Study on 3D Scene Reconstruction in Robot Navigation using Stereo Vision

NURNAJMIN QASRINA ANN[1], M.S. HENDRIYAWAN ACHMAD[3], LUHUR BAYUAJI[2], M. RAZALI DAUD[1], DWI PEBRIANTI[1]

[1]Faculty of Electrical and Electronics Engineering, Universiti Malaysia Pahang, Malaysia
[2]Faculty of Computer Science and Software Engineering, Universiti Malaysia Pahang, Malaysia
[3]Department of Electrical Engineering, Universitas Teknologi Yogyakarta

qasrinaann@gmail.com, hendriyawanachmad@uty.ac.id, luhurbayuaji@ump.edu.my, mrazali@ump.edu.my, dwipebrianti@ump.edu.my

***Abstract*-In this paper, a 3D scene reconstruction by using stereo vision is presented. Stereo camera parameters from the camera calibration process and disparity map are two important parameters to obtain an accurate result for the 3D reconstruction. 3D reconstruction process generates the coordinates of world points (point cloud). From the information of the point cloud, the interested object is segmented out. Additionally, the noises left in the image is eliminated. The experimental result obtained shows that only the background and the interested object from the overall scene appeared in the processed image. The center of gravity is also determined as the reference value for the robot navigation. An estimation error model is also introduced in this study to increase the accuracy of the distance measurement by using the developed stereo vision system from the mobile robot. This experiment provides a foundation for navigation and object tracking for a mobile robot.**

***Index Term*-stereo vision, robot navigation, 3D scene reconstruction, stereo camera calibration**

I. INTRODUCTION

In the past few years, a growing interest in mobile robots research has been registered; interest that has been in part motivated for the high development reached in the sensors and mechanisms areas. This development allowed the implementation of systems with a high degree of interaction between the robot and the environment so the robot can obtain a more confident description of the context, which can be used to develop its tasks precisely [1]. The advantages of the robot are the advanced intelligent on the robot; digital or hybrid minds will be far more robust and specialized than those humans. A robotic brain could be designed to excel at any number of arbitrary tasks including able to calculate better and faster than humans, have superior pattern recognition skills, unlimited attention spans, and infallible memory [2]. Other than that, it will be easy for engineers to upgrade robots. Robots will also improve access to their internal physical and cognitive states.

Stereo vision is a good choice for several reasons: it is low power, low cost, and can register dense range information from close objects [3]. A group of researcher from Wuhan University, China introduced the 3D reconstruction workflow using Microsoft Kinect Sensor, a new instrument utilized in surface measurement [4]. Kinect can reach 0.4 to 1.5 meter for the distance range, which is only suitable for room size environment [4]. The features of the Kinect mentioned above are not suitable for too high accuracy capture such as 3D virtual displays of indoor environment or human body reconstruction and detection.

Camera calibration is considered as an important issue in computer vision. Accurate cameras calibration is especially crucial for applications that involve quantitative measurements, depth from stereoscopy or motion from images [5]. The accuracy of the camera calibration is assessed by comparing the difference between the known 3D world-coordinate of the test points and their reconstruction from measured data [6]. The correctness of the result of all the functions is heavily dependent on an accurate calibration of the vision system [7].The problem of camera calibration is to compute the camera extrinsic and intrinsic parameters [5]. The extrinsic parameters of a camera indicate the position and the orientation of the camera respect to the coordinate system, and the intrinsic parameters characterize the inherent properties of the camera optics, including the focal length, the image center, the image-scaling factor and the lens distortion coefficients. The techniques found in the literature for camera calibration can be broadly divides into three types: linear methods, nonlinear methods and two-step techniques. The first technique, linear methods is assumed as a simple pinhole camera model and incorporate no distortion effects. The limitation in this case is the camera distortion cannot be incorporated and the lens distortion effects cannot be corporated. In non-linear techniques, the relationship between parameters is established and then an iterative solution is found by minimizing some error term. Lastly, two-step techniques involve a direct solution of some camera parameters and an iterative solution for the other parameters. Iterative solution is also used to reduce the errors in the direct solution.

