Monterey Bay Aquarium
Research Institute

Stereo camera calibration analysis for seafloor mapping

José Antonio Eyzaguirre, Pontificia Universidad Católica de Chile

Mentors: David Caress, Eric Martin & Giancarlo Troni

Summer 2020

Keywords: Stereo camera calibration, 3D reconstruction, seafloor mapping, image type analysis for camera calibration.

ABSTRACT

In this paper a stereo camera calibration analysis was developed for the MBARI's Low Altitude Survey System. An image type contribution approach was taken, splitting the input images for the calibration in different groups. The reprojection error and row alignment error were used as measures for the different grouped combinations from calibrations. This procedure allowed the researchers to improve object 3D reconstructions and 2D mosaics generation.

INTRODUCTION

This project is part of MBARI's ongoing development of the Low Altitude Survey System for 1-cm scale seafloor mapping and imaging in the deep ocean. This system is located in a sled attached to the bottom of a remotely operated vehicle (ROV). The sled has a tilting platform, pointing to the seafloor, where different sensors take data with the purpose to map a region of interest of the ocean. Stereo cameras are one of these sensors and collect visual data sets. Data from the stereo cameras have two main applications for this project: (i)

to create colored two-dimensional mosaics; and (ii) to create colored three-dimensional reconstructions. For this purpose, it is very important to calibrate the cameras.

Geometric camera calibration is the process of estimating the camera intrinsics and extrinsics parameters. These parameters relate the world and camera coordinate systems, taking into account the focal length of the lens, the size of the pixels, the position of the image center, and the position and orientation of the camera. The calibration can be modeled as an optimization process, where the discrepancy between the observed image features and their theoretical positions is minimized with respect to the camera's intrinsic and extrinsic parameters (Forsyth & Ponce, 2002).

Stereo camera calibration, is the process of estimating the translation and rotation of one camera as seen by the other one. The binocular camera geometry defines a restriction, in which a point seen in one camera is restricted to lie on a line in the other camera, this is called the epipolar restriction (Forsyth & Ponce, 2002). Figure 1 shows the essential geometry of stereo imaging.

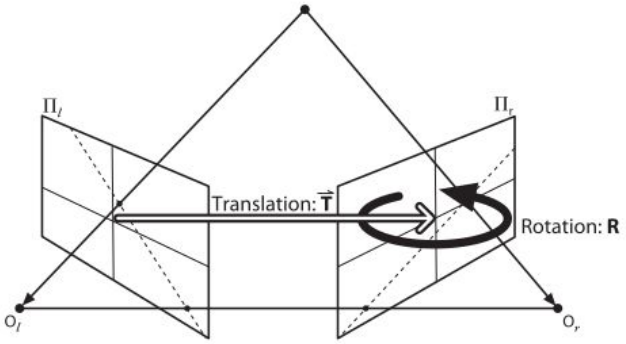

Figure 1: The essential geometry of stereo imaging (Bradski & Kaehler, 2008).

One common way to find these parameters is using Zhang's planar pattern technique. This method only requires the camera to observe a planar pattern shown at a few (at least two) different orientations (Zhang, 2000). This technique is implemented in the open source computer vision library OpenCV. There are different types of planar patterns, a common one is to use a checkerboard as the Figure 2 shows.

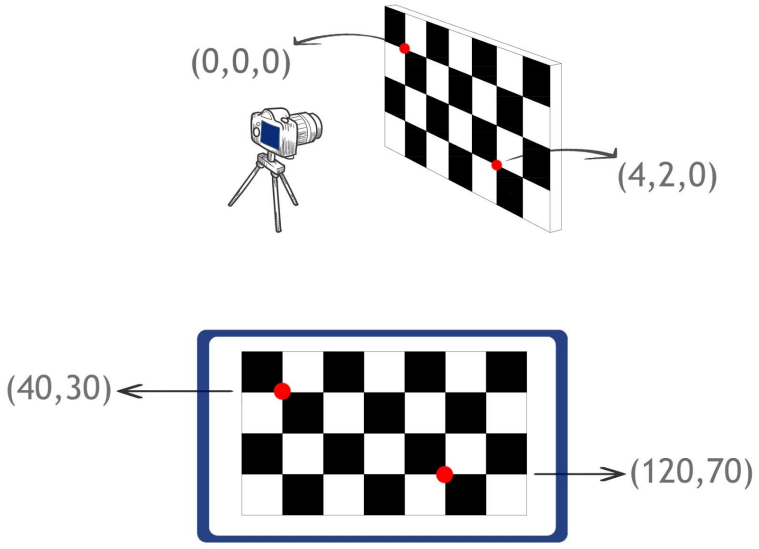

Figure 2: Relation between the world coordinate system and the camera coordinate system.

