Research on Intelligent Speech Guide Robot Control Method Based on Machine Vision

Zhengbo Wang, Xing Ma and Chunyang Mu
North Minzu University, Yinchuan, Ningxia, China

*Corresponding author: maxingsky@126.com

Abstract. Robotic navigation via speech guidance gain potential attention on the field of artificial intelligence and it stat-of-art. This paper highlights speech guide for robot toward machine vision were designed. The robot navigation can determine and identify the objects around the blind area. The process of determining the direction and distance of the blind people to consider for deciding and predicating the next step. The information of the sensors is fused together, and a path information suitable for the blind person is transmitted to the user via speech voice broadcast are the key elements for identification part. Moreover, deep learning algorithm involve for accurate classifiers target into categories, to archive the high recognition accuracy. The direction and distance judgment part adopt the binocular ranging, angle measurement algorithm with a consideration of errors rate less than 2%. The system can correct, guide blind people to move forward, advance, stop, turn, which greatly compensates for the visual defects of the blind, with respect of constraint of improving the indoor safety of the blind people.

1. Introduction

According to 2010 data released by the world health organization, there are nearly 45 million blind peoples in the world, and yearly around 7 million become blind[1]. In our daily life, most of the information are obtained by visual sense. However, the physical defects of the blind and the influence of the complex social environment have brought great inconvenience to the life of the blind people[2]. Some countries have given more support to this category of people, moreover it has more institutions for training centers. The integration of this category into social demands creation more job opportunities, the emergence of guide sticks and assistance guide such as dogs were has greatly impact for helping in trips. But unfortunate these have not really helped to solve all the inconveniences of life issues.

Carnegie Mellon university of Pittsburgh had developed a robot suitable for guiding the blind[3]. The robot emphasizes the stability of movement, providing support for the blind or elderly, but requires a handle. Zhang Zhimei et al. developed a crawler guide robot, which uses various sensors to track the black trajectory preset on the ground to realize navigation and obstacle avoidance, so it cannot work in an environment without black trajectory[4]. In this paper, present conditions and constraints of process (the guide robot and find trajectory) however is designed a novel approach of intelligence for navigating and guide robot based on voice with machine vision. The robot is mainly used in indoor (household), office environment, to detect and identify the blind around objects, perception object distance of the blind angle and distance, in dynamic area, variety of sensor information fusion together, passed to the user in the form of speech.
2. Design of intelligent speech guide robot

In terms of applied intelligent speech guide robot used to design object detection and recognition[5] based on deep learning, binocular calibration including ranging, motion control, ultrasonic and short infrared ranging, wireless communication and speech broadcast system[6]. Among variety and complex elements, the design is mainly consider the robot's perception of the external environment. (in door) The detection and recognition of external objects is held by machine vision, as with perception of the distance and angle between objects and users, to calculate the optimal path suitable for blind people.

In terms of physical component, four dc motors are used to control McNamee wheels, and the wheeled chassis as the carrier to realize 360 degree rotation of the car, optimize obstacle avoidance effect, and equipped with binocular camera, controller, ultrasonic and short infrared sensor, wireless transmission equipment, to complete information acquisition, information processing, communication and other functions.

![Figure 1. Schematic diagram of hardware structure](image)

As shown in figure 1, this design includes 1 as the main controller, 2 as the robot chassis, 3 as the voice broadcaster, 4 as the binocular camera, and 5 as the infrared sensor.

3. Intelligent guide robot

3.1 Object detection and recognition based on neural network

The robot's perception of the surrounding environment with accurate detection and recognition of objects are the focus of this system. Therefore, it is necessary to detect multiple objects in the image signals received by the camera and accurately identify object categories. Joseph Redmon[7] mentioned in literature that YOLO algorithm can ensure the accuracy and detection rate of detection. Standard YOLO algorithm can detect 45 frames/s, while Fast YOLO algorithm can detect 155 frames/s. Author Wang yuning analyzed in literature[8]. By comparing several different target detection algorithms, it is concluded that YOLO algorithm has the fastest detection speed.

In addition, YOLO algorithm is most suitable for the requirements of simultaneous detection of multiple obstacles in this paper, so the object detection part of this paper adopts YOLO V2[9] detection system. YOLO algorithm belongs to CNN, which also consists of convolution layer, pooling layer and full connection layer. However, the difference is that in output layer is a tensor, and its training samples need not be specially cut out from the sample image, but the whole image is trained and detected, which improves the stability of the system and reduces the false detection on the background. Input image is divided into S * S small grid, each grid is responsible for the inspection is to be in the center of the object on the grid, the grid inspection when produce B a frame, each frame contains 5 D information (x, y, w, h, C_obj), respectively, represents the abscissa and ordinate of border center, width and height of the whole image, boxes of confidence, confidence level as shown in formula (1), (2):

\[
C_{\text{Obj}} = Pr(\text{Object}) \times \text{IOU}_{\text{truth}}
\]  
(1)
In equation (1), Pr(Object) is the probability of the Object in the grid, and if there is, it is 1, otherwise, it is 0. IOU represents the ratio of the intersection area between the reference standard box and the detection box. YOLO algorithm extracts features in the convolution layer, and target prediction is carried out in the full connection layer. When Pr(Object)=1, the confidence of some kind in the whole picture is shown in formula (3).

