COST-EFFECTIVE STEREO VISION SYSTEM FOR MOBILE ROBOT NAVIGATION AND 3D MAP RECONSTRUCTION

Arjun B Krishnan and Jayaram Kollipara

Electronics and Communication Dept.,
Amrita Vishwa Vidyapeetham, Kerala, India

abkrishna39@gmail.com
kollipara.jayaram@gmail.com

ABSTRACT

The key component of a mobile robot system is the ability to localize itself accurately in an unknown environment and simultaneously build the map of the environment. Majority of the existing navigation systems are based on laser range finders, sonar sensors or artificial landmarks. Navigation systems using stereo vision are rapidly developing technique in the field of autonomous mobile robots. But they are less advisable in replacing the conventional approaches to build small scale autonomous robot because of their high implementation cost. This paper describes an experimental approach to build a cost- effective stereo vision system for autonomous mobile robots that avoid obstacles and navigate through indoor environments. The mechanical as well as the programming aspects of stereo vision system are documented in this paper. Stereo vision system adjunctively with ultrasound sensors was implemented on the mobile robot, which successfully navigated through different types of cluttered environments with static and dynamic obstacles. The robot was able to create two dimensional topological maps of unknown environments using the sensor data and three dimensional model of the same using stereo vision system.

KEYWORDS

Arduino, Disparity maps, Point clouds, Stereo vision, Triangulation

1. INTRODUCTION

The future will see the deployment of robots in the areas of indoor automation, transportation and unknown environment exploration. Implementation of Robotic systems in such tasks is widely appreciated technique as they handle these tasks more efficiently and reliably. Currently, a growing community of researchers are focusing on the scientific and engineering challenges of these kinds of robotic systems.

This project tries to address the main challenges in the field of autonomous robots – Autonomous Navigation. There are several techniques for effective autonomous navigation, among which Vision based navigation is the most significant and popular technique which experiences rapid

Natarajan Meghanathan et al. (Eds) : ICCSEA, SPPR, VLSI, WiMoA, SCAI, CNSA, WeST - 2014
pp. 75–86, 2014. © CS & IT-CSCP 2014 DOI : 10.5121/csit.2014.4716

developments. Other techniques include navigation using ultrasound sensors, LIDAR (Light Detection and Ranging) systems, preloaded maps, landmarks etc. Navigation which uses ultrasound sensors will not detect narrow obstacles such as legs of tables and chairs properly, and hence leads to collision. LIDAR systems are perfect tools for Indoor navigation because of their accuracy and speed but they are less impressive for large scale implementation due to their high cost [1]. Navigation based on landmarks and preloaded maps become valid options only when there is prior information about the environment and thus, it does not give a generic solution to the problem of autonomous navigation. Vision can detect objects just as in the case of human vision and it gives the sense of intelligence to the robots. Out of all vision based techniques, stereo vision is the most adoptable technique because of its ability to give the three dimensional information about how the environment looks like and decide how obstacles can be avoided to safely navigate through that environment. Commercially available stereo cameras are expensive and require special drivers and software to interface with processing platforms which again adds up the cost of implementation. In this scenario, building a cost-effective stereo vision system using regular webcams which is able to meet the performance of commercially available alternatives is highly appreciable and this fact makes the theme of this project.

2. RELATED WORKS

Several autonomous mobile robots equipped with stereo vision, were realized in the past few years and deployed both industrially and domestically. They serve humans in various tasks such as tour guidance, food serving, transportation of materials during manufacturing processes, hospital automation and military surveillance. The robots Rhino [2] and Minerva [3], developed by the researchers from Carnegie Mellon University (USA) and University of Bonn (Germany) are famous examples of fully operational tour guide robots used in museums. These robots use stereo vision along with sonar sensors for navigate and building the map. The robot Jose [4], developed in University of British Columbia (Canada), uses Trinocular stereo vision - which is a combination of vertical and horizontal binocular stereo vision - to accurately map the environment in all three dimensions. PR2 robot designed by Willow Garage research laboratory is one of the most developed home automation robot [5]. This uses a combination of stereo vision and laser range finders for navigation and grasping of objects.

