Improving disparity map of a specific object in a stereo image using camera calibration, Image rectification, and object segmentation

4 authors, including:

Hermawan Syahputra
State University of Medan
8 PUBLICATIONS   14 CITATIONS

SEE PROFILE

Agus Harjoko
Universitas Gadjah Mada
148 PUBLICATIONS   549 CITATIONS

SEE PROFILE

Reza Pulungan
Universitas Gadjah Mada
82 PUBLICATIONS   299 CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Identifikasi landmark sephalometri menggunakan algoritma projected principal edge distribution (PPED) View project

Image processing View project

Improving Disparity Map of A Specific Object in A Stereo Image Using Camera Calibration, Image Rectification, and Object Segmentation

Hermawan Syahputra[1*], AgusHarjoko[2], Retantyo Wardoyo[2], Reza Pulungan[2]

[1]*Department of Mathematics, Faculty of Mathematics and Natural Sciences,State University of Medan, Medan, Indonesia*
[2]*Department of Computer Science and Electronics, Faculty of Mathematics and Natural Sciences, Universitas GadjahMada, Yogyakarta, Indonesia*
*Tel. +6281375040550; e-mail: ma1_unimed@yahoo.co.id

Abstract

The objective of the study was to improve the disparity map of a certain object in a stereo image. The steps required to complete this are first, to perform camera calibration in order to extract camera parameters, and then to use these parameters to rectify the stereo image. Second, to perform segmentation to separate the object from the background of the image. Separation of the object from the background was performed to produce a good disparity for the object. This had not been done in previous studies. Using segmentation, bias and error of stereo correspondence had been reduced, since the correspondence was focus on the intensity of the left and right objects in the stereo image. Using rectified and segmented stereo images, better disparity maps had been obtained.

Keywords: Disparity map, image rectification, camera calibration, segmentation object, stereo image.

Introduction

Vision provides a very important meaning for human beings. Stereo vision is a reliable tool to find the depth of a scene. Stereo vision has been proved to provide an extraordinary result in estimating the disparity map or the depth[1].Stereo vision is the process of transforming the information from two areas of images into a 3D description of a scene and then finding depth information in terms of the proper distance. With the information depth, one can make a model of an area and other

natural environments; and this model can be used in a variety of applications, from virtual reality, flight and robot simulation, to feature extraction and navigation [2,3].

In addition, in stereo vision applications, especially in 3D reconstructions with disparity, both intrinsic and extrinsic parameters of the camera should be obtained in the beginning. The accuracy of the 3D reconstruction data depends on the accuracy of the disparity, the calibration of stereo system, image rectification, and the overall construction of the stereo system [3].

Several previous studies have developed the estimation of the disparity map, including the disparity estimation on a particular object [4]. The disparity map estimation of stereo images is also performed by using a hybrid method, *i.e.*, a combination of belief propagation and mean shift [5]. The disparity map estimation that uses the back propagation and pyramid segmentation, which divides images into several segments, has also been proposed [6].

In this research, a segmentation step is added to focus on a particular object in the image and to ignore the background. Therefore, the presence of the segmentation and the rectification should quicken the matching process and reduce the intensity bias between the object and its background. To get an accurate disparity map, rectification and good preprocessing are required. Besides, to reduce the focus on an object in the image, the preprocessing and image segmentation are required. Segmentation produces images displaying objects on white backgrounds. With the construction process of this stereo system, the resulting accuracy of the disparity map and the depth is expected to improve.

Materials and Method

The proposed model in this study is illustrated in Figure 1.
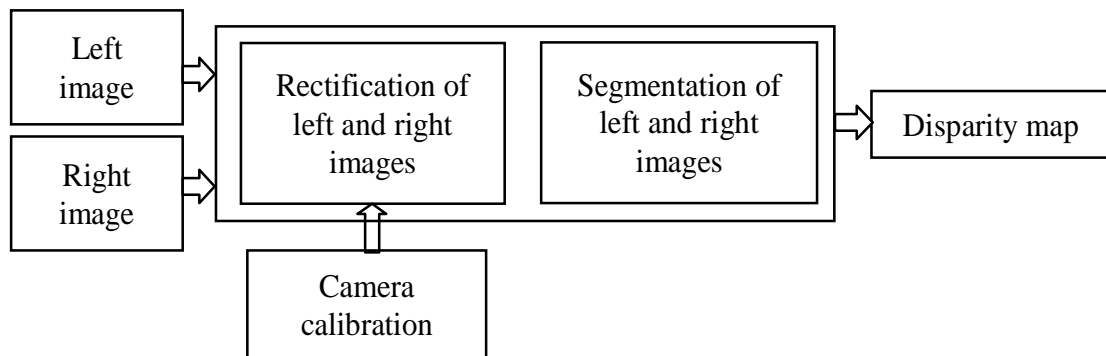
Figure 1. The proposed model of the disparity map computation

