Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2020)
IEEE Xplore Part Number:CFP20K74-ART; ISBN: 978-1-7281-4876-2

Voice Assisted and Gesture Controlled Companion Robot

Ms. Quanitah Shaikh
Assistant Professor
EXTC Department
SFIT
Mumbai, India
quanitahshaikh@sfit.ac.in

Mr. Rohit Halankar
System Engineer
Tata Consultancy
Services
Mumbai, India

Mr. Akshay Kadlay
System Engineer
Tata Consultancy
Services
Mumbai, India

Abstract— **This paper implements the design for a robot that can be controlled simply by using interactive inputs from the operator such as voice and gesture along with object tracking. The system aims to create a prototype of a futuristic automated personal assistant for domestic as well as industrial purposes. Google text to speech API and Grassfire algorithm is used to control the basic locomotion of the system. The robot consists of a gripper arm which is used to pick and hold objects as desired by the operator.**

Keywords—image processing; Raspberry Pi; Arduino; gripper arm

I. INTRODUCTION

As the world moves towards automation, it is beneficial to have a robotic assistant to facilitate human tasks. The robot described in this paper, has been built considering the fulfillment of this necessity as the fundamental objective. There are several day-to-day tasks performed by humans for which they need an assistant to help themselves with some equipment, tool, or any object, in general. This robot can be controlled by the human operator using speech or by an object tracking mechanism. Thus, the robot could navigate to the location as per the operator's wish. The human would then use gestures to have the gripper arm of the robot pick objects, and then the robot could return to the operator as it did while proceeding towards the location earlier.

The field of robotics is relatively new in the progress of technology. With the improvement of living standards and technological development, more and more household appliances and service robots will enter our daily life [1]. In its basic form, it can be defined as the use of machines to simplify the task of man. In its essence, robotics involves precise and intricate communication between various sets of electronic and mechanical components, to perform a particular task about a particular application. Tracing back to the very roots of the subject, the principle of robotics and related automata was given a comprehensive push by the set of theorems proposed by the scientist Isaac Asimov [2]. Both the philosophy [3] and AI communities have discussed ethical considerations of robots in society using the three laws as a reference, with a discussion in 'IEEE Intelligent Systems' [4]. The three theorems as proposed by Asimov would eventually be the foundation stone for the development of the field of robotics. A robot consists of an electronic component such as a microcontroller and/or a mechanical component such as a shaft or crank. The electronic component functions similar to the brain of the robot, directing tasks and scheduling algorithms while the motion of the mechanical components is a consequence of the actions as directed by the electronic component and thus, this interplay between the electronic and mechanical domains results in a physical action performed by the overall system. Robots were made popular in the entertainment industry and gradually have worked their way into the different aspects of human life.

As globalization is spreading in almost every part of the world, the need for machines is increasingly being felt by humans. This is large because utilizing such machines, we can achieve greater efficiency of the result while reducing the human effort for the same operation. Besides, robots can also be used in areas wherein there lies a fatal danger to the life of humans, e.g., in radioactive environments, to regulate the generation of Uranium atoms. The need of robots with artificial intelligence is needed the most in the space research for e.g., the control systems used onboard the Hubble space telescope, the Voyager missions etc. Also, robots are indispensable in security systems such as biometric identification, such as fingerprint technology, iris detection and face detection etc. The four-wheeled robot described in this paper accepts speech input and performs the corresponding tasks as specified by user speech.

Speech processing is nothing but performing certain operations on a speech input to utilize them for a particular application. The speech processing technique used in this system comprises of two parts i.e. speech to text and text to speech. In speech to text, the conversion of speech analog signals into a plaintext format is done which can enable easy understanding of human speech by the machine. Whereas, in the text to speech, the text is converted into the corresponding speech signal by the aid of speech synthesizing algorithms. It is the artificial generation of human speech by the robot to respond to a particular command given by the user.

The use of speech processing avoids the use of other complex communication processes to be used in human-robot interaction as speech is the simplest, easier, and convenient way for humans to communicate. Also, speech processing is used in various applications like car systems, telephony and

other domains etc. In this paper, the speech processing is used as a means of communication to the robot. That is, the desired commands to be given to the robot will be given via speech by the user [5]. Following this, the speech to text conversion is performed by the robot and the corresponding intended response is given via text to speech conversion by the robot.

