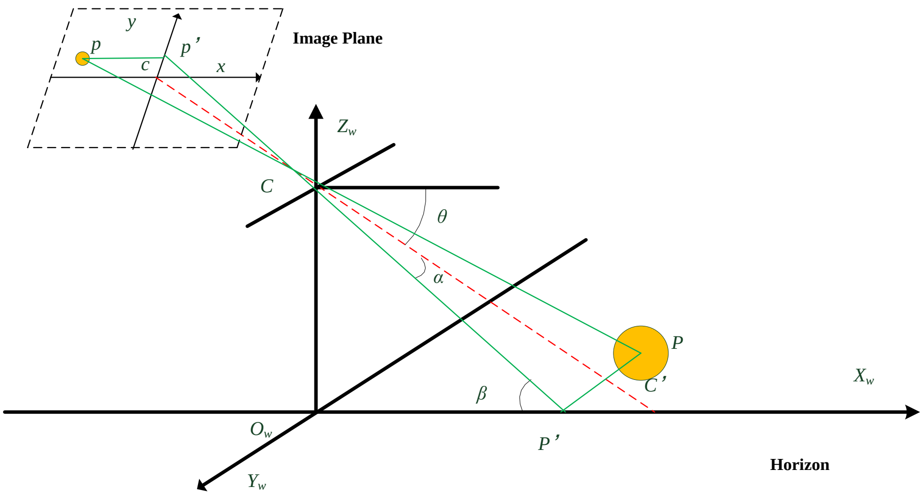


Fig 3.3 Ranging and the localization model based on single image by horizon constraint (c)

As illustrated in Fig 3.3, when the P' coordinate of the center of centroid P is less than the C' coordinate of the optical center C and the center of centroid P is on the left side of the Xw, $\beta=\theta+\alpha$, and:

$$Y_w = \frac{(h-r)\times(x_{pix}-x_0)\times f_y}{\sin(\theta+\arctan(\frac{y_{pix}-y_0}{f_y}))\times\sqrt{f_y{}^2+(y_{pix}-y_0)^2}\times f_x} \tag{4}$$

(3) Compute the distance between table tennis centroid P and camera base O_wP.

From the above conclusion,

$O_wP=\sqrt{(O_wP')^2+(PP')^2}=\sqrt{(Y_w(P))^2+(X_w(P))^2}$.

4 Experimental Results and discussion

4.1 Initial construction of the robot model

In order to verify the proposed table tennis ball recognition algorithm and the monocular vision ranging and positioning model, a robot was initially constructed, as shown in Fig. 4.1. The 360° all-round aspirating table tennis intelligent collection robot model consists of six main modules: master terminal, monocular vision collector, calculation control software, communication module, omnidirectional motion controller and auxiliary module (including Drive motor, power management circuit, battery.) The logical structure of the robot system is shown in Fig.4.2

