

Object Detection and Distance Estimation of a Mobile Robot with Stereo Vision

^{*1}Susmita Guin, ²Dr.R.Hablani, ³Rahul Gupta

^{1, 2, 3} Dept. of Computer Science and Engineering, Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India.

*Email: susmita.guin198@gmail.com

Received: 09th July 2018, Accepted: 14th August 2018, Published: 31st August 2018

Abstract

In day to day life, the studies on robotics are increasing very rapidly. Especially, for obstacle detection and estimating distance in dynamic and static environments. Therefore, researchers contribute more of their interest towards robotics as well as artificial intelligence. Hence, this paper mainly focused on detecting object and how we are estimating distance of a mobile robot. There are many solutions already given by the numerous researchers such as laser sensor, sonar sensor or camera sensor for estimating distance process. The main goal of our proposed approach is to fuse information gained from the stereo vision, infrared and ultrasonic sensors to enhance the accuracy. Thus, we proposed a system with a hybrid approach system.

Keywords: Obstacle Detection, Artificial Intelligence, Mobile Robot, Stereo Vision, Hybrid Approach.

Introduction

Being a humans, we see the 3D structure of the world around us. There are numerous studies encountered in this subject. Mobile robot is widely used in these kinds of fields. For any mobile robot, navigation and accurate information of its position is very important research area. The function of a mobile robot is commanded by the program which is running on the mobile robot ('job'), the area (surrounding) and various sensors available on the mobile robot (hardware). These three essential elements are as displayed in Figure 1.

Robots are equipped with many sensors. Such as laser, sonar, radar, infrared, ultra-sonic and stereo camera etc. These sensors have their own benefits and features. In recent years, some researchers provide keen interest towards stereo vision. The Greek word "stereos" which means solid or firm, with stereo vision the object are seen solid in 3D range. The stereo vision contains two different systems that are: active system and passive system. The active stereo system provides depth directly by using ultra sound and laser whereas passive stereo system includes algorithms such as SAD (sum of absolute difference) algorithm and RANSAC (random sample consensus) algorithm used for

images stereo system to estimate the accurate depth.

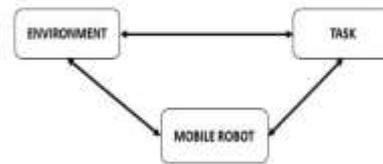


Figure 1. Obstacle Detection Essential Component

A scene is clicked with horizontally and fixed distance lenses in stereo camera will obtain two slightly different projections of a scene. If comparing these two images we will find additional information, the depth of a scene. This process of extraction of 3D structure of a scene from stereo image pairs is called computational stereo.

Computational stereo includes these following calculations on images:

1. Image Rectification
2. Image Matching
3. Calculation of Disparity Map
4. Collision Handling.

For successful accomplishment of our proposed approach, we use the information gained from the stereo vision camera, infrared and ultrasonic sensors to enhance the accuracy. Thus, we proposed a system with a hybrid approach system. This system will detect the obstacles in front of our mobile robot and estimate the distance, with short response time. The work presented here is mostly focus on mobile robot and stereo vision camera to perform smartly on roads. Here we are detecting and estimating distance using stereo vision (ZED camera) method which is fixed at centre and on the front side of robot. The ultrasonic and infrared sensors are mounted on our mobile robot (Figure 3).

Stereo Vision

Stereo vision is widely measured to sense the environment because it always provide us revealing advantage of data information, which makes the 3D reconstruction work easier. The stereo vision-based techniques provide, advantage of directly estimating the 3D coordinates on an image feature. It is a strategy that analyzes the foreground and background information and extracts objects by means of computer without human supervision, and also a closer assumption to the human perceptivity.

Stereo vision-based object detection techniques need to find correspondences from the left to the right image.

A stereo camera, “ZED” is used in our proposed approach. This ZED stereo camera is shown in Figure 2. This camera duplicate the way, how human vision works. Like, the person sense the objects using data taken from the real world with their two eyes. Then, the brain combines the captured two images together by matching the similarities and incorporate the differences to get a 3D model for the captured objects.



