International Journal of Soft Computing 12 (1): 59-65, 2017
ISSN: 1816-9503


# A 3D Mapping of the Surrounding Object using Stereo-Vision Technique

Hairol Nizam Mohd Shah, Marizan Sulaiman, Mohd Zamzuri AB Rashid, Zalina Kamis,
Nursabillilah Mohd Ali, Mohd Shahrieel Mohd Aras, Faizil Wasbari and Zakarya Omar
Faculty of Electrical Engineering, Center for Robotics and Industrial Automation,
University Teknikal Malaysia Melaka, Durian Tanggal, Melaka, Malaysia

**Abstract:** In this study, a 3D mapping of the surrounding object using stereo-vision technique is proposed. A stereo-vision system is constructed and a navigation algorithm is developed to extract 3D features of objects in the surrounding. Those 3D features are back projected onto a mesh grid to reconstruct the 3D module of that object. Building 3D maps and autonomous localization is a fundamental characteristic of an autonomous operating robot in unknown environment. In addition, rescue activities in hazardous places require robots with such capabilities. In the past, building 3D maps were based on the use of laser, sonar sensor or a combination of a single camera with either laser or sonar sensor. Stereo vision system is the technique used to construct the 3D modules. Two cameras are mounted horizontally on the top of the robot and are directed to scan the forward path of the robot. 3D maps containing the location of objects in the surrounding provide the robot with a proper guidance necessary for its navigation. Those maps will enhance the vision-based robot's ability to navigate and localize themselves autonomously in the future.

**Key words:** Image processing, 3D map reconstruction, feature extraction, stereo camera, building

## INTRODUCTION

Many robotic applications need the 3D model of the surrounding in order for robots to able to detect location, navigate towards destination and avoid colliding with obstacles. For this purpose stereo-vision was first introduced few decades ago. It extracts features from the captured images and represents them in terms of plane location and depth in relation to the current robot position. Stereo-vision approach uses two cameras (Sulaiman *et al*., 2013; Sulaiman *et al*., 2014) mounted horizontally and identically on top of the robot. They are directed towards the front path of the robot to acquire the path's directories and store it in the robot memory. These two cameras capture identical images for the same object from two different locations. Those images can be further processed to provide reliable and useful features. Applying disparity approach on the two images the depth (z-axis) of an object can be obtain besides the planar (x, y-axis) coordinates. For the full 3D model, the map models the environment without any assumptions about it. The representation models the occupied areas as well as free path. In the case of unknown areas, the navigation information about the areas is important for the autonomous exploration and for future navigation.

**Literature review:** Recently, there has been a great interest in utilizing mobile robots in building 3D maps of the surroundings. There are many approaches in building 3D maps. In laser range finder sensors which have been developed in order to obtain more compact and detailed 3D modules to be used in semi-structured environment such as partially destroyed buildings where robots are to do rescue activity (Puente *et al*., 2009). It's emphasized that for a mobile robot to be capable of sensing its location and navigate successfully towards its required destination avoiding obstacles a featured-based representation of the environment must be used. These futures are extracted are to be as reliable and precise as possible. However, it's still difficult to apply this technique due to noise accompanying the gathered data.

3D feature of a scene can be recovered by the use of stereo vision (Sun *et al*., 2010). The difference between the two images taken by an artificial vision system can be used for the extraction of 3D characteristics such as object position, depth and the surface normal. Using the disparity between both left and right images along with the focal length of the cameras and the baseline between them can enhance the extraction of the 3D details and the real world coordinate of an object. Scharstein and Szeliski

**Corresponding Author:** Hairol Nizam Mohd Shah, Faculty of Electrical Engineering,
Center for Robotics and Industrial Automation, University Teknikal Malaysia Melaka, Durian Tanggal,
Malaysia

(2002) it's shown that any algorithm of vision system makes some assumption for the real world captured by an image. An example of these assumptions is how the algorithm measures the evidence that points on both images match each other and they represent the same point on the scene.

