



VISUAL ATTENTION MODEL FOR MOBILE ROBOT NAVIGATION IN DOMESTIC ENVIRONMENT

E.D.G. Sanjeewa, K.K.L. Herath, B.G.D.A. Madhusanka, H.D.N.S. Priyankara

Department of Mechanical Engineering, Faculty of Engineering Technology, The Open University of Sri Lanka, Nawala, Sri Lanka

KeyWords

Visual Attention, Kinect Camera, Robot Navigation, Hand Gesture, Mobile Robot, Domestic Environment

ABSTRACT

Paying more attention to the elderly and disabled people has become an essential part of a developed society. It means they need to be looked after when the responsible people are not around. That means using an automated system is compulsory in fulfilling this objective. The concept of visual attention model for mobile robot navigation in domestic environment is used in this research paper. This solution is mainly aimed on elderly and disabled people. This helps them to obtain a desired object by action recognition using computer vision. In this paper it proposes a method to deliver an object that is required by the elderly or disabled user. First, the robot identifies the hand gesture shown by the user to identify the direction where the objects are used to complete this task. Then it identifies the hand gesture shown by the user, which is representing the required object number, by using the convexity defects detection method. After that robot navigates to the target object by avoiding obstacles IR sensor is used to detect and avoid obstacles. In this process, objects are placed in predefined places in the room. Then the robot platform reaches the object shown by the user and then returns to the user. To measure the angles rotated by the platform, a gyro sensor is used. The main process is calculating the distance values to the required objects. Kinect V2 camera is used to calculate the depth values.

INTRODUCTION

One of the major problems the country is facing is the growth of elderly population at present. Currently elderly population is 12.4% of population [1]. This figure is continuously increasing, and it requires paying attention to elderly people, especially when they are unable to do their own work. Not only the elderly people, but also disabled people need more attention because they face difficulties in day to day life. Around one billion individuals worldwide face difficulties, according to the world health Organization. There are thousands of individuals in Sri Lanka who suffer from several disabilities. The Ministry of Health of Sri Lanka estimates that by 2025, the number of disabled people in Sri Lanka will rise by 24.2% [2].

One of the major problems they face is inability to reach and grab a required object. There is a need for human assistance for assisting them. It requires more effort and is not economical. Also, it is not practical to acquire human assistance just to reach and get/grab the required object. As a solution there are many solutions proposed such as automated wheelchairs and voice recognition models. The major disadvantage of these models is if the user is fully disabled or cannot speak, it will be useless. This paper proposes a method that can overcome above problems. The proposed system is Visual attention model for mobile robot navigation in the domestic environment. This system is mainly aimed at navigating the robot to the desired object by the user and returning. Haar cascade method [3] is used for human identification and the object number shown by the person is identified by using the convexity defect [4] method. The main origins of the computer vision applications are visual photographs and recordings. Noise from the picture must be removed before starting the creation of the algorithm. Pre-processing procedures help remove noise from the original source [5].

A Kinect V2 camera is used for visual processing of the system. The Kinect V2 camera is rotated around the mobile robot platform to search for predefined objects, represented by the hand gesture number. After that system calculates the depth values to the desired object using Kinect V2 camera and navigates towards the object. Three IR sensors are placed in front of the platform to detect obstacles on the way. If an object is detected, the platform rotates and navigates to the right side and then again rides forward. A gyro sensor is used to measure the angles rotated by the platform. Navigation process is controlled by an ATmega 2560 microprocessor.

LITREATURE REVIEW

There are many approaches made by many researchers to present a solution for this problem. Luo et al. have been proposed a combining method for hand gesture recognition based on combinatorial approach recognizer equation [6]. It was consisted with two recognizers. They are hand skeleton recognizer and recognizer based on support vector machines. This system can identify the hand gestures from no sign, to one to five. Their results have proven that the used algorithm was successful in harsh environments as cluttered backgrounds and angle changes in hand appearance.