Geometric camera calibration and stereo camera calibration have been studied extensively. Several books and papers have been written about these topics since the 1990s. A new problem arises when attempting to calibrate the cameras underwater, since most studies have been made considering air as the medium between the scene and the camera. Water adds more distortion to the image, making the problem more complex.

The objective of this internship was to improve the calibration parameters results, for the specific application of seafloor mapping. For that, the approach to address this problem was to analyze the contribution of different types of images to the calibration. These types are categorized using the parameters of distance and rotation from the checkerboard to the camera.

MATERIALS AND METHODS

MATERIALS

Checkerboard images taken at test tanks from the years 2014, 2018, and 2019 are the primary data. Also, more data are available from a test tank calibration conducted in July 2020, targeting different views of the checkerboard images and a uniform distribution of them across the camera's field of view.

CALIBRATION

The calibration of the cameras was developed using python and OpenCV. The analysis has 5 steps: (i) image points recognition from the checkerboard; (ii) image clusterization into the different types of images; (iii) image selection for each cluster; (iv) calibration of the different combinations; and (v) testing and selection of the best calibration. These are discussed in more detail below.

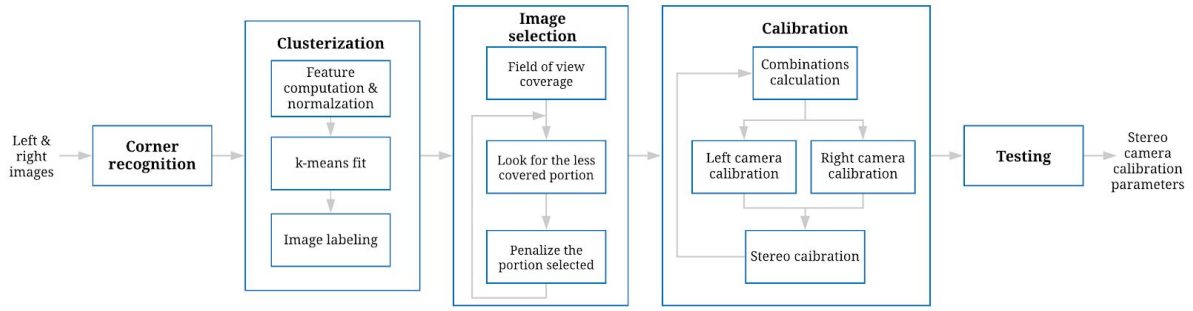

Figure 3: Chart of the calibration process proposed.

RESULTS AND DISCUSSION

IMAGE POINTS RECOGNITION FROM THE CHECKERBOARD

The first step to calibrate the cameras is to acquire the checkerboard images for each camera simultaneously. In this case, the images processed were taken from previous years, so this step was omitted until the end of the internship.

The geometric calibration of each camera relates the three-dimensional scene coordinate system to the camera coordinates, and for that it needs correspondence information from both coordinate systems as input. That is why the checkerboard is so useful, each corner is assigned a world coordinate starting at <0,0,0> and as the sizes of the squares inside of it are known, the next corner in the x-axis is <1,0,0> times the size of the squares. This process is applied to each corner maintaining the z-axis as zero, so that the world coordinate origin is at one of the corners of the checkerboard, and the pattern lies on the xy-plane.

The coordinates of each corner in the image plane are found recognizing the checkerboard in the image. This is accomplished through an OpenCV function, called

findChessboardCorners, that searches for a specific checkerboard size. Once found, the world and camera point coordinates of the checkerboard are stored.

IMAGE K-MEANS CLUSTERIZATION

As it was stated, the objective of the internship was to improve the calibration parameters results. The approach taken was to analyse the images and their contribution to the calibration of the stereo rig. To do the analysis, the images were split into groups, according to the position of the camera as seen by the checkerboard. These groups or clusters were computed using a k-means algorithm. The features used were: (i) The area in pixels covered by the checkerboard; (ii) the height and (iii) pitch rotation from the checkerboard to the camera; and in the cases where available (iv) the yaw rotation from the checkerboard to the camera. Each column of the features matrix were normalized, so that each column has a mean of zero and a standard deviation of one.