Stereo vision approach must carry out some sort of details extraction instead of the 2D maps (Saez and Escolano, 2004) which is often represented in a condensed 2D grid module with no realistic or scalable 3D generalization. For instance, stereo data is used to find the 3D plane's Hough transform in the surrounding and are extracted through a voting scheme. Although, some manual guidance is used when there is a lack of input data, a stereo vision is fused with information in order to recover 3D features. On the other hand, 3D landmarks in the captured image are used to construct the map. By using a stereo vision the depth information of an image can be obtained which makes it possible to get the geometric features of the detected objects (Shah *et al*., 2013; Shah *et al*., 2016). The core of the algorithm developed to process the stereo images and to combine the region based and edge-based elements. So, the analysis is done by the use of three images processing techniques: segmentation process in the second stage, feature detection in the third stage and finally a feature classification in the fourth stage of the extractor (Maragos and Schafer, 1987):

$$Z = \frac{fb}{d} \quad (1)$$

$$X = \frac{u - u_c}{f} Z \quad (2)$$

$$Y = \frac{v - v_c}{f} Z \quad (3)$$

Where:

b (m) = The baseline of the stereo camera
f (pixel) = The focal lengt
d = Xr-Xl
(u, v) = The pixel position of the feature point from the center of the image ($u_c$, $v_c$) on the 2D disparity image (Moghadam *et al*., 2008)

The scene structure is segmented by a set of planar disparity planes (Li and Wang, 2012). Disparity planes are determined by the three parameters $C_1$, $C_2$ and $C_3$ that define the disparity d for each image pixel (x, y): $d = C_1x+C_2y+C_3$. However, due to the variety number of disparity planes the number is reduced by extracting a set of disparity planes that is enough to represent the scene. It's done by applying local matching in the pixel domain followed by a disparity plane estimation step. The most popular dissimilarity measures are Squared intensity Differences (SD) and Absolute intensity Differences (AD) which are directly assuming the constant colour constraint (Klaus *et al*., 2006). In a very simple visual grid mesh is generated together with a set of segmented colour images. In particular, each colour image is pre-segmented into an object and background (Nghiem *et al*., 2010). The intersection of all the mesh cones forms a backbone for our 3D reconstruction. The visual mesh is then triangulated to obtain the surface representation. On the other hand, Li and Zhou (2010) says that due to the negative impact of noise in extracting or matching straight line, critical edging and line distinguishing algorithm cannot be used. Thus, a novel triangulation algorithm is suggested to recognize those not extracted or matched edges and lines easier.

Table 1: 3D mapping techniques comparison

| Criteria/Method | Extracted information | Accuracy | Cost |
|---|---|---|---|
| Laser range finder | Less compact | Less assumption | Expensive |
| Camera+depth sensor | Less compact | Many assumption | Affordable |
| Stereo vision | Compact | Less assumption | Affordable |

In have proved that the disparity errors and uncertainty of a stereo vision system are Gaussian distributed. There is a constant standard deviation of the disparity map (Matthies and Grandjean, 1994). However; it's always assumed a varying disparity standard deviation interrelated with the estimated disparity. Faraway objects have a smaller standard deviation since the disparity is smaller than the disparity of the close objects and vice versa.

To sum up there are three main methods have been used for the purpose of building 3D maps of the surrounding. First method is by using the laser range finder sensor. However, this technique give a less compact maps on which the information extracted are not sufficient enough to build the 3D maps. The second method is by the combination of a single camera and a depth finder sensor (e.g., sonar sensor). However, this technique required a sensor that scans a wider area or a rotating sensor to scan the whole range of view which exaggerates the price. The third technique is by using a stereo vision approach (Lim *et al*., 2009, 2010). The following table gives a summarize on the three techniques. Table 1 presents some of the differences and similarities phases among the three different techniques.