According to [6] there are two essential algorithms for every stereo vision systems: Stereo Calibration algorithm and Stereo Correspondence algorithm. Calibration algorithm is used to extract the parameters of the image sensors and stereo rig, hence has to be executed at least once before using the system for depth calculation. Stereo correspondence algorithm gives the range information by using method of triangulation on matched features. A stereo correspondence algorithm based on global matching is described in [7] and [8], [9], [10] are using correspondence search based on block matching. Considering these techniques as a background, an algorithm is designed for this project, which uses horizontal stereo vision system by block matching for obtaining stereo correspondence. Low cost ultrasound sensors and infrared sensors are chosen for overlapping with visual information.

3. ROBOT PLATFORM

Our experimental platform is a six wheeled differential drive rover that can carry a portable personal computer. Two wheels are free rotating wheels with optical encoders attached for keeping track of the distance travelled. Other four wheels are powered by high torque geared motors of 45 RPM each, which gives the robot a velocity of 20cm/sec. Three HC-SR04 ultrasound sensors are attached in the front for searching obstacles in 4m range. Two Infrared range finders are employed to monitor the vertical depth information of the surface on which

robot operates and hence avoid falling from elevated surfaces. A digital compass – HMC5883L – is used to find the direction of robot's movement. The core elements of the embedded system of this robot are two 8 bit ATmega328 Microcontroller based Arduino boards. One Arduino collects information from optical wheel encoders based on interrupt based counting technique, whereas the second Arduino collects data from all other sensors used in the platform and also controls the motion of motors through a motor driver. Heading from compass and distance data from wheel encoders gives reliable odometric feedback to the control system. A PID algorithm has been developed and implemented in order to keep the robot in straight line motion in an obstacle free region. Both the Arduino boards continuously transfer the collected data from the sensors to the on-board PC for storage and receive decisions from vision system implemented in on-board PC. Figure 1 shows the overall architecture of the mobile robot used in this project.

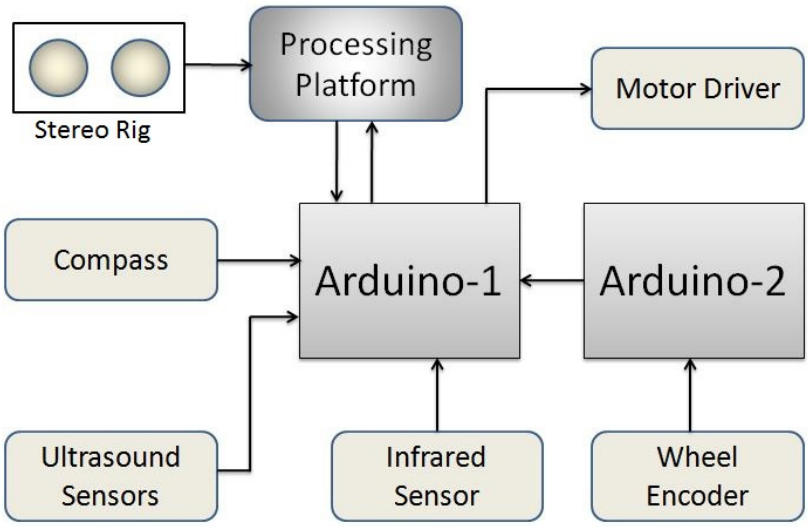

Figure 1. The architecture of the mobile robot

4. STEREO VISION SYSTEM

Stereo vision is a technique for extracting the 3D position of objects from two or more simultaneous views of a scene. Stereo vision systems are extensively used in object classification and object grasping applications because of its ability to understand the three dimensional structure of objects. Mobile robots can use a stereo vision system as a reliable and effective primary sensor to extract range information from the environment.

In ideal case, the two image sensors used in a stereo vision system has to be perfectly aligned along a horizontal or vertical straight line passes through the principle points of both images. Cameras are prone to lens distortions, which are responsible for introducing convexity or concavity to the image projections. The process called Stereo-pair rectification is adopted to remap distorted projection to undistorted plane. The obtained rectified images from both the sensors are passed to an algorithm which searches for the matches in the images along each pixel line. The difference in relative positions of an identified feature is called as the disparity associated with that feature. Disparity map of a scene is used to understand the depth of objects in the scene with respect to the position of the image sensors through the process called Triangulation. Figure 2.a shows the arrangement of image planes and Figure 2.b describes the Pinhole model [11] of two cameras to illustrate the projection of a real world object is formed in left and right image planes. The formation of disparity is shown in Figure 2.c

Accuracy of depth perception from a robust stereo vision system is sufficient for segmenting out objects based on their depth, in order to avoid collisions during navigation in real time. The following sections describe the details of hardware and software implementations of stereo vision system in this project.