Figure 2. ZED Stereo Camera

The ZED is a camera with dual lenses. Utilizing its two "eyes" and through triangulation, the ZED gives a three-dimensional comprehension of the scene it watches, enabling your application to wind up space and movement mindful. Our stereo camera (ZED) visualize the whole world in 3D. This camera utilize high-resolution sensors and field glasses and, at the rate of 100FPS from 0.5 upto 20m can tell how far the object around you, inside and outside environments.

3D depth is calculated using the triangulation formula below. The triangulation method determine depth from disparity d . For stereo vision camera, depth is inversely proportional to disparity.

$$\text{Depth (Z)} = \frac{f \times b}{xl - xr} = \frac{f \times b}{d} \quad (1)$$

$$X = \frac{xl \times Z}{f} \quad \text{OR} \quad \frac{xr \times Z}{f} \quad (2)$$

$$Y = \frac{yl \times Z}{f} \quad \text{OR} \quad \frac{yr \times Z}{f} \quad (3)$$

Where, f is the focal length and b is the and (xl, yl) , (xr, yr) are the corresponding points of left image and right image respectively. After this ZED will store a value of distance (Z) for every (X, Y) point of the world coordinate system. The distance is expressed in metric units (for e.g. in meters) and compute from the back of the left eye of the camera to the scene object.

A Mobile Robot

A mobile robot, “FIRE BIRD VI” is used in our proposed approach. Fire Bird VI platforms has already set the standard for intelligent mobile robots. Because it contains all the components for navigation and sensing in real-world environment. It

comes in fully assembled and ready to use form. It has high rigor DC gear motors with the high resolution position encoders. We can also select correct locomotion drive, so that robot can be used in both the environment (indoor and outdoor). Fire Bird VI is equipped with 8 ultrasonic range sensor and 8 IR distance range sensor interfaces. Our mobile robot have a ultrasonic range of 6 meters which has sensing capability to sense all the object type and can work in any critical conditions whereas infrared distance range sensor can communicate due to line of sight or point-to-point mode of communication and can also detect motion in daytime and night time too. The IR depth sensor [15] obtains the depth image data of the actual environment which is sent to the processing unit. FIRE BIRD VI mobile robot is shown with two of its inbuilt sensors in Figure 3.

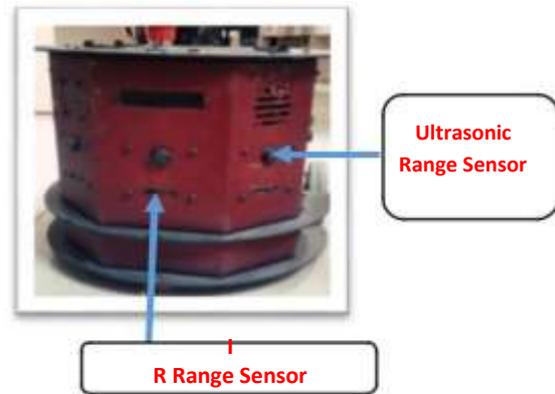


Figure 3. FIRE BIRD VI Mobile Robot

Literature Review

In this section, we introduce some of the research works done by many researchers. Some approaches depends on using multiple images, while some use filters on the images or by the use of small-baselines which is also termed as micro-baseline. Prashanth et al., [1] have used web camera with no auxiliary sensors. A novel segmentation algorithm was used to detect the object which has same color as that of the road by eliminating shadows. However, this system works in outdoor environment only at day time or in bright light areas. This system fails due to poor condition of light, especially at night. Rostam et al., [2] used Block matching algorithm which solve the issue occurred in comparing stereo images, where left image works as a reference block to the right image. Serdar Solak et al., [3] calculated the distance between the cameras and the colored object using two cameras situated in the same axis. But the main disadvantage of their approach is that their camera only detect the red colored object. Jian-Hong Sun et al., [4] used Triclops library for building there stereo vision system where, at first two raw images are captured by the stereo camera. After getting raw images, Low pass filter is used to smooth it and reduce the aliasing effects. Secondly, rectification is

