The Detection of 3D Object Using a Method of a Harris Corner Detector and Lucas-Kanade Tracker Based on Stereo Image

W. Prawira, E. Nasrullah, S. R. Sulistiyanti, F. X. A. Setyawan
Department of Electrical Engineering, Faculty of Engineering
University of Lampung
Bandar Lampung, Indonesia
wiraprawinal@gmail.com, sr_sulistiyanti@eng.unila.ac.id, fx.arinto@eng.unila.ac.id

Abstract— **This research proposes the use of Harris Corner Detector and Lucas-Kanade Tracker methods for the detection of 3D objects based on stereo image. The test image obtained from the results of capturing of the camera to the object of the form of tubes, balls, cubes, and 2D images. This research is the early step in the development of the ability of a computer vision to be able to mimic the performance of eye organs in humans in detecting an object. The detection step of the proposed method begins by determining the feature point on the image of the taking results of two cameras using the Harris Corner Detector. After the feature point of the two images obtained, then performed tracking feature point using the Lucas-Kanade Tracker method. In this research, the distance between the cameras used 10, 20, and 50 cm. The Effectiveness of the detection result of the proposed method is measured using the recall and precision parameter values obtained in the merged of the image. The proposed method gives a Recall value above 90% and a precision value above 50% for a distance of the cameras 10cm and 20cm for ball and tube objects. In the box object, the Recall value is 60% for a distance between the cameras 50cm and below 25% for a distance of the cameras 10cm and 20cm. The precision value for detection of the box object is very low, i.e. less than 25%.**

Keywords—object detection; 3D object; feature point; stereo image

I. INTRODUCTION

The very rapid progress of digital image processing technology is expected to be used to simplify human life. This technology can be applied in various fields, ranging from military and medical field to the home appliance fields. Automation technology uses it to eliminate the human role in decision making.

This research developed a system that would replace the performance of the human eye organ as a decision maker to detect the dimensions of the observed object. This research is expected to support the progress of image processing technology and can be applied as part of computer vision such as an automatic parking process, object detection in an industrial area, surveillance system, and others. It is also hoped that with this detection system of the object dimension can improve the optimization of human performance in completing the activity.

Previous research on object detection has been done by many researchers. The research about detection of a moving object on video taken by the moving camera using background reduction method has been done before [1,2,3]. The research about detection of 3D objects done previously is to reconstruct a three-dimensional object from a two-dimensional image using epipolar geometry [4]. The difference with the proposed method in this research is on the use of cameras. If the previous research using 1 camera then this research uses 2 cameras so that object detection is done by comparing between the images results of the right camera with the left camera. The other research that has been done is the introduction of 3D objects based on image features [5], canonical angles between the shape subspace [6], and the use of an artificial neural network method [7,8]. These researches used a mono image that obtained using a single camera while the proposed method used the stereo image. In a research using mono cameras, it is not possible to use the feature point tracking method because only has a single frame for each taking an image. In this proposed method used a grayscale image to reduce computational loads in order to shorten the processing time.

The research on the object detection using stereo cameras has also been done [9]. The difference between the research that has been done with this research is the method used. This research proposes the use of Harris corner detector and Lucas Kanade tracker methods while the research that has been done using a disparity depth method.

II. THE PROPOSED METHOD

A. Preprocessing

The image is taken using two cameras mounted car dashboard within 10, 20 and 50 centimeters. The distance of the midpoint between two cameras with the object is 150 cm. Fig. 1 shows how to capture the picture using 2 cameras. After the image is obtained, then the image is converted into a grayscale image.

978-1-4799-7675-1/17/$31.00 ©2017 IEEE

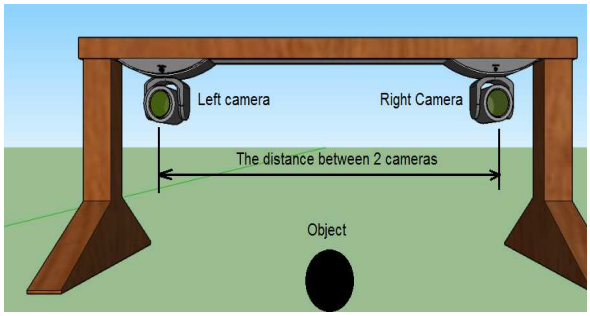


Fig. 1. Equipment to take the image.

Here, I is the intensity of the grayscale pixel, R is the intensity of red, G is the intensity of green, and B is the intensity of the blue color.