Though the robot can interpret the voice commands provided by the operators, it is also necessary to guide the robot concerning the specific direction of its locomotion. Image processing is a technique which makes use of an image as an input, performs certain operations on it, and yields the output as required for a specific application. In the robotic system described in this paper, an image of the signaling object, provided as an input to the robot, can, in turn, be used to help the robot identify the intended direction. The camera would be subjected to motion and it would capture a video of the surroundings. In order to perform the required operations, a software aid is essential, that is, a program would be written which will have the details of the signaling object predefined. This program will also accept as inputs, the frames captured by the camera, and accordingly make a comparison of these captured frames with the details of the object predefined in the program so that the frame consisting of the signaling object is identified by the program being executed.

The process of image segmentation, and in particular, region-based segmentation, is required in this robotic system. Region-based image segmentation is a method in which an image is partitioned into distinct, non-overlapping regions. The partitioning into regions is such that the pixels in the individual regions have similar gray level values, which form a connected region, whereas the pixels which lie in adjacent regions have dissimilar values. For the robotic system, this methodology would be used to divide each of the individual frames into distinct regions, and then identify the region of interest which would correspond to the signaling object.

Gesture recognition means interpreting human gestures via mathematical algorithms. Gestures can originate from the movement of any body part particularly from the face or hand. The most common technique of this type is hand gesture recognition, facial recognition etc. By using gesture recognition, the user can communicate with the machine or robot. This human-robot interaction does not include any mechanical devices and hence the communication is quite simpler and the efficiency of the system is increased. These gestures can act as input commands for the machine and by recognizing these commands with the help of some technology like software programs, modules the corresponding output can be generated by the machine [6]. Due to this reason, gesture recognition can be used in the field of robotics. Over the past few years, it has found many applications such as in gaming, virtual reality etc.

The important part of any gesture recognition system comprises of the accelerometer. An accelerometer is a device that can be used to sense tilting, rotation, and acceleration of any object. Miniature three-axis accelerometers with element size smaller than 1mm2 and good in-plane sensor performances have been presented [7], [8].

The gesture-controlled robotic arm can be controlled with the help of hand gestures using Arduino microcontroller, accelerometer, and a Bluetooth module. The use of gesture recognition will avoid controlling the robot with a remote or a switch which is quite complicated.

II. DEVELOPMENT METHODOLOGY

A. Task Analysis

First, the mechanism for the fundamental operation of the robot, that is, the wheeled locomotion, needs to be taken care of. According to the system design, the speech input from the operator will be used to trigger locomotion. For the robot to proceed in a specified direction, the strategy utilized in this system is to make use of a signaling object. To help the robot identify this particular object, the image processing technique needs to be employed. The control of the gripper arm would be accomplished by using a gesture control mechanism.

B. Formal Design

In Fig.1, the Raspberry Pi is the central unit of the system. The Raspberry Pi is a credit-card-sized personal computer, running open-source Linux as an operating system [9]. An Arduino board is connected to it, which together, perform the locomotion operation using the voice input from the operator. The DC motors are the components that cause the movement of the wheels, which are driven by using the Motor driver ICs, IC L293D. An electrical circuit is to be mounted on a breadboard, for enabling the motor driver IC to perform its function properly. Since the voice input is provided by the operator using a smartphone at a certain distance from the Arduino board, wireless transmission of data is necessary, which is accomplished using the Bluetooth Module. The Raspberry Pi module also serves the purpose of image processing. A couple of other Arduino Boards are required for control of the gripper arm using gestures.

