Matching algorithm performance analysis for autocalibration method of stereo vision

Raden Arief Setyawan[1], Rudy Soenoko[2], Moch Agus Choiron[3], Panca Mudjirahardjo[4]
[1,4]Electrical Engineering Department, Brawijaya University, Indonesia
[2,3]Mechanical Engineering Department, Brawijaya University, Indonesia

Article Info

ABSTRACT

Stereo vision is one of the interesting research topics in the computer vision field. Two cameras are used to generate a disparity map, resulting in the depth estimation. Camera calibration is the most important step in stereo vision. The calibration step is used to generate an intrinsic parameter of each camera to get a better disparity map. In general, the calibration process is done manually by using a chessboard pattern, but this process is an exhausting task. Self-calibration is an important ability required to overcome this problem. Self-calibration required a robust and good matching algorithm to find the key feature between images as reference. The purpose of this paper is to analyze the performance of three matching algorithms for the autocalibration process. The matching algorithms used in this research are SIFT, SURF, and ORB. The result shows that SIFT performs better than other methods.

Corresponding Author:

Raden Arief Setyawan,
Department Electrical Engineering,
Brawijaya University,
167 MT Haryono St., Malang 65145 Indonesia.
Email: rarief@ub.ac.id

1. INTRODUCTION

Vision-based measurement has been one of the most interesting research topics in the last decades. Many applications have been developed using vision-based measurement [1]. The two major methods of 3D measurement can be categorized into active and passive methods. Structured illumination or laser is used in the active measurement. This method is not applicable in many cases. The passive 3D measurement is based on stereo vision and provides more advantages than active measurement. It requires simpler instrumentation, offering higher applicability in many environments. However, the major issue for passive measurement is the difficulty in finding accurate correspondence between stereo images [2].

Stereo calibration is the most important step to find a correspondence point. Camera calibration is required to ensure that both cameras are in perfect position and to remove distortion. Traditionally, camera calibration is performed using the standard chess-board picture [3]. However, much work is required in the self-calibration methods. Stereo self-calibration refers to the automatic determination of stereo camera parameters from image sequences.

Self-calibration is an important ability required for the introduction of stereo cameras into the market. Many works have been published with this method [4-11]. It can guarantee maintenance-free and the long-term operation, as the environmental conditions may change the camera position. Special expertise is required to do the offline calibration. Self-calibration may reduce regular offline calibration time

and cost. Even if human eyes have different characteristics with minus/plus/cylindrical properties, the human brain can automatically adjust. Consequently, the human being will have no difficulties in merging two visions from the left and right cameras. In designing a self-calibration method, a matching algorithm is an important tool to find a correspondence point between images of two cameras.

The main objective of this paper is to analyze the performance of three matching algorithms for the autocalibration process. Two of the most common techniques for stereo correspondence are the sum of absolute differences (SAD) and the sum of squared differences (SSD). The corresponding points between images have been obtained by minimizing SAD or SSD in area-based block matching [12]. However, these two techniques result in low accuracy as their major drawback. An improvement by using sub-pixel block matching techniques has been explored in [4], but the obtained accuracy was still not enough. Recently, there have been many algorithms proposed on image matching using various techniques [13]. In this work, a set of experiments demonstrates that the stereo vision system employing the proposed technique can measure 3D surfaces of free-form objects with sub-mm accuracy. Three matching techniques used in this research are SIFT, SURF, and ORB. The matching algorithm provides the characteristics of each camera [14]. It used to transform the second image to perform automatic stereo calibration. The explanation of each algorithm is explained as follows.