\[
\text{Pr}(\text{Class} \mid \text{Object}) \times \text{Pr}(\text{Object}) \times \text{IOU}^{\text{truth}}_{\text{pred}} = \text{Pr}(\text{Class}_i) \times \text{IOU}^{\text{truth}}_{\text{pred}}
\]  

(3)

Where \( \text{Pr}(\text{Class} \mid \text{Object}) \) is the classification conditional probability of the target to be detected. \( \text{Pr}(\text{Class}_i) \) is predicting the probability of a certain class. After the above calculation, set the threshold value of the detection border, filter out the surrounding border whose score is lower than the threshold value, and conduct non-maximum suppression treatment on the reserved border to obtain the detection result.

3.2 Stereo matching and ranging based on binocular vision

3.2.1 Camera calibration. The calibration and balance of objects in this paper are two logitech C270 cameras with fixed relative positions. A 9*6 checkerboard plane was used as the calibration template, and the side length of the positive square was 19mm. In the experiment, the template plane was placed in the field of view of two cameras, and images with pixel size of 640×480 were collected. Then adjust the angle and position of the template, so that the left and right cameras can successfully detect the corner coordinates of 20 pictures, and then implement calibration program to solve projection matrix, internal parameter matrix and external parameter matrix.

3.2.2 Stereo correction. In real binocular stereo vision, there is no complete coplanar alignment between two camera planes. Therefore, we need to carry out stereo correction to correct the two images that are actually aligned with non-coplanar rows into coplanar rows. OpenCV provides us with a function “cvStereoRectify” for the left and right image alignments. Figure 2 shows the left and right image pairs without stereo correction. Obviously, the same points in the left and right images are not on the same line. Figure 3 shows the image pair after stereo correction to eliminate the distortion.

3.2.3 Stereo matching and ranging. SGBM stereo matching is performed on the corrected left and right images to find the corresponding coordinates of the same object in the two images for obtaining the distance between them -- parallax. In fact, stereo matching is to match the same points in the left and right camera image pairs to get a parallax map.

The main steps are as follows:

- preprocess and filter to normalize the brightness and enhance the image texture.
- conduct matching search.
- remove the false matching points.

Analysis of the above table shows that when the target object to be measured is relatively close to the camera, the depth information accuracy of the target measured by the system is relatively high, and
the measurement accuracy decreases as the distance gets farther. In this paper, we do not pursue the accuracy of distance, and the experimental results meet the requirements of broadcasting environmental information for blind people.

3.2.4 Calculation of migration angle. According to the actual needs, we only need to guide the blind people in the general direction and distance, so it is very important to choose a simple and convenient path with fewer obstacles. We propose a method to judge the number of obstacles (whether the path is complex or not) in front based on the depth map information. In the depth map, the area where obstacles may appear. And split into left and right parts on average, and the pixel points with known depth information are counted. The side with more points is identified as the side with more obstacles and more complex paths, so the other side is selected as the area where the robot passes.

If the selected area on the left side of the image, to detect objects with the latest information of the pixel area of the leftmost pixel position, the position and the center of the image pixels, using trigonometric function, calculate the robot to offset angle: according to the edge of the object in the position of image pixels, and according to the pixel location and image pixel location x direction difference among pixel values divided by the x direction to take the perspective of the whole image (image represents the view angle), obtains the real object and robot angle deviation. The angle is obtained and then transferred to the robot chassis control. The motor controls the universal wheel rotation angle, which is the angle that the blind person should move to the left. Similarly, if the selected area is on the right half of the image, that is, the angle that the blind person should move to the right.

3.3 Voice broadcasting system
In this experiment, espeak speech synthesis tool under Linux system is used to convert the text information collected from images into speech information, so that the robot can broadcast the distance between the current obstacle and the robot and the angle formed directly in front of the person.
Espeak speech synthesizer is simple, fast, small, can read text from standard input, speech output can be saved as .wav. For the voice broadcasting system, there are three parts: the type of obstacles, the distance of obstacles and pedestrians' obstacle avoidance.

3.4 Robot motion control and sensor signal receiving system
Robot control program with raspberry PI, python programming. Figure 4 shows the flow chart of the control program. The control of the chassis walking device is carried out by judging which direction sensor has signal.

As shown in Figure 4, when the signal is received by the sensor in a certain direction, the robot will walk 0.2s in the opposite direction, with a speed of 0.5m/s and a distance of 10cm under control. The motion time and speed can be modified according to the situation to change the travel distance.

4. Experimental analysis

4.1 Analysis of experimental results of hardware system
This novel approach can control the robot chassis motion speed and eight direction motion, but also can achieve autonomous obstacle avoidance function. In the design, 8 near-infrared sensors are mainly used to realize obstacle avoidance of the robot, which are respectively installed one 45 degrees apart, to detect whether the object in front is less than 20cm from the robot.