B. Feature Point Detection

The next step is the detection of feature points conducted using Harris Corner detector method [10]. Using this method, features of a corner point in the image will be detected and characterized by a particular colored pixel. Fig. 2 shows the results of a corner feature detection using the Harris Corner Detector method. The detection result from the corner feature point of the image captured from the left camera is different from the right camera this is because this method only shows the best corner point only.

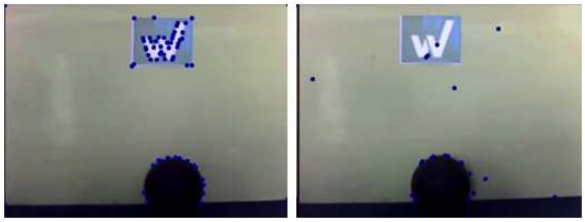
C. The correspondence of Corner point

After the corner points obtained, then the next step is to find correspondence between corner points. This correspondence is obtained by tracking the corner point between of the image obtained from the left and right cameras. The used method for tracking this corner point is the Lucas Kanade Tracker method [11]. Figure 3 shows the result of tracking the corner correspondence of the left and right images.

There is an inappropriate corner correspondence between the left and right images, shown in Fig. 3. The correspondence of the incorrect point corners between the left and right images is omitted using the Random Sample Consensus (RANSAC) method [12]. This method iteratively estimates the parameters of mathematical models a set of observed data containing outliers. The more correspondence of a corner point is detected then the detection process is also more accurate. The result of correspondence is used to obtain the homography matrix. After the homography matrix is obtained, the two images can be merged to determine the dimensions of the objects in the image.

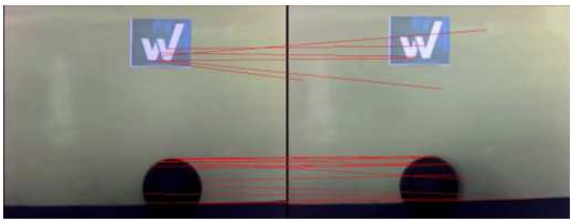
After converting to a grayscale image, the region of interest (ROI) is determined so that the right image and the left image are not too much different. Converting the RGB image to grayscale images is done using the equation (1).

$$I = 0.2989R + 0.5870G + 0.1141B \quad (1)$$



(a) Left camera (b) Right camera

Fig. 2. The detection result of a corner feature detection.



(a) Left camera (b) Right camera

Fig. 3. The corner correspondence of the left and right images.

D. Homographies (2D Projective Transforms)

The homography transformation between a left image and a right image is calculated after correspondence of the corner point is obtained. The homography transformation H between a left image and a right image is defined by Equation 2.

$$P_L = HP_R \quad (2)$$

Here P_L is a left image and P_R is a right image that contains corresponding points between a left image and a right image. Then H is calculated from the Equation 3 and 4 using the 8-point algorithm.

$$\begin{pmatrix} x_1 & y_1 & 1 & 0 & 0 & 0 & -x_1'x_1 & -x_1'y_1 & -x_1' \\ 0 & 0 & 0 & x_1 & y_1 & 1 & -y_1'x_1 & -y_1'y_1 & -y_1' \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ x_n & y_n & 1 & 0 & 0 & 0 & -x_n'x_n & -x_n'y_n & -x_n' \\ 0 & 0 & 0 & x_n & y_n & 1 & -y_n'x_n & -y_n'y_n & -y_n' \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ h_3 \\ h_4 \\ h_5 \\ h_6 \\ h_7 \\ h_8 \\ 1 \end{pmatrix} = 0 \quad (3)$$

$$H = \begin{pmatrix} h_1 & h_2 & h_3 \\ h_4 & h_5 & h_6 \\ h_7 & h_8 & 1 \end{pmatrix} \quad (4)$$

Here x', y' denotes the pixel coordinates of the left image and x, y denoting the pixel coordinates of the right image.

E. Determination of objects

The object determination is done by merging a left and a right image. If the difference in pixel intensity between the left

images with the right image is equal or less than the given threshold, then it is considered a background or 2D object and marked in white. Otherwise then pixels are considered part of 3D objects and are marked in black. In this research, the threshold value used is 30. The object determination is done using Equation 5.

$$P_{(x,y)m} = \begin{cases} P_{(x,y)} - P_{(x',y')} \leq 30 & P_{(x,y)m} = white \\ P_{(x,y)} - P_{(x',y')} > 30 & P_{(x,y)m} = black \end{cases} \quad (5)$$

Fig. 4 shows the pixel determination results of the object in the image with a spherical object with a distance between 10cm camera. The black pixel represents the portion of the detected object.

The closing operation is done in order for the object is more visible so that more easily detected. The detected object from the closing operation result marked the bounding box form. Fig. 5 shows the result of marking the location of the object after the closing operation.