## MATERIALS AND METHODS

As shown in Fig. 1, the following flow chart of the stereo vision processing starts by capturing and a pair of images 'left and right'. Those images are then preprocessed in which a color conversion from RGB to

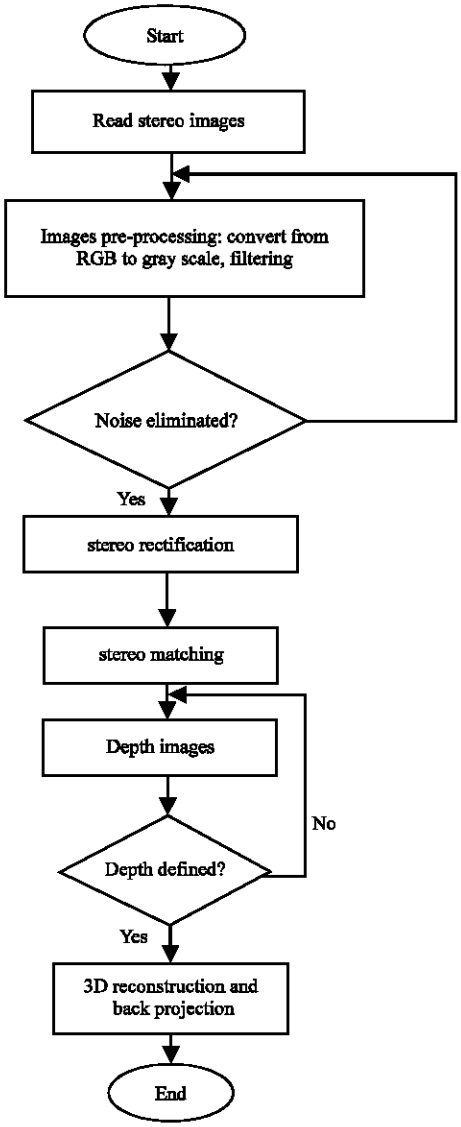


Fig. 1: Stereo vision system flow chart

grayscale is carried on (Hanif *et al*., 2009; Lim *et al*., 2009). Furthermore, filter is added to eliminate/minimize the effect of noise of the images. Then, the stereo rectification, block matching and depth estimation are take part.

**Depth recovering:** If P = (XA, YA, ZA) is a point in a coordinate space based on the optical centre of the left camera from its projections, Pl and Pr as shown in Fig. 2. Where f is the focal length and T is the base line 'horizontal distance between left and right cameras. Z refers to the depth "distance between point P and the camera:

$$X_A = f\frac{x_l}{z_l} \Rightarrow X_A = \frac{x_l z_l}{f} \quad (4)$$

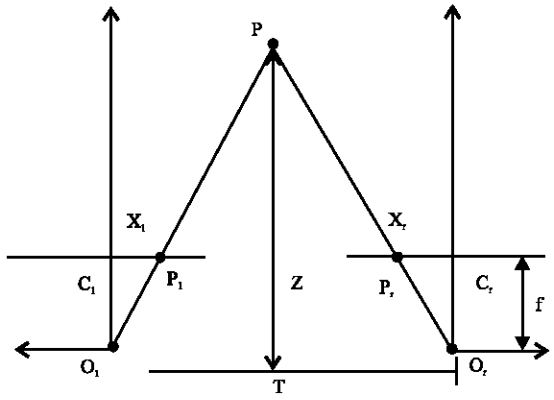


Fig. 2: Stereo triangulation and geometry

$$X_A = \frac{x_r z_r}{f} \Rightarrow X_A = f\frac{x_r}{z_r} \quad (5)$$

Generally, the two cameras are related to each other by a Rotation R and a Translation T. Based on the parallel camera optical axes:

$$Z_r = Z_i = Z \text{ and } X_r = X_i = X - T \quad (6)$$

So, we have:

$$P_r = R(P_l - T) \quad (7)$$

$$\frac{x_l Z}{f} - T = \frac{x_r Z}{f} \quad (8)$$

Finally:

$$Z = \frac{Tf}{d} \quad (9)$$

Where the disparity $d = X_i - X_r$ is the difference in position between the corresponding points in the two image planes, commonly measured in pixels.