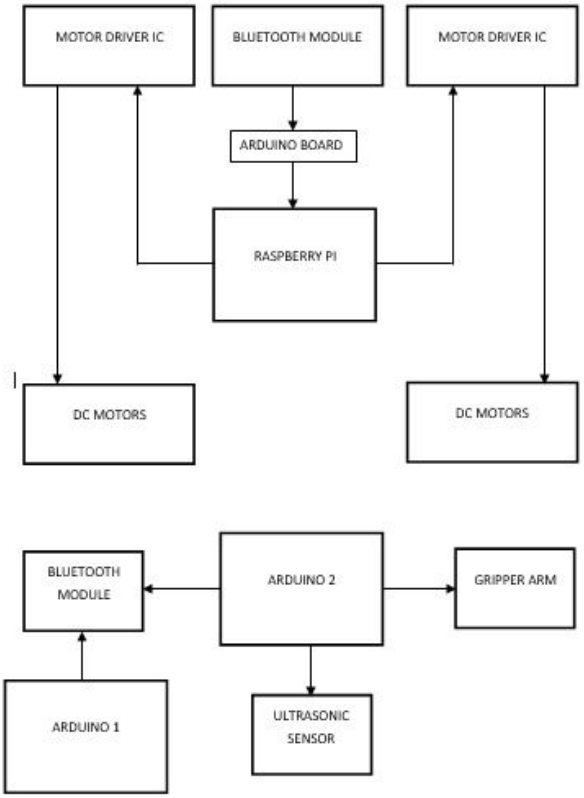

Fig. 1: Block Diagram of Robotic system

III. WORKING OF THE ROBOTIC SYSTEM

The entire working of the system can be divided into three categories: 1) Speech Processing, 2) Image Processing, 3) Gesture Recognition and processing.

Of these techniques, speech processing is performed at the beginning. The Raspberry Pi microcomputer, equipped with Python version 2, is considered as the brain of the robot. Speech processing involves the user speaking into a smartphone equipped with Google Speech API [10], which is common in all Android smartphones. Speech recognition is done via this API, which converts the recognized speech into a text format. Transmission of this recognized text to the robotic system can be done by using an Arduino microcontroller, which involves using a Bluetooth module connected with the speaker's smartphone, and the recognized speech is sent from the Arduino to the Raspberry Pi via a serial USB cable. This speech is converted into text format and can be processed by treating this text as a string variable. This string variable is coded as per the requirement of the user. E.g. if the user says 'forward', then the robot is programmed to provide torque to its wheels in such a way that the system goes in the forward direction. Similarly, to accurately simulate the actions of a fully-fledged personal assistant, the robot is made to understand and therefore perform several tasks.

The speech processing technique is used to make the system more interactive by including the text-to-speech feature that is, the system communicates with the user, thereby simulating the response generation abilities of a personal assistant. The Raspberry Pi is used to generate speech responses, specifically, by making use of its internal audio file player. The responses to every action are stored as text within the internal memory of the Pi. Text-to-speech generation is performed by several open-source libraries that are interfaced with Python 2. The audio can be heard by connecting a speaker to the Pi's 3.5mm audio jack.

The image processing section involves using the Raspberry Pi and a USB webcam which acts as the 'eye' of the robot. The Python 2 version, along with the open-source-SimpleCV image processing software is used for image processing. For the system, image processing is incorporated in the form of object tracking. The user, for operating the robot's image processing capability, is to wear a blue-colored circle affixed to his leg. Then, using an appropriate speech command, the robot is instructed to identify the object and then perform a physical activity. For e.g. when the user says 'Come Here', the robot rotated around its axis until the object affixed to the user is detected. When the object is detected, the robot is programmed using image segmentation algorithms [11] to differentiate between the desired object and the background. It is necessary to select a unique shape of the chosen object by the user, to avoid any ambiguity in detection. The image processing techniques can be further enhanced by such that the robot can be used for several domestic applications, say, a butler. The Python processing language is one of the most powerful languages when it comes to image processing applications. This is one of the several major advantages of using the Raspberry Pi as compared to other microcontrollers in this robot, as Python is the default coding language for the Pi.

The robotic system also involves gripper arms that are constructed in a certain manner such that the whole robotic arm is connected onto a base, in which the driver and the mechanics for the first degree of freedom of motion are mounted. The second, third, and fourth-degree of motion are implemented inside the arm itself [12]. To enhance the operability, the gripper arm is designed independently of the other blocks in the robot. Instead of the Raspberry Pi, controlling the gripper arm, the Arduino microcontroller is chosen to act as the controlling device. In order to make the entire gripper mechanism wireless, we make use of another Bluetooth module, connected to the Arduino. The Bluetooth module is configured to act as transmitter and receiver respectively. Thus, two Arduino boards are used at the transmitter and the receiver. The transmitter is fixed to a glove on the user's hand. For the purpose of gesture recognition, we make use of a 3-axis accelerometer, which constantly gives values of x, y and z axes depending upon the relative position of the user's hand. Thus if the user flexes in hand in a particular direction, there is a corresponding change in one of the values of coordinates. These values are serially recorded by the transmitter Arduino and wirelessly transmitted by Bluetooth to the receiver Arduino. The values are then coded appropriately to ensure a particular movement of the arm. In

order to concentrate resources to enhance the computational efficiency, this program contains some movement algorithm [13], [14]. The locomotion of the robot is controlled by giving power to the rear wheels via an L293d integrated circuit.

