

Sum of Absolute Differences Algorithm in Stereo Correspondence Problem for Stereo Matching in Computer Vision Application

Rostam Affendi Hamzah

FKEKK, UTeM, Melaka, Malaysia.
rostamaffendi@utem.edu.my

Rosman Abd Rahim

FKEKK, UTeM, Melaka, Malaysia.
rosmanrahim@utem.edu.my

Zarina Mohd Noh

FKEKK, UTeM, Melaka, Malaysia.
zarinamohdnoh@utem.edu.my

Abstract—This paper presents a method to solve the correspondence problem in matching the stereo image using Sum of Absolute Differences (SAD) algorithm. The computer vision application in this paper is using an autonomous vehicle which has a stereo pair on top of it. The estimation of range is using curve fitting tool (cftool) for each detected object or obstacles. This tool is provided by Matlab software. The disparity mapping is produced by block matching algorithm Sum of Absolute Differences (SAD). The range determined by cftool will be used as a reference for navigation of autonomous vehicle. The process of camera calibration, image rectification and obstacles detection is also discussed in this paper.

Keywords- *curve fitting tool; rectification; stereo correspondence; stereo vision, sum of absolute differences.*

I. INTRODUCTION

The cftool is a graphical user interface (GUI) that allows to visually exploring data and fits as scatter plots. Cftool accesses GUIs for importing, preprocessing, and fitting data, and for plotting and analyzing fits to the data. In this paper, cftool is a tool to estimate the distance of each obstacle detected by stereo camera. In mapping the disparity values, cftool produces a graph which consist of range versus disparity values each pixel that match up in stereo image (left and right image).

II. HARDWARE IMPLEMENTATION

Vision-based sensing methods usually employ two cameras using stereo vision technique to acquire images surrounding the robot. In the case of an AGV navigation in this project, the stereo cameras used to acquire the image sequence is in motionlessness. Each frame that is captured by the cameras represents an image of the scene at a particular instant in time. As the robot navigates by avoiding the obstacles in its path, tremendous numbers of frames are acquired, and they should be processed quickly in order to achieve real-time motion. Figure 1 shows the basic system components required for this project. An AGV navigation system consists of;

- static stereo cameras to acquire images of the scene,
- a laptop with core 2 duo processor, 3 gigabyte RAM, 3 USB ports and image processing software to digitize the images, and
- a motor controller from National Instrument (Data Acquisition) to process the relevant input signal from notebook to generate the signals for motors.



Figure 1. Autonomous vehicle

A machine vision system is very similar to human vision system in a sense that the stereo cameras and the image processor act as the eyes, and the motor controller acts as the brain to control the AGV. As the AGV navigates, the onboard stereo cameras capture the images and send them to the image processor, which extracts the relevant visual information such as range to nearby objects. The connection between NI-DAQ and laptop is using USB port. It uses 5 volt to operate. The stereo cameras installation is not in the same serial port location. It means that each stereo camera USB port has to be in separate location or hub. This procedure is to avoid an error from Matlab. The software of Matlab is only capable to capture the image with one port at a time.

III. TSAI'S METHOD FOR STEREO CAMERA CALIBRATION

Construction of a full model from the stereo pair requires calibration of the camera system using software. From a stereo camera system (stereo rig) has accurately specified intrinsic and extrinsic parameters for both cameras. According to [1] the intrinsic camera parameters specify a pinhole camera model with radial distortion. The pinhole model is characterized by its focal length, image centre, pixel spacing in two dimensions and the radial distortion is characterized by a single parameter. The extrinsic parameters describe the relative position and orientation of the two cameras. Intrinsic parameters for a given camera are constant, assuming the physical parameters of the optics do not change over time, and thus may be pre-calculated. Extrinsic parameters depend on the relative camera poses and will be constant if the cameras are fixed relative to one another [1]. Both intrinsic and extrinsic calibration parameters are calculated using the Tsai's method from a tool by Matlab provided by [2]. There are images of a calibration target, see Figure 5(a), consisting images of left and right scene. It is totally uncalibrated images due to unaligned each comparing image left to right.