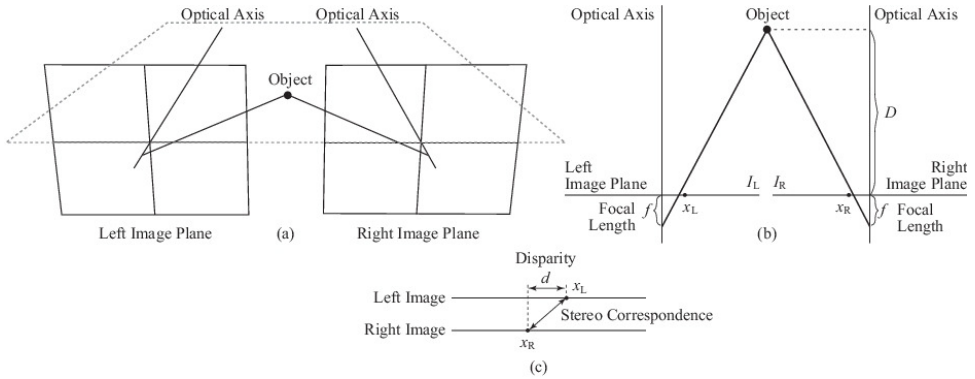

Figure 2. Modelling of stereo rig and disparity formation using pinhole model of cameras

4.1. Building the Stereo Rig

A stereo camera is a type of camera with two or more lenses with separate image sensors for each lens. This allows the camera to simulate human binocular vision, therefore giving it the ability to capture three-dimensional images, a process known as stereo imaging. In this project two CMOS VGA web Cameras (6 USD per camera) of resolution 640×480 with USB2.0 high speed (UVC) interface are used to build the Stereo Rig.

The distance of separation between two cameras which is also known as baseline length, is a crucial parameter of a stereo vision system, which decides the range of reliable depth perception. A longer baseline length increases both the minimum and maximum bounds of this range whereas a shorter baseline length deceases the bounds [12]. Hence the choice of baseline length of a stereo rig is mostly application dependent and limits the usable information available from the rig. Since a mobile robot in an indoor environment is similar to a human navigating in indoor, the most adoptable option for baseline length is the distance of separation between the eyes. A detailed study on human binocular vision system was conducted and the results were recorded. The typical interpupillary distance of humans varies between 50-75mm. The mean interpupillary distance for a human is found out to be 63.2mm [13] and hence the distance of 63mm is selected for the stereo rig used in this project. The mechanical setup was designed using CAD tool and the design is manufactured on acrylic sheet using CNC machine. The cameras were fixed on the rig precisely by monitoring the collinearity of the left and right images obtained. The manufactured stereo rig is shown in Figure 3.

Figure 3. Stereo camera rig made from two webcams.

4.2. The software for stereo vision system

Software required for this project has been developed in C++ language using Microsoft Visual C++ IDE. OpenCV, which is a popular open source computer vision library, is used to implement image processing algorithms. The stereo vision system modelled with Pinhole model is described with the help of two entities, Essential matrix E and Fundamental matrix F. The matrix E includes information about relative translation and rotation between two cameras in physical space whereas matrix F contains additional information related to the intrinsic parameters of a both cameras. Hence Essential matrix relates two cameras with their orientation and Fundamental matrix relates them in pixel coordinates.

The stereo camera will provide simultaneously taken left and right image pairs as an input to the processing unit. The initial task for a stereo vision system implementation is to find above mentioned fundamental and essential matrices. OpenCV provides predefined functions to find these matrices using RANSAC algorithm [14] and hence calibrate cameras and the rig. Calibration requires a calibration object which is regular in shape and with easily detectable features. In this project, the stereo camera calibration is performed using a regular chessboard as it gives high contrast images which contain easily detectable sharp corners which are separated with equal distances. Several left and right image pairs were taken at different orientations of the chessboard and the corners were detected as shown in Figure 4.

Figure 4. Stereo camera and rig calibration using chessboard as a calibrating object. Detected chessboard corners are marked in simultaneously taken left and right images.