IV. IMPLEMENTATION OF ROBOTIC SYSTEM

C. Speech Processing (Google text to speech API)

The first task necessary for the implementation of the system is the set up of the speech to the text recognition system. The method for achieving this is to use an Android phone's internal Google speech to text API. The phone is connected by Bluetooth to a Bluetooth module that is connected to an Arduino microcontroller. The Arduino, then, by using a serial USB cable sends this data to the Raspberry Pi. On the Raspberry Pi, we open up a python sketch (version 2) and write a code which will then further control the action of the motors. Similarly, the Raspberry Pi is connected to the internet and can be used as a personal assistant. The task which we want the robot to perform is included in the code.

D. Image Processing

The image processing task is implemented after the speech processing block. SimpleCV is an open-source image processing computer vision library for python. It is a lite version of the more exhaustive OpenCV library which makes it useful for operation on the Raspberry Pi, since the Pi has lesser processing power. Here, a webcam is used to simulate the action of an eye. The webcam is connected to the Pi and its main function is to capture video in the form of frames. The user or the target is to wear a blue-colored circular object. The SimpleCV library contains a function "findBlobs" which internally uses BLOB extraction algorithms like a recursive grass fire and sequential grass fire algorithms. By applying these algorithms, the object worn by the user is detected and a command is written into the system to move towards the said object after detection.

E. Gesture Recognition

Finally, the robotic gripper arm is implemented. Initially, the two Bluetooth modules need to be configured by connecting each of them by setting one module as a transmitter and the other as a receiver. Then, by using a three-axis accelerometer integrated circuit connected to the transmitter Arduino, data can be sent using Bluetooth to the receiver. At the receiver, the incoming data can be processed to drive the pan and tilt servo motors connected to the arm.