Also, Anjaly et al. have proposed a method that can track a moving target [7]. This was implemented for indoor environments. The major concern in this method is the localization of the tracker and the target. Tracking system was developed using the modified proportional navigation guidance law. They confirmed the accuracy of the system by using snapshots at different instants for a curved and straight path traced by the target.

Further to that, Zehui Meng has been introduced a semantic mapping system along with path creating system [8]. It can process high level semantic beliefs to improve robot navigation in complex environments with moving objects. A deep neural network is used object detection, scene understanding and object recognition. Scene beliefs and 3D object beliefs were used to build the semantic map. This system worked in high efficiency with moving objects by semantic beliefs and creating paths. The stereo vision and color to detect and track people implemented a new system has been developed by Rafael [9]. This system was implemented using a height map that was used to extract foreground objects. This system was very fast so that it has a very low computing time, results in fast detection and tracking of objects. Also, this system could mostly adapt to illumination variants in the environment.

Lijun Zhao has been proposed an intuitional human-robot interaction algorithm for 3D mapping based on real time features computations to navigate the robot [10]. Few gesture instructions were chosen as user's intentions. A real sensor is used to collect the hand node data that is used for processing. A 3D camera for depth sensing, motion tracking and real-time 3D mapping was used as a binocular vision sensor with Oculus Rift helmet to achieve the operator's perspective stereoscopic observation and operation. The 3D environment reconstruction of dense cloud which enabled to move the robot independently was realized by using depth calculations. A Kinect camera was used to capture depth information which was used to avoid obstacles.

Niko [11] has been proposed a transferable and expandable categorization and sematic mapping system. It mainly adapted Bayesian filter framework that allows the synchronization of previous information and enforces temporal coherence of the classification results. Humans describe places, goals, and objects using semantic categories. They have showed that how semantic information can influence robotic navigation tasks in a workplace, making robot operations more compliant with human needs. Also, they have showed that how semantic information supports robotic object detection and increases the performance.

Finally, Sahak et al [12] has been introduced this method of detecting objects using Kinect sensor. It was able to detect the objects from the distance of 4-5 feet of the Kinect camera. Some of the objects were shiny and curved, so they showed incorrect detections due to reflections.

METHODOLOGY

Firstly, the signal from the user is taken to the system as an input. The first is the direction signal. Then the direction is identified the robot rotates to the direction and again to the user. Then the second input is taken which is the object number. This signal would be either 1, 2 or 3 which stands for mug, bottle, and first aid box as shown in Figure 1. This data is later used to identify the required object. After the signal is received, Kinect sensor is used to measure the lengths and angles to the objects required. That recorded data is sent to the system. After receiving object number, a Haar cascade classifier is used to identify the objects according to the input signal. After comparing the input and the classified objects, the robot starts to navigate to the object, using previously recorded measurements. This platform is equipped with three sharp IR sensors to avoid obstacles along the way. If an obstacle is detected, then the robot moves 50cm to the right-hand side and again starts the measuring process. A gyro sensor is used to give feedback on how much angle the robot has rotated. Finally, after reaching the location of the object the robot rotates back 180 de-

grees back. Then again it searches for the human using cascade classifier. Then the distance and the angle to the user is measured using the Kinect camera and the previous process is repeated until the robot platform reaches the user.

To the actuation part of the mobile robot is shown in Figure 2, four DC motors with 12V are used which has 120rpm. To control this four motors dual H-bridge motor driver is used. When the robot platform is moving on the calculated path it should avoid obstacles. As the microcontroller for this ATmega 1280 Arduino mega board is used. It has 16MHz crystal oscillator and can provide 14 PWM output pins which can be used to control the motion of the motors.