used to correcting raw images for the distortions of lenses. Further, rectified images is corrected so that the rows of images digitized from horizontally displaced cameras are aligned, similarly that the columns of images obtained from vertically displaced cameras are aligned. Thirdly, edge detection is used for detection of object edge then stereo matching is done by the absolute differences correlation method and depth of the image is calculated. Marshall et al., [5] have used two sonars coupled with a single camera. To detect obstacles color based segmentation is used and to find the direction of movement sonars are used if segmentation algorithm fails. However the cost is low, but it works only in indoor environments. Detection of obstacles that are of the same color as that of the road are difficult, as color based segmentation is being used. Tanuja Sheorey et al. have used two ultrasonic sensors. They have used ultrasonic sensor to measure the distance between the host vehicle and the preceding vehicle. However, this system does not work in outdoor environment due to the low range of sensor. Ishan Kumar Shahu and Dr. R. Hablani [13] has done Velocity estimation using digital video. Levinson et al., [8] have used four cameras along with a LIDAR in their automation system. They achieved the overall accuracy of about 98% while testing on automobile. LIDAR requires the high processing power, which is the main disadvantage and due to this, the cost is also high. Such high processing would require too much power to work and if the automobile is battery operated, it will starve for power early due to decrease in mileage. Lategahn et al., [9] applied grid-mapping technique that uses both dense and sparse point clouds from stereo vision and SLAM (structure and motion). But the weakness of this approach is, while encountering outliers, as no multi-view, multi-baseline stereo matching technique is utilized. Ishankumar Shahu et al., [10] illustrates about the video processing, where they find how object is detected in the video sequence. They only surveyed about the various techniques of object classification on the basis of object's characteristics. Mohammed Faisal et al., fuse the multi-baseline stereo-vision and laser-range reading data to enhance the accuracy of the point cloud, to reduce the ambiguity of correspondence matching, and to extend the field of view of the proposed mapping system to 180°.

Proposed System

Hybrid approach enhances the mapping operation by building a real-time 3D object detection of unstructured and in unknown environment. This approach fuses the three sensors that are stereo vision sensor, infrared sensor as well as ultrasonic sensor to estimate the distance between the camera (which is placed on our mobile robot) and object. Therefore it combines the advantages of these three

sensors. Like, any of the sensor or stereo camera fails then other two sensor will continue to operate.

A general block diagram of our proposed approach is shown in Figure 4.

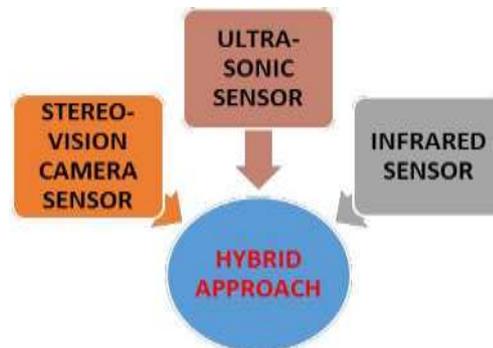


Figure 4. Proposed System

These study includes three modules as image pre-processing, object-detection and finally the distance calculation. Here, we have chosen OpenCV[6] library written in C++ and runs under Windows. OpenCV automatically uses the suitable instructions from IPP at runtime. The GPU module also provides CUDA-accelerated[14] versions of many routines (for NVidia GPUs). OpenCV's motive is to provide a simple-to-use computer vision infrastructure that helps people to build fairly sophisticated vision applications quickly. The OpenCV library contains over 500 functions in many areas, including factory product inspection, medical imaging, security, user interface, camera calibration, stereo vision, and robotics.

Figure 5 shows the three major modules to succeed in our proposed system.



Figure 5. Three Major Modules of Proposed System

Implementation

The Firebird VI mobile robot used in our proposed system which is DC motor driven, wheels 100mm long and 20mm wide. As per real time testing, essential requirement for all the sensors, stereo camera and actuators (present on Firebird VI mobile robot) were used. The ZED camera and eight (gp2y0a41sk0f) IR range sensors and ultrasonic sensors. For driving the robot, 12v DC motor was used which gives 100 rpm.

ZED stereo vision camera is used throughout the real time testing. Using its two visual lenses and through triangulation, it get the static and dynamic positions. The ZED camera have been placed in front of the

mobile robot. For communication between the Firebird VI mobile robot and laptop, USB serial port was used. The resolution of the camera have four different resolution modes[12] but we have used 3840x1080 with 30 frames per second (fps).The very important knowledge about the ZED is that it only connects on USB 3.0 which is present in the laptop. All the images captured by the stereo ZED camera are processed by computer.