V. RESULTS

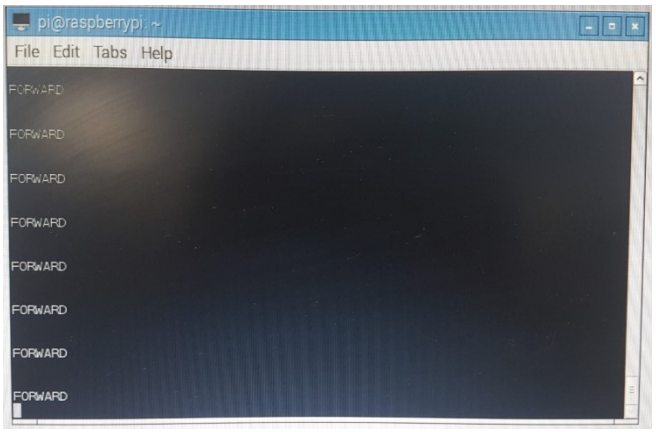

Figure 1: Output when speech command 'forward' is given

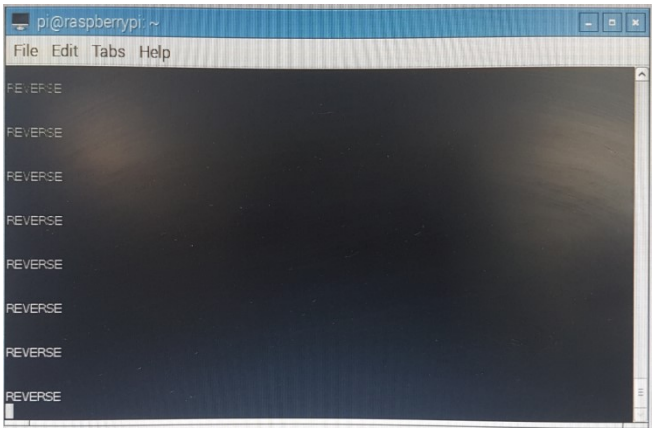

Figure 2: Output when speech command 'back' is given

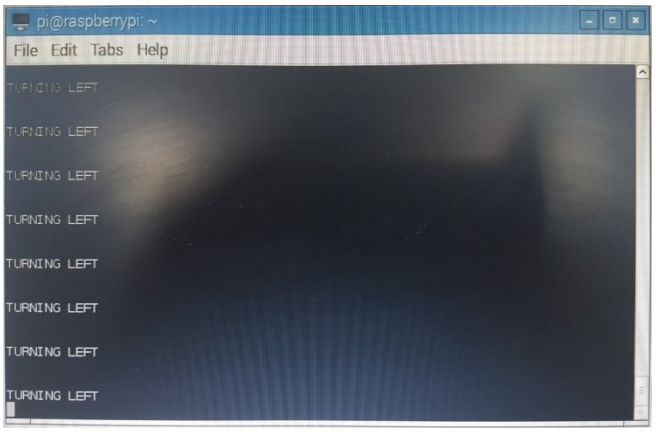

Figure 3: Output when speech command 'turn left' is given

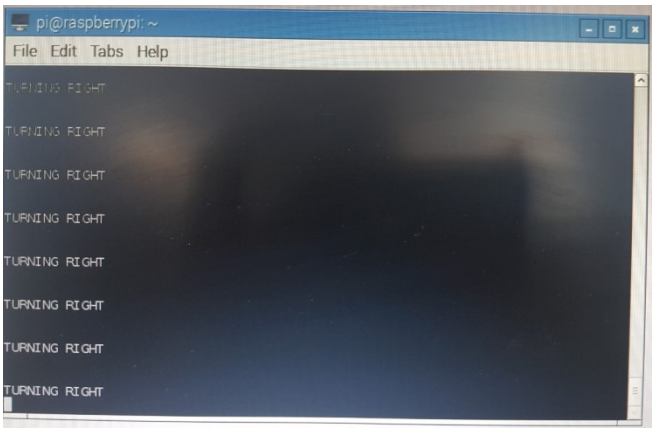

Figure 4: Output when speech command 'turn right' is given

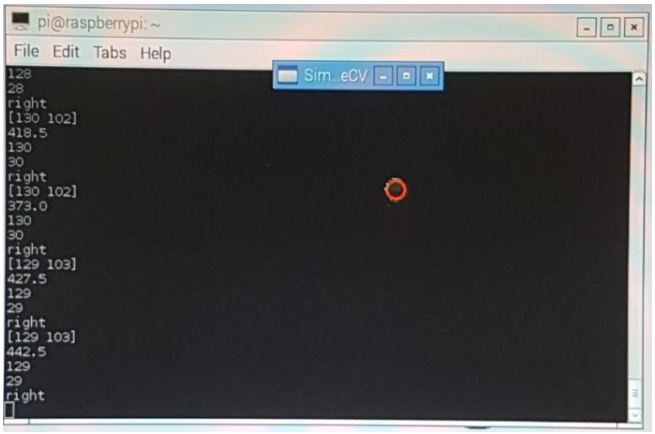

Figure 5: Window showing detected blob and desired direction of movement

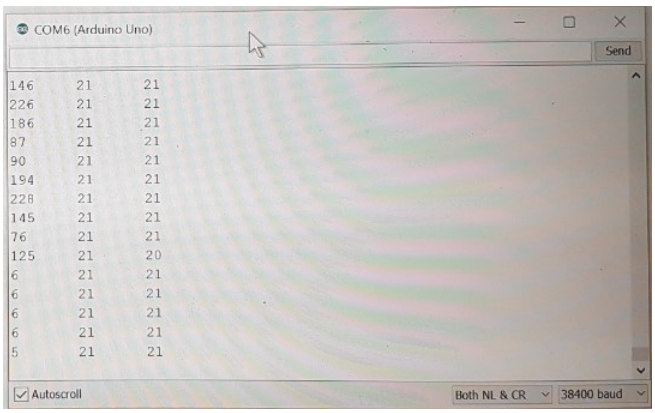

Figure 6: Serial Monitor displaying values corresponding to outputs on 3 pins [Gesture Control]

Figures 1 to 4 are a visual console representation of the actual motion of the robot. When the user speaks the appropriate speech command, the robot moves in the corresponding direction as displayed in the aforementioned figures. Figure number 5 denotes the output of the image processing unit of the robot. The system determines the position of the blob relative to the orientation of the camera and then based on this, makes a decision on which direction the robot should move to. The figure consists of the coordinates of the blob as captured by the camera and the direction in which the robot moved. Figure 6 displays the console output of the accelerometer which is used in the gesture control unit of the robot. As the user moves his hand, the 3D coordinates calculated by the accelerometer are sent by Bluetooth to the Arduino which moves the gripper arm accordingly.