III. Experimental Result

In this research used 9 pairs of video scenes that are grouped by object observed and the distance between cameras. There are 3 forms of objects will be observed, namely balls, tubes, and boxes. Meanwhile, based on the distance between the cameras, the image was taken from the distance between the cameras 10, 20, and 50cm respectively. Image data capture is done indoors with video frame rate are 30 fps and image frame size is 640 x 480 pixels with 10 seconds of video duration. In this research used a static camera (not moving). The experimental environment is as follows: The operating system is Windows 8, the processor is Intel® core™ i3, 2GB RAM, and the used software is Ms Visual Studio 2010 and OpenCv 2.4.10.

The video capture results for each object are extracted become sequences of images with a BMP format. Table I shows the number of frames for each video after being extracted. There is a difference in the number of frames for each video because of the inaccuracy of the video duration for each video.

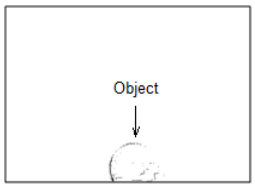


Fig. 4. The pixels determination result of the object.

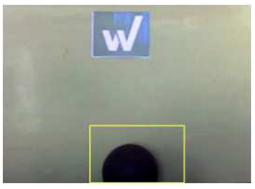


Fig. 5. Location of the 3D objects.

Not all video frames are used for the object detection. Each frame has an almost identical image so is used in part to represent it. The number of frames used for each object is not the same. This number represents the number of frames of the result of frame merging between the right and the left image. Table II shows the amount of each frame is used for the detection of each object. The effectiveness of the proposed method is measured using the recall and precision parameters. Recall and precision are parameters that are widely used in the measurement performance of the detection system using image processing. The equations used to measure the value of recall and precision are as follows:

$$Recall : \frac{N_{TP}}{N_{TP}+N_{FN}} x100\% \quad (6)$$

$$Precision : \frac{N_{TP}}{N_{TP}+N_{FP}} x100\% \quad (7)$$

Here N_{TP} is the number of pixels in the true positive area; N_{FP} is the number of pixels in the false positive area; N_{FN} is the number of pixels in the false negative area. True positive (TP) is part of the overlapping region between ground truth and detection results. False negative (*FN*) is the part included in the ground truth but is not included in the detection results of the proposed method. False positive (*FP*) is the part that is included in the detection result of the proposed method, but not the actual object region.

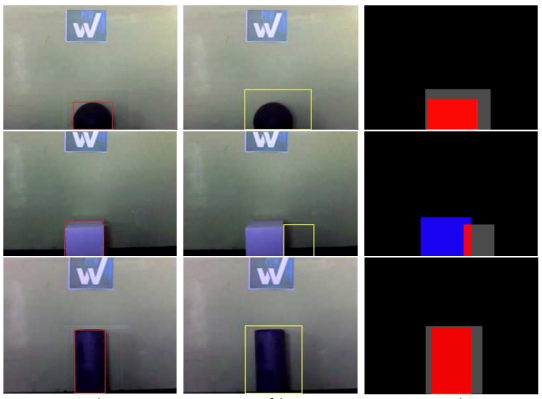
Fig. 6 shows the resultant images of comparison between the ground truth and the detection result of the proposed method. Fig. 6 is an example of the detection results of the proposed method for a distance between 2 cameras is 10cm. The true positive region is marked in red, a false negative is marked in blue, and a false positive is marked in gray. The results of the detection of the proposed method are shown in Table III. This result is the average of the number of frames for each object as shown in Table II. The recall value of the ball and tube objects gives a value above 50%. As for the box object, the value of recall and precision is less than 50%. The distance between 2 cameras affects the success of the proposed detection method. On the object of the ball and tube, the farther the distance between the 2 cameras will give the value of recall and precision are low. On the box objects, the recall value is higher for the distance between the 2 cameras is 50cm.

TABLE I. THE NUMBER OF EXTRACTION OF AN IMAGE FRAME FOR EACH CAMERA.

Videos	Right camera (Image)			Left camera (Image)		
	10cm	20cm	50cm	10cm	20cm	50cm
Ball	320	324	314	319	323	314
Box	319	319	314	319	319	314
Tube	313	313	323	313	312	323

TABLE II. THE NUMBER OF EACH FRAME USED IN THE RESEARCH.

Video	The distance between 2 camera		
	10cm	20cm	50cm
Ball	101	113	100
Box	100	100	100
Tube	100	100	100



a) b) c)

Fig. 6. The result of the 3D objects detection using the proposed method. Time elapses from left to right: (a) Ground truth, (b) The location of the 3D object from the proposed method result, (c) the result of the evaluation.