The calibration algorithm computes intrinsic parameters of both the cameras and extrinsic parameters of the stereo rig and stores the fundamental and essential matrixes in a file. This information is used to align image pairs perfectly along the same plane by a process called Stereo Rectification. Rectification enhances both reliability and computational efficiency in depth perception. This is a prime step in the routine if the cameras are misaligned or with an infirm mechanical setup. The custom made stereo setup used in this project showed a negligible misalignment which suggested no requirement of rectification of image pairs for reliable results needed for safe indoor navigation. An example of rectified image pair obtained from the algorithm is shown in Figure 5.

The image pair is passed through a block-matching stereo algorithm which works by using small Sum of Absolute Difference (SAD) windows to find matching blocks between the left and right images. This algorithm detects only strongly matching features between two images. Hence the algorithm produces better results for scenes with high texture content and often fails to find correspondence in low textured scenes such as an image of a plane wall. The stereo correspondence algorithm contains three main steps: Pre-filtering of images to normalize their brightness levels and to enhance the texture content, Correspondence search using sliding SAD window of user defined size along horizontal epipolar lines, and post-filtering of detected matches to eliminate bad correspondences.

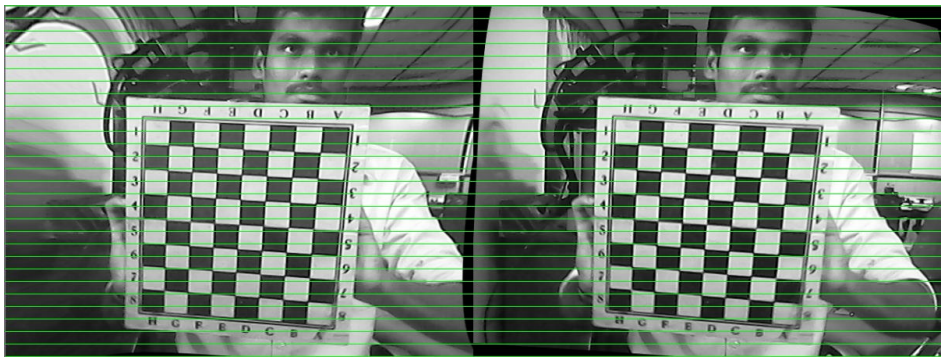

Figure 5. The rectified image pairs

The speed of the algorithm depends on the size of SAD window and the post-filtering threshold used in the algorithm. Larger SAD windows produce poorer results but elapses less time and vice versa. The choice of window size exhibits a trade-off between quality of the results and algorithm execution time, which leads to the conclusion that this parameter is completely application specific. The window size of 9x9 was selected empirically for the algorithms used in this project. Other parameters associated with the correspondence search algorithm are minimum and maximum disparities of searching. These two values establish the Horopter, the 3D volume that is covered by the search of the stereo algorithm. If these values are fixed, the algorithm limits the search for a match in the range between these two values which indirectly confines the real world depth perception between two well defined distances. The formation of horopter is shown in Figure 6 [15]. Each horizontal line in the Figure 6 represents a plane of constant disparity in integer pixels 20 to 12. A disparity search range of five pixels will cover different horopter ranges, as shown by vertical arrows. Considering the velocity of the robot the disparity limits are chosen such that a horopter is formed from 40cm to 120 cm from the frontal plane of the camera.

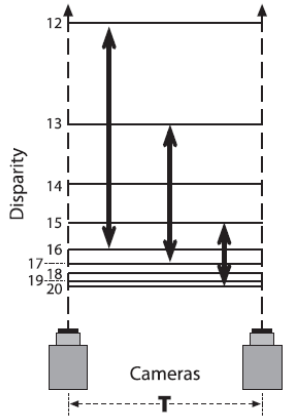

Figure 6. The formation of horopter for different disparity limits

The stereo correspondence algorithm generates a greyscale image in which intensity of a pixel is proportional to disparity associated with corresponding pixel location. The obtained disparity values in the image are mapped to real world distances according to the triangulation equation 1.

$$Z = \frac{f \times T}{d} \qquad (1)$$

Where f is the known focal length, T is the distance of separation between cameras, d is the disparity obtained.