With the data available, each of clusters computed were tagged as: (i) near and no rotation; (ii) medium distance and no rotation; (iii) far and no rotation; and (iv) pitch rotation. Depending on the year the images were acquired, the group (v) yaw rotation was incorporated.

IMAGE SELECTION

From each cluster it is necessary to take a selection of the images. The calibration process can be computationally heavy if all images are being processed, requiring extended time without guarantee of a good result.

A simple heuristic was used: select the *n* images that the sum of the total area covered by the checkerboards covers a maximized field of view. The process begins by creating a matrix of the covered area by the checkerboards. In equation 1 this process can be seen, the initial covered area COV_0 is equal to the sum of the matrices M_i that contains ones where the checkerboard is, and zero elsewhere.

$$COV_0 = \sum_{i=0}^{Totalimg} M_i \qquad (1)$$

The algorithm then selects the image where its checkerboard is covering the less covered area. The selection is done computing the mean of the non zero values of a mask of COV_k and each M_i matrix. The lowest mean value is selected. Then, the algorithm runs again but penalizes the area of the selected checkerboard as seen in equation 2, and removes the index of the selected value from the search list. A visualization of the algorithm can be seen in Figure 4.

$$COV_k = COV_{k-1} + M_{selected} \cdot Totalimg \qquad (2)$$

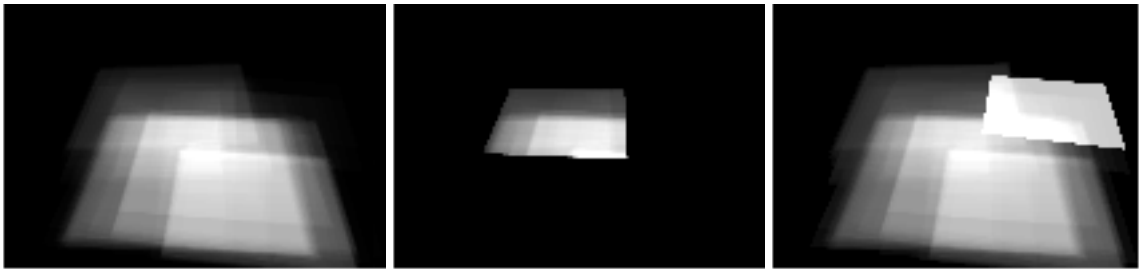

Figure 4: From left to right: (i) initial covered area; (ii) masking of the covered area by one of the checkerboards; and (iii) new covered area being penalized by the selected checkerboard.

CALIBRATION

With the goal aiming to analyse the contribution of each group to the calibration, all the combinations of the groups were computed. For each combination a fixed number of images were selected with the algorithm previously explained, maintaining the same amount of images per cluster. Table 1 explains this procedure for a two clusters analysis.

Combination	Images used from group 1	Images used from group 2
Group 1	36	-
Group 2	-	36
Group 1 & 2	18	18

Table 1: Amount of images used per combination, for a two cluster analysis, selecting the 36 images that covered more field of view.

After selection of the images each camera is calibrated using the function *calibrateCamera* from OpenCV, using the rational model for the distortion. The result are the intrinsics, extrinsics and distortion parameters of each camera. These results are given as a start point to the function *StereoCalibrate*, also from OpenCV.

TESTING AND SELECTION OF THE BEST CALIBRATION

To test each calibration two measurements were taken: (i) the reprojection error; and (ii) the row alignment error. The first computes the difference in magnitude between the reprojected three-dimensional coordinate of the corners of the checkerboard into the image, and the actual coordinate of the corners in the image. Figure 5 illustrates the error in magnitude of the reprojection of one corner into the image plane. The error is calculated as stated in equation 3, where IC_i and RC_i represent the *image corner* and *reprojected corner* of the *i* corner of the checkerboard.

$$R_{error} = \frac{1}{Totalcorners} \sum_{i=0}^{Totalcorners} \|IC_i - RC_i\| \qquad (3)$$