A stereo vision normally takes two or more images obtained using parallel cameras as inputs. With the images and the stereo camera parameters, disparity map can be created. A desired disparity map should be smooth and should contain sufficient detail for subsequent processing. The spatial displacements of corresponding points between the images of a stereo pair (forming the stereo disparity map) are finite, in principle and unconstrained [8]. The stereo matching problem is considered as an optimization method. There is a way to improve the accuracy of the disparity map by removing the mismatches caused by both occlusions and false targets. A dynamic programming algorithm is used to estimate a disparity field between each image pair [9]. Depth is then estimated and occlusions are optionally detected, based on the estimated disparity fields. Spatial interpolation techniques are examined based on the disparity or depth information and the detection of occluded regions using either stereoscopic or trinocular camera configurations. Various disparity field and depth map coding techniques are then proposed and evaluated, with emphasis given to the quality of the resulting intermediate images at the receiver site.

The outline of the paper is as follows. Section II is about the experimental setup for the experiment. The details of the experiment methodology is described in Section III. Section IV is the result and discussion. The estimate error model is introduced in Section V, which is followed by conclusion and future works in Section VI.

II. EXPERIMENTAL SETUP

The Blackfly stereo vision camera system from Point Grey Research is used throughout the experiment. It consists of two different cameras, which are mounted in parallel to get the stationary position. Then, the cameras have been placed on the mobile robot. The maximum resolution of the camera is 1200x800 pixels. All the images captured by the stereo camera are displayed in the Image Acquisition Toolbox, Matlab. D-Link USB3.0 to Gigabit Ethernet Adapter is used to make sure all the packets sent can be displayed without any error. The packet sent by stereo vision camera is too big and that is why, the real-time analysis is not encouraged to handle the system [10]. Figure 1 shows the mobile robot with sensors. The figure shows a laptop is also placed on the robot for image processing and controls the sensors.

Figure 1 Mobile Robot

III. METHODOLOGY

A simple procedure for 3D scene reconstruction from the developed stereo camera system is shown below.

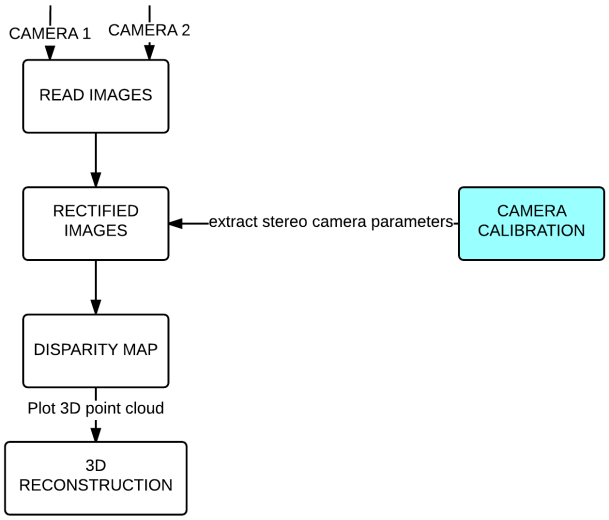


Figure 2 System Overview

A. Stereo Camera Calibration

A stereo system consists of two cameras, camera 1 and camera 2 which are placed horizontally with a certain distance called baseline. Calibration process estimates the parameters of the two cameras. It also calculates the position and orientation of the camera 1 relative to the camera 2.

To begin the calibration process, checkerboard images are captured by the camera with various styles, distances and angles. Ten to twenty images of checkerboard will produce the accurate result for camera calibration. Figure 3 shows one example of the fourteen-checkerboard images during the stereo camera calibration within 50cm to 200cm distance range. The distance of the checkerboard is 80cm away from the stereo camera. The ambient light is the best exposure for the calibration technique compared to the light from the bulb. This is because the calibration process had been done at the car porch (outdoor).