Figure 7 shows the disparity map of a testing image taken during the camera calibration process using chessboard. The low intensity (dark) portions are distant objects whereas high intensity (light) portions are objects which are closer to the camera.

Figure 7: Image from the left camera (left), Computed disparity map (right)

The disparity results are obtained as expected only in particular range of distances from the Stereo Rig because of the nonlinear relationship between disparity and distance [15].

4.3. Depth based Image segmentation and obstacle avoidance

The disparity maps generated by above mentioned algorithm plays a vital role in obstacle avoidance during navigation. The segmentation based on the intensity levels is same as segmentation based on depth. A segmentation algorithm is used to detect near objects which isolates regions which are having high intensity range and searches for connected areas that can form blobs within the segmented regions. The contours of these blobs are detected and bounding

box coordinates for each blob are calculated. The centres of the bounding boxes as well as the bounding boxes are marked on the image. The input image from left camera is divided into two halves to classify the position of the detected object to left or right. The centre of the contour is tracked and if it is found out to be in the left half of the image, algorithm takes a decision to turn the robot to the right side and vice versa. If no obstacles are found in the search region robot will continue in its motion along the forward path. In case of multiple object occurrences in both halves, robot is instructed to take a 90 degree turn and continue the operation. Figure 8 shows the disparity map of several obstacle conditions and the corresponding decisions taken by the processing unit in each case.

Instruction from processing unit is communicated with robot's embedded system through UART communication. Instruction to move forward will evoke the PID algorithm implemented and robot follows exact straight line path unless the presence of an obstacle is detected by the vision system. Our algorithm elapses 200 ms for a single decision making. Dynamic obstacles such as moving humans may not be properly detected by the stereo vision. But this issue is handled by giving high priority for ultrasound sensors and the robot is able to stop instantly.

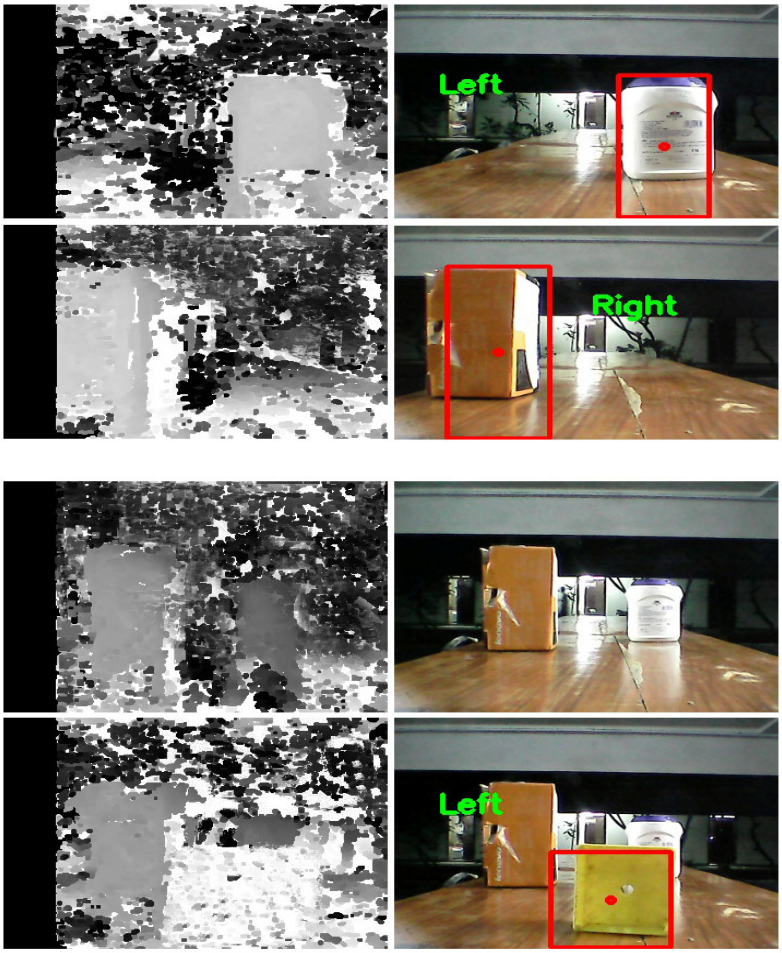

Figure 8. Disparity map of several obstacle conditions in an indoor environment (left). Detected obstacles in the specified distance range and corresponding decisions taken by the processing unit are shown (right)