**3D map reconstruction (back projection):** Reconstructing the 3D models depicts a realistic object model through connecting points altogether to get the surface of the object. With the stereo depth map and knowledge of camera's intrinsic parameters, it becomes possible to back-project image pixels into 3D points on a map. The camera intrinsic parameters have been found from the calibration step and the results are plugged into the following matrix:

$$k = \begin{bmatrix} \text{focal length\_x} & \text{skew\_x} & \text{camera center\_x} \\ 0 & \text{focal length\_y} & \text{camera center\_y} \\ 0 & 0 & 1 \end{bmatrix}$$

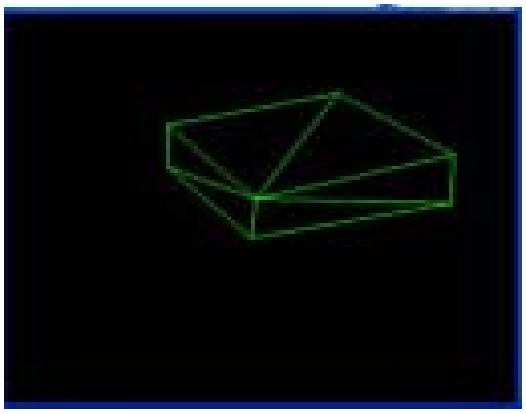

Fig. 3: Corner detection and triangulation result

which relates the 3D world coordinate to the camera coordinate as:

$$\left[x_{camera}\ y_{camera} 1\right]^T = k \times \left[x_{world}\ y_{world}\ z_{world}\right]^T \quad (11)$$

$$z_{world} = \text{focal length} \times \frac{1 + \text{stereo baseline}}{\text{disparity}} \quad (12)$$

As shown in Fig. 3 the first step is to perform some pre-processing on the stereo images. This preprocessing can speed up the execution time of the algorithm and provides more accurate and reliable model of the environment. Approach to be used for feature extraction is based on the observation of their particular geometric form and their color. As a result, regions on the basis of geometric are extracted form and color information. On the other hand, edge detection is employed which is used for two reasons as in Fig. 3. Firstly, the edge information is used to detect the border to the contour of a surface. Secondly, it is used to define the geometric model of the connected component to classify them as an accepted feature or not.

## RESULTS AND DISCUSSION

**3D mapping with stereo (baseline length = 15 cm):** Figure 4a-c shows the left image, right image and their resulted 3D model. The 3D model is lays in the space of x,y and z coordinates. This 3D model contains the information of the various dimensions of that object including the depth (distance of an object from the camera plane). Table 2 summarizes few points on that 3D model and compare it to the actual dimensions where z represents the depth.

Those eight points were randomly picked up along the object. After measuring the actual dimension the cursor is used to point at the corresponding point on the 3D model to read the exact values of x, y and z. The estimation errors of all x, y and z are the difference between the actual dimension and the model estimation of each corresponding points. For point 1:

$$e_x = x_{actual} - x_{model} \quad e_x = 0.55 - 0.24 = 0.31 \text{ cm}$$
$$e_z = z_{actual} - z_{model} \quad e_z = (-5.85) - (-515) = 0.7 \text{ cm}$$
$$e_y = y_{actual} - y_{model} \quad e_y = (-0.8) - (-0.8) = 0 \text{ cm}$$

Figure 5 shows that there is an inaccuracy in estimating the x dimension 'width' of an object. There is a mean error of 0.32375×20 cm which can be reasoned to the different in angles of the stereo camera with respect to the object. Usually there is a translation distance between the object and the camera "called depth",; however, if there is an angle difference between the camera axis of view and the object a rotational difference occurs. This rotation is very difficult to estimate accurately.

On the other hand, Fig. 6 shows the y dimension 'height' of the object with less error in compare to the x dimension. The mean error is equal to 0.12875×20 cm. Both cameras are place at an identical height with respect to the object. This error can be eliminated by a trial and error method to improve the values of camera's intrinsic matrix so that estimation be further improved and accurately represent the object.