Figure 3 Example of Checkerboard

B. Image Rectification

Stereo image rectification projects images onto a common image plane in such a way that the corresponding points have the same row coordinates. This image projection makes the

image appear as though the two cameras are parallel. Internal camera parameters and information about mutual camera positions and orientations are used in the transformations [11].

After the calibration process, the parameters such as focal length, width and height of the images obtained are used in the rectification process. The algorithm of the image rectification is decomposed to three steps as shown in Figure 4. The first step is the image planes become parallel to CC'. Then, the images rotate in their own plane to have their epipolar lines also parallel to CC'. The last step is a rotation of one of the image planes around CC' aligns corresponding epipolar lines in both images.

It might be argued that one-step of the rotation is enough to rectify the images instead of three steps. However, the three steps procedure makes the algorithm more robust and the residual distortion is equal or only slightly higher [12].

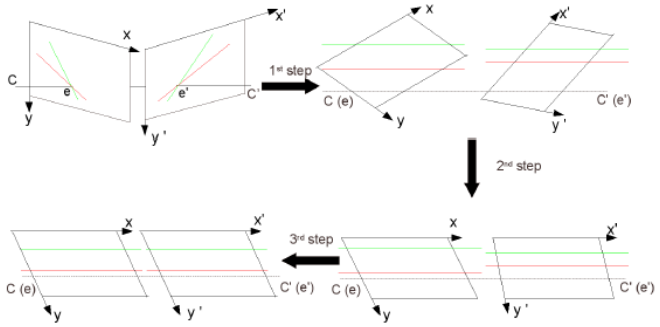


Figure 4 Three-step Rectification [12]

C. Disparity Map

Disparity map is created from a pair of stereo images, I1 and I2 obtained from left and right cameras respectively. In this paper, the disparity estimation algorithm used is semi-global block matching method. In this matching method, the function additionally forces similar disparity on neighboring blocks. This additional constraint results in a more complete disparity estimate than in the basic block matching. Figure 5 shows the steps that performed by the algorithm.

Semi global matching method used a slightly different global cost function for penalizing small disparity steps that are often part of slanted surfaces, less than real discontinuities [13].

$$E(D) = \sum_{p}(C(p, D_p) + \sum_{q=N_p} P_1 T[|D_p - D_q| = 1] + \sum_{q=N_p} P_2 T[|D_p - D_q| > 1] \quad (1)$$

where the first term of the function is sums all pixel-wise matching costs over the whole image, while the second and third terms add a penalty for all pixels with neighbours that have a different disparity.

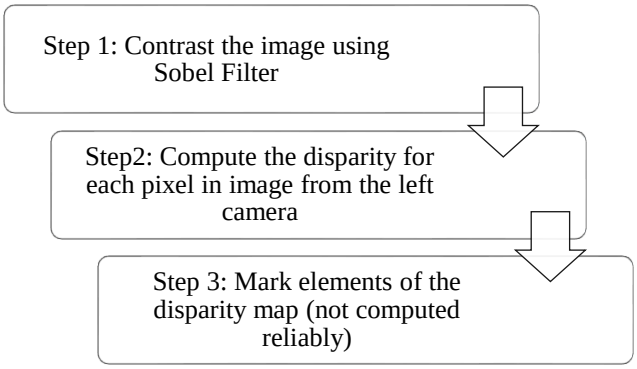


Figure 5 Semi-global Block Matching algorithm

D. 3D Scene Reconstruction

The scene reconstruction process is the relationship between disparity value and stereo camera system parameters. From the process, the point cloud (also known as coordinates of world points) is generated. The 3-D world coordinate are relative to the optical center of left camera in the stereo system.