4.4 3D Reconstruction

Three Dimensional reconstruction is the process of generating the real world model of the scene observed by multiple views. Generated disparity maps from each scene can be converted into corresponding point clouds with real world X, Y and Z coordinates. The process of reconstruction of 3D points requires certain parameters obtained from the calibration of Stereo rig. An entity called Re-projection matrix is formed from the intrinsic and extrinsic parameters and it denotes the relation between real-world coordinates and pixel coordinates. The entries of re-projection matrix are shown in Figure 9.

$$Q = \begin{bmatrix} 1 & 0 & 0 & -c_x \\ 0 & 1 & 0 & -c_y \\ 0 & 0 & 0 & f \\ 0 & 0 & -1/T_x & (c_x - c'_x)/T_x \end{bmatrix}$$

Figure 9. Re-projection matrix of a Stereo Rig

(c_x, c_y) – is the principal point of the camera. The point at which, the image plane coincides with the middle point of the lens.

f – Focal length of the camera, as the cameras in the stereo rig are set to same focal length thus the Re-projection matrix has a single focal length parameter.

T_x – Translation coefficient in x –direction.

The Re-projection matrix thus generated converts a disparity map into a 3D point cloud by using the matrix computation shown in equation 2.

$$Q \begin{bmatrix} x \\ y \\ d \\ 1 \end{bmatrix} = \begin{bmatrix} X \\ Y \\ Z \\ W \end{bmatrix} \quad (2)$$

Where x and y are the coordinates of a pixel in the left image, d is the corresponding disparity associated with that pixel and Q is the re-projection matrix. The real world coordinates can be computed by dividing X, Y and Z by W present in the output matrix.

The calculated 3D point clouds along with corresponding RGB pixel values are stored in the memory in a text file along with the odometric references at each instance of point cloud generation. After the successful completion of an exploration run in an unknown environment, the stored point cloud is retrieved and filtered using Point Cloud Library integrated with C++. Point clouds groups which are having a cluster size above a particular threshold level only is used for 3D reconstruction and thus inherently removes the noisy point clusters. Since error of projection increases with increasing real world distance, point clouds which lie beyond a threshold distance is also removed. 3D reconstructions of each scene are generated and stored according to the alignment of robot at that corresponding time. The visualised 3D reconstruction examples are shown in Figure 10. The overlapped re-projection of continuous scenes can be done to obtain the

complete 3D mapping of the environment. This 3D map can be used as a powerful tool in further navigations in the same environment. It can also be used to plan the path if a destination point in the environment is given to the robot.