Lastly on Fig. 7, it shows the actual and model depth of the object with a mean estimation error of -0.6975×20 cm. The cameras are principally placed at the origin of the z-axis so the negative sign merely indicates that an object is place in the negative direction of the z-axis.

**Combination of uncertainties along the 3 coordinates:** Fractional or percentage uncertainty can often be determined in the final result simply by adding the uncertainties. Table 2 point 8 is taken as an example to explain how uncertainties are accumulated.

The 3D features of this objects has uncertainties that accumulate uncertainties/errors in the length width and depth. If we assume that point to be the smallest change (differential) in volume ΔV then:

$$\Delta V,\ \text{volume} = X \times Y \times Z \quad (13)$$

In this situation, each measurement 'X and Y and Z' enters the calculation as a multiple to the first power; we can find the percentage uncertainty in the result by adding together the percentage uncertainties in each individual measurement:

$$\text{Uncertainty in volume} = (\text{uncertainty in X}) + (\text{uncertainty in Y}) + (\text{uncertainty in Z}) \quad (14)$$

Table 2: Data comparison and 3D estimation error (baseline = 15 cm)

| Items | Model (x, y, z)×20 cm | Actual (x, y, z)×20 cm | Error (e x) | Error (e y) | Error (e z) |
|---|---|---|---|---|---|
| 1 | (0.24, - 0.80, -5.15) | (0.55, -0.8, -5.85) | 0.31 | 0 | -0.7 |
| 2 | (0.20, -0.54, -5.27) | (0.48, -0.55, -5. 90) | 0.28 | -0.01 | -0.63 |
| 3 | (0.49, -0.08, -5.17) | (0.75, -0.10, -5.90) | 0.26 | -0.02 | -0.73 |
| 4 | (-0.06, -0.12, -5.18) | (0.03, -0.14, -5.93) | 0.09 | -0.02 | -0.75 |
| 5 | (- 0.47, -0.60, -5.18) | (-0. 21, 0.58, -5.95) | 0.26 | 1.18 | -0.77 |
| 6 | (-1.23, -0.71, -5.20) | (-0.85, -0. 68, -5.98) | 0.38 | 0.03 | -0.78 |
| 7 | (0.02, 0.94, -5.48) | (0.21, 0.90, -6.00) | 0.19 | -0.04 | -0.52 |
| 8 | (-0.82, 1.09, -5.18) | (0.02, 1.00, -5.88) | 0.80 | -0.09 | -0.7 |
| | | Mean error = Σe/8 | 0.32374 | 0.12875 | -0.6975 |