As shown in Figure 6, when two images are acquired by a stereo camera system, every physical point *M* yields a pair of 2D projections m_1 and m_2 on two images. If both the intrinsic and extrinsic parameters of the stereo system are known, it is possible to reconstruct the 3D location of the point *M* from m_1 and m_2 [14].

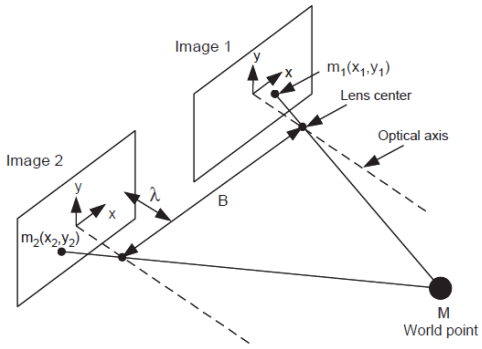


Figure 6 Parallel Stereo Camera Geometry

In the simple case of a parallel camera system, as shown in figure below, the depth of a point *M* can be simply calculated by

$$Z = f - \frac{fB}{x_2 - x_1} = f - \frac{fB}{d} \quad (2)$$

where *B* is the baseline distance between the two cameras, *f* is the focal length of the camera and *d* is the difference distance of *x* in left image and right image.

The main objective of the paper is to capture the interest object in one complete scene by using stereo camera. The process of image analysis is discussed in the next section and the analysis is done off-line.

IV. RESULT AND DISCUSSION

From the previous section, there are details about the methodology of the experiment that had been conducted. The results of every stage are discussed in this section.

Extrinsic parameter from both of the stereo cameras had been visualized in the Figure 7. The figure shows the coordinate of the stereo camera and the position of the checkerboard during the calibration process. All the important parameters such as focal length, rotation and translational vectors had been displayed and saved as *Stereo Parameters*. Figure 8 shows the Mean Projection Errors Bar Graph. The projection errors are the distances in pixels between the detected and the projection points. The overall mean error is 0.40 pixels. The projection errors of less than one pixel are acceptable and showed the calibration process is accurate.

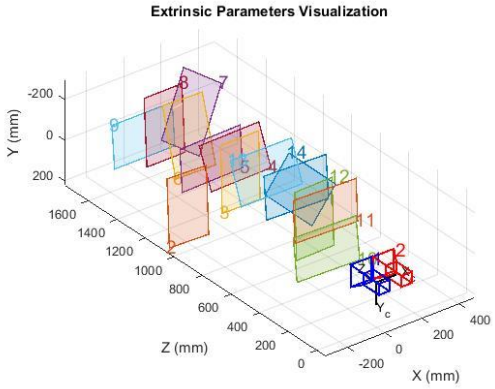


Figure 7 Extrinsic Parameter

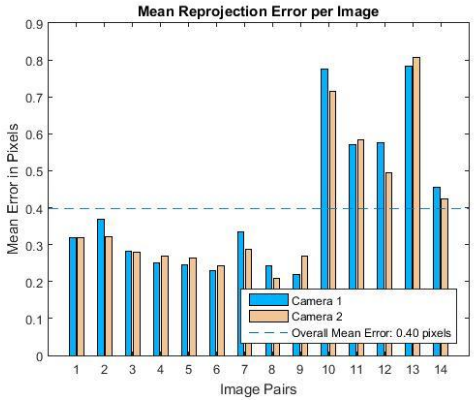


Figure 8 Mean Projection Error

The analysis of the image is using the same image as for calibration. The first image is from left camera and the second image capture from the right camera (refer Figure 3). The position of the camera is showed in Figure 1. Figure 11(a) shows the undistorted and rectified image versions of the original image versions of the original images obtained earlier.

The disparity map appeared after the rectification process is shown in Figure 9. The Sobel filter is used to compute a measure of contrast of the image. The distance of object and camera are quite close to each other and the smaller disparity range is needed. The range of disparity map used is within -6 to 10. With the smaller range of disparity, accurate distance is obtained compared to the larger range.