- SIFT

Scale invariant feature transform (SIFT) is a matching algorithm proposed by Lowe [15]. This algorithm works very well in finding a correspondence point of the image which is rotated and transformed. This algorithm consists of four steps. The first step is the estimation of scale-space extrema using the Difference of Gaussian method, being express using (1) and described in Figure 1.

$$
\begin{aligned}
D(x,y,\sigma) &= (G(x,y,k\sigma) - G(x,y,\sigma)) * I(x,y) \\
&= L(x,y,k\sigma) - L(x,y,\sigma).
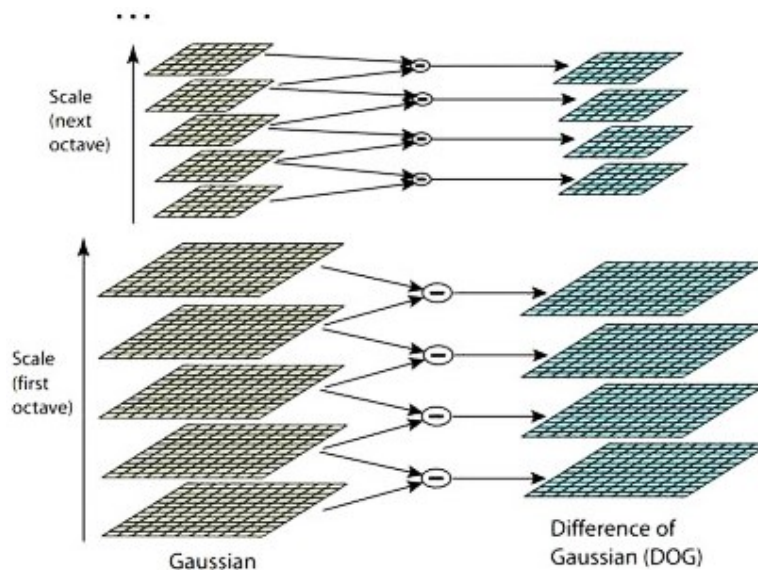\end{aligned}
\tag{1}
$$

Figure 1. The estimation of scale-space extrema using Difference of Gaussian method

In the next step, the key point candidates are refined by the elimination of low value. Laplacian of Gaussian $\sigma2\nabla2G$ is used since it produces the most stable image feature than others. The correlation between the Difference of Gaussian and the Laplacian of gaussian can be expressed using (2) and (3).

$$
\sigma\nabla^2 G = \frac{\partial G}{\partial \sigma} \approx \frac{G(x,y,k\sigma) - G(x,y,\sigma)}{k\sigma - \sigma}
\tag{2}
$$

$$
G(x,y,k\sigma) - G(x,y,\sigma) \approx (k-1)\sigma^2\nabla^2 G.
\tag{3}
$$

The key point orientation is assigned by using an image gradient. The final step is the computation of the local image descriptor based on the gradient and orientation of the key point. Because of its algorithm complexity, SIFT requires a large computational capacity, even though it is very suitable for object recognition applications [16, 17].

- SURF

Speed up robust feature (SURF) technique performs faster than SIFT [18]. In some cases, it performs with equal quality to SIFT. SURF technique is based on a descriptor and a detector, which is equal to SIFT. Instead of using the gaussian average of the images, SURF uses squares for approximation. It employs the Hessian matrix-based Blob detector to find the point of interest. Wavelet response is used for orientation assignment by applying gaussian weight. SURF feature descriptor is generated by the wavelet response of the subregion. The subregion is the division of the neighbor around the key point. Two points will form a correspondence (match) if they the same contrast, generated from Laplacian.

- ORB

Oriented FAST and rotated BRIEF (ORB) has been proposed by Rublee, et al. [19]. It is another alternative for SIFT. ORB is a combination of the FAST key point and the BRIEF descriptor. The FAST is used to determine the key point [20]. In the next step, Harris corner is used to find the top N point. FAST computes the intensity-weighted centroid, located at the center. The orientation is obtained by the vector direction to the centroid.

2. RESEARCH METHOD

The purpose of this research is to find the best algorithm for the auto-calibration of stereo vision. The first step of calibration is the finding of the corresponding points between two images. The accuracy of this step determines the accuracy of stereo vision. The object of this research is a microscopic object with the size of a few millimeters. The disparity of the points is converted into the intrinsic parameter of the camera.