VI. CONCLUSIONS

This paper implements the design of a robotic system that can be controlled using voice and gestures provided as inputs by the user. The robot will be efficient and useful in reducing the human efforts in various applications and hence will improve the overall efficiency of the system. The operation of the robot is distributed into three parts i.e. speech processing, gesture control, and image processing.

Using natural language processing (NLP), the scope of this robot, in terms of communicating with humans, can be broadened. Such robots can then be used as waiters in hotels or for domestic purposes in household activities. A sturdy robotic arm can be built and thereby upgrading the hardware of the robot, it can be used to carry heavy loads or objects where the involvement of humans is not possible and is dangerous. It can also be used as a robotic assistant in hospitals for doctors and nurses during treatments of patients or performing surgeries.

ACKNOWLEDGMENT

The authors would like to thank Mr. Vaqar Ansari for his suggestions regarding the use of Raspberry Pi for implementing image processing techniques.

REFERENCES

[1] Chunjie Chen, Xinyu Wu, Long Han, Yongsheng Ou,"Butler Robot," IEEE International Conference on Information and Automation, pp. 731-737, June 2011.

[2] Robin R. Murphy and David D. Woods, "Beyond Asimov: The Three Laws of Responsible Robotics," IEEE Intelligent Systems, vol. 24, no.4, 2009, pp. 14-20,

[3] S.L. Anderson, "Asimov's 'Three Laws of Robotics' and Machine Metaethics," AI and Society, vol. 22, no. 4, 2008, pp. 477–493.

[4] C. Allen, W. Wallach, and I. Smit, "Why Machine Ethics?" IEEE Intelligent Systems, vol. 21, no. 4, 2006, pp. 12–17.

[5] Martin Urban and Peter Bajcsy. "Fusion of Voice, Gesture, and Human-Computer Interface Controls for Remotely Operated Robot," 2005 International Conference on Information Fusion, vol.2, pp. 1644-1651, July 2005.

[6] Ze Lei, Zhaohui Gan, Min Jiang, and Ke Dong, "Artificial Robot Navigation based on Gesture and Speech Recognition," 2014 IEEE International Conference on Security, Pattern Analysis, and Cybernetics, pp. 323-327, Oct. 2014.

[7] M. A. Lemkin, B. E. Boser, D. Auslander, and J. H. Smith, "A 3-Axis force balanced accelerometer using a single proof mass," 1997 International Conference on Solid State Sensors and Actuators, pp. 1185-1188, 1997.

[8] M. Lemkin and B.E. Boser, "A three-axis micromachined accelerometer with a CMOS position-sense interface and digital offset-trim electronics," in IEEE J. Solid-State Circuits, vol. 34, pp. 456-468, Apr. 1999.

[9] Nikolaos K. Ioannou, George S. Ioannidis, George D. Papadopoulos, Athanasios E. Tapeinos, "A novel Educational Platform, based on the Raspberry-Pi," in International Conference on Interactive Colloborative Learning, pp. 517-524, Dec. 2014.

[10] Prashant G. Ahire, Kshitija B. Tilekar, Tejaswini A. Jawake, Pramod B. Warale, "Two Way Communicator between Deaf and Dumb People and Normal People," in International Conference on Computing Communication Control and Automation, pp. 641-644, Feb. 2015.

[11] Xuesong Le, Ruben Gonzalez, "A Consistent, Real-Time Image Segmentation for Object Tracking," International Conference on Digital Image Computing: Techniques and Applications (DICTA), pp. 1-7, December 2016.

[12] Pavol Krasnansky, Filip Toth, Vladimir Villaverde Huertas, and Boris Rohal-Ilkiv, "Basic Laboratory Experiments with an Educational Robotic Arm,"in International Conference on Process Control (PC)

[13] D. Fox, W. Burgard, S. Thurnand, and Cremers A, "The dynamic window approach to collision avoidance," in M. IEEE Robo. Auto., vol. 4(1), 1997, pp. 23-33.

[14] S. S. Ge and Y. J. Cui, "New potential functions for robot path planning," in IEEE Trans. Robot. Auto., vol. 16(5), 2000, pp. 56-60.