The proposed model consists in the following steps: first, the left and right images are taken using a stereo camera. Then, each image is corrected by using both the intrinsic and extrinsic parameters of the camera obtained from the calibration process. The rectified left and right images are then segmented to obtain the object of interest in the image. Finally, the rectified and already segmented left and right

images are used to determine the disparity map. For detail proposed model can be described as following step.

Input of Stereo Images

The input of proposed model were left and right images (Figure 1). The images was captured using FujiFinepix W3 3D stereo camera. The camera have a two lenses 10 MP, and resolution until 3584 x 2016pixel.

Camera Calibration

In this research, the first step was to calibrate the camera.Zang[7] explained that camera calibration is a process of finding the intrinsic and extrinsic parameters of a camera or a subset of these. A camera is usually described using the *pinhole model*.A camera model consists of a set of intrinsic parameters that determine how the camera forms images, and a set of extrinsic parameters that determine the position and direction of the camera in the world.The image of a 3D point $M = [X; Y; Z]^T$, denoted by $m = [u; v]^T$ is formed by an optical ray from M passing through the optical center *C* and intersecting the image plane. The three points M, m, and *C* are collinear. We use $\tilde{x}$ to denote the augmented vector by adding 1 as the last element: $\tilde{m} = [u; v; 1]^T$ and $\tilde{M} = [X; Y; Z; 1]^T$. The relationship between the 3D point M and its image projection m is given by

$$P\tilde{M} = A[R \mid t]\tilde{M} \tag{1}$$

with

$$A = \begin{bmatrix} \alpha & \gamma & u_0 \\ 0 & \beta & v_0 \\ 0 & 0 & 1 \end{bmatrix} \text{ and } P = A[R \mid t]$$

where (R; t), called the extrinsic parameters, is the rotation and translation which relates the world coordinate system to the camera coordinate system, and A is called the camera intrinsic matrix, with $(u_0; v_0)$ the coordinates of the principal point, α and β the scale factors in image u and v axes, and γ the parameter describing the skew of the two image axes. The 3 x 4 matrix P is called the camera projection matrix, which mixes both intrinsic and extrinsic parameters.

After the camera's perspective projection matrix P is obtained, the intrinsic and extrinsic parameters of the camera can be uniquely determined. First, by denoting the submatrix 3x3 of P by B and the last column of P by b, *i.e.*, P= [B | b]. Since P = A[R | t], we obtain:

B = AR and b = At (2)

Once the intrinsic parameters or the equivalent matrix A is known, the extrinsic parameters can be determined from Equation (2) as:

$$R = A^{-1}B \text{ and } t = A^{-1}b. \tag{3}$$

Images Rectification

The second step is to rectify the stereo image. Once the intrinsic and extrinsic parameters of the camera have been resulted using the camera calibration, then these parameters are used to rectify the image. Images rectification was performed by using the concept of epipolar geometry [8].

Stereo image rectification is a process of image transformations in such a way that the corresponding epipolar lines in all images become collinear with each other and with the image scanning lines. In rectified images all optical axes are parallel as well.

The other interesting feature inherent to the rectified stereo system is a shift of the epipoles to infinity. Thus, rectification of images can be thought of as a process of changing positions of epipoles to infinity. The rectification process is limited to the search for the transformation of the planes Π_{l0} and Π_{r0} to the planes Π_{l1} and Π_{r1}, respectively (Figure 2). The transformation sought can be described as a composition of the following transformations [8].

1. Rotation of the left and right camera planes in such a way that the epipoles go to infinity (and thus the epipolar lines become parallel). This rotation is described by a rotation matrix **Q**.
2. Rotation of the right camera according to the transformation described by a matrix **R** from **Pr** = **R**(**Pl** − **T**) where Pl and Pr are two vector that pointing at the same point P from 3D space, translation T = Or–Ol and rotation determined by an orthogonal matrix R.