The flowchart Figure 2 shows the steps of calibration using the Tsai's method in Matlab toolbox. The first step is to get a set of images in digital form and start to evaluate the

error between the images of left and right. If the images are not converge each other then the system will adjust the value

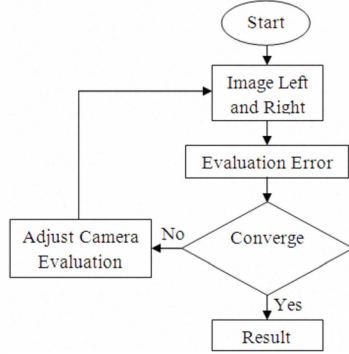


Figure 2. Flowchart of Tsai's method

of the camera evaluation until they converge. The adjusted value or parameters will be used as a result for calibration process to be used in rectifying process [3]. The result contains of intrinsic parameters TABLE I and TABLE II and extrinsic parameters in TABLE III. These values are represented in pixel intensity values.

TABLE I. INTRINSIC PARAMETERS OF LEFT CAMERA

Focal length	[372.90575 373.58482] ± [5.14011 5.07494]
Principal point	[159.76547 91.23921] ± [5.46331 5.76730]
Skew	[0.00000] ± [0.00000]
Distortion	[0.19492 -1.07074 -0.01118 -0.00091 0.00000]

TABLE II. INTRINSIC PARAMETERS OF RIGHT CAMERA

Focal length	[373.83866 373.83178] ± [4.82569 4.85865]
Principal point	[167.07304 139.96128] ± [5.64311 5.82166]
Skew	[0.00000] ± [0.00000]
Distortion	[0.19645 -0.84041 -0.01428 0.00668 0.00000]

TABLE III. EXTRINSIC PARAMETERS (LEFT AND RIGHT CAMERA)

Rotation Vector	[-0.00757 0.16506 -0.00930] ± [0.01161 0.01458 0.00162]
Translation Vector	[-47.82594 1.48281 3.10607] ± [0.77822 0.53278 4.14571]

IV. DEPTH ESTIMATION IN DISPARITY MAPPING

To get the depth value, the rectified images have to run through several processes starting from stereo correspondence until the disparity mapping. According to Figure 3, the stereo correspondence or stereo matching have using SAD algorithm as described earlier in abstract. Range estimation is equivalent to depth estimation in mapping the disparity values using intensity of pixel value for each matching point.

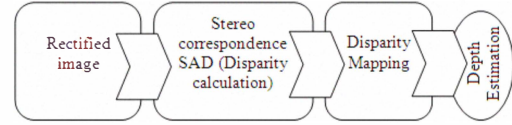


Figure 3. Range estimation process

A. Image Rectification

The rectification of stereo image pairs can be carry out under the condition of calibrated camera. To quickly and accurately search the corresponding points along the scan-lines, rectification of stereo pairs are performed so that corresponding epipolar lines are parallel to the horizontal scan-lines and the difference in vertical direction is zero. Image rectification is the undistortion according to the calibration parameters calculated in the camera calibration. After all intrinsic and extrinsic camera parameters are calculated they can be used to rectify images according to the epipolar constraint [4]. The rectification process is shown by Figure 4. The process above starting with acquire stereo images after that the image programming software Matlab will enhance the images using histogram equalization method. The next step is finding the matching point to be rectified. This problem faces a correspondence problem. Then the matched point and camera calibration information are applied to reconstruct the stereo images to form a rectified images.

The equation below is used to rectify the images in Matlab.

$$I_{new}(x_0, y_0) = a_1 I_{old}(x_1, y_1) + a_2 I_{old}(x_2, y_2) + a_3 I_{old}(x_3, y_3) + a_4 I_{old}(x_4, y_4) \quad (4)$$

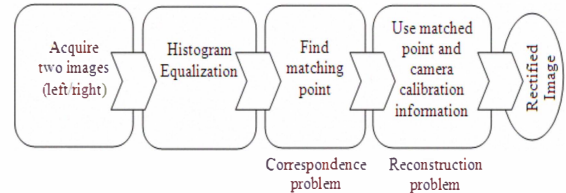
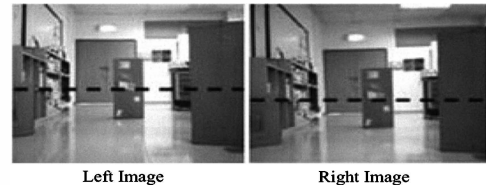
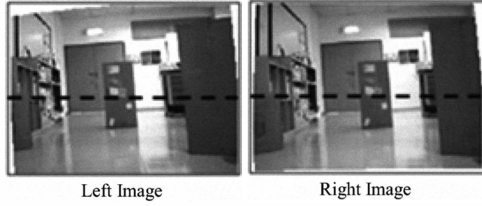


Figure 4. Rectification process



(a)



(b)

Figure 5. Original image (a) and after rectification process (b)

With I_{new} and I_{old} as the original and the rectified image and the blending coefficients a_i separate for each camera. Figure 5(a)(b) are the original image before rectification and after rectification. The output size of rectified stereo image is 320x240. The horizontal line for both images indicates the left and right image is horizontally aligned compared to image Figure 5(a).