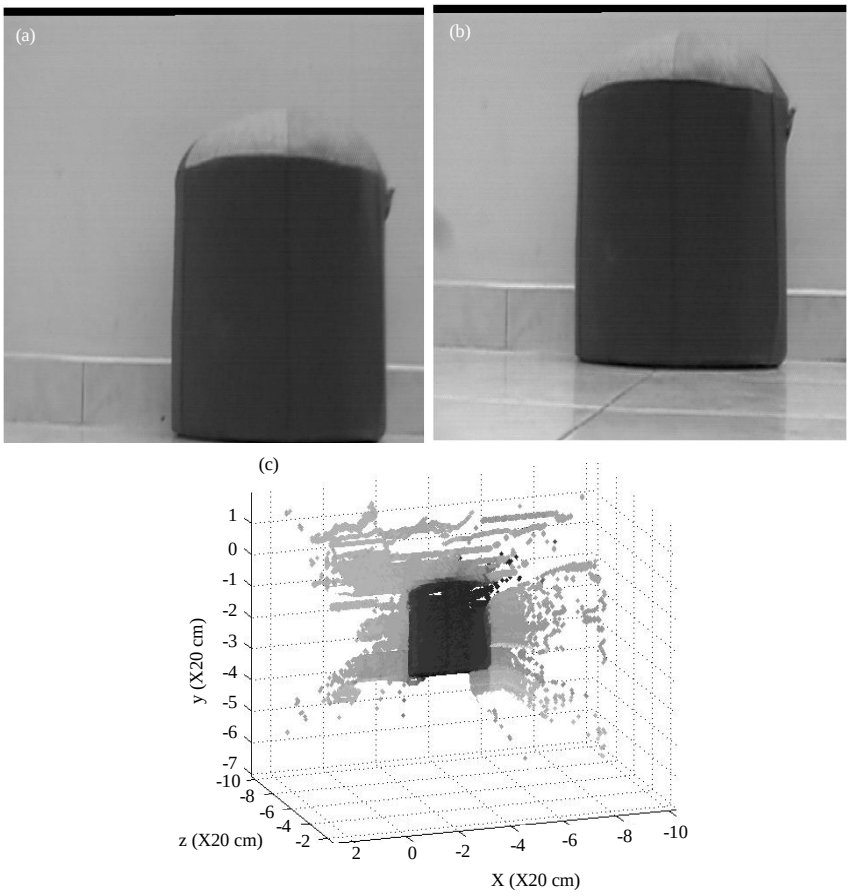


Fig. 4: Stereo images and their corresponding 3D model: a) Left image; b) Right image and c) 3D model

As shown on Fig. 4, reading stereo images from file. These two images were taken by a two camera translated about 15cm from each other. The color space conversion from RGB to gray and the image filtering process is to get the images ready for the steps to come. Filtering, on the other hand, can be repeated as many times as needed once there's a noise that needs to be eliminated. The color composition shows the disparity between booth images. However, the rectification is needed to ensure that similar pixels are aligned in the same row of pixels. Last but not least, depth estimation and 3D reconstruction algorithm development is the task to be continued to achieve the objectives of this project.