As shown in Figure 5, the actual obstacle avoidance effect is shown as follows: (a) there is an obstacle within a distance of 20cm behind the right, and the robot drives ahead to the left for 0.2s, (b) where there are obstacles within a distance of 20cm behind, the robot drives ahead for 0.2s. (c), (d) have the same effect as (a) and (b).

![Figure 5. Hardware experimental results](image)

According to the analysis in Figure 5, when the sensor in a certain direction has a signal, the robot will drive 0.2s in the opposite direction to avoid obstacles.

4.2 Analysis of experimental results of the software system

4.2.1 Analysis of stereo matching and ranging results. According to the system design, we selected the object closest to the robot in the original picture, and displayed the distance between it and the robot and the angle of the object's most right end to the robot. The actual nearest distance is 32.3cm with an angle of 21 degrees to the left, as shown in Figure 6. The system can automatically find out the nearest object and display the distance and angle information, and the values are basically consistent with the actual values. SGBM algorithm is used to realize the matching of right and left corrected images, and the parallax map is obtained. Object depth can be implemented by the function cvReprojectImageTo3D, and the experimental results are shown in table 1.

| number | actual distance/mm | 20 mean measurements/mm | relative error/% |
|--------|---------------------|-------------------------|-----------------|
| 1      | 400                 | 404.55                  | 1.138           |
| 2      | 500                 | 506.55                  | 1.310           |
| 3      | 600                 | 609.65                  | 1.608           |
| 4      | 700                 | 711.87                  | 1.696           |
| 5      | 800                 | 814.20                  | 1.775           |
| 6      | 900                 | 918.28                  | 2.031           |
| 7      | 1000                | 1021.57                 | 2.157           |
| 8      | 1100                | 1125.77                 | 2.343           |

![Figure 6. Binocular ranging results](image)

4.2.2 Object detection and recognition result analysis. Object detection and recognition based on neural network can detect multiple objects in the same image and display the object category and
recognition accuracy. As shown in Figure 7, the network accurately identifies cups, laptops, potted plants and chairs in the real scene.

![Figure 7. object detection and recognition](image)

4.2.3 Result analysis of voice broadcasting system. The voice broadcast system collects the output information of binocular ranging, object detection and recognition, and can carry out intelligent voice broadcast. The robot will broadcast "there are obstacles in front of you -- laptop, potted plant, chair and cup. Please pay attention to the distance of the nearest obstacle in front is 32 centimeters, please walk 21 degrees to the left".

5. Conclusion
This paper mainly integrates motion control technology, sensor technology and image processing technology to complete the design of the entire intelligent speech guide robot system, including hardware circuit design, car control code writing, image recognition code writing and so on. According to the experimental results, the guide robot basically realizes obstacle avoidance, accurate object recognition, accurate and reasonable ranging and azimuth measurement, voice broadcast and other functions.

Acknowledgments
This work was financially supported by Ningxia natural science foundation project (2019AAC03121), University Science and Technology Innovation Platform of Ningxia Province (Industry-University Cooperative Research Base for Key parts of advanced equipment and system innovation), Open project of Ningxia key laboratory of intelligent information and big data processing(2019KLBD002), Scientific research project of North Minzu university (2019XYZDX01), the Master graduate student innovation project of North Minzu University(YCX19087).

References
[1] Z Li, G Xu, M Guo. Design and research of audio-visual fusion guide robot [J]. Progress in laser and optoelectronics, 2017, 54(12).
[2] Monteiro-Junior R S, Figueiredo L F D S, Maciel-Pinheiro P D T, et al. Acute effects of exergames on cognitive function of institutionalized older persons: a single-blinded, randomized and controlled pilot study[J]. Aging Clinical & Experimental Research, 2016, 29(3):1-8.
[3] Morris A, Donamukkala R, Kapuria A, et al. A robotic walker that provides guidance[C]// IEEE International Conference on Robotics & Automation. 2003.
[4] Z Zhang, L Cheng, Y Zhao, et al. Fuzzy-PID arithmetic for blind guiding robots.Journal of Shenyang Normal University(Natural Science Edition) 2015,33(1):81-8.
[5] He K, Zhang X, Ren S, et al. Deep Residual Learning for Image Recognition[J]. 2015:770-778.
[6] Kastrati R, Hamiti M, Abazi L. The opportunity of using eSpeak as text-to-speech synthesizer for Albanian language[J]. 2014.
[7] Redmon J, Divvala S, Girshick R, et al. You Only Look Once: Unified, real-time Object Detection[C]// Computer Vision and Pattern Recognition. IEEE, 2016:779-788.
[8] Y Wang,Z Pang, D Yuan. Real-time vehicle detection based on YOLO algorithm [J]. Journal of wuhan university of technology, 2016, 38(10):41-46.
[9] C Cai, K Wu, Y Yan. Airport runway target detection based on optimized YOLO method [J]. Command information system and technology, 2018(3).