Abstract—This paper presents an aligned epipolar line for stereo images with analysis of multiple size region of interest in selected area or segment of depth maps in application of computer vision. Epipolar line works as a reference line for two same views of images during the process of pixel matching. Then the depth maps (disparity mapping) could be produced with almost similar position to each other. The mapping is about to deal with the intensities of pixels which allocate the depths of an image and can be extracted to get the distance values. This region is a reference view of the stereo camera and stereo vision baseline is based on horizontal configuration. The block matching technique is briefly described with the performance of its output. The disparity mapping is generated by the algorithm with the reference to the left image coordinate. The algorithm uses Sum of Absolute Differences (SAD) which is developed using Matlab software. The rectification and block matching processes are also briefly described in this paper.

Index Terms—Stereo camera; disparity mapping; region of interest; rectification; block matching algorithm

I. INTRODUCTION

The origin of the word “stereo” is the Greek word “stereos” which means firm or solid, with stereo vision, the objects are seen solid in three dimensions with range [1]. In stereo vision, the same seen is captured using two sensors from two different angles. The captured two images have a lot of similarities and smaller number of differences. In human sensitivity, the brain combines the captured to images together by matching the similarities and integrating the differences to get a three dimension model for the seen objects. In machine vision, the three dimension model for the captured objects is obtained finding the similarities between the stereo images and using projective geometry to process these matches. The difficulties of reconstruction using stereo is finding matching correspondences between the stereo pair.

Stereo Vision Application

A stereo vision application such as a stereo vision mobile robot navigation requires a number of various capabilities, including the ability to execute uncomplicated goal-achieving actions, like reaching a given location; to react in real time to unexpected events, like the sudden appearance of an obstacle; to build, use and maintain a map of the environment; to determine the robot's position with respect to this map; to form plans that pursue specific goals or avoid undesired situations; and to adapt to changes in the environment [2]. In a navigation of a stereo vision mobile robot research, the goal is to build up a vehicle that can navigate at a certain speeds using stereo camera whether in outdoor or indoor environments such as fields or building. These vehicles require massive computational power and suitable algorithm in order to adapt their sensitivity and control capabilities to the certain speed of motion and in avoidance of obstacles.

Basically, the stereo camera in Figure 1 takes pictures of a same object but with different view. The two dimension images on the plane of projection represent the object from camera view. These two images contain some encrypted information, the image-depth of each other. This information is the third dimension of two dimension images. Therefore, the object distance and its depth can be determined by using the stereo cameras [3]. With referring to Figure 2 the distance between two points of view is called the baseline [4]. The baseline’s distance end to end affects on the range resolution, which establish the range of depth that can be calculated. A difference between scenes of the same object on the two images is called disparity [5]. The stereo system uses horizontal baseline. The cameras are placed at the same elevation. The process of stereo vision is then typically defined as finding a match between features in left and right images [6]. The stereo baseline b is the distance between the centres of the projection Ol and Or. The xl and xr are the coordinates of pl and pr with respect to the principal points cl and cr. Where f is a focal length of the camera. The depth of P is uniquely determined by the following equation [4]:

\[ \frac{b+x_r-x_l}{z-f} = \frac{b}{Z} \]  

And let the disparity \( d = x_r - x_l \), then the depth of point P is:

\[ Z = f \frac{b}{d} \]
The corresponding two images representing the same point of the scene is called disparity matching. The set of displacements between matched pixels is usually indicated as disparity map [6].

![Figure 2. Basic principle of stereo vision](image1)

**A. Camera Calibration**

The calibration process is using Tsai’s method. This method produces the values of extrinsic parameters such as the rotation and translation of images. Another result from Tsai’s method is intrinsic parameters that generate focal length, principal point, skew, distortion coefficient and pixel size [7][8].

**II. SOFTWARE ARCHITECTURE**

The Matlab software is used to implement the software part. From Figure 3 images are captured by using the calibrated stereo camera. Then it will be changed to grey level scale to enhance the images using histogram equalization method [9]. The equation is:

\[ q = T(w) = \max \sum_{i} P(w) \]  

(3)

The data from camera calibration using Tsai’s method will be used to rectify the stereo images to obtain corresponding epipolar lines that parallel to the horizontal scan-lines for both images. The rectification process using equation (4) is to align the epipolar line for each image (left and right). With Iold and Inew are the original images and rectified image respectively with the combination of coefficient ai for each camera.

\[ I_{new}(x_0, y_0) = a_1 I_{old}(x_1, y_1) + a_2 I_{old}(x_2, y_2) + a_3 I_{old}(x_3, y_3) + a_4 I_{old}(x_4, y_4) \]  

(4)

For example with referring to the dotted line left hand upper thumb of the model in Figure 4(a) shows stereo image before rectification process and Figure 4(b) is the image with rectification process. The output rectification process is look like a distortion image. But in fact that the image has been improved to be aligned each other with reference to epipolar line. The Sum of Absolute Differences (SAD) is applied to solve the stereo correspondence problem between two images. The SAD algorithm from the equation (5) shows an area-based correspondence algorithm [10][11][12]. It computes the intensity differences for each center pixel of (i, j) as follows:

\[ \text{SAD}(x, y) = \sum_{i,j} |\text{gt}(x+i, y+j) - \text{gt_{1}}(x+i, y+j)| \]  