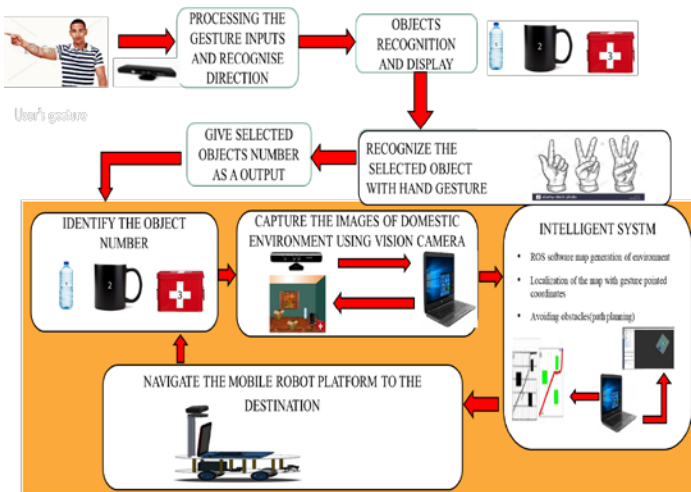


Figure 1: System overview



Figure 2: Mechanical design

RESULTS

Object identification using Haar feature-based cascade classifiers is an efficient form of identification of artifacts has suggested by Paul Viola et al. [13]. This is a machine-based learning technique, in which a cascade function is learned from both positive and negative Images. It's then used in other images to identify objects. For this process, objects are taken from both positive and negative pictures. Positive image is an image that only contains the correct object. Negative image is any image that doesn't include the entity needed. A positive picture is taken in cascade processing, and it is put on harmful pictures. In these process positive images are resized and put on negative images in various angles rotated. For this process it was used 100 positive images from each object category. For negative images 500 images were taken.

Figure 3 shows the object identification process. In this research, three objects were trained to be identified using the Haar cascade method. The lightning condition was used in normal indoor environment. Figure 4 indicates the human identification process. The same method of Haar cascade is used for this task. When implementing this method on humans, it was used both standing and sitting samples of human in various angles to train the model with high accuracy.



Figure 3: Object identification



Figure 4: Human identification



Figure 5: Input direction from the user

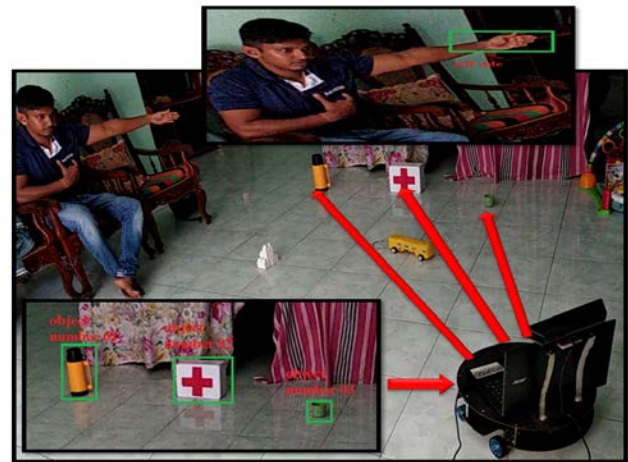


Figure 6: Identifying the direction and all objects

Figure 5 shows the process of input direction input from the user and the identifying process of the direction of the objects. Three objects are fed to the system to be identified as object number 1, 2 and 3. First the robot is always looking at the user, by identifying him as the human. After the person shows the direction of the objects as shown in Figure 6, the system identifies it as left of right direction. Then the robot rotates to the desired direction and identifies available objects. Also, it calculates the distance values to each object by using the Kinect camera. Then it turns back to the user again.



Figure 7: Hand gestured object number identification



Figure 8: Navigating by obstacle avoiding

Figure 7 shows the system identification process of the detect hand gestured number and identify the hand gesture is used the convexity defects detection method. Figure 8 shows the navigate by obstacle avoiding to the object. One object was placed directly in front of the robot and other two objects at 45° degree angle to the robot. Then robot navigates to the required object by avoiding obstacles. To detect obstacles, sharp IR sensors are used. It detects obstacles within 15cm and rotates and moves 50cm to the right side and then again starts the distance measuring process. A gyro sensor is used to give feedback on how much angle the robot has rotated.

Hand gesture algorithm

Hand gesture showed number = no of defect + 1



Figure 9: After drawing lines to get hand shape

After the hand gesture was shown the above map was drawn by using OpenCV with proposed algorithm. Blue dots represent the defects of the binary image of the hand as shown in Figure 9. Blue dots of the hand image used to count the number of defects, plus one.