B. Stereo Correspondence

With assume from now on that the images are rectified. That is, the epipolar lines are parallel to the rows of the images. By plotting the rectified images indicate the horizontal coordinate by x and the vertical coordinate by y . With that geometry, given a pixel at coordinate x_1 , the problem of stereo matching is to find the coordinate x_r of the corresponding pixel in the same row in the right image. The difference $d = x_r - x_1$ is called the disparity at that pixel. The basic matching approach is to take a window W centered at the left pixel, translate that window by d and compare the intensity values in W in the left image and W translated in the right image [4][5]. The comparison metric typically has the form:

$$SAD : \sum (I_l(x, y), I_r(x+d, y)) = \sum |I_l(x, y) - I_r(x+d, y)|$$

The function of SAD measures the difference between the pixel values. The disparity is computed at every pixel in the image and for every possible disparity. It sums up the intensities of all surrounding pixels in the neighborhood for each pixel in the left image. The absolute difference between this sum and the sum of the pixel, and its surrounding, in the right image is calculated [6]. The minimum over the row in the right image is chosen to be the best matching pixel. The disparity then is calculated as the actual horizontal pixel difference. The output is a disparity image. Those images can be interpreted as disparity being the inverse of the depth (larger disparity for points closer to the cameras) [4].

To calculate stereo correspondence of stereo images, there are some simple standard algorithms by using block matching and matching criteria. The blocks are usually

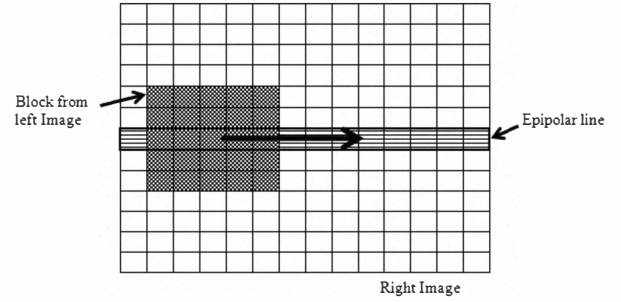


Figure 6. SAD block matching process

defined on epipolar line for matching ease. 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 6. At each shift, the sum of comparing parameter such as the intensity or color of the two blocks is computed and saved. The sum parameter is called "match strength". The shift which gives a best result of the matching criteria is considered as the best match or correspondence [7].

According to [8] the SAD algorithm works on 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 6. Ideally, for every pixel mask within the original image there should be a single mask within a second image that is nearly identical to the original and thus the SAD for this comparison should be zero [9].

C. Disparity Mapping

Together with the stereo camera parameters from calibration and the disparity between corresponding stereo points, the stereo images 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 [10]. Absolute differences of pixel intensities are used in the algorithm to compute stereo similarities between points. By computing the sum of the absolute differences 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 [4] Figure 7 shows the disparity mapping using SAD block matching algorithm.

D. Range Estimation using Curve Fitting Tool

The estimation of the obstacle's range in this paper is using curve fitting tool in Matlab to determine the range according to the pixel values. Each pixel in the mapping of disparity will be calculated through the curve fitting tool and the coordinate of horizontal is referring to left image.

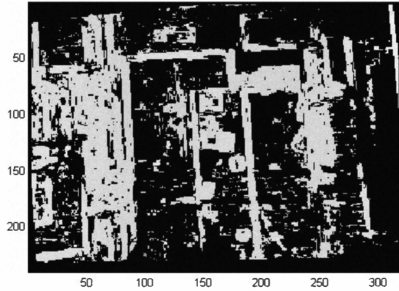


Figure 7. Disparity mapping of calibrated cameras using Tsai's method

The equation of the distance estimation is:

$$\text{Range} = a \cdot \exp(b \cdot x) + c \cdot \exp(d \cdot x)$$

- $a = 0.339$
- $b = -3.525$
- $c = 0.9817$
- $d = -0.4048$

Where the value of a , b , c and d is a constant value produced by curve fitting tool. The value of x represents the value of pixels in the disparity mapping. The curve can be explained as Figure 8. X axis represents disparity value in pixel density and y axis shows the distance or range in meter for every disparity values. TABLE IV shows the result of pixel values comparing with range in meter using cftool. The data can be analyze with some pixel values taken during the experiment of autonomous vehicle navigation.