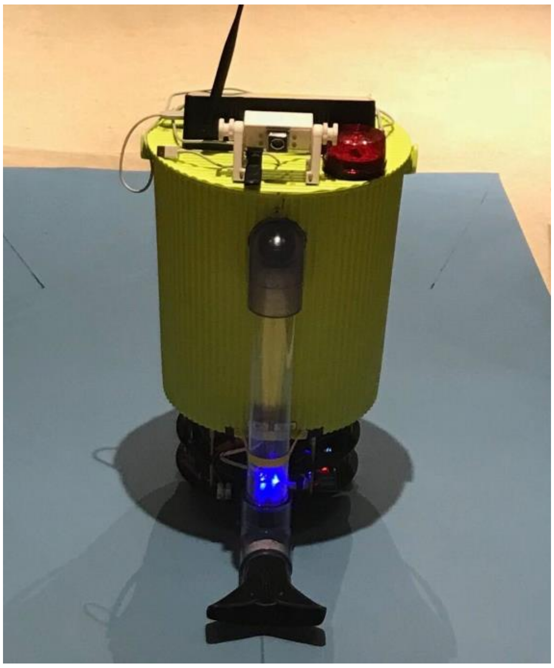


Fig 4.1 the robot model

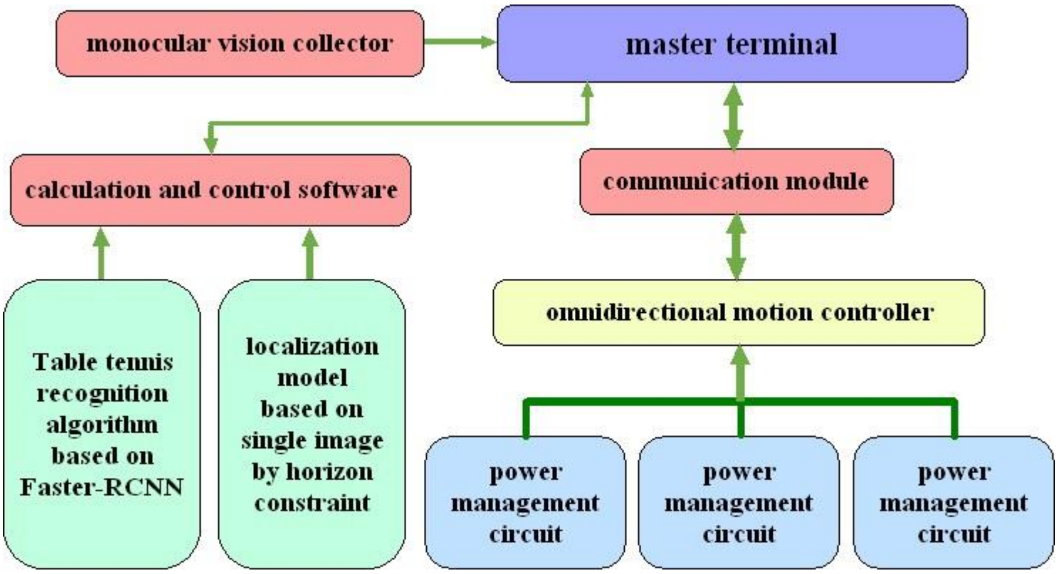


Fig 4.2 the logical structure of the robot system

（1）Master terminal

This module is the core control module. It is responsible for receiving the real-time image data of the “Monocular vision collector”, and processing the image data and then submitting it to the “Calculation Control Software” for analysis and identification, and using the “Communication Module” to send data such as positioning and distance to “omnidirectional motion controller, which controls the "drive motor" for accurate displacement, Drawing table tennis, counting and other functions.

（2）Monocular vision collector

The monocular vision sensor system uses a CMOS chip camera, and its imaging principle can be approximated to small aperture imaging. The principle of small aperture imaging states that the smaller the distance between the object and the camera, the larger the image in the camera coordinate system, and vice versa. I used the S103 MoFang camera in the experiment. The performance parameters of the camera are shown in Table 4.1. The camera has a focal length f = 3.5mm while the resolution is 640 × 480. The camera internal parameters were calibrated using the calibration toolbox in Matlab R2015b software. During the experiment, the height of the camera remains the same and keeps the camera's top view angle fixed.

Table 4.1 Camera performance index

Type	Pixel	Resolution	Interface	Sensor	Frame rate	Focal length
S103	1300000	640*480	USB2.0	CMOS	30 f/s	3.5mm

（3）Calculation and control software

After the image data sent by the "monocular vision collector" is processed by the Faster-RCNN detection and recognition algorithm, the centroid and radius of the table tennis ball are calculated, and then the monocular vision ranging and positioning model is setup to calculate the distance and the positioning Data such as spatial position coordinates.

（4）Communication Module

"Communication module" is the data exchange module of the "master terminal" and "omnidirectional motion controller". It adopts serial data communication protocol. The baud rate is 115200, 8 data bits, 1 stop bit, and including parity bit.

（5）omnidirectional motion controller

The "omnidirectional motion controller" receives the positioning, distance and other data sent by the "master terminal" through the "communication module", which in turn converts the data into drive commands and controls the angle, number of revolutions, direction, etc. of the "drive motor". Then it will drive the whole move to the position of table tennis, activate the “negative pressure” fan, and draw the table tennis ball into the “capacity bucket”. After each successful draws of table tennis balls, the counter is incremented by one and the result is transmitted to the “master terminal”.

The controller consists of STM32F103C8T6 series chips from STMicroelectronics, as shown in Fig. 4.3.

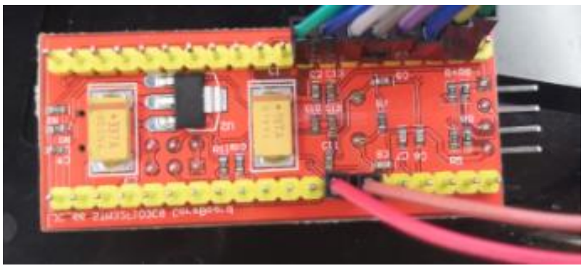


Fig 4.3 omnidirectional motion controller

（6）Auxiliary module

The auxiliary module consists of the drive motor, power management circuit, battery, etc., as shown in Figure 4.4-4.6.