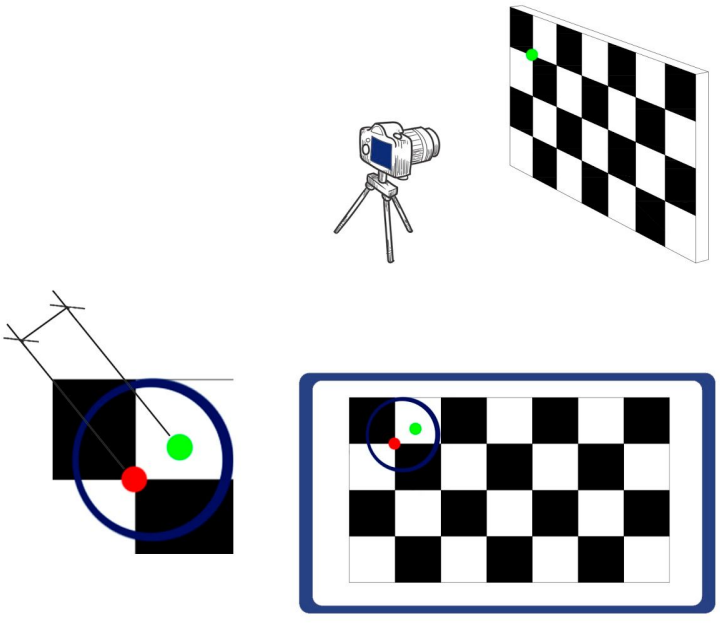

Figure 5: Reprojection error illustration of one corner.

The row alignment error is computed after a process called rectification. The idea of this process is to force the epipolar lines to be horizontal, and by doing that both image planes are forced to be row aligned and lie in the same plane. If the stereo calibration parameters were estimated correctly each corresponding point in the left image should match the same *y-coordinate* in the right camera. Using the checkerboard images, this error can be computed comparing the coordinates of each corresponding corner. Equation 4 shows how this error is computed, where LIC_i and RIC_i represents the *left image corner y-coordinate* and the *right image corner y coordinate* of the *i*-corner of the checkerboard.

$$RAE = \frac{1}{Totalcorners} \sum_{i=0}^{Totalcorners} | IC_i - RC_i | \qquad (4)$$

Each checkerboard from the selected images of each group combination was tested by these two measures, storing its mean and standard deviation. While using the reprojection error, a visualization of this error was developed. The direction of the error could be computed, from the image corner to the reprojected one, as Figure 6 shows.

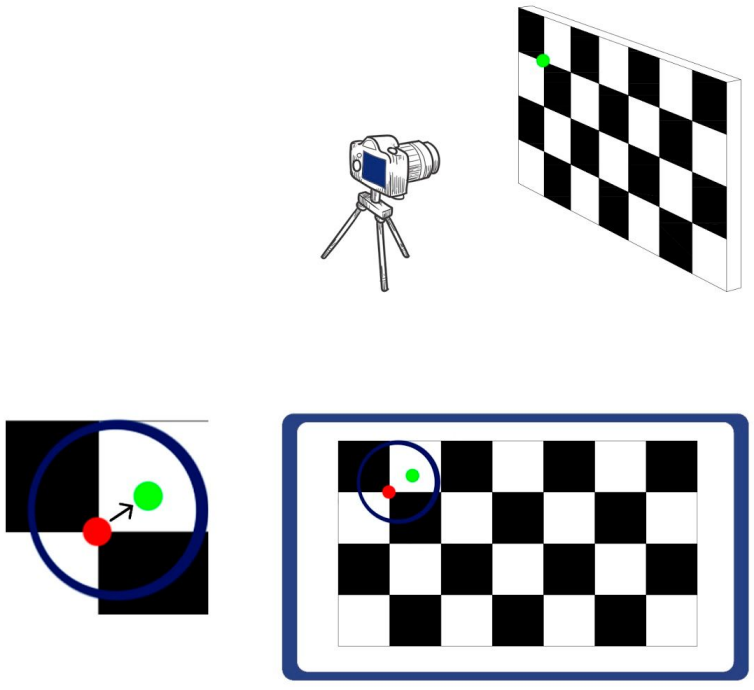

Figure 6: Reprojection error direction of one corner.

This discovery led to the creation of a new representation, called a *reprojection arrow error map*. This map gathers the direction and magnitude of the reprojection error from different views of the checkerboard, filling the field of view of each camera. In Figure 7 the

arrow error maps of the left and right camera after calibration can be seen. The tiny colored dots seen in each image are actually arrows, their magnitude is represented by the length of the arrow and its color.