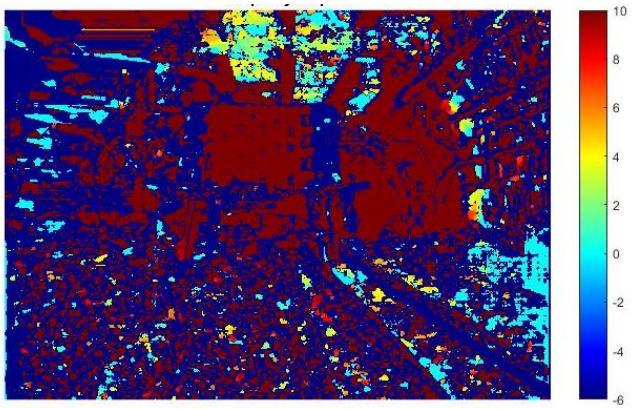


Figure 9 Disparity Map

The next process is to analyze the accuracy of the distance measurement using the developed stereo cameras. This is done by taking a certain object in the image called as Region of Interest (ROI). In this study, checkerboard will be used as the target. Since there are many objects located in the captured image from the stereo camera, other objects need to be taken out.

The ROI in the study is the checkerboard position. The detected checkerboard in the whole scene is in the red circle as shown in Figure 10. The 3D point cloud is plotted to show coordinates of world points.

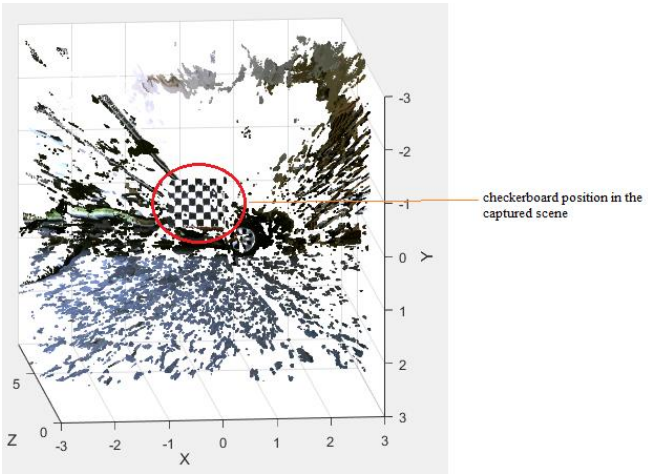


Figure 10 3D Point Cloud

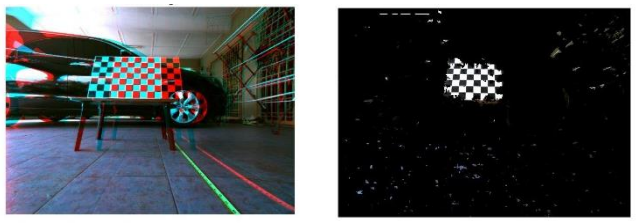
The first step of the algorithm is by segmenting out the checkerboard between 230 and 252.6 centimeters away from the camera. The distance mentioned earlier is obtained from the point cloud data. By referring to Figure 11(b), there are other parts of the scene included after the segmentation process. These parts are called outliers. In order to remove outliers, a

few additional steps are conducted to obtain accurate object extraction.

Figure 11(c) shows the binary image. The binary image has only two values of pixel which are 0 and 1 and it is the simpler version of image to be analyzed for the robot navigation. The RGB (red-green-blue) image is converted to BW (black and white) image. The next step of the algorithm is by filling all the holes with white (binary 1) color as shown in Figure 11(d). The entire detected checkerboard and a few dots in the scene are changed to white. The purpose of the step is to differentiate the detected object and the background of the image. The checkerboard consists of white and black color and it is similar to the binary image. That is why, the black part of the checkerboard is changed or filled with white.