The matrix **Q** can be found by considering three mutually orthogonal unit vectors: **q**1, **q**2 and **q**3. The vector **q**1 is collinear with the translation vector **T** between the focus points of the two cameras and is given as

$$q_1 = \frac{T}{\|T\|} \tag{4}$$

The vector **q**2 is orthogonal to the vector **q**1. Because

$$[-T2, T1, 0] \cdot [T1, T2, T3]^T = 0 \tag{5}$$

then **q**2 takes the form

$$q_2 = \frac{[-T2, T1, 0]^T}{\sqrt{T_2^2 + T_1^2}} \tag{6}$$

The third vector **q**3 has to be simultaneously orthogonal to the vectors **q**1 and **q**2. Therefore it can be set to the vector product

$$q3 = q1 \times q2. \tag{7}$$

The vectors q1, q2 and q3 determine the following rotation matrix Q:

$$Q = \begin{bmatrix} q_1^T \\ q_2^T \\ q_3^T \end{bmatrix}_{3x3}$$

In practice, to obtain integer values of coordinates in the new (i.e. rectified) camera setup, the rectification process should be performed backwards, i.e. starting

from the new coordinates and applying the inverse transformation Q^{-1}. This way, the new intensity values in the 'new' system can be determined, for example, by the bilinear interpolation of the original values from the 'old' setup.
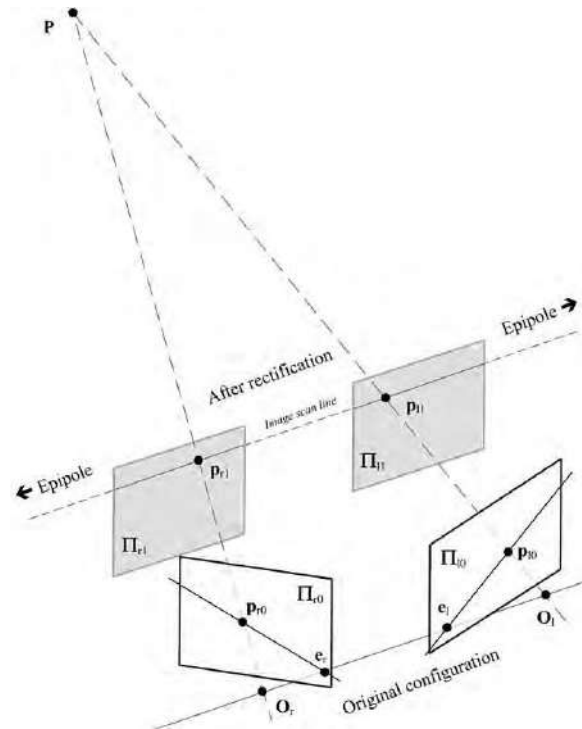
Figure2. Stereo image rectification. The epipolar lines become collinear and parallel to the image scanning lines [8]

Fusiello*et al.*[9] described that positions (*i.e.*, optical centers) of the new PPMs are the same as the old cameras, whereas the new orientation (the same for both cameras) differs from the old ones by suitable rotations, since intrinsic parameters are the same for both cameras. Therefore, the two resulting PPMs will differ only in their optical centers, and they can be thought as a single camera translated along the *X*-axis of its reference system. The new PPMs can be written in terms of their factorization. From Equations (2) and (3):

$$\tilde{P}_{n1} = A[R| - RC_1], \ \tilde{P}_{n2} = A[R| - RC_2] \qquad (8)$$

The intrinsic parameters matrix A is the same for both PPMs, and can be chosen arbitrarily. The optical centers C1 and C2 are given by the old optical centers, computed with Equation (3). The matrix R, which gives the camera's pose, is the same for both PPMs. More information on image rectification can be found in [9,10].