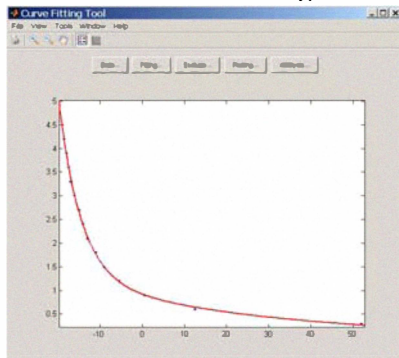


Figure 8. Curve fitting tool window

TABLE IV. DATA FROM CFTOOL

Pixel Value	52	12.5	0.5	-5.5	-9	-11	-13	-14	-15	-16	-17	-17.5	-18	-18.5	-19
Range (m)	0.3	0.6	0.9	1.2	1.5	1.8	2.1	2.4	2.7	3	3.3	3.6	3.9	4.2	4.5

V. NAVIGATION OF STEREO VISION AUTONOMOUS VEHICLE

To perform complete autonomous vehicle navigation in this paper using stereo vision, an experiment is prepared in Mechatronic Lab. The experiment's setup is shown by Figure 9. There are labels for the obstacles P1 until P6.

TABLE V. PIXEL AND RANGE DATA FROM THE SCENES V1-V4

Pixel vs Range	Obstacle P1		Obstacle P2		Obstacle P3		Obstacle P4		Obstacle P5		Obstacle P6	
	Pixel	Range	Pixel	Range	Pixel	Range	Pixel	Range	Pixel	Range	Pixel	Range
V1	-8	1.4	-11	1.8	-15.00	2.7	-17	3.3	-	-	-18	3.9
V2	-	-	-	-	15.00	0.7	-	-	-4.8	1	-14	2.4
V3	-	-	-	-	-	-	-	-	-	-	-8	1.4
V4	-	-	-	-	-	-	-	-	-	-	8	0.7

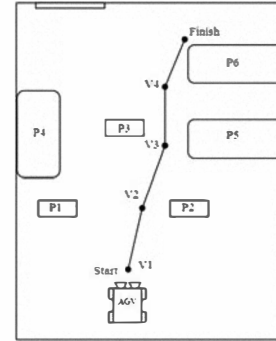


Figure 9. An experiment setup for AGV navigation in Mechatronic Lab

After the navigation, the results TABLE V of every autonomous processing scene V1 until V4 are plotted. So, it has moved successfully to avoid the obstacles until the finish point. Figure 10 shows the flowchart of controlling the autonomous surveillance vehicle. The value of D_s represents the main distance of red line in Figure 11. The θ_s represents the value of autonomous surveillance vehicle turning in degrees and the value of d_s is minimum detected range of obstacle. After the main program activated, the process makes a decision by comparing the range of obstacle (D) if the value below than 0.5 meter then the autonomous surveillance vehicle reversed and turn 90 degrees to the right. And start to run the program again. If the distance of the obstacle more than 0.5 meter, it turns to a certain degrees (θ) to the left or right depends on which area is the faraway object detected. After that the autonomous surveillance vehicle moves forward to d distance and start to run the program again.

The disparity mapping from Figure 7 can be plotted to the digital values. Therefore it can be used as a relationship with the distance estimation [9]. The result is a graph with the horizontal coordinate of left image Figure 12 (a)(b) as a reference. The images from Figure 9 of each scene V2-V4 are shown in Figure 13.