The last stage is to eliminate the noise present in the scene by using flat morphological element and used diamond-shaped structuring element, which is an essential part of morphological dilation and erosion operations. The segmented image produced is shown in Figure 11(e). Only two different objects, the background and the interested object appeared in the processed image.

The distance between the cameras and the checkerboard is calculated by considering the center of gravity (COG) of the checkerboard. The COG of an image is fixed in the relation to the body, and if the body has uniform density, it will be located at the centroid. As shown in Figure 11(f), the red * mark is located at the centroid of the white region and the estimate coordinate value of the centroid is (835, 449). The depth value of the centroid is 173.5 cm corresponding to the point cloud data. The value of COG is very important because it will be the reference value for the robot navigation.



(a) (b)

(c) (d)

(e) (f)

Figure 11 (a) Rectified Image captured from both Cameras (b) Mask the Target in the Captured Scene(c) Binary Image (d) Binary Image with Filled Holes (e) Segmented Image (f) COG of Segmented Image

V. ESTIMATE ERROR MODEL

The raw and experimental data are measured to obtain the error of each corresponding distance image from the stereo camera to the object. Error model is constructed to increase the accuracy of object distance captured by the stereo camera. The fourth order polynomial equation is the best mathematical model that produces the highest regression value (R^2) compared to other type of polynomial equation to improve the accuracy of the stereo camera distance measurement. The regression value obtained is 0.99. Equation (3) shows the equation of the estimate error model.

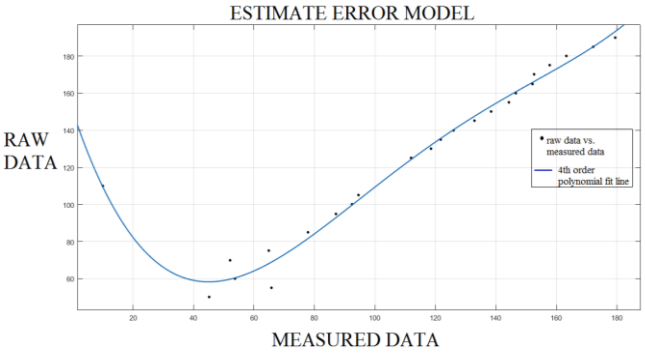


Figure 12 Estimate Error Model

$$y = 9.379 \times 10^{-7}x^4 + 0.0004651x^3 + 0.08175x^2 - 4.875x + 150.63 \quad (3)$$

Table 1 contains the real distance of the object from the stereo cameras, the distance generated by the developed stereo vision system, accuracy from the stereo cameras, the distance after applying the correction and the improved accuracy. The mean of accuracy of the distance measurement accuracy before applying the correction is 89.95% while the mean of accuracy after applying the model is 96.09%. 6.14% of error had been increased by applying the error estimate model as explained earlier in this section.

Table 1 Analysis of the Distance Measurement

Real Distance (cm)	Distance using stereo camera (cm)	Accuracy (%)	Distance after applying correction (cm)	Accuracy (%)
50	45.28	90.56	58.2645	83.5
55	65.86	80.25	68.937	74.66
60	53.79	89.65	60.4025	99.63
70	52.06	74.37	59.6658	85.3
75	64.9	86.53	68.0741	90.77
85	77.91	91.66	81.6441	96.05
95	87.19	91.78	92.9724	97.87
100	92.42	92.42	99.6218	99.62
105	94.48	89.98	102.2598	97.39
110	10.09	91.73	109.296	99.36
125	112	89.6	124.2504	99.4
130	118.5	91.15	131.9066	98.53
135	121.8	90.22	135.6488	99.52
140	126	90	140.2634	99.81
145	132.9	91.66	147.4873	98.28
150	138.4	92.27	152.9522	98.03
155	144.3	93.1	158.5779	97.69
160	146.6	91.62	160.7208	99.55
165	152.1	92.18	165.7777	99.53
170	152.6	89.76	166.2351	97.79
175	157.8	90.17	171.0034	97.72
180	163.3	90.72	176.1506	97.86
185	172.1	93.03	184.9531	99.97
190	179.4	94.42	193.2053	98.31
Mean of Accuracy		**89.95**		**96.09**