Figure 10: Reaching the object



Figure 11: Returning to the user



Figure 12: Reaching the User

Figure 10 indicates the target touching and the frame breaks from the target at 15cm distance. It was called at this point as the framework has reached the appropriate target and selected it. Robot spins 180° degrees backwards. Once again it uses cascade classifier to check for the person. The distance and angle to the user are then measured using the Kinect camera and the preceding process is repeated until returning and reaching to the user of robot platform as shown in Figure 11 and Figure 12 consequently.

CONCLUSION

In this research was discussed the identification of events using Kinect sensor. The target can be accurately located at the optimal distance of 4 feet to 5 feet from the Kinect, from the tests received. This may also be found that the reflection caused the shiny and curved target to be identified incorrectly. Specific forms of artifacts shall be tested at various distance ranges for further study. The results, it can be concluded that the robot has been successful in navigating by user input and reaching the object and returning to the user showing the work.

REFERENCES

- [1] Maduwage, S., 2019. Sri Lankan 'silver-aged' population. Journal of the College of Community Physicians of Sri Lanka, 25(1), p.1.
- [2] Herath, Kasun & de Mel, W.R. (2019). ELECTROENCEPHALOGRAM (EEG) CONTROLLED ANATOMICAL ROBOT HAND. 10.6084/m9.figshare.9759248.
- [3] Ismail, A.P. & Tahir, N.M., 2017. Human Gait Silhouettes Extraction Using Haar Cascade Classifier on OpenCV. 2017 UKSim-AMSS 19th International Conference on Computer Modelling & Simulation (UKSim).
- [4] Anon, 2019. Hand Gesture Recognition using Convexity Defect. International Journal of Innovative Technology and Exploring Engineering Regular Issue, 9(1), pp.1161-1165.
- [5] Herath, Kasun. (2016). RICE GRAINS CLASSIFICATION USING IMAGE PROCESSING TECHNIQUES.

- [6] Luo, R.C. & Wu, Y.- C., 2012. Hand gesture recognition for Human-Robot Interaction for service robot. 2012 IEEE International Conference on Multisensor Fusion and Integration for Intelligent Systems (MFI).
- [7] Anjaly, P., Devanand, A. & Jisha, V.R., 2014. Moving target tracking by mobile robot. 2014 International Conference on Power Signals Control and Computations (EPSCICON).
- [8] Sun, H., Meng, Z. & Ang, M.H., 2017. Semantic mapping and semantics-boosted navigation with path creation on a mobile robot. 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM).
- [9] Muñoz-Salinas, R., Aguirre, E. & García-Silvente, M., 2007. People detection and tracking using stereo vision and color. *Image and Vision Computing*, 25(6), pp.995–1007.
- [10] L. Wang, L. Zhao, G. Huo, R. Li, Z. Hou, P. Luo, Z. Sun, K. Wang, and C. Yang, "Visual Semantic Navigation Based on Deep Learning for Indoor Mobile Robots," *Complexity*, vol. 2018, pp. 1–12, 2018.
- [11] N. Sünderhauf, F. Dayoub, S. McMahon, B. Talbot, R. Schulz, P. Corke, G. Wyeth, B. Upcroft, and M. Milford, "Place categorization and semantic mapping on a mobile robot," in *Robotics and Automation (ICRA)*, 2016 IEEE International Conference on. IEEE, 2016, pp. 5729-5736
- [12] M. S. A. Manap, R. Sahak, A. Zabidi, I. Yassin, and N. M. Tahir, "Object detection using depth information from Kinect sensor," 2015 IEEE 11th International Colloquium on Signal Processing & Its Applications (CSPA), 2015.
- [13] Anon, Face Detection using Haar Cascades. OpenCV. Available at: https://opencv-python-tutorials.readthedocs.io/en/latest/py_tutorials/py_objdetect/py_face_detection/py_face_detection.html [Accessed July 13, 2020].