Segmentation

The third step was to segment theobjects from the background. Segmentation is carried out to obtain the image with a white background, and to remove the noise of the image. The following stages are performed in image segmentation (see Figure 3):

- Reading and adjust the color image (acquire background color information and minimum pixels area). Converting the colored image into a black and white image and removing noises.
- Extracting the image boundary.
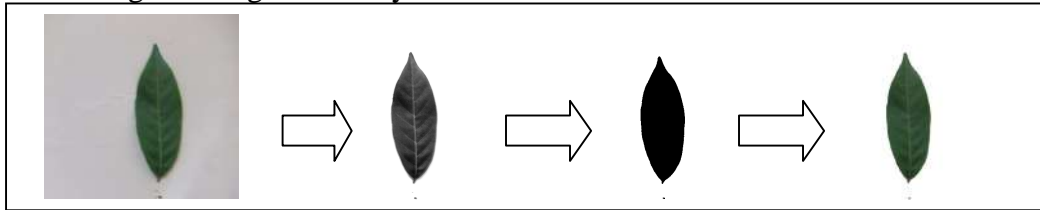
Figure 3. Stages in images segmentation

The color setting is performed to improve the color difference between the background and the object. The image manipulation is performed in black and white version of the images, *i.e.*, edge, boundaries, and so on. While converting the image from colored into black and white, if the object's color is similar to the background, we will not be able to obtain a profile overview of the overall object.

Disparity Map

The final step was construct the disparity map. The disparity map is generated by using a stereo correspondence algorithm [11]. The virtue of the stereo matching algorithm can be described by using the same structure groups. The main steps are (see Figure 4):
1. The computation of the matching values for each pixel of both input images.
2. The aggregation of the computed matching values, in addition to the supporting area for each pixel of each image.
3. To find the optimum disparity value for each pixel of an image.
4. To refine the resultant disparity maps.
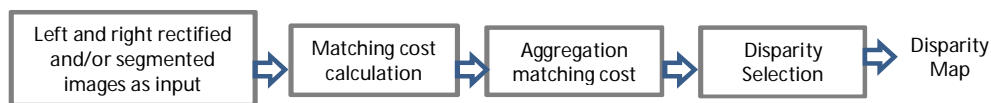
Figure4. Disparity map computation

Sum of absolute intensity differences (SAD) is the simplest and most frequently used measure in the matching. SAD involves a simple subtraction and the computation of the absolute values, as follows:

$$SAD(x, y, d) = \sum_{(i,j) \in W(x,y)}^{N} |I_L(i,j) - I_R(i-d,j)|$$

where I_l, I_r are intensity value of left and right images, (x, y) are pixel coordinates, W is window size and d is the disparity value.

Results and Discussion

Figure 6 is the left and right images, which are captured by a Fujifilm W3 stereo camera, taken from two checkerboard patterns attached on the corner of the wall with an angle of 90^0 to each other. These are the images used in calibration.
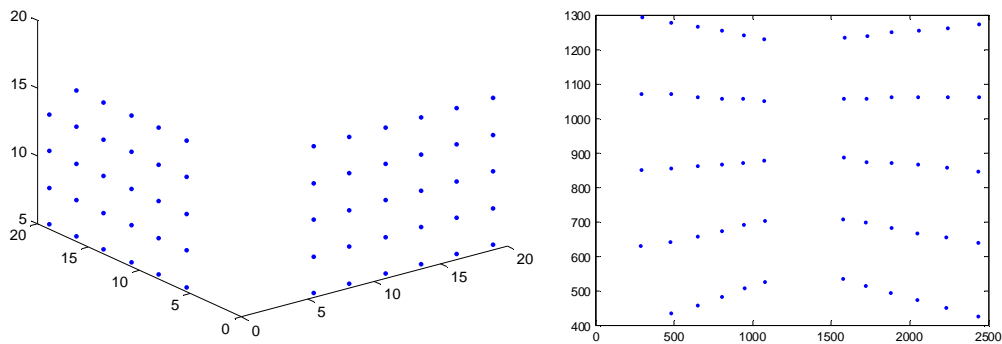