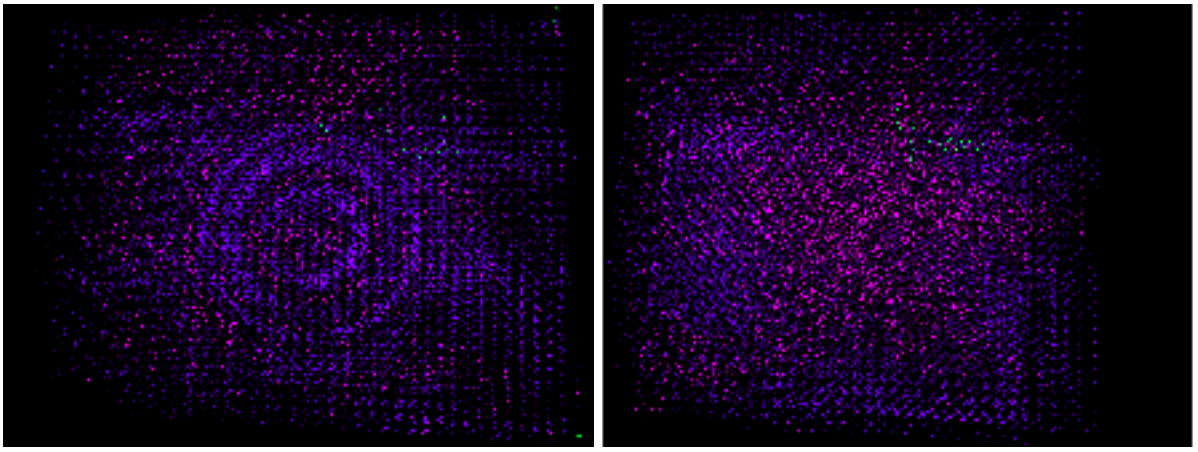

Figure 7: From left to right: (i) reprojection arrow error map of the left camera; and (ii) reprojection arrow error map of the right camera.

In the left camera's reprojection arrow error map, harmonic radial distortion can be seen in the center. In the right camera's reprojection arrow error map this radial distortion is not distinguishable. Both cameras are theoretically equal, so the result should be the same for each. In practice it is impossible to build two identical things, but the difference is still very interesting. The explanation to this phenomena is not clear yet, as not enough time was invested, and there is more than one possible explanation.

RESULTS

Each group combination was tested with the reprojection error and row alignment error measures. The mean and standard deviation was computed for the total checkerboard used. In Figure 8 and Figure 9 the results are shown for the 2018 data.