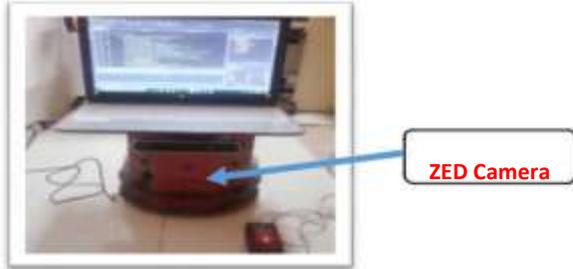


Fig. 6 Fire Bird VI Robot with Sensors and ZED Stereo Camera

The figure 6 shows a laptop is placed on the Firebird VI mobile robot for storing the images captured by the stereo camera placed in front of the robot and for compiling code in software. ZED SDK, CUDA[15] include with OpenCV and Visual Studio was used for object detection and distance estimation. CMake was used to build libraries and linkers to release as well as in debug mode for the software’s mentioned above. Real time testing has been achieved by three methods that are:

- i. ZED Stereo Camera
- ii. Ultrasonic Range Sensor
- iii. Infrared Range Sensor

The distance of an object with respect to left eye of the camera and computed by the Euclidean distance formula.

$$Distance = \sqrt{point3D.x * point3D.x + point3D.y * point3D.y + point3D.z * point3D.z} \tag{4}$$

The distance evaluated is denoted in meters (m) and 1080p video mode was used with 30 frame per second (fps) and 3840x1080 resolution was selected for calculations.

As we can see ultrasonic and infrared range sensors are situated in a mobile robot and these sensors also estimating distance of an object which is detected in front of the mobile robot.

Now, finally we compare each of the results given by the three sensors that are selected for our proposed approach.

Results and Discussion

Here, we have tested with our two sensors simultaneously for detecting and estimating

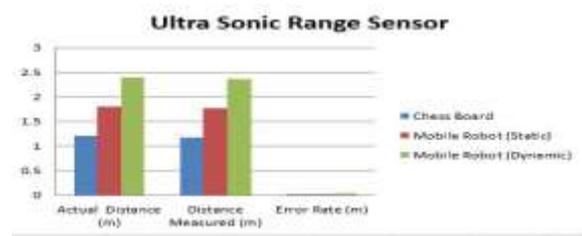
distance. Where, three different object were selected with different size, environment and shape for object detection. After, this we calculate the distance between the camera and object. Objects were placed at certain distance and then also tested with moving object.

A.



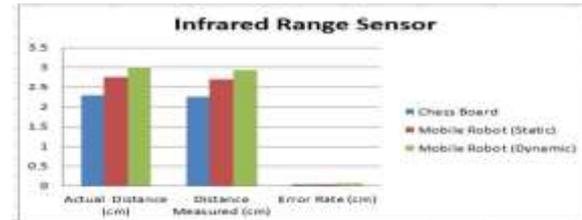
Name of the obstacles	Actual Distance (m)	Distance Measured (m)	Error Rate (m)
CHESS BOARD	1.2	1.17	0.03
MOBILE ROBOT (STATIC)	1.8	1.78	0.02
MOBILE ROBOT (DYNAMIC)	2.4	2.34	0.08

B.



Name of the obstacles	Actual Distance (m)	Distance Measured (m)	Error Rate (m)
CHESS BOARD	1.2	1.18	0.02
MOBILE ROBOT (STATIC)	1.8	1.77	0.03
MOBILE ROBOT (DYNAMIC)	2.4	2.37	0.03

C.



Name of the obstacles	Actual Distance (m)	Distance Measured (m)	Error Rate (m)
CHESS BOARD	2.3	2.25	0.05
MOBILE ROBOT (STATIC)	2.75	2.7	0.05
MOBILE ROBOT (DYNAMIC)	3	2.93	0.07

Conclusion

In this study, we have enhance the object detection and estimation of distance operation by fusing stereo-vision sensor, ultra-sonic sensor and infrared sensor. This fusion raise real-time mapping and accuracy.