Figure 5. The left and right images used in calibration

The result of camera calibration based on the left and right images

Camera calibration is performed on both the left and right images. The real world coordinate and image coordinate points are manually determined in advance (see Figure 6 and 7).
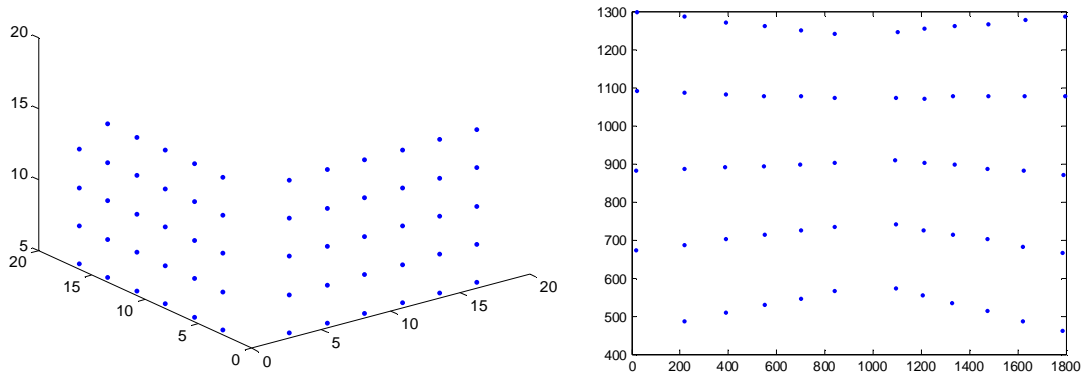
- *Calibration of the left image*

Figure 6. Coordinate points of left image on 3D and 2D plane
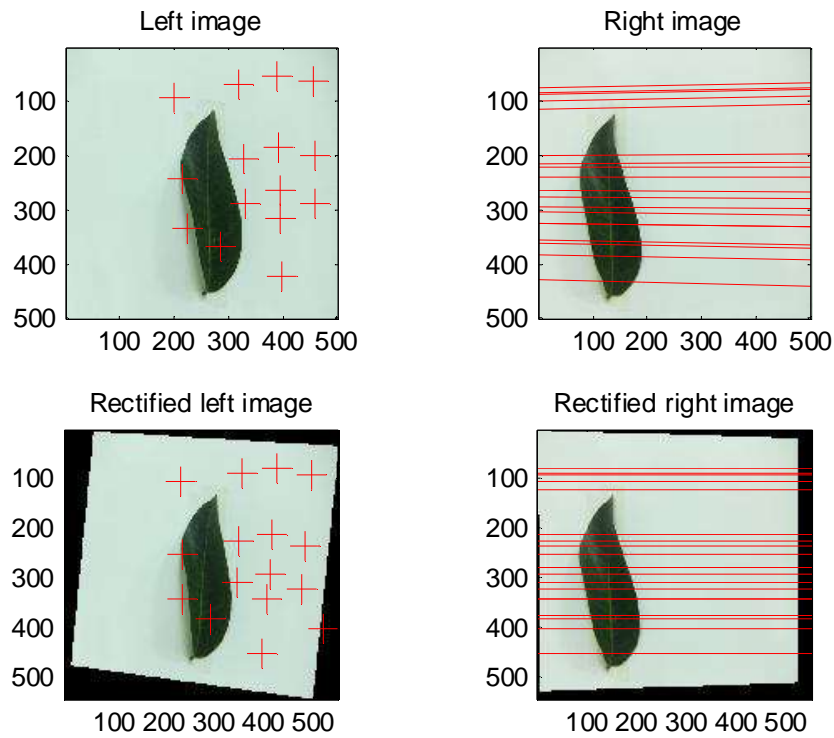
- *Calibration of the right image*

Figure7. Coordinate points of right image on 3D and 2D plane

Table 1. Extrinsic and intrinsic parameters resulting from the calibration of cameras

| Parameters	Calibration of the right image	Calibration of the left image
Calibration Matrix	$\begin{pmatrix} 0{,}015228 & -0{,}02853 & 0{,}000127 & 0{,}645008 \\ -0{,}006 & -0{,}00685 & -0{,}02988 & 0{,}762852 \\ -5{,}86E-06 & -6{,}76E-06 & 2{,}21E-07 & 0{,}000492 \end{pmatrix}$	$\begin{pmatrix} 0{,}015245 & -0{,}03217 & 0{,}000101 & 0{,}527837 \\ -0{,}00708 & -0{,}00706 & -0{,}03271 & 0{,}84791 \\ -6{,}70E-06 & -6{,}83E-06 & 3{,}00E-07 & 0{,}00054 \end{pmatrix}$
Rotation Matrix	$\begin{pmatrix} -0{,}75568 & 0{,}654928 & 0{,}003452 \\ 0{,}01875 & 0{,}016374 & 0{,}999689 \\ -0{,}65466 & -0{,}755514 & 0{,}024659 \end{pmatrix}$	$\begin{pmatrix} -0{,}71347 & 0{,}7006567 & 0{,}005895 \\ 0{,}02614 & 0{,}018209 & 0{,}999492 \\ -0{,}70019 & -0{,}71327 & 0{,}031307 \end{pmatrix}$
Translation Matrix	$\begin{pmatrix} -0{,}25379 \\ -10{,}0607 \\ 54{,}98734 \end{pmatrix}$	$\begin{pmatrix} 4{,}9007 \\ -10{,}3723 \\ 56{,}3552 \end{pmatrix}$
Scaling parameter in the direction of X	3373,920	3490,079
Scaling parameter in the direction of Y	3362,329	3447,735
Optical center of the image	(1295; 935)	(1281,78; 936,94)
Average pixel error in the direction X	0,0025	0,00123
Average pixel error in the direction Y	-0,00026	0,0088