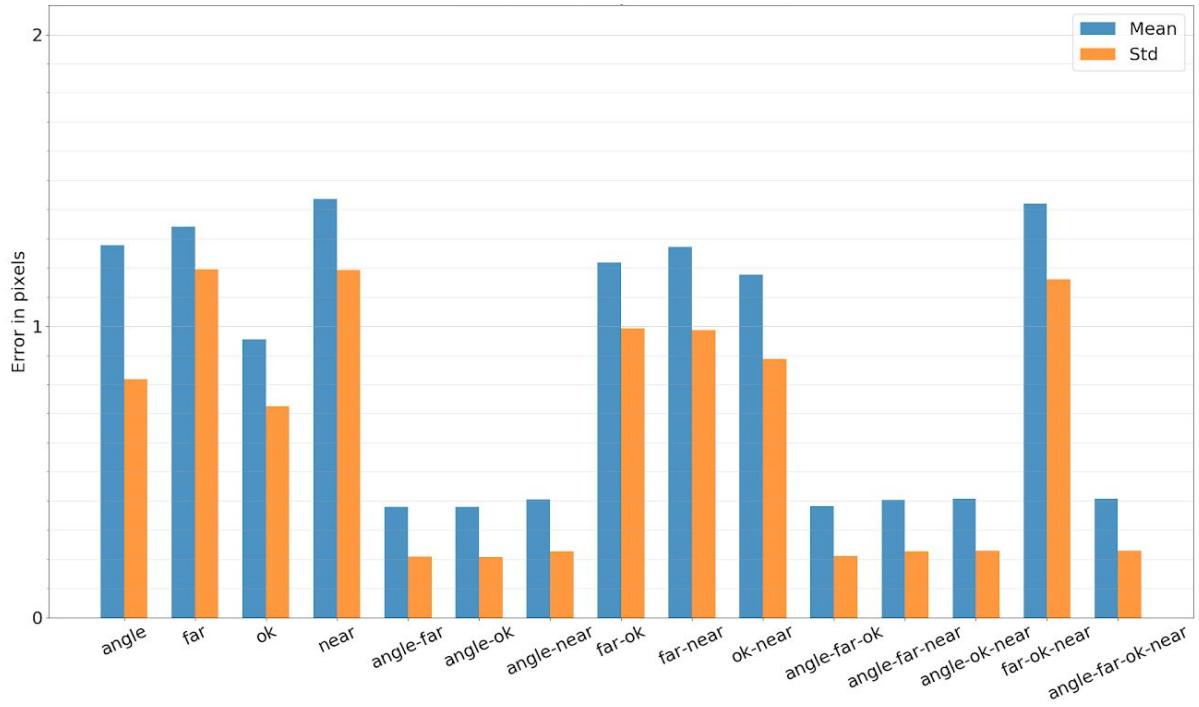

Figure 8: Reprojection error mean and standard deviation versus group combination.

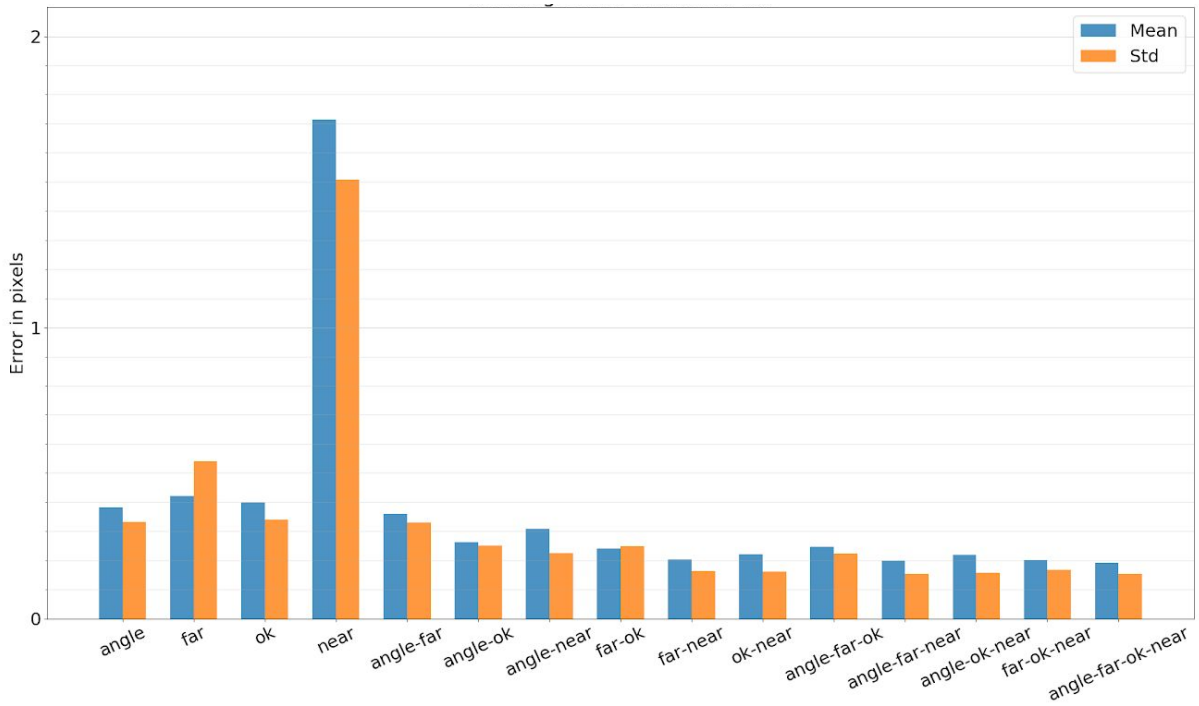

Figure 9: Row alignment error mean and standard deviation versus group combination.

Each group was labeled with a name, in order to observe its contribution. As seen in both bar plots, each group by itself does not give a good result. By using different combinations, a better result can be achieved. Mixing the information of both measurements one can select the best combination and by doing so the best stereo calibration parameters.

Appreciably, there is no clear correlation between the graphs shown. When there is no clear group combination winner, the selected one was chosen by the minor row alignment error.