VI. CONCLUSION AND FUTURE WORKS

In this paper, an experimental hardware system is developed which consist of mobile robot and stereo vision sensor for scene capture. An estimate error model is constructed to overcome the accuracy problem of images obtained from the stereo camera with 6.14% increase of the accuracy. This experiment provides a foundation for navigation and obstacle avoidance for a mobile robot.

In future work, many experiments will be conducted to study the effect of the illumination to the images captured by the stereo camera. The algorithm of disparity map also should be amended to obtain the accurate and precise result.

VII. ACKNOWLEDGEMENT

The work presented in the paper has been supported by Ministry of Higher Education under Exploratory Research Grant Scheme (ERGS) RDU 130604 and Fundamental Research Grant Scheme (FRGS) RDU 140137.

VIII. REFERENCES

[1] S. U.-G. Horacio Martinez-Alfaro, "Designing, Making and Using a Mobile Robot," Monterrey, Mexico, 1998.

[2] G. Dvorsky, December 2014. [Online]. Available: http://io9.gizmodo.com/12-reasons-robots-will-always-have-an-advantage-over-hu-1671721194.

[3] K. K. R. C. B. Motilal Agrawal, "Localization and Mapping for Autonomous Navigation in Outdoor Terrains: A Stereo Vision Approach," in *IEEE Workshop on Applications of Computer Vision*, 2007.

[4] W. J. H. J. S. T. B. Y. J. Z. Wan Yi, "A Study in 3D-Reconstruction Using Kinect Sensor," 2012.

[5] Q. M. &. S. Khan, "Camera Calibration and three-dimensional world reconstruction of stereo vision using neural network," *International Journal of System Science*, pp. 1155-1159.

[6] F. T. M. L. J.J. Aguilar, "Stereo Vision for 3D measurement: Accuracy analysis, calibration and industrial applications," *Elsevier*, vol. 18, pp. 193-200, 1996.

[7] M. B. A. F. Alberto Broggi, "Self-Calibration of a Stereo Vision System for Automotive Applications," in *International Conference on Robotics and Automation*, Seoul, Korea, 2001.

[8] A. V. Emanuele Trucco, "Chapter 8: Motion," in *Introductory Technique for 3D Computer Vision*, Upper Sadder River, New Jersey, Prentice Hall, 1998, pp. 178-194.

[9] N. G. M. G. S. Dimitrios Tzovaras, "Disparity Field and Depth MapCoding for Multiview 3D Image Generation," *Elsivier*, vol. II, pp. 205-230, 1996.

[10] [Online]. Available: https://www.ptgrey.com/.

[11] M. B. R. J. P. L. M. Z. Patrik Kamencay, "Improved Depth Map Estimation from Stereo Images Based on Hybrid Method," *Radioengineering*, vol. 21, pp. 70-78, 2012.

[12] J.-M. M. a. Z. T. Pascal Monasse, "Three-step Image Rectification," in *British Machine Vision Conference 2010*, 2010.

[13] H. Hirschmuller, "Semi-Global Matching - Motivation, Developments and Applications," *Dieter Fritscn (Ed.)*, pp. 173-184, 2011.

[14] K. S. Hansung Kim, "3D Reconstruction from stereo images for interactions between real and virtual objects," *Elsevier*, pp. 61-75, 2005.

[15] M. G. a. Y.-H. Yang, "Generic-Based Stereo Algorithm and Disparity Map Evaluation," in *International Journal of Computer Vision 47*, 2002.

[16] R. Klette, Concise Computer Vision, London: Springer-Verlag, 2014.