Image rectification

The resulting image obtained after the rectification performed on the left and right images can be seen in Figure 8.

Figure 8. Before and after rectification

Segmentation of the left and right images

The results obtained after image segmentation on the left and right images can be seen in Figure 9.
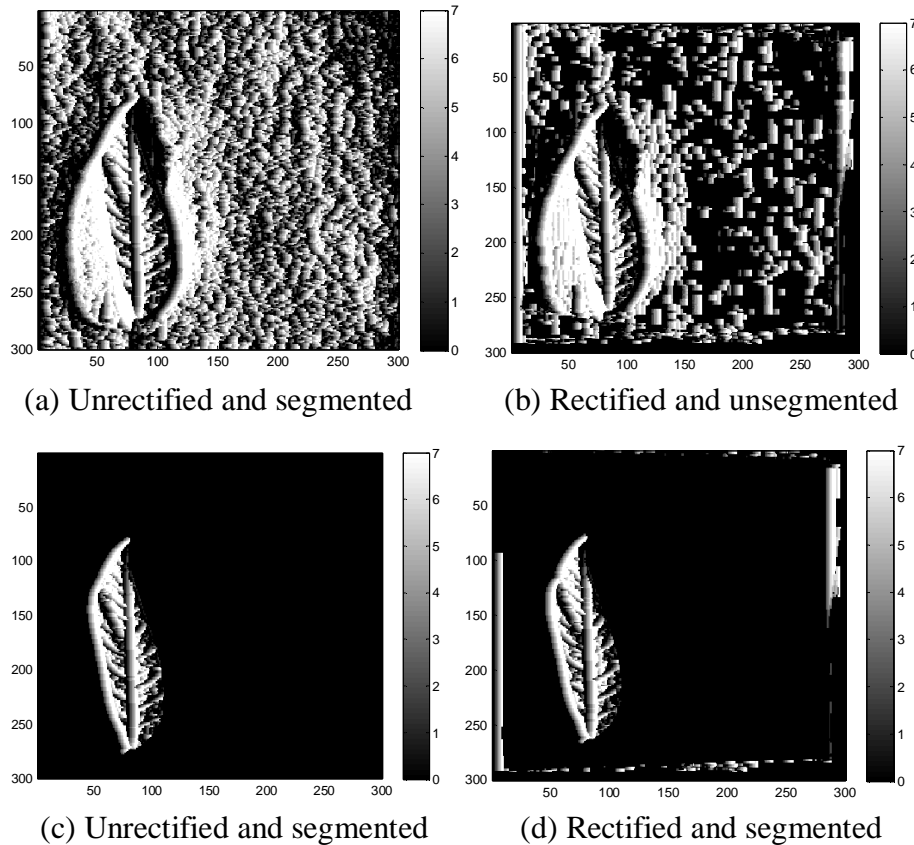

(a)   (b)
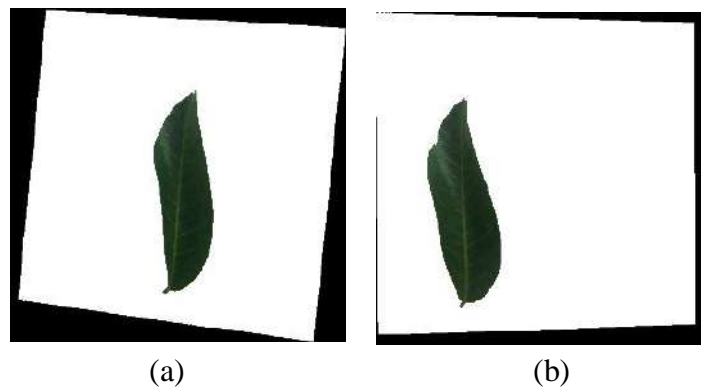
Figure 9. Segmentation (a) image segmentation of left rectified image, (b) image segmentation of left rectified image

Disparity of the stereo image

The disparity map is obtained by using the SAD (Sum Absolute Difference) matching algorithm. In this algorithm, the setting is done by adjusting the window value w=3x3 and the maximum disparity dmax=7. The results of several experiments

to obtain the disparity map of stereo images can be seen in the Figure 10.