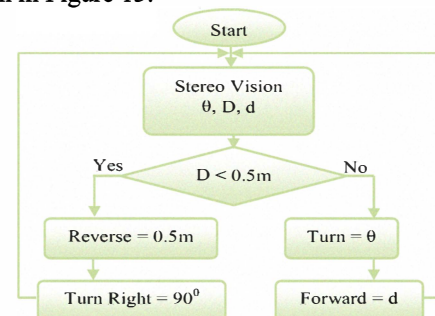


Figure 10. Flowchart of autonomous surveillance navigation

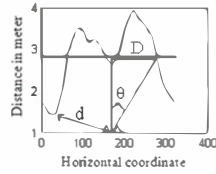


Figure 11. Stereo vision parameters

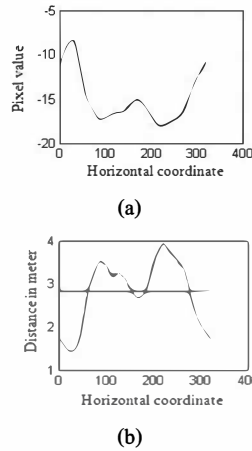


Figure 12. Distance estimation works opposite to the pixel value

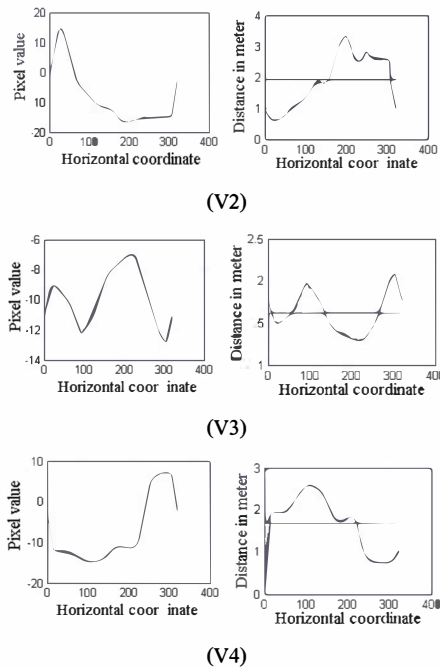


Figure 13. Results from V2-V4

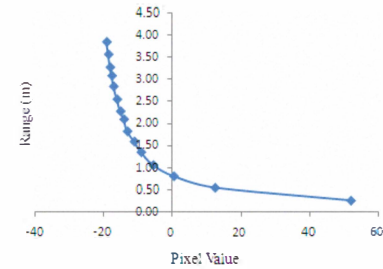


Figure 14. Relation of pixel value and range

The result could be explained when the pixel values are big, the range is close to the autonomous surveillance vehicle and if the pixel values are small, the range located faraway from autonomous surveillance vehicle. So the pixel works contrary to the range shown by Figure 14.

VI. CONCLUSION

Sum of Absolute Differences (SAD) algorithm is capable to solve the stereo correspondence images especially in computer vision application. Then the result from matching process is called disparity mapping or depth maps that enable other process using the same result for example distance estimation and AGV navigation. The curve fitting tool (cftool) is a tool that trustworthy in Matlab software. It gives a reference data to the autonomous vehicle to navigate and avoiding an obstacle. In grayscale color of disparity mapping, the darker color of an obstacle represents the object located far away from autonomous vehicle compared to a lighter color. And the value of pixel intensity for lighter color bigger than a dark color. It's proved by the output of cftool. The effective range of this paper presents a range about 0.3 meter until 4.5 meter.

REFERENCES

- [1] J. Steele, C.D., T. Vincent and M. Whitehorn, 2001. Developing stereovision and 3D modelling for LHD automation. 6th International Symposium on Mine Mechanization and Automation, South African Institute of Mining and Metallurgy.
- [2] Tsai, R.Y., An Efficient and Accurate Camera Calibration Technique for 3D Machine Vision. Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Miami Beach FL, 1986, pp. 364-374.
- [3] Zhang, Z., A flexible new technique for camera calibration. IEEE transactions on Pattern Analysis and Machine Intelligence, 2000.
- [4] Johan C. van den Heuvel, J.C.M.K., Obstacle Detection For People Movers Using Vision And Radar. TNO Physics and Electronics Laboratory Oude Waalsdorperweg 63, The Netherlands, 2003.
- [5] Manduchi, C.T.a.R., Stereo matching as a nearest neighbour problem. IEEE Trans. on PAMI, 20(3):333-340. 1998.
- [6] Teerapat Chinapirom, U.W., and Ulrich Rückert, Steroscopic Camera for Autonomous Mini-Robots Applied in KheperaSot League. System and Circuit Technology, Heinz Nixdorf Institute, University of Paderborn, 2001.
- [7] L. D. Stefano, M.M., and S. Mattoccia, 2004. A fast area-based stereo matching algorithm. Journal of Image and Vision Computing, Vol. 22, No. 12, pp.983-1005.
- [8] Kuhl, A., Comparison of Stereo Matching Algorithms for Mobile Robots. The University of Western Australia Faculty of Engineering, Computing and Mathematics, 2005.

- [9] Sharkasi, A.T., Stereo Vision Based Aerial Mapping Using GPS and Inertial Sensors. Blacksburg, Virginia, 2008.
- [10] Pan, A.K.a.J., Purdue Experiments in Model-based Vision for Hallway Navigation. Proceedings of workshop on Vision for Robots in IROS'95 conference, Pillsburgh, PA, 1995, pp. 87-96.