(5)

The reference point of the left image gt is minus with the right image gt1 at the same epipolar plane. It sums up the intensities of all surrounding pixels in the neighbourhoods for each pixel in the left image. To calculate stereo correspondence of stereo images, this paper is using block matching technique. Each block from the left image is matched into a block in the right image by shifting the left block over the searching area of pixels in right image as shown in Figure 5. In order to find corresponding pairs of stereo points, they first have to be compared for different disparities, after which the best matching pairs can be determined. The maximum range at which the stereo vision can be used for detecting obstacles depends on the image and depth resolution. Absolute differences of pixel intensities are used in the algorithm to compute stereo similarities between points. By computing the sum of the absolute differences SAD for pixels in a window surrounding the points, the difference between similarity values for stereo points can be calculated. The disparity associated with the smallest SAD value is selected as best match [13].

![Original Left Image](image2) ![Original Right Image](image3)

![Rectified Left Image](image4) ![Rectified Right Image](image5)

**Figure 3. Software implementation for continuous process**

**Figure 4. Rectification process of stereo image**

**Figure 5. SAD block matching algorithm**

**III. RANGE DETECTION AND FILTERING OBSTACLES**

Together with the stereo camera parameters from software calibration and the disparity between corresponding stereo points, real-world distances can be retrieved. In order to find
corresponding pairs of stereo points, they first have to be compared for different disparities, after which the best matching pairs can be determined. The maximum range at which the stereo vision can be used for detecting obstacles depends on the image and depth resolution [13]. The equation below shows the distance or depth of obstacles that has been used in this paper.

\[ \text{Distance} = ae^{bx} + ce^{bx} \]  

The value of a, b, c and d is a constant value from curve fitting tool in Matlab. The value of x represents a pixel value in disparity mapping. The calculated distance has an effective range of detection for the stereo vision application. In the obstacle detection step, the points belonging to obstacles must be found. For data combination in image processing software, it is important to cluster or group the obstacle points in such a way that individual obstacles can be identified. If there are obstacles at closest range, the stereo program will filter using Gaussian lowpass filter [9] (7) to determine the object and cluster the pixels location.

\[ H(u, v) = e^{-B^2(u,v)/2\sigma^2} \]  

IV. ROI AND PIXEL INTENSITIES

The disparity values from the SAD algorithm are mapped to disparity mapping with the reference to the left image coordinate. The region of interest is the area within the green line box that search for objects or obstacles Figure 7. This area is a reference box for stereo camera to navigate if the application is applied on the stereo vision application such as a mobile robot. So, this paper is presenting the pixel intensities for every selected region of interest in graph. The idea is the pixels can be manipulated to certain applications such as the distance of the object detected and the shape of the object. The size of the region of interest is 320X20. From Figure 8 shows the pixels spread in the region of interest in pixel value with the horizontal coordinate.

A. Multiple ROI

With some illustrations of basic stereo theory shown by Figure 8, the obstacles detection for close object and far object can be concluded by Figure 9. Which is the baseline range are permanently setup and the left plane works as a reference plane. The result is the disparity values at a close object bigger than the disparity of a distant object. In disparity mapping the brighter pixels in the disparity map, the more close the corresponding pixel in the reference image. As a result Figure 10 (a), an obstacle could be identified as a group of pixels, which are generally brighter than their neighbourhood. The original images are gray level scale but not aligned each other (left and right). The rectification process produces aligned images with same image size about
At the border of both images the rectification process unable to allocate the pixel values because of the original images has been reconstructed. So, the SAD algorithm cannot identify the disparity values at the border of rectified images. Figure 10 (b) is a collection of multiple selected region. The range of disparity values about 1 until 40 and the size of selected region of interest from 320x20 until 320x50. The graphs show the total point of pixels with the coordinate systems detected each of selected green boxes on grey level scale image. The maximum or minimum value of pixel intensities can be manipulated according to the stereo vision application.

![Disparity map]

**Figure 10. Multiple selected region of interest for a single image with the maximum disparity value 40 and minimum value 1**

**V. CONCLUSION**

An epipolar line works as a reference to compare stereo images. The process of comparing describes as a rectification process which this process would arrange the new coordinates of pixels to a standard coordinates for two images. Then the pre-matching process takes place. This process produces disparity mapping or depth maps. By changing the size region of interest on disparity mapping, the pixel intensities could be manipulated depends on the application. A disparity map image is an efficient method for storing the depth or distance of each pixel in an image. Each pixel in the map corresponds to the same pixel in an image, but the grey
level corresponds to the depth at that point rather than the gray-shade or color. From the colour mapping itself, the lighter are closer and darker are farther away. Viewing a gray level histogram equalized disparity map image, the objects also lighter are closer, and darker object are farther away. Stereo vision is also a type of passive sensor, meaning that it uses the radiation available from its environment. It is non-intrusive as it does not need to transmit anything for its readings.

ACKNOWLEDGEMENT

The authors would like to acknowledge the funding from Short Term Grant of Universiti Teknikal Malaysia Melaka.

REFERENCES