International Journal of Latest Engineering Research and Applications (IJLERA) ISSN: 2455-7137

Volume – 07, Issue – 11, November 2022, PP – 26-34

Real-time Image Segmentation and Multi Object Detection Based on Stereo-vision

Jotje Rantung, Romels Cresano Apelles Lumintang, Yan Tondok

Mechanical Engineering Department, Faculty of Engineering, Sam Ratulangi University, Manado, Indonesia

Abstract : Image segmentation and stereo vision are crucial to many video processing applications, including object detection and recognition. The main problems of an image captured by the video camera results in excessive information, complex disparities, the significant change of the shape and appearance of regions. Therefore, a method to extract the foreground object image is needed. A real-time image segmentation method must be proposed to solve this problem.To do this work the followings are done. Firstly, the stereo-vision is calibrated to correct its intrinsic parameters and distortion parameters rightly. Secondly, real-time image segmentation is done by combining HSV color space, threshold, mathematical morphological transformation, and contour detection techniques to extract objects in a graph-based image. Thirdly, multi object detection of various common objects the object and fish object. Experiment results showed that real-time image segmentation of the proposed method had a good result. execution times of image segmentation and feature extraction in the real-time method were 23 milliseconds and 0.0019 millisecond, respectively.

Keywords: Real-time, image segmentation, stereo-vision, object detection

I. INTRODUCTION

Image segmentation and stereo-vision plays an important role in application of video processing such as object detection and recognition. To recognize an object, firstly, the image must be segmented into the object background and the object to be recognized. Because the results of image segmentation are often not very appropriate for some visual tasks, the shape and appearance of the area change significantly for landscape images obtained from different viewpoints or different lighting conditions. Therefore, matching a segmented region in such an image is indispensable. There are a lot of researches in relation to image segmentation and object handling task with camera as a vision system. Implementation of real-time vision for tracking task and robot guidance has been introduced [1]. This method was a basic theoretical approach and simply described a typical architecture of 3D scene reconstruction. S. Helmer et al. [2] proposed object recognition using stereo-vision, where a model utilized a chamfer-type silhouette classifier. However, it was easy to lose stereo depth information. A method using segmentation-based stereo-vision for 3D object recognition was proposed [3]. In this work, segmentation-based stereo-vision was employed for 3D sensing, and matched with computer-aided design (CAD). This method is not suitable for real-time processing due to the computational complexity to combine between pre-processing with CAD matching.

The quality of the segmented image always depends on color, distance, plan information, and light. Nevertheless, the user did not stop finding the way to get good image segmentation methods. A method for image segmentation using mathematical morphology was proposed [4]. This method used approach based on the watershed transformation. However, this method was applied to the static image. An image segmentation algorithm based on threshold segmentation was proposed [5]. In this work, the segmentation algorithm assumed that each pixel in the image has its own threshold. However, this method was applied to the offline method, and threshold-based segmentation is difficult to be performed if the object has low contrast with its background. Automatic threshold selection for image segmentation based on genetic algorithm was proposed [6]. This paper focused on the issue of automatic selection for multi-level threshold for image segmentation. However, this method in implementation was applied to the offline method.

Image segmentation is done before estimating the actual features of object recognition. The main problems of an image captured by the video camera results in excessive information, complex disparities, the significant change of the shape and appearance of regions. Therefore, a method to extract the foreground object image is needed. A real-time image segmentation method must be proposed to solve this problem.

II. PROPOSEDMETHOD

This sectiondescribes the proposed real-time image segmentationand multi object detection method based on stereo-vision. To do this work the followings are done. Firstly, the stereo-vision is calibrated to correct its

intrinsic parameters and distortion parameters rightly. Secondly, real-time image segmentation is done by combining HSV color space, threshold, mathematical morphological transformation, and contour detection techniques to extract objects in a graph-based image. Thirdly, multi object detection of various common objects the object and fish object.

3.1 Camera calibration

Camera calibration is an important step in 3D computer vision to extract metric information from 2D images. Focal length is the important parameter in a measurement algorithm using stereo-vision. This parameter is obtained through the stereo-vision calibration. This parameter determines the strength or weakness of a camera lens to focus on the object by distorted images. There are four intrinsic parameters:f_x and f_y as thefocal lengths of the camera in terms of pixel dimensions in the x and y direction, and (u_0,v_0) is the principal point. The camera usually represents lens distortion that is a radial distortion given by

$$\mathbf{u}_d = \begin{cases} u_d = (u_u - u_0)(1 + k_1 r_u^2 + k_1 r_u^4) \\ v_d = (v_u - v_0)(1 + k_1 r_u^2 + k_1 r_u^4) \end{cases} \quad (1)$$

where $p(u_u,v_u)$ and $p(u_d,v_d)$ are distortion-free and distortion-normalized image coordinates, respectively. k_1 and k_2 are the radial distortion coefficients, and $r_u^2 = u_u^2 + v_u^2$ [7,8]. According to the camera model, focal length is given by

$$f = \frac{1}{2}(f_x / m_x + f_y / m_y) \quad (2)$$

where f_x and f_y are defined as the focal length of the camera in terms of pixels dimensions in the x and y direction, respectively.

The important task of the camera calibration is to determine the four intrinsic parameters and the two distortion coefficients. The camera calibration follows the procedure proposed in [9]. The recommended calibration procedure as follows:

1. Setup camera, print a pattern and attach it to a planar surface.
2. Take a few images of the model plane under different orientations by moving either the plane or the camera.
3. Detect the feature points in the images.
4. Estimate the intrinsic and extrinsic parameters of the camera.

The camera system used in this dissertation is a single-head stereo-vision put on the table as shown in Fig. 1(a). The camera is connected to personal computer using USB cable. The personal computer is used to process acquired images. The frames obtained from the camera system are captured and stored on the computer system using open source computer vision (OpenCV C++) [10]. The calibration object used in the computation is a planar checkerboard pattern. The checkerboard pattern is obtained by a laser printer paper. The use of the checkerboard pattern allows sub-pixel corner detection of checkerboard corners. The calibration pattern is shown in Fig. 1(b). Each square box over the calibration pattern has dimensions 25mm × 25mm. Fig. 2 shows a set of pairs of close-up RGB images of the checkerboard placed in different positions and orientations. The RGB images are corrected by taking a variety of focus images that can cover the entire picture.

(a) (b)

Fig. 1Camera calibration setup (a), and image pattern for calibrating intrinsic parameter (b)

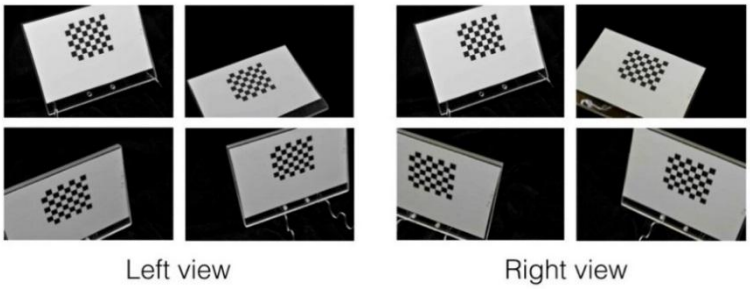

Fig. 2Set of pairs of close-up RGB images of the checkerboard

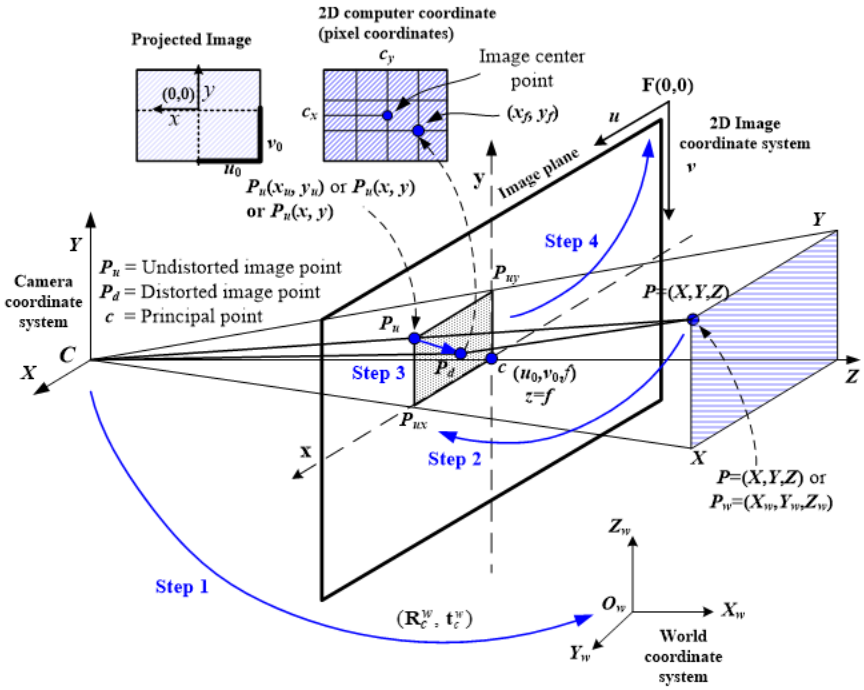

Fig. 3 Steps of camera calibration.

Fig. 3 shows the four stepsof camera calibration.The four steps to convert a point from world coordinate to the computer memory image coordinate as follows [11]:

Step 1 (Transformation from world to camera): Transformation from a point Let $P_w=(X_w,Y_w,Z_w)$ in 3D world space to the point $P=(X,Y,Z)$ in the 3D camera frame using homogeneous transformation is represented as follows:

$$\mathbf{x} = \begin{bmatrix} x_k \\ y_k \\ z_k \end{bmatrix} = \begin{bmatrix} r_1 & r_2 & r_3 \\ r_4 & r_5 & r_6 \\ r_7 & r_8 & r_9 \end{bmatrix} \begin{bmatrix} x_w \\ y_w \\ z_w \end{bmatrix} + \begin{bmatrix} t_x \\ t_y \\ t_z \end{bmatrix} = \mathbf{R}\mathbf{x}_w + \mathbf{t} \tag{3}$$

where **R** is 3x3 rotation matrix, and **t** is the 3x1 translation vector.

Step 2 (Projection): Transformation from 3D coordinate to the undistortion image coordinate (x_u, y_u) using perspective projection of pinhole camera geometry and focal length*f*is represented as

$$\mathbf{x}_u = \begin{bmatrix} x_u \\ y_u \end{bmatrix} = \begin{bmatrix} f\dfrac{X}{Z} \\ f\dfrac{Y}{Z} \end{bmatrix} \tag{4}$$

Step 3 (Lens distortion): Transformation from undistorted image coordinate to distorted coordinate (x_d,y_d) with distorted function f_d and lens distortion coefficients k_i, for i=1~3, is represented as

$$\mathbf{x}_d = \begin{bmatrix} x_d \\ y_d \end{bmatrix} = f_d \mathbf{x}_u = \left(1 + k_1 r^2 + k_2 r^4 + k_3 r^6\right)\mathbf{x}_u \tag{5}$$

where $r^2 = \sqrt{x_u^2 + y_u^2}$ .There are two kinds of distortions: radial and tangential. However only radial distortion k_i is to be considered as shown in Eq. (5).

Step 4 (Camera to image): Transformation from distorted image coordinate to (x_d,y_d) to computer image coordinate (x_f,y_f) is presented as

$$\mathbf{x}_f = \begin{bmatrix} x_f \\ y_f \end{bmatrix} = \begin{bmatrix} f_x & 0 \\ 0 & f_y \end{bmatrix} \mathbf{x}_d + \begin{bmatrix} c_x \\ c_y \end{bmatrix} \tag{6}$$

$$f_x = s_x d_x^{'-1},\ f_y = d_y^{-1},\ \mathrm{d}_x^{'} = d_x \frac{N_{cx}}{N_{fx}} \tag{7}$$

where s_x is an uncertainty scale factor, d_x is the center to center distance between adjacent sensor elements in X direction, d_y is the center to center distance between adjacent sensor elements in Y direction, N_{cx} is number of sensor elements in X direction, and N_{fx} is number of pixels in one scan line of the image captured by computer.

3.2 Proposed real-time image segmentation method

In this section, an image segmentation method in real-time is proposed. The segmentation process is done in real-time on the video frame by using stereo-vision. To reduce complexity and computation time, hue and value feature spaces are segmented separately before they are combined. Stereo-vision is applied to capture the images of the fish. In this work, a contour based segmentation and mathematical morphological method are used for real-time segmentation. Image segmentation is done by combining HSV color space, threshold, mathematical morphological transformation, and contour detection techniques to extract objects in a graph-based image. Software design is implemented by using C++ programming language. For implementing a real-time image segmentation, a library available in open source computer vision (OpenCV, C++) is used Library functions are used for loading image, creating windows to hold an image in real-time, and saving image.

Algorithm 1 explains the real-time image segmentation used for fish surface area and volume calculations.

Algorithm 1. Algorithm to perform real-time image segmentation used for fish surface area and volume calculations

1. Creates window user interface on OpenCV image processing tool to adjust the value of parameters in real-time.
2. Get the centroid of the object by creating a two dimensional OpenCV image processing tool point as follows:

   pt = *cv*Point (*x*,*y*) (1)

   (Obtaining XY coordinates from (x,y) from *cv*Point)
3. Segment color images using features color of images extracted from the HSV space. To convert from RGB to HSV use standard transformation, (assuming normalized RGB values) first find the maximum and minimum values from the RGB triplet.

4. Set a saturation level. Based on human vision perception, the intensity domination is at saturation level th_{sat} below 0.2. Color is defined as black at $th_{sat}(v)$=1 for white=0. v=value(V)

$$th_{sat}(v) = 1 - \frac{0.8 \cdot v}{255} \tag{2}$$

The saturation S is defined as

$$S_i = \begin{cases} (G\max - G\min)/G\max, & \text{for } G\max \neq 0 \\ 0, & \text{otherwise} \end{cases} \tag{10}$$

If saturation S is 0 (zero), hue is undefined (i.e. the color has no hue, therefore, it is monochrome)

5. Create a structuring element. The structuring element B is a 3×3 square, that is, B={(-1,-1), (-1,0),(-1,1), (0,-1), (0,0), (0,1), (1,-1), (1,0),(1,1)}.
6. Select the centroid based on its magnitude. The centroid is calculated as follows:

$$x_c = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} x \times g(x,y)}{\sum_{x=1}^{N}\sum_{y=1}^{N} g(x,y)}; \quad y_c = \frac{\sum_{x=1}^{N}\sum_{y=1}^{N} y \times g(x,y)}{\sum_{x=1}^{N}\sum_{y=1}^{N} g(x,y)} \tag{11}$$

where x_c and y_c are the centroid of a fish, N is the image length in the pixel, and $g(x, y)$ is grey level.

7. Apply boundary region by erosion and dilation.
8. Repeat step 3 to 6 until a point (x, y) converge to object image centroid.
9. Output a binary image.

3.3 Scene setup

Scene settings play an important role in real-time image segmentation. Scene settings must be explained in advance to evaluate the functionality of the algorithm and its components. Four different objects are tested before being done on fish objects to test the ability of the proposed real-time image segmentation method as follows:

1. Different simple objects to test and demonstrate the basic functionality.
2. Different size of the objects to test the upper and lower bounds of objects which can be detected.
3. An object with different shapes like sharp-edged, elongated and hairy objects to test the limits of the fixation-based approach.
4. Textured objects to test the ability to handle multi-colored objects.

III. EXPERIMENTAL RESULTS

3.1 Calibration intrinsic parameter of camera

Camera intrinsic parameters are calibrated at working distance about 1000 mm. Chessboard image is used as calibration pattern, and 19″ LCD monitor is used to display image pattern. Fig. 4 shows an image pattern for calibrating intrinsic parameters. Table 1 shows the calibration results of the left camera and right camera, respectively.

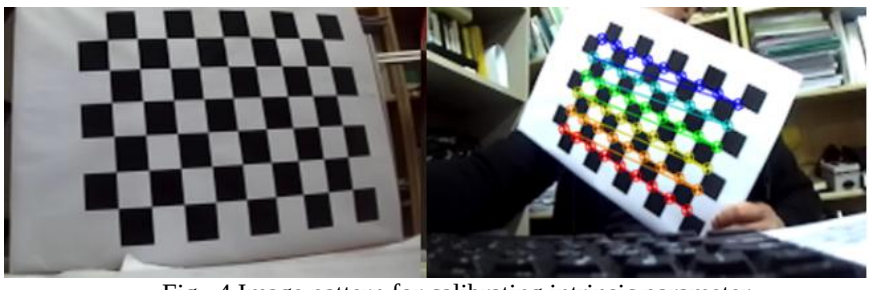

Fig. 4 Image pattern for calibrating intrinsic parameter.

Table 1 Calibration results of the intrinsic camera parameter.

Camera	Focal length (pixel)		Principal point (pixel)	
	f_x	f_y	u_0	v_0
Left	940.49	769.00	674.65	180.35
Right	941.49	769.50	675.65	181.25

Camera	Distortioncoefficients		pixel/mm		Focal length (mm)
	k_1	k_2	*Pux*	*Puy*	
Left	-98.91	11.34	674	180	2.8289
Right	-99.01	11.84	675	181	2.8311

3.2 Real-time image segmentation results of various common objects

The segmentation of simple objects is the basis for testing the capability of the algorithm. If it fails at this stage, it will probably also fail for more complex objects. The simple object, in this case, is defined as objects with constant color. Their shape is convex without having sharp edges. A segmentation of simple objects can be shown in Fig. 5. Two simple objects with different sizes are all segmented very well. The location of the contours is optimal. On the other hand, the quality of the disparity map gets worse for relatively small objects as there is less depth information available. Only the contour on the cup shifts slightly from the edge of the object because the color difference between the object and background is almost the same.