(a) Unrectified and segmented   (b) Rectified and unsegmented

(c) Unrectified and segmented   (d) Rectified and segmented

Figure 10. Segmentation and rectification

Discussion

Based on Figure 10, it can be shown that visually, the disparity map resulted from segmentation andunrectification gives better results than those unsegmentation and rectification. Meanwhile, the disparity results obtained by using image rectification do not provide better results than those obtained by using images without rectification. It means that the original images without rectification produced by stereo cameras do not undergo significant distortion or significant shifts. Finally, the disparity results will look better than previous researches such as San Yong and Hon [4],Kamencay *et al.*[5,11], Baha and Larabi [6]if the segmentation is performed on both images, the left and right image.This is due to the presence of the segmentation and rectification that can reduce the matching process and reduce the intensity bias between the objects of which background will be recognized. Therefore, the new schame of disparity map calculation based on images rectification and segmentation provided better but not optimum result, so for future work, the schame is necessary to develop in the future by taking the smoothing or post processing of original images or disparity map resulted.

Conclusion

In this paper, we had done images rectification and segmentation to improve quality of disparity maps an object in the stereo images. Images rectification and segmentation is required to reduce the complexity of the computation for the left and right correspondences of pixel images. For images rectification, camera calibration will be done earlier. The experimental results show that visually, the disparity map resulted from images rectified gives better results than those unrectified. Furthermore, the disparity map results show a better appearance by conducting left and right image segmentation beforehand.The disparity map is obtained by using the SAD (Sum Absolute Difference) matching algorithm. In this algorithm, the setting is done by adjusting the window value w=3x3 and the maximum disparity dmax=7.

Reference

[1] Malik, A. S., Choi, T. S., Nisar, H. & Global, I, 2012.*Depth map and 3D imaging applications: Algorithms and technologies*:Information Science Reference.

[2] Islam, M. S., Hannan, M. & Basri, H., 2012. An Application of Stereo Matching Algorithm for Waste Bin Level Estimation. *Journal of Asian Scientific Research,* 2(11): 731-736.

[3] Lipnickas, A. & Knyš, A., 2009. A Stereovision System for 3-D Perception. *Electronics & Electrical Engineering,* 3(91): 99-102.

[4] San Yong, Y. & Hon, H. W., 2008. Disparity Estimation for Objects of Interest. *Proceedings of World Academy of Science: Engineering & Technology,* 45(19): 533-536.

[5] Kamencay, P., Breznan, M., Jarina, R., Lukac, P. & Zachariasova, M., 2012a. Improved Depth Map Estimation from Stereo Images Based on Hybrid Method. *Radioengineering,* 21(1): 70-78.

[6] Baha, N. & Larabi, S., 2012. Neural Disparity Map Estimation from Stereo Image. *International Arab Journal of Information Technology (IAJIT),* 9(3): 217-224.

[7] Zang, Z., 2004.*Emerging Topics in Computer Vision*:Prentice Hall Profesional Technical Reference

[8] Cyganek, B. & J., P. S., 2009.*An Introduction to 3D Computer Vision Techniques and Algorithms*:John Wiley & Sons, Ltd.

[9] Fusiello, A., Trucco, E. & Verri, A., 2000. A compact algorithm for rectification of stereo pairs. *Machine Vision and Applications,* 12(1): 16-22.

[10] Shimizu, M. & Okutomi, M. (Year) Calibration and rectification for reflection stereo. *In:* Computer Vision and Pattern Recognition, 2008. CVPR 2008. IEEE Conference on, 2008. IEEE, 1-8.

[11] Kamencay, P., Zachariasova, M., Breznan, M., Jarina, R., Hudec, R., Benco, M. & Matuska, S., 2012b. A New Approach for Disparity Map Estimation from Stereo Image Sequences using Hybrid Segmentation Algorithm. *International Journal of Modern Engineering Research (IJMER),* 2(5): 3201-3206.