Fig 4.4 Drive motor

Fig 4.5 Omnidirectional sports chassis

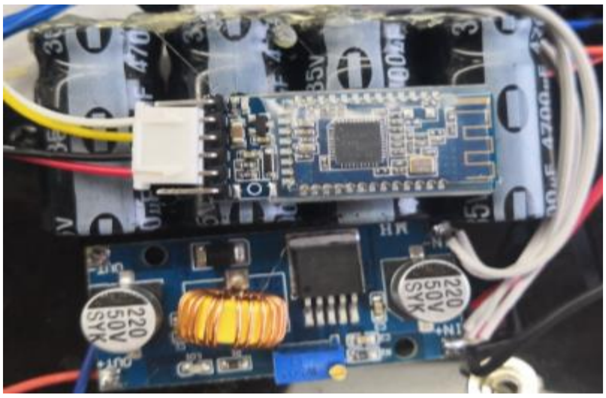


Fig 4.6 power management circuit

4.2 Table tennis detection and recognition experiment

As presented in Figure 4.7-4.9, the identification of a single ball and five balls is relatively simple and accurate. The background of the chair, the water bottle and the baffle are not affected in the background; however, when there are many balls, there is a missed detection.

After many experiments, I found that when the robot faces a lot of table tennis balls, it can't be finished all at once. Therefore, any missed table tennis ball can be photographed again, identified, and positioned. Therefore, the missed detection does not affect the using effect, but misdetection is not allowed.

（1）**Single ball recognition result**

Fig 4.7 Single ball image and recognition result

（2）Five balls recognition result

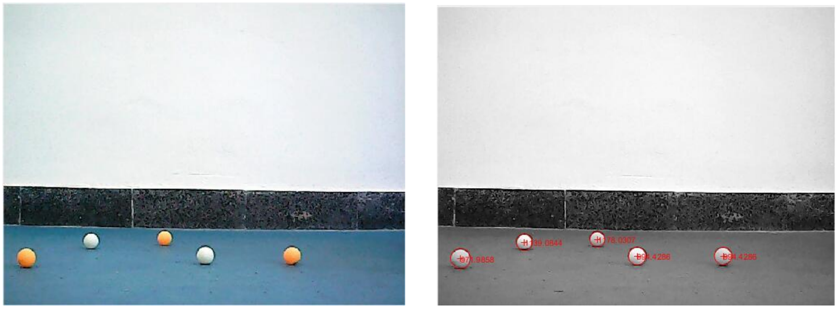


Fig 4.8 Five balls image and recognition result

（3）Multiple balls recognition results

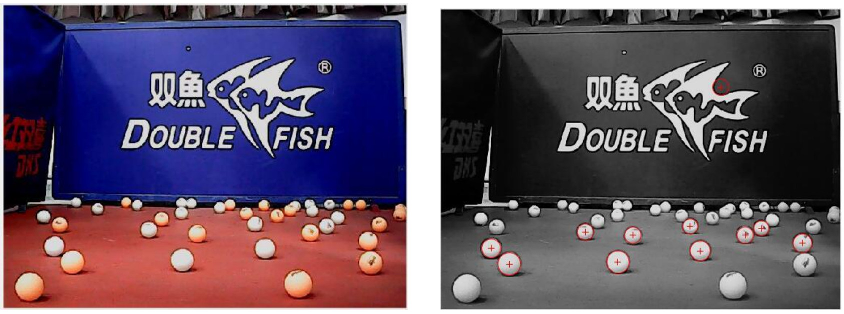


Fig 4.9 Multiple balls image and recognition result

4.3 Monocular vision ranging experiment

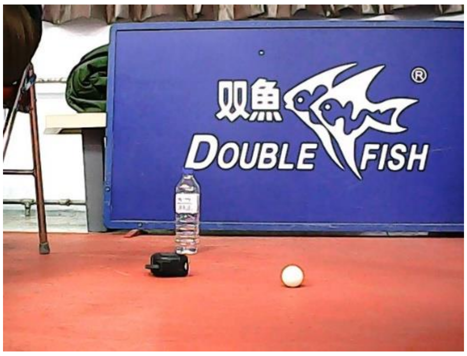
The images from series of experiments are shown in Figure 4.10. The results are shown in Table 4.2.