The proposed system that is hybrid approach system, acquire for mobile robot applications provide us with highly accurate results. Here, the open source software tools are utilized with intelligent hardware tools system. The speed, computational cost and efficiency were considered as the major design target. The main desire was to develop and design an intelligent visual supervision systems to help the human operators to detect unusual events and respond rapidly to them. Therefore, it can be employed in any kind of mobile robot applications, just as automated transportation systems, material handling, manufacturing, housework, search-and-rescue operations, or security patrolling. The proposed approach system has the succeeding attributes:

- 1) 3D vision.
- 2) Accurate point clouds.
- 3) Works under any unfamiliar light conditions.
- 4) Any of the sensor or stereo camera fails then other two sensor will continue to operate.

Future Work

The robot has been successfully detecting obstacle and estimating distance as well. Moreover, we can improve this by adding more features like avoidance of obstacles and also we can improve by additional sensors on the mobile robot. Except that, computer vision can include with more camera features which can be implemented for monitoring applications. In the future, we can implement an obstacle avoidance with LIDAR sensor. The LIDAR sensor system is powerful and efficient especially in off-road outdoor environments.

Acknowledgements

The authors would like to acknowledge the funding and support from the department of Computer Science and Engineering of Shri Ramdeobaba College of Engineering and Management, Nagpur, Maharashtra, India.

References

1. Prashanth C R, *et al.*, "Obstacle Detection & Elimination of Shadows for an Image Processing Based Automated Vehicle," *International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, 2013.
2. Rostam Affendi Hamzah, *et al.*, "An Obstacle Detection and Avoidance of A Mobile Robot with Stereo Vision Camera," *International Conference on Electronic Devices, Systems and Applications (ICEDSA)*, 2011.
3. Serdar Solak and Emine Dogru Bolat, "Distance Estimation using Stereo Vision for Indoor Mobile Robot Applications," *Electrical and Electronics Engineering International Conference on (ELECO)*, 2015.

4. Jian-Hong Sun, *et al.*, "Stereo Vision based 3D Modeling System For Mobile Robot," *International Conference on Control, Automation and systems (ICCAS)*, 2010.
5. Ian W. Marshall, *et al.*, "Vision-based obstacle avoidance for a small, Low-cost robot," *International Conference on Advanced Robotics (ICAR)*, pp 275-279, 2007.
6. Nidhi, "Image Processing and Object Detection." *International Journal of Applied Research* 2015; 1(9): 396-399
7. Sheorey, *et al.*, "Development of sensor based front end collision avoidance system for highways." *In Information and Automation, 2015 IEEE International Conference on*, pp. 594-598, 2015.
8. Jesse Levinson, *et al.*, "Towards fully autonomous driving: Systems and algorithms." *(IEEE)-In Intelligent Vehicles Symposium (IV)*, pp. 163-168, 2011.
9. Lategahn, *et al.*, "Visual SLAM for autonomous ground vehicles." *International Conference on Robotics and Automation (ICRA)*, pp. 1732-1737, 2011.
10. Ishankumar Shahu and Prof.R. Hablani. "An object detection, tracking and parametric classification." *International Research Journal of Engineering and Technology (IRJET)*, Volume: 02 Issue: 08, Nov-2015.
11. Faisal, *et al.*, "Multi-sensors multi-baseline mapping system for mobile robot using stereovision camera and laser-range device." *Advances in Mechanical Engineering* 8, no. 6 (2016): 1687814016654634.
12. "Stereo Lab's ZED Camera," [Online]. Available: <https://www.stereolabs.com/zed/specs/>, March 2018.
13. Ishankumar Shahu and Ramchand Hablani. "An Object Detection, and Technique for Velocity Estimation along With Classification on the Basis of Speed." *Int. Journal of Advanced Research in Computer Science and Software Engineering (IJARCSSE)*, vol.6- Issue 5, May 2016
14. Al-Shehabi, *et al.*, "An Obstacle Detection and Guidance System for Mobility of Visually Impaired in Unfamiliar Indoor Environments." *International Journal of Computer and Electrical Engineering* 6, no. 4 (2014): 337.
15. Kim, *et al.*, "Effective object recognition based on CUDA." *International SoC Design Conference (ISOCC)*, pp. 159-160. IEEE, 2015.