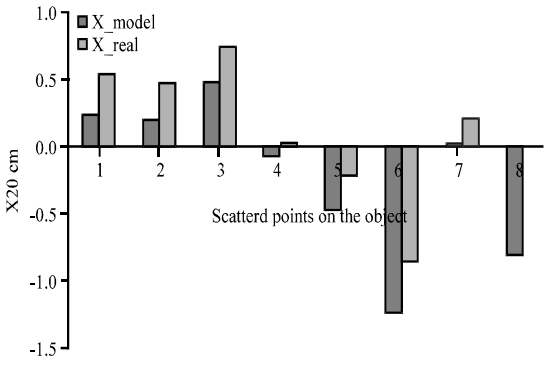


Fig. 5: Real and model estimation of the x dimension

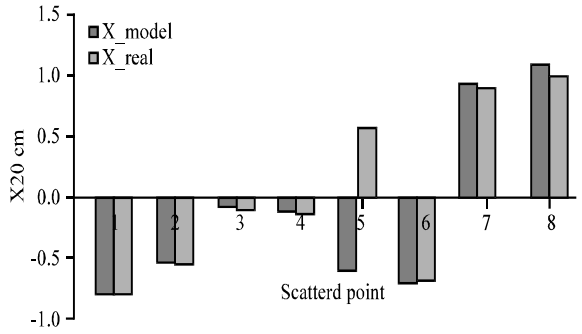


Fig. 6: Real and model estimation of the y dimension

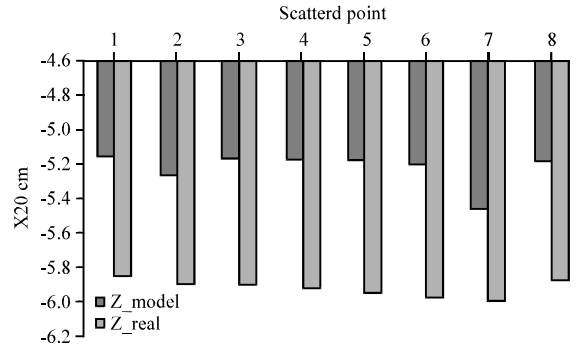


Fig. 7: Real and model estimation of the Z dimension

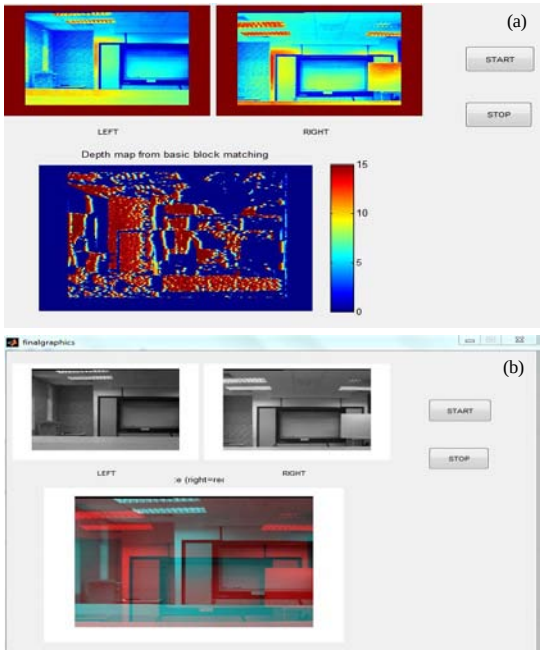


Fig. 8: a) GUI layout and b) GUI when the process is ON

**Graphic User Interface (GUI):** Figure 8a and b shows the developed GUI of the system. Mainly, there are two push buttons that start and terminate the process. In addition, there are three displays to show the processed stereo images and the resulting 3D model. This enables users to deal with the system without interfering with the source code program.

## CONCLUSION

3D mapping algorithm has been developed and implemented in MATLAB. The algorithm was developed and tested for a numerous of times changing the baseline length, depth variation, environment background and light intensity. However, the algorithm has been proved to be efficient to some extent if and only if the background color and light intensity sufficient. The performance of the stereo vision system was analyzed thoroughly. In addition, the 3D maps of the surrounding were successfully built. The result section of this report shows how the model of an object varies in dimension with compare to the actual object as a result of accumulated errors. However, the 3D reconstruction of the surrounding was done with a minimal distortion and a tangible estimation error. Estimation error and distortion may affect badly future application of the built maps.

## RECOMMENDATIONS

For the future research of this project, it's recommended that a higher resolution camera to be used to enhance the extraction of features as it is going to provide a great deal of information. Furthermore, the execution time of the program could be noticeably reduced when using a higher speed processor. Furthermore, different algorithm should be implemented and compared to the existing one to ensure better results to be implemented in a real application. In addition, there are many other problems which need to be addressed in the future. The algorithms developed here can be extended to enable the building of global 3D maps.

## ACKNOWLEDGEMENTS

Researchers are grateful for the support granted by Universiti Teknikal Malaysia Melaka (UTeM) in conducting this research through grant FRGS/2/2014/FKE/01/F00238 and Ministry of Higher Education.

## REFERENCES


Haniff, H.M., M. Sulaiman, H.N.M. Shah and L.W. Teck, 2011. Shape-based matching: Defect inspection of glue process in vision system. Proceedings of the 2011 IEEE Symposium on Industrial Electronics and Applications (ISIEA), September 25-28, 2011, IEEE, Malaysia, Asia, ISBN:978-1-4577-1418-4, pp: 53-57.

Kalus, A., M. Sormann and K. Karner, 2006. Segment-based stereo matching using belief propagation and a self-adapting dissimilarity measure. Proceedings of the 18th International Conference on Pattern Recognition, Volume 3, August 20-24, 2006, Hong Kong, pp: 15-18.

Li, C. and Y. Zhou, 2010. 3D auto-reconstruction for street elevation based on line and plane feature. Proceedings of the 2nd International Conference on Computer and Automation Engineering (ICCAE) 2010, Vol. 1, February 26-28, 2010, IEEE, Wuhan, China, ISBN:978-1-4244-5585-0, pp: 460-466.

Li, X. and J. Wang, 2012. Image matching techniques for vision-based indoor navigation systems: Performance analysis for 3D map based approach. Proceedings of the 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), November 13-15, 2012, IEEE, Sydney, Australia, ISBN:978-1-4673-1955-3, pp: 1-8.