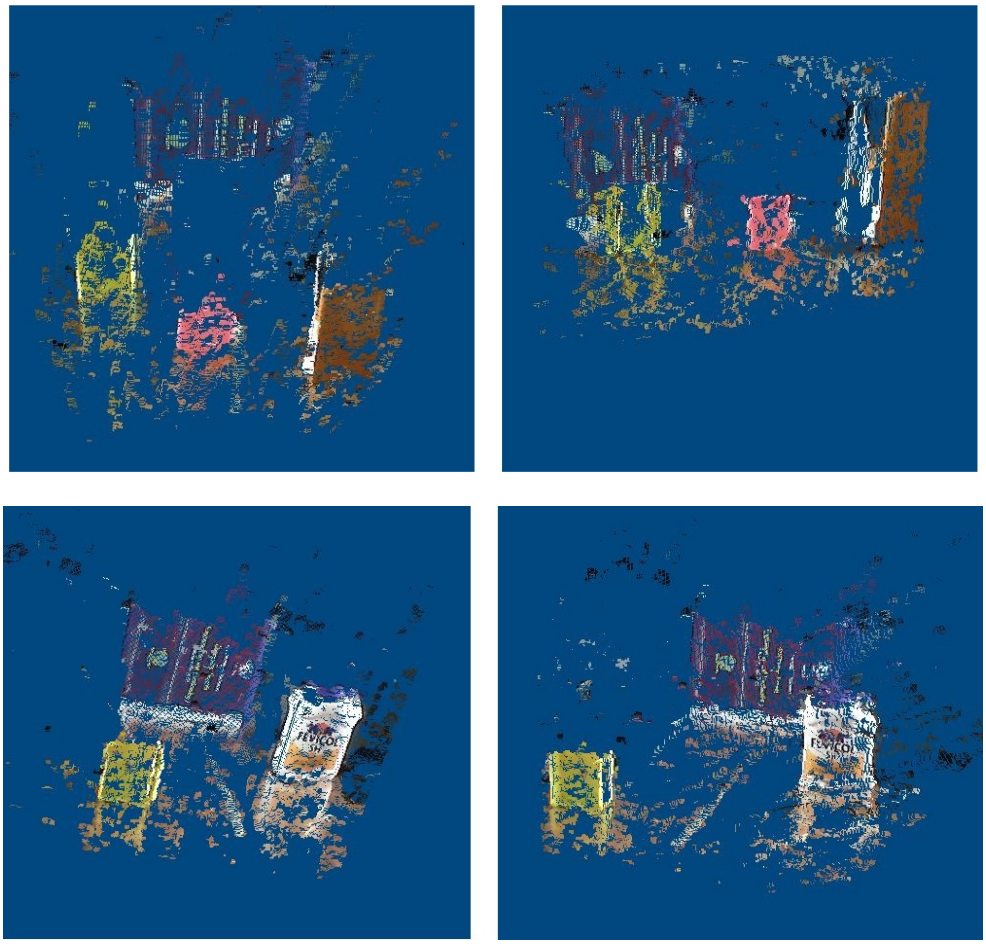

Figure 10. 3D Reconstructions of filtered Point clouds

5. RESULTS

Stereo vision based SLAM architecture is one of the least pondered but rapidly developing research area which has been dealt in this project and we have successfully implemented a cost effective prototype of the stereo camera and robotic platform. The Stereo Vision System produces comparable outputs with that of commercially available alternatives. The total stereo matching program is able to process five frames per second in a 1.6Ghz Intel Atom processor board equipped with 2GB RAM. This performance is adequate for safe indoor navigation for slowly moving robots. The overlapping of vision perception with other information from sensors ensures a nearly error-proof navigation for robot in indoor environments. Accurate 2D mapping of the environment based on the ultrasound data is implemented along with the 3D mapping using the stereo vision. 3D reconstruction elapses 25 to 80ms per frame whereas 2D mapping requires less than 50ms for a sample data collected from a test run timed four minutes. Vision can detect objects just as in the case of human vision and gives the sense of intelligence to the robot. The choice of mechanical parameters of stereo rig, range of the horopter, stereo correspondence

algorithm parameters and filter parameters used for reconstruction were proved to be sufficient for the successful accomplishment of tasks identified during project proposal. The images of robot navigating in the indoor environment are shown in Figure 11.

Figure 11. Robot operates in cluttered indoor environment

6. CONCLUSION AND FUTURE WORK

This paper outlines the implementation of a cost-effective stereo vision system for a slowly moving robot in an indoor environment. The detailed descriptions of algorithms used for stereo vision, obstacle avoidance, navigation and three dimensional map reconstruction are included in this paper. The robot described in this paper is able to navigate through a completely unknown environment without any manual control. The robot can be deployed to explore an unknown environment such as collapsed buildings and inaccessible environments for soldiers during war. Vision based navigation allows robot to actively interact with the environment. Even though vision based navigation systems are having certain drawbacks when compared with other techniques. Stereo vision fails when it is being subjected to surfaces with less textures and features, such as single colour walls and glass surfaces. The illumination level of environment is another factor which considerably affects the performance of stereo vision. The choice of processing platform is crucial in the case of processor intense algorithms used in disparity map generation. Point clouds generated are huge amount of data which has to be properly handled and saved for better performances.

The future works related to this project are developing of a stereo camera which has reliable disparity range over longer distance, implementing the stereo vision algorithm in a dedicated processor board and further development of the robot for outdoor navigation with the aid of Global Positioning System.

REFERENCES

[1] Borenstein, J., Everett, B., and Feng, (1996) Navigating Mobile Robots: Systems and Techniques, A.K. Peters, Ltd.: Wellesley, MA.

[2] J. Buhmann,W. Burgard, A.B. Cremers, D. Fox, T. Hofmann, F. Schneider, J. Strikos, and S. Thrun, (1995) “The mobile robot Rhino,” AI Magazine, Vol. 16, No. 1.