Objection distance=80cm

Objection distance=90cm

Objection distance=100cm

Objection distance=110cm

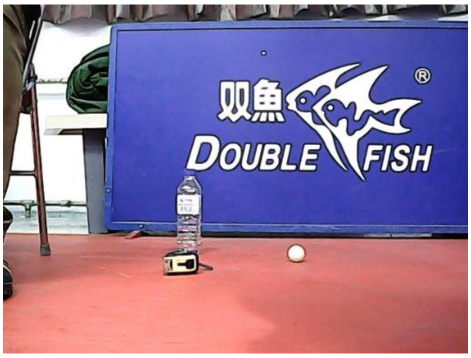


Objection distance=120cm

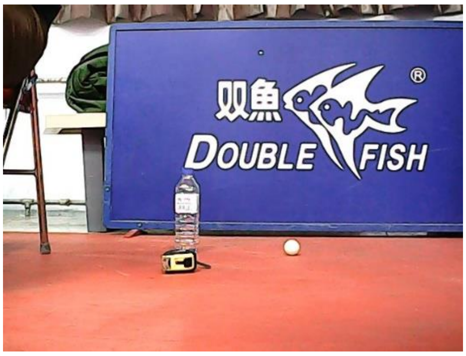
Objection distance=130cm

Objection distance=140cm

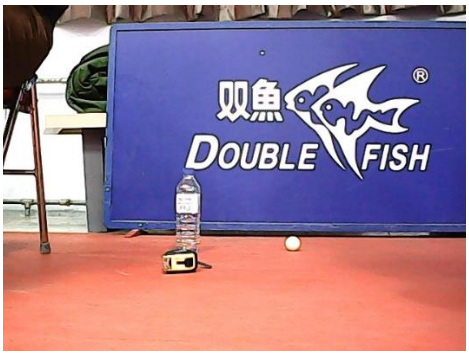
Objection distance=150cm



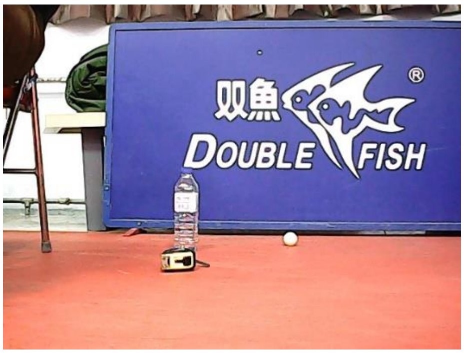
Objection distance=160cm



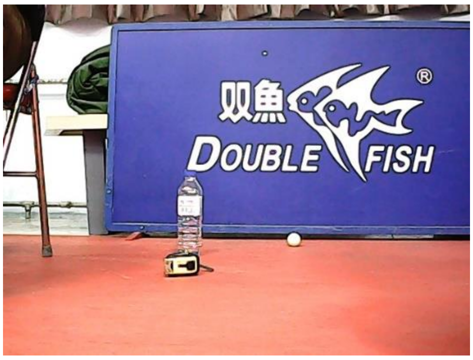
Objection distance=170cm



Objection distance=180cm



objection distance=190cm



Objection distance=200cm

Objection distance=210cm

Fig 4.10 Experimental images of different object distances

Table4.2 Monocular vision ranging result

Measurement distance (m)	Calculated distance (m)	Absolute error（m）	Relative error
0.9	0.936	0.036	0.0400
1	1.038	0.038	0.0380
1.1	1.128	0.028	0.0255
1.2	1.219	0.019	0.0158
1.3	1.316	0.016	0.0123
1.4	1.408	0.008	0.0057
1.5	1.488	0.012	0.0080
1.6	1.579	0.021	0.0131
1.7	1.681	0.019	0.0112
1.8	1.762	0.038	0.0211
1.9	1.852	0.048	0.0259
2	1.941	0.059	0.0304
2.1	2.046	0.064	0.0313