CONCLUSIONS/RECOMMENDATIONS

From the procedure taken to analyze the image's contribution to the calibration, one could possibly expect that the best combinations should be the same across calibrations from different years. However, this is not happening, and we cannot conclude this as the checkerboard images from different years vary significantly in distance, rotation and field-of-view coverage. For this year's calibration, a more complete acquisition was made, resulting in a better and more diverse image type contribution..

From previous years results, this procedure allowed the Low Altitude Survey System to generate better three-dimensional reconstructions of different objects and two-dimensional mosaics as the ones shown in Figure 10 and Figure 11.

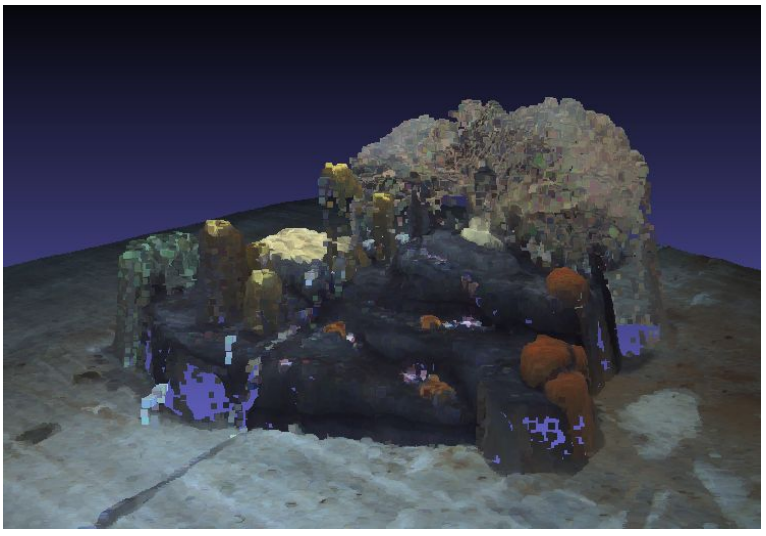

Figure 10: Three dimensional reconstruction of an artificial reef.

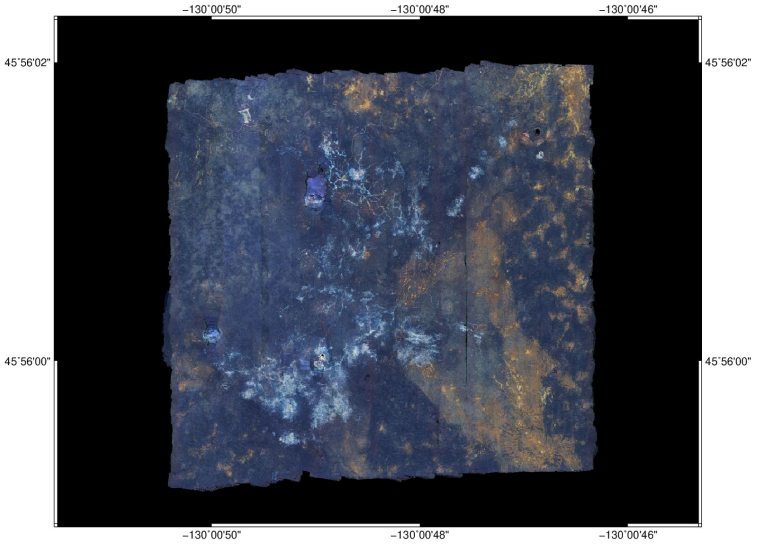

Figure 11: Two dimensional mosaic of the ASHES Hydrothermal Vent Field.

The next steps to continue with this analysis are to process the new data acquired for the 2020 calibration, and select the best stereo parameters. After that, the parameters can be applied to future and past surveys, giving more insights of the unknown regions of the deep ocean.

ACKNOWLEDGEMENTS

The MBARI Summer Internship Program is generously supported through a gift from the Dean and Helen Witter Family Fund and the Rentschler Family Fund in memory of former MBARI board member Frank Roberts (1920-2019) and by the David and Lucile Packard Foundation.

References:

Bradski, G., Kaehler, A. (2008). Learning OpenCV. (1st ed.) United States of America; O'Reilly Media, Inc.

Forsyth, D., Ponce, J. (2002). *Computer Vision a modern approach*. (2nd ed.) United States of America; Pearson Education, Inc.

OpenCV (version 4.0.0) [Software]. Retrieved from https://opencv.org/

Zhang, Z. (2000). A Flexible New Technique for Camera Calibration. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 22* (11), 1330-1334.