[3] S. Thrun, M. Bennewitz, W. Burgard, A.B. Cremers, F. Dellaert, D. Fox, D. Hähnel, C. Rosenberg, N. Roy, J. Schulte and D. Schulz, (1999) “MINERVA: A second generation mobile tour-guide robot,” in Proc. IEEE International Conference on Robotics and Automation (ICRA), vol.3, No., pp.1999.

[4] Don Murray, and Jim Little, (2000) “Using real-time stereo vision for mobile robot navigation,” Autonomous Robots, Vol. 8, No. 2, pp.161-171.

[5] Pitzer, B., Osentoski, S., Jay, G., Crick, C., and Jenkins, O.C., (2012) "PR2 Remote Lab: An environment for remote development and experimentation," Robotics and Automation (ICRA), vol., no., pp.3200 – 3205.

[6] Kumar S., (2009) "Binocular Stereo Vision Based Obstacle Avoidance Algorithm for Autonomous Mobile Robots," Advance Computing Conference, IACC 2009. IEEE International, vol., no., pp.254-259.
[7] H. Tao, H. Sawhney, and R. Kumar. (2001) "A global matching framework for stereo computation," In Proc. International Conference on Computer Vision, Vol. 1.
[8] Iocchi, Luca, and Kurt Konolige. (1998) "A multiresolution stereo vision system for mobile robots," AIIA (Italian AI Association) Workshop, Padova, Italy.
[9] Schreer, O., (1998) "Stereo vision-based navigation in unknown indoor environment," In Proc. 5th European Conference on Computer Vision, Vol. 1, pp. 203-217.
[10] Yoon, K.J., and Kweon, I.S, (2006) "Adaptive support-weight approach for correspondence search," IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 28, No. 4, pp.650-656.
[11] Z. Zhang, G. Medioni and S.B. Kang, (2004) "Camera Calibration", Emerging Topics in Computer Vision, Prentice Hall Professional Technical Reference, Ch. 2, pp.443.
[12] M. O kutomi and T . K anade, (1993) "A multiple-baseline stereo," IEEE Transactions on Pattern Analysys and Machine Intelligence, Vol. 15, No. 4, pp.353-363.
[13] Dodgson, N. A, (2004) "Variation and extrema of human interpupillary distance," In A. J. Woods, J. O. Merritt, S. A. Benton and M. T. Bolas (eds.), Proceedings of SPIE: Stereoscopic Displays and Virtual Reality Systems XI, Vol. 5291, pp.36–46.
[14] M.A. Fischler and R.C. Bolles, (1981) "Random sample consensus: a paradigm for model fitting with application to image analysis and automated cartography". Communication of ACM, Vol. 24, No. 6, pp.381–95.
[15] G. Bradski and A. Kaehler, (2008) "Learning OpenCV: Computer Vision with the OpenCV Library," O'Reilly Media, Inc.
[16] Martin Humenberger, Christian Zinner, Michael Weber, Wilfried Kubinger, and Markus Vincze, (2010) "A fast stereo matching algorithm suitable for embedded real-time systems", Computer Vision and Image Understanding, Vol. 114, No. 11, pp.1180-1202.
[17] Murray, D. and Jennings, C., "Stereo vision based mapping and navigation for mobile robots," in Proc. 1997 IEEE International Conference on Robotics and Automation, Vol. 2, pp.1694-1699.
[18] Z. Zhang, and G. Xu, (1996) "Epipolar Geometry in Stereo, Motion and Object Recognition," Kluwer Academic Publisher, Netherlands.

AUTHORS

Arjun B Krishnan received Bachelor of Technology degree in Electronics and Communication Engineering from Amrita Vishwa Vidyapeetham, Kollam, India in 2014. Currently, he is working as a researcher in Mechatronics and Intelligent Systems Research Laboratory under Mechanical Dept. of Amrita Vishwa Vidyapeetham. His research interests include Autonomous mobile robotics, Computer vision and Machine learning.

Jayaram Kollipara received Bachelor of Technology degree in Electronics and Communication Engineering from Amrita Vishwa Vidyapeetham, Kollam, India in 2014. He joined as a Program Analyst in Cognizant Technology Solutions, India. His research interests are Image and Signal processing, Pattern recognition and Artificial intelligence.