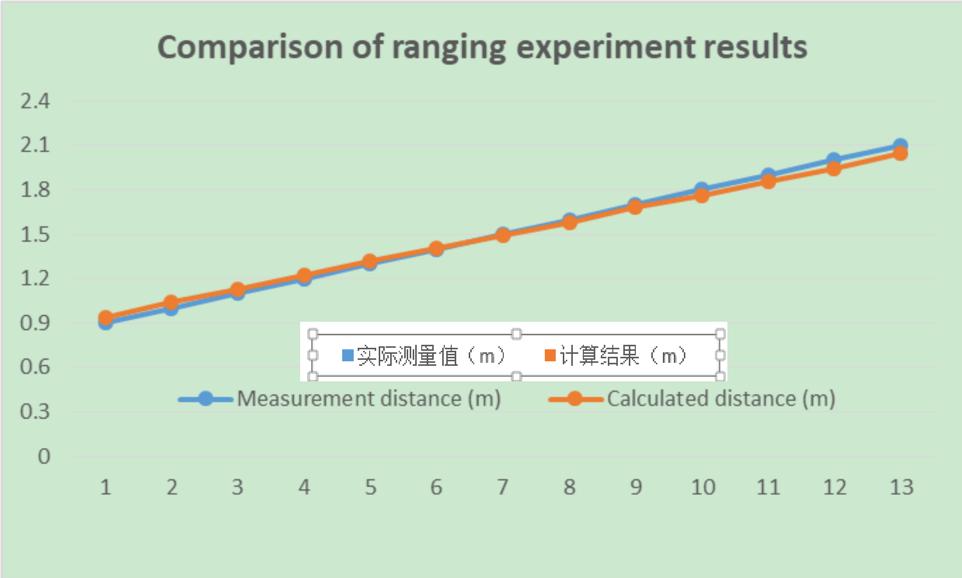


Fig 4.11 Comparison of ranging experiment results

4.4 Monocular vision positioning experiment

The camera's pitch angle was set at 15 degrees. The camera focal length is 3.50 mm; the image size is 640px x 480px. And the camera height is 538mm. The Y coordinate is fixed. The X coordinate calculated by the positioning model and the actual measured value are shown in Table 4.3. The maximum absolute error was 4.9 cm, the average relative error was 0.0237. When the actual distance is 70 cm, the error was smaller. The fixed Y coordinate is 70 cm, and different X coordinates are tested in the field of view. The experimental results are shown in Table 4.4.

Table 4.3 X coordinate experiment results (pitch angle =15 degrees)

Measurement distance(m)	Calculated distance(m)	Absolute error（m）	Relative error
53	53.2	0.2	0.0038
55	56.2	1.2	0.0214
60	59.0	-1.0	0.0169
65	66.0	1.0	0.0152
70	69.9	-0.1	0.0014
75	77.4	2.4	0.0310
80	84.2	4.2	0.0499
85	89.6	4.6	0.0513
90	94.9	4.9	0.0516

Table 4.4 Y coordinate experiment results (the pitch angle =15 degrees, X=70cm)

Measurement distance（cm）	Calculated distance（cm）	Absolute error（cm）
-20	-18.2	1.8
-10	-9.3	0.7
0	-0.4	0.4
10	9.3	0.8
20	18.1	1.9

（2）the camera's pitch angle was set at 20 degrees. The camera focal length is 3.50 mm. The image size is 640px x 480px. And the camera height is 538mm. The Y coordinate is fixed. The X coordinate calculated by the positioning model and the actual measured value are shown in Table 4.5. The maximum absolute error was 4.9 cm, the average relative error was 0.0301.

Table 4.5 X coordinate experiment results (pitch angle =20 degrees)

Measurement distance (m)	Calculated distance (m)	Absolute error（m）	Relative error
50	48.0	-2.0	0.0400
55	53.2	-1.8	0.0327
60	58.5	-1.5	0.0250
65	62.5	-2.5	0.0385
70	67.8	-2.2	0.0314
75	73.3	-1.7	0.0227
80	78.5	-1.5	0.0188
85	85.2	0.2	0.0024
90	90.9	0.9	0.0100
95	96.7	1.7	0.0179
100	102.8	2.8	0.0280
105	109.8	4.8	0.0457
110	115.7	5.7	0.0518
120	126.8	6.8	0.0567

When the actual distance was 85 cm, the error was smaller. The X coordinate is fixed to 85 cm, and different Y coordinates are tested in the field of view. The experimental results are shown in Table 4.6.

Table 4.6 Coordinate experiment results (the pitch angle =20 degrees, X=85cm)

Measurement distance（cm）	Calculated distance（cm）	Absolute error（cm）
-25	-23.1	1.9
-15	-13.5	1.5
-5	-4.5	0.5
5	6.2	1.2
15	16.3	1.3
25	26.4	1.4

It can be seen from the above experimental results that, at fixed Y coordinate and pitch angle, the farther the object distance, the larger the error of X experimental results. When the X coordinate is fixed, the farther the ball is from the optical center position, the greater the error.