Lim, W.T., M. Sulaiman and H.N.M. Shah, 2009. Flexible approach for region of interest creation for shape-based matching in vision system. Proceedings of the 2009 Conference on Innovative Technologies in Intelligent Systems and Industrial Applications (CITISIA 2009), July 25-26, 2009, IEEE, Malaysia, Asia, ISBN:978-1-4244-2886-1, pp: 205-208.

Lim, W.T., M. Sulaiman, M. Shah, H. Nizam and R. Omar, 2010. Implementation of shape-based matching vision system in flexible manufacturing system. J. Eng. Sci. Technol. Rev., 3: 128-135.

Maragos, P. and R.W. Schafer, 1987. Morphological filters-Part II: Their relations to median order-statistic and stack filters. IEEE. Trans. Acoust. Speech Signal Process., 35: 1170-1184.

Matthies, L. and P. Grandjean, 1994. Stochastic performance, modeling and evaluation of obstacle detectability with imaging range sensors. IEEE. Trans. Robo. Autom., 10: 783-792.

Moghadam, P., W.S. Wijesoma and D.J. Feng, 2008. Improving path planning and mapping based on stereo vision and lidar. Proceedings of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV 2008), December 17-20, 2008, IEEE, Singapore, Asia, ISBN:978-1-4244-2286-9, pp: 384-389.

Nghiem, V.N., J. Cai and J. Zheng, 2010. Rate-distortion optimized progressive 3d reconstruction from multi-view images. Proceedings of the 18th Pacific Conference on Computer Graphics and Applications (PG) 2010, September 25-27, 2010, IEEE, Singapore, Asia, ISBN:978-1-4244-8288-7, pp: 70-77.

Puente, P.D.L., D.R. Losada, A. Valero and F. Matia, 2009. 3D feature based mapping towards mobile robots enhanced performance in rescue missions. Proceedings of the 2009 IEEE International Conference on Intelligent Robots and Systems (IROS 2009), October 10-15, 2009, IEEE, Madrid, Spain, ISBN:978-1-4244-3803-7, pp: 1138-1143.

Saez, J.M. and F. Escolano, 2004. A global 3D map-building approach using stereo vision. Proceedings of the 2004 IEEE International Conference on Robotics and Automation (ICRA'04), Vol. 2, April 26-May 1, 2004, IEEE, Madrid, Spain, ISBN:0-7803-8232-3, pp: 1197-1202.

Scharstein, D. and R. Szeliski, 2002. A taxonomy and evaluation of dense two-frame stereo correspondence algorithms. Int. J. Comput. Vision, 47: 7-42.

Shah, H.N.M., M.Z.A. Rashid, M.F. Abdollah, M.N. Kamarudin and Z. Kamis *et al*., 2016. Detection of mobile object in workspace area. Intl. J. Signal Process. Image Pattern Recognit., 9: 225-232.

Shah, M., H. Nizam, A. Rashid and M. Zamzuri, 2013. Develop and implementation of autonomous vision based mobile robot following human. Intl. J. Adv. Sci. Tecnol., 51: 1-6.

Sulaiman, M., H.N.M. Shah, M.H. Harun and M.N. Fakhzan, 2014. Defect inspection system for shape-based matching using two cameras. J. Theor. Appl. Inf. Technol., 61: 288-297.

Sulman, M., M. Shah, H. Nizam, M.H. Harun and T.L. Wee *et al*., 2013. 3D Gluing defect inspection system using shape-based matching application from two cameras. Intl. Rev. Comput. Software, 8: 1997-2004.

Sun, J.H., B.S. Jeon, J.W. Lim and M.T. Lim, 2010. Stereo vision based 3D modeling system for mobile robot. Proceedings of the 2010 International Conference on Control Automation and Systems (ICCAS), October 27-30, 2010, IEEE, Seoul, South Korea, ISBN:978-1-4244-7453-0, pp: 71-75.