TABLE III. THE EVALUATION RESULTS OF THE EFFECTIVENESS OF THE PROPOSED METHOD.

Video		Evaluation values	
		Recalll	**Precision**
Ball	10cm	95%	52%
	20cm	95%	57%
	50cm	75%	17%
Box	10cm	25%	23%
	20cm	32%	6%
	50cm	60%	12%
Tube	10cm	95%	60%
	20cm	95%	57%
	50cm	60%	12%

IV. CONCLUSION

This paper proposes a technique for detecting 3D objects using a method of a Harris corner detector and Lucas-Kanade tracker from two images taken using two cameras. Recall and precision parameters are used to determine the effectiveness of this proposed method. Table 3 shows that the evaluation results in 3D object detection of three objects give the recall result more than 60% for a ball and tube objects, while for box objects under 60%. This means that the proposed method provides satisfactory performance for detection a ball and tube objects while for the box object is not satisfactory. The precision values above 50% are achieved for the distance between cameras 10cm and 20cm on ball and tube objects, while for box objects less than 50%.

In the object of the box, the precision value is very low. This is because when converted to grayscale mode, the color of the object becomes similar to the background so it is difficult to do the detection. The challenge in the future is to use an RGB image without converting to grayscale so that the object's color can be more different from the background. This will cause the computation load to become heavier and the processing time of each frame will be longer.

The other challenge is trying to apply this proposed method for the outdoor image. In the outdoor image, many disorders encountered in addition to changes in the intensity, such as the background movement due to the wind or because of the number of 3D objects that exist. In addition, the proposed method can be tried on a microcomputer so that it can be applied to the automatic control.

ACKNOWLEDGMENT

This work was supported by DIPA Research Grant from Department of Electrical Engineering, Faculty of Engineering, Lampung University.

REFERENCES

[1] F. X. A. Setyawan, J. K. Tan, H. Kim, S. Ishikawa, "Detecting foreground objects by sequential background inference in a video captured by a moving camera," Proceedings of the SICE Annual Conference, Nagoya - Japan, 2013, pp. 1699-1702.

[2] F. X. A. Setyawan, J. K. Tan, H. Kim, S. Ishikawa, "Detecting moving objects from a video taken by a moving camera using sequential inference of background images," Artif life Robotics, Vol. 19, 2014, pp. 291-298, doi:10.1007/s10015-014-0168-7.

[3] F. X. A. Setyawan, J. K. Tan, H. Kim, S. Ishikawa, "Moving objects detection employing iterative update of the background," Artif life Robotics, Vol. 22, 2017, pp. 168-174, doi:10.1007/s10015-016-0347-9.

[4] Rachmawati, R. Hidayat, S. Wibirama, "Rekonstruksi Obyek Tiga Dimensi Dari Citra Dua Dimensi Menggunakan Epipolar Geometry," Jurnal Teknologi, Vol. 5 No. 2, 2012, pp. 98-103.

[5] A. Khaled, M. Elmogy, S. Barakat, "3D Object Recognition Based on Image Features: A Survey," International Journal of Computer and Information Technology, Vol. 03, Issue 03, 2014, pp. 651-660.

[6] Y. Igarashi and K. Fukui, "3D Object Recognition Based on Canonical Angles between Shape Subspaces," Asian Conference on Computer Vision, 2011, pp. 580-591, doi: 10.1007/978-3-642-19282-1_46.

[7] L. A, Alexandre, "3D Object Recognition using Convolutional Neural Networks with Transfer Learning between input Channels," (eds) Intelligent Autonomous Systems 13. Advances in Intelligent Systems and Computing, vol 302. Springer, 2016.

[8] M. Y. Mashor, M.K. Osman, M. R. Arshad,"3D Object Recognition Using Multiple Views and Neural Networks," Proceedings of International Conference on Man-Machine Systems, 2006.

[9] R. Holonec, R. Copindean, F. Dragan, V. D.Zaharia, "Object Tracking System Using Stereo Vision and LabView Algorithm", Acta Electrotehnica, Vol. 55, No. 1-2, 2014, pp. 71-76.

[10] C.Harris and M. Stephens, "A combined edge and corner detector," Proc. 4th Alvey Vision Conf., 1988, pp. 147-151.

[11] B. D. Lucas and T. Kanade, "An iterative image registration technique with an application to stereo vision," Proc. 7th Int. Joint Conf. on Artificial Intelligence, 1981, pp. 647-679.

[12] M. A. Fischler, R. C. Bolles, "Random sample consensus: A paradigm for model fitting with applications to image analysis and automated cartography", Communication of ACM, Vol. 24, issue 6, 1981, pp. 381-395.