When the pitch angle was 15 degrees and the object distance was 70 cm, the error is smaller. The maximum absolute error recorded was 4.9 cm in the range of 1 m, and the average relative error was 0.0237. When the pitch angle was 20 degrees and the object distance was 85 cm, the error was smaller. In the range of 1.2m, the maximum absolute error was 6.8cm, and the average relative error was 0.0301. The experimental results revealed that it can meet the needs of the ping-pong pickup robot.

5 Conclusion and Future work

In this paper, the application of the recognition algorithm based on the Faster-RCNN framework combined with the ground plane constrained positioning model of single frame image was used for the visual recognition of the ping-pong ball by a pickup robot. It can be concluded thus:

（1）Aiming at the problem of misidentification based on edge recognition, an improved algorithm based on Faster-RCNN framework for accurate recognition of table tennis ball is proposed. First, the table tennis ball image in various scenes is learned through the training depth network, and then the learned parameter model is applied to the actual table tennis image. The experiment proves the superiority of the method and solves the problem of accurate recognition of table tennis in various scenarios.

（2）Aiming at the problem of table tennis positioning in monocular vision environment, a single frame image table tennis positioning and distance model based on ground plane constraints is proposed. On the basis of obtaining the horizontal and vertical coordinates of the table tennis ball centroid in the pixel coordinate system, the model converts the pixel coordinates and image coordinates to obtain the coordinates of the table tennis centroid image, and effectively combines the three-dimensional geometry and the perspective projection principle to calculate the coordinate values the ball centroid in the world coordinate system, in turn, realize the positioning and distance function of the robot.

（3）The robot model was initially built, and the table tennis recognition and ranging and positioning experiments were carried out. The experimental results revealed that the system can detect and identify table tennis balls correctly, measure and locate it. The average relative errors recorded were 0.0223 and 0.0301, which proves that the algorithm improvement and positioning model for the application scenario can obtain good experimental results.

The monocular vision system of the table tennis robot has been achieved, and the ping-pong ball pickup robot is preliminarily built. The next step is to study the path planning algorithm of the mobile robot to realize the optimal path planning of the robot. In the end, the robot can automatically identify, locate and efficiently pick up a large number of table tennis balls scattered on the ground in the table tennis training venue.

6 Acknowledgments

I am grateful to Li Jinhua, My teacher in Experimental School Affiliated to Jilin University， for his guidance on the road of scientific research.

I would like to thank Prof. Hu Liang, from college of Computer Science and Technology in Jilin University, for his guidance and help.

I would like to thank my parents and friends for their strong support during the research period.

7 References

【1】 LingBao. The automatic table tennis picking up system was developed based on embedded technology. [J]. Automation and instrumentation, 2016(5):118-119.

【2】 Ji Yunfeng, Zhu Ling，Shen Yanni. Table Tennis Recognition Based on OpenCV[J]. Microcomputer Application , 2016, 32(4):68-70.

【3】 Wang Xiaolong, Mu Chunyang, Ma xing. Automatic picking up table tennis embedded system based on OpenCV[J]. Computer Measurement & Control, 2015, 23(11):3861-3863.

【4】 Wang Zhongcheng, Zhao Xuezeng.Research on Space Target Orientation of NAO Based on Monocular Vision. Machinery and electronics, 2015(09):63-67.

【5】 Zhang Zhiguo. Study on monocular vision location system[D]. Huazhong University of Science and Technology,2009.

【6】 Liu Xi. Researches of the Target Detective and Locative Methods for the Surface Mobile Robot[D]. Shenyang Ligong University, 2013.

【7】 Liu Hongwei. Target Recognition and Location of Mobile Robot Based on Monocular Vision[D].Shandong University, 2011.

【8】 XU Dongwei, Liu Jianqun, Lin Gan. Design and Realization of Table Tennis Picking Robot. [J]. Machine and Hydraulics，2014,42(03):16-19.

【9】 Yuan Yutong. Intelligent vehicle front obstacle recognition and ranging based on monocular vision[D]. Jilin University,2016

【10】 Xu Min. Research on Target Recognition and Location of Mobile Robot System Based on Monocular Vision [D]. South China University of Technology, 2014.

【11】 https://github.com/tzutalin/labelImg.