The method used in this research is described in Figure 2. The stereo image has been produced using two cameras. In order to handle the very narrow view area caused by the small-size objects, the converged camera setup is used. It is hard to put objects in the overlapped area if parallel cameras are used. The histogram equalization steps are required since the illumination of each image or camera color character possibly different [21]. To reduce the noises, the combination of Gaussian and medium filter applied. Both filters are proposed to improve the image quality [22]. Gaussian can be expressed in (4). While the median filter expressed in (5). A combination of both filter expressed using (6).

$$G(x,y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}} \tag{4}$$

$$M(x,y) = med\{f(x-i, y-j), i, j \in W\} \tag{5}$$

$$f_o(x,y) = G(x,y) + M(x,y) \tag{6}$$

The result of histogram equalization is processed using a feature extraction algorithm. Three feature extraction algorithms are used to find the match correspondence point on each set of images [20]. The match correspondence point used for the rectification process [23]. Distance between each corresponding point is used to extract the stereo parameter. The output of this method is the stereo calibration parameter [24]. The result of this process can be transformed into a 3D surface.
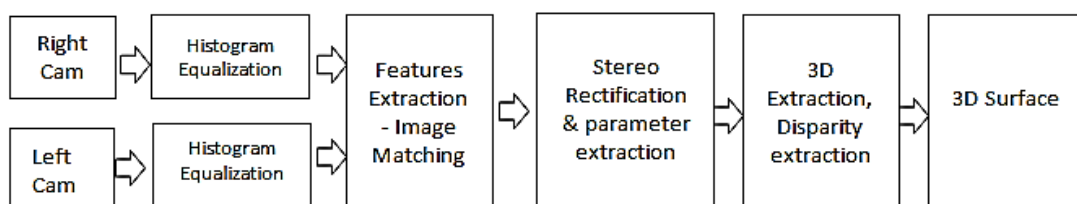

Figure 2. The distance measurement procedure

Two industrial standard HD camera is used in this research. These cameras are equipped with a 100x lens to enlarge the object size. Two captured images from both cameras are then compared and evaluated using the matching algorithm to find the corresponding point. Figure 3 shows the camera setup and the object size.

Figure 3. The cameras set-up and the object size

A millimeter template is used to measure the size of the object and as a reference of the auto-calibration. The dimension of the object is shown in Figure 4 (a), whereas Figure 4 (b) represents five pairs of image sets which are generated using the system for testing purpose. Each set is compared using three matching algorithms: SIFT, SURF, and ORB.
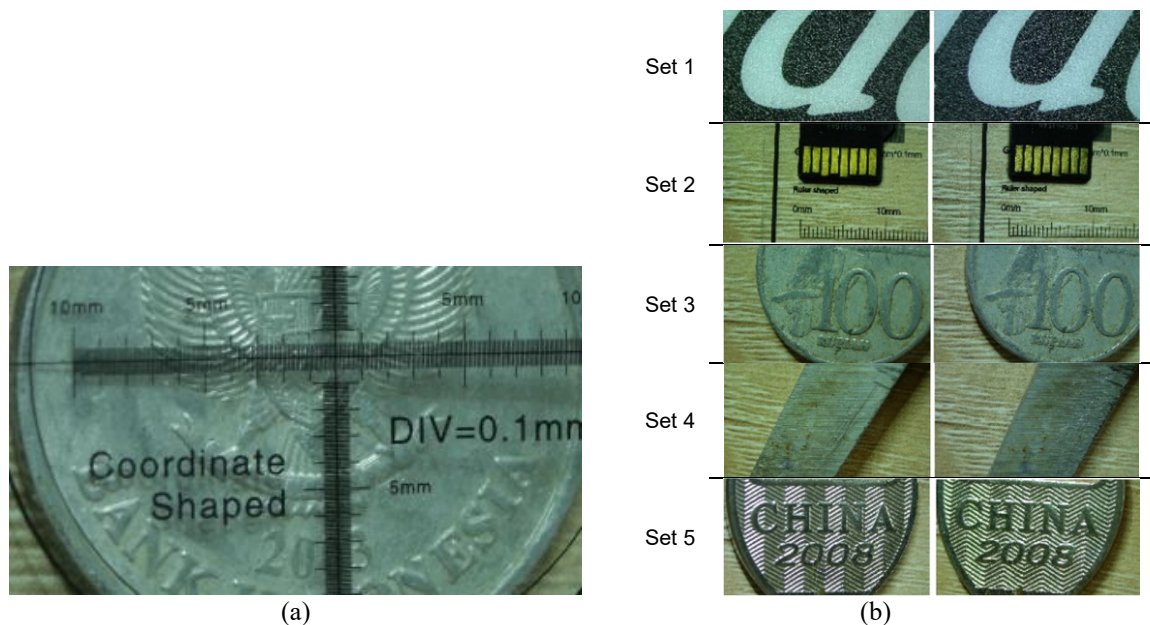

(a)

(b)

Figure 4. (a) The dimension of the object and (b) the datasets used in the research

3. RESULTS AND ANALYSIS

The execution of SIFT, SURF, and ORB on each pair of image sets has been performed to find the best method for image matching. In the obtained results, the green line indicates the correspondence point between the left and right images. The number of connected lines shows the number of matched points. How

ever, each algorithm still resulted in an error if the algorithm failed to match the correct points. The result of this matching is used to generate the calibration parameter of stereo vision.

3.1. Matching results using SIFT, SURF, and ORB

The result of implementing the SIFT, SURF and ORB algorithms on the captured object are given respectively in Figures 5 (a-c). As seen in the image set 1 and 5 of Figure 5 (a), only a few lines have been generated by the SIFT algorithm. The background has very high similarities between images. The result of SURF algorithm implementation given in Figure 5 (b) indicates that on the image set 1 there have been only a few lines generated by an algorithm and some lines indicated a major error. The rest of the image sets shows the correct corresponding points. The result of implementing the ORB algorithm shown in Figure 5 (c) also indicates that there have been only a few lines generated by the algorithm on the image set 1, with some lines indicated major error. The four other image sets indicated the correct corresponding points.

The comparison of the matching results using SIFT, SURF, and ORB techniques is presented in Table 1. It indicates matching accuracy of the three algorithms SIFT, SURF, and ORB. It can be known from the table that the SIFT algorithm gives the highest average percentage accuracy. However, the percentage of correct lines varies depending on the image characteristics. For the image with high similarities, SURF failed to give a good result, whereas ORB could generate many lines, but with high error rates.
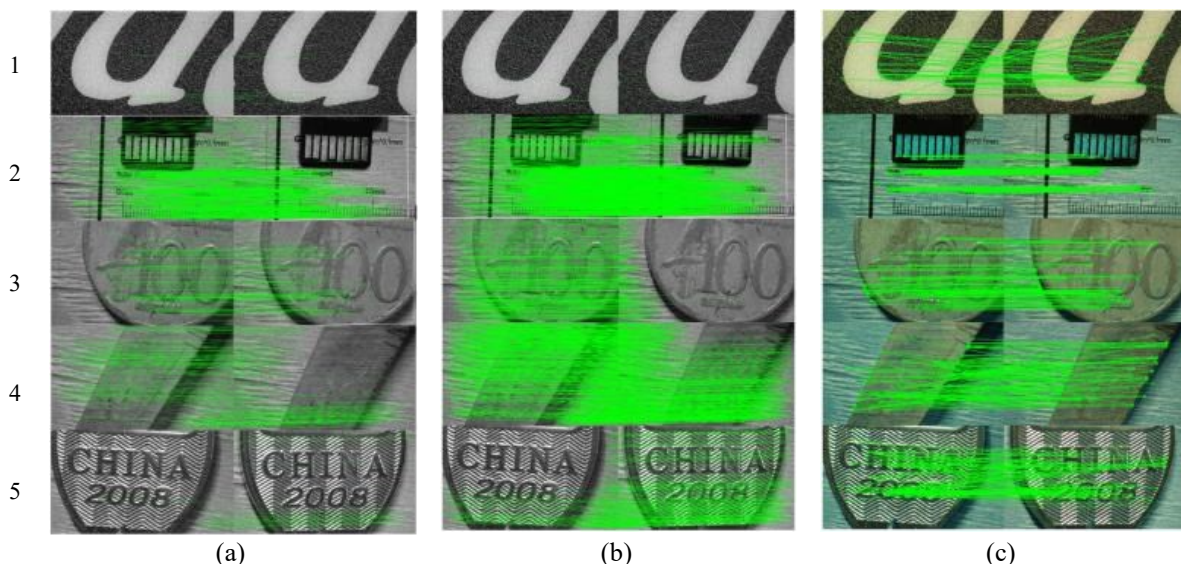

Figure 5. Experiment results of matching algorithm using: (a) SIFT, (b) SURF, and (c) ORB

As seen in Figure 5 (a), the image set 1 and set 5 only have a few lines have been generated by the SIFT algorithm. The background has very high similarities between images. In contrast, the results of the SURF algorithm (Figure 5 (b)), which applied to the image set 1, only a few lines generated by the algorithm. The rest of the image sets shows the correct corresponding points. The result of implementing the ORB algorithm shown in Figure 5 (c) also indicates that there have been only a few lines generated by the algorithm on the image set 1, with some lines indicated major error. Parallel lines, which group together to form a thicker image symbolize accuracy. In contrast, line out of parallel, crisscrossing each other creating a dispersed image signify inaccuracy.

The comparison of the matching results using SIFT, SURF, and ORB techniques is presented in Table 1. It indicates matching accuracy of the three algorithms SIFT, SURF, and ORB. As can be seen in the table, the SIFT algorithm gives the highest average accuracy percentage. However, the percentage of correct lines varies depending on the image characteristics. For the image with high similarities, SURF failed to give a good result, whereas ORB could generate many lines, but with high error rates.

The result in Table 1 compared with the result from Karami et.all [13] with the case of varying intensity shown in Table 2. It shows that in both works, SIFT performs better than other methods. Table 3 shows the comparison of the computational time of each algorithm. It shows that the SIFT method required a longer time than the others due to the complex algorithm computation. SIFT required a longer time when the image had high similarities in its texture. Figure 6 indicates that the ORB algorithm has the fastest computation time for all images sets. It takes less than 0.5s processing time. However, the ORB algorithm

gives less matching rates compared with other methods. The line chart in Figure 6 also indicates that the complexity of the images linear with the computation time. Image set 1 and 5 give the longest computation time than other images set because of their complexity.

Table 1. Comparison of the matching results using the SIFT, SURF, and ORB techniques

No.	SIFT			SURF			ORB		
	Lines	Correct Point	% Correct	Lines	Correct Point	% Correct	Lines	Correct Point	% Correct
1	15	14	93.33%	14	2	14.29%	78	24	30.77%
2	150	120	80.00%	443	430	97.07%	89	70	78.65%
3	400	356	89.00%	278	256	92.09%	254	224	88.19%
4	345	321	93.04%	600	467	77.83%	345	156	45.22%
5	20	16	80.00%	125	112	89.60%	375	153	40.80%
	Average		87.08%	Average		74.17%	Average		56.73%

Table 2. Comparison of the matching results between Karami and this work

Method	Match Rate (%)	
	Karami	This Work
SIFT	76.7	87.08
SURF	72.6	74.17
ORB	63.6	56.73

Table 3. Computational time using the SIFT, SURF, and ORB techniques

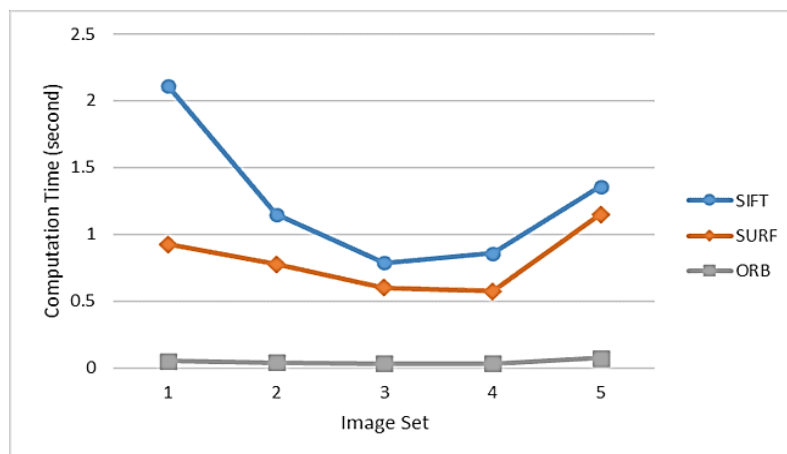| Image Set	Computational Time		
	SIFT	SURF	ORB
1	2.114	0.926	0.052
2	1.149	0.777	0.039
3	0.788	0.6	0.033
4	0.858	0.576	0.033
5	1.36	1.149	0.075

Figure 6. Computational time comparison chart

3.2. Image rectification

The matching point from previous steps is used for rectifying the images. The difference position between source and destination point used as a reference for transformation. Figures 7a and 7b show a distorted image from left and right camera. Figure 7a used as the reference, while the Figure 7b is the object of transformation. The result of the image transformation of Figure 7b displayed in Figure 7c. This transformation based on the homography equation to reduce distortion [25].
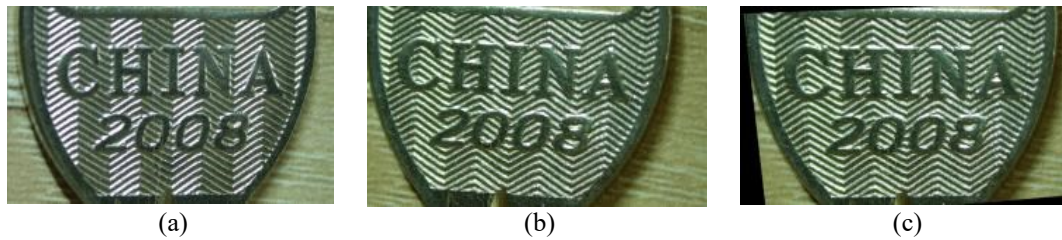
(a)   (b)   (c)

Figure 7. Rectification result (a) left Image as reference (b) right image (c) the result of rectification

3.3. 3D surface generation

The matching process results in the distance between points. Using the distance values, a 3D surface object can be generated by projecting them onto the z-axis [26, 27]. Distance value between both images assigned as the depth value. If the distance is small, the object is closer to the camera, and vice versa. Depth value for each pixel than converted to grayscale to distinguish the depth of point. Figure 8 shows the generated disparity map of the dataset using SIFT Adjustment. Correlated point produces by SIFT is used to calculate the stereo camera parameters. The result shows that the algorithm successfully generates match stereo, however, the noisy output is a bit challenging. Using the depth value as z-axis produce 3d view as shown in Figure 9 the algorithm successfully produces 3D reconstruction, but the noises reduce image quality.
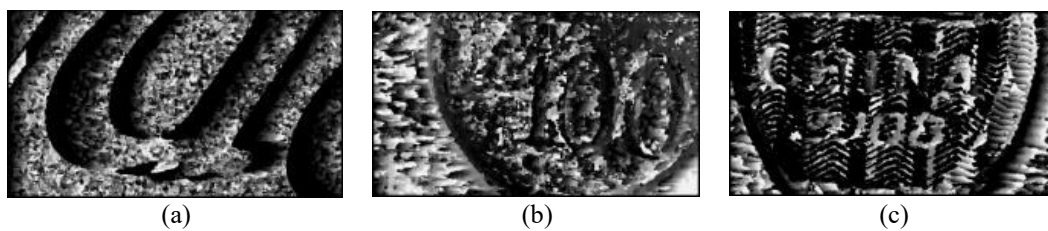

(a)   (b)   (c)

Figure 8. Generated depth value based on SIFT matching algorithm

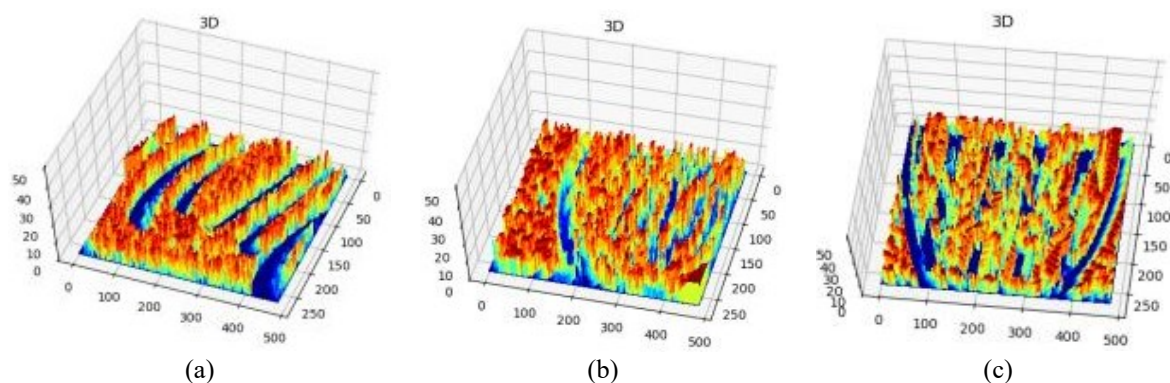(a)   (b)   (c)

Figure 9. 3D surface reconstruction

4. CONCLUSION

In this paper, three different image matching techniques, SIFT, SURF, and ORB, for stereo autocalibration system have been compared. SIFT indicates the best performance in most scenarios under consideration. In the special case, when the images contain multiple high similarities texture, SURF failed to give good results. In the ORB implementation, the features are mostly concentrated in objects at the center of the image. While SIFT and SURF, the features are distributed over the image. The 3D reconstruction image has successfully generated, but the noise reduces the quality of the images. For future work, a good filtering algorithm required for a better result, without scarifying the details of images.

REFERENCES

[1] J. C. H. Londono, "Surface roughness estimation by 3D stereo SEM reconstruction," *Universidad Nacional de Colombia*, pp. 70, 2015.

[2] T. Shibahara and T. Aoki, "A Sub-Pixel Stereo Correspondence Technique Based On 1d Phase-Only Correlation," *2007 IEEE International Conference on Image Processing, San Antonio*, TX, pp. V-221-V–224, 2007.

[3] R. A. Setyawan, R. Sunoko, M. A. Choiron, and P. M. Rahardjo, "Implementation of stereo vision semi-global block matching methods for distance measurement," *Indones. J. Electr. Eng. Comput. Sci*., vol. 12, no. 2, pp. 585–591, 2018.

[4] T. Dang, C. Hoffmann, and C. Stiller, "Continuous stereo self-calibration by camera parameter tracking," *IEEE Transactions Image Process.*, vol. 18, no. 7, pp. 1536–1550, July 2009.

[5] H. Wang et al., "An automatic self-calibration approach for wide baseline stereo cameras using sea surface images," *2015 IEEE International Workshop on Advanced Robotics and its Social Impacts (ARSO),* Lyon, pp. 1-6, 2015.

[6] Z. Xiao, J. Liang, D. Yu, Z. Tang, and A. Asundi, "An accurate stereo vision system using cross-shaped target self-calibration method based on photogrammetry," *Optics and Lasers in En*gineering, vol. 48, no. 12, pp. 1252–1261, December 2010.

[7] S. Shah, N. Bhuiyan, and O. O. Khalifa, "Efficient 3D stereo vision stabilization for multi-camera viewpoints," *Bulletin of Electrical Engineering and Informatics*, vol. 8, no. 3, pp. 882–889, September 2019.

[8] H. Yu and Y. Wang, "An improved self-calibration method for active stereo camera," *2006 6th World Congress on Intelligent Control and Automation, Dalian,* pp. 5186-5190, 2006.

[9] B. H. Yang and F. L. Lian, "Continuous self-calibration of stereo camera based on particle filter framework for endoscopic videos," *2016 IEEE International Conference on Industrial Technology (ICIT),* Taipei, pp. 1476-1481, 2016.

[10] I. El Batteoui, K. Satori, and A. Saaidi, "A Stereo Planar Self-Calibration Method of a Camera with Variable Intrinsic Parameters," *2016 13th International Conference on Computer Graphics, Imaging and Visualization (CGiV),* Beni Mellal, pp. 6-11, 2016.

[11] Y. Hu, Q. Chen, S. Feng, T. Tao, A. Asundi, and C. Zuo, "A new microscopic telecentric stereo vision system Calibration, rectification, and three-dimensional reconstruction," *Optics and Lasers in Engineering*, vol. 113, pp. 14–22, February 2019.

[12] D. Scharstein and R. Szeliski, "A Taxonomy and Evaluation of Dense Two-Frame Stereo Correspondence Algorithms," *International Journal of Computer Vision,* vol. 47, no. 1-3, pp. 7–42, April 2002.

[13] E. Karami, S. Prasad, and M. Shehata, "Image Matching Using SIFT, SURF, BRIEF and ORB: Performance Comparison for Distorted Images," *Conference: 2015 Newfoundland Electrical and Computer Engineering Conference*, 2015.

[14] M. Oktiana, F. Arnia, Y. Away, and K. Munadi, "Features for cross spectral image matching: A survey," *Bulletin of Electrical Engineering Informatics*, vol. 7, no. 4, pp. 552–560, December 2018.

[15] S. Keypoints and D. G. Lowe, "Distinctive Image Features from," *International Journal Computer Vision,* vol. 60, no. 2, pp. 91–110, November 2004.

[16] I. Rey Otero and M. Delbracio, "Anatomy of the SIFT Method," *Image Processing on Line*, vol. 4, pp. 370–396, 2014.

[17] C. Je and H. M. Park, "Optimized hierarchical block matching for fast and accurate image registration," *Signal Processing: Image Commun.,* vol. 28, no. 7, pp. 779–791, August 2013.

[18] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool, "Speeded-Up Robust Features (SURF)," *Computer Vision and Image Understanding*, vol. 110, no. 3, pp. 346–359, June 2008.

[19] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski, "ORB: An efficient alternative to SIFT or SURF 2011 *International Conference on Computer Vision, Barcelona*, pp. 2564-2571, 2011.

[20] O. Andersson and S. R. Marquez, "A comparison of object detection algorithms using unmanipulated testing images: Comparing SIFT, KAZE, AKAZE and ORB," 2016.

[21] Y. Chang, C. Jung, P. Ke, H. Song, and J. Hwang, "Automatic Contrast-Limited Adaptive Histogram Equalization with Dual Gamma Correction," *IEEE Access*, vol. 6, pp. 11782–11792, 2018.

[22] B. Chan, K. H. Lim, L. Gopal, A. A. Gopalai, and D. Gouwanda, "Stereo vision human motion detection and tracking in uncontrolled environment," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 16, no. 3, pp. 955–964, June 2018.

[23] X. Li, W. Liu, W. Fan, J. Sun, and N. Satoshi, "Perspective correction using camera intrinsic parameters," *2016 IEEE 13th International Conference on Signal Processing (ICSP),* Chengdu, pp. 854-858, 2016.

[24] C. Loop and Z. Zhang, "Computing rectifying homographies for stereo vision," *Proceedings. 1999 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (Cat. No PR00149), Fort Collins, CO, USA,* vol. 1, pp. 125-131, 1999.

[25] Kukelova Z., Heller J., Bujnak M., Pajdla T., "Radial Distortion Homography," *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* Boston, MA, pp. 639-647, 2015.

[26] K. Peng, X. Chen, D. Zhou and Y. Liu, "3D reconstruction based on SIFT and Harris feature points," *2009 IEEE International Conference on Robotics and Biomimetics (ROBIO)*, Guilin, pp. 960-964, 2009.

[27] K. Yamada and A. Kimura, "A performance evaluation of keypoints detection methods SIFT and AKAZE for 3D reconstruction," *2018 International Workshop on Advanced Image Technology (IWAIT)*, Chiang Mai, pp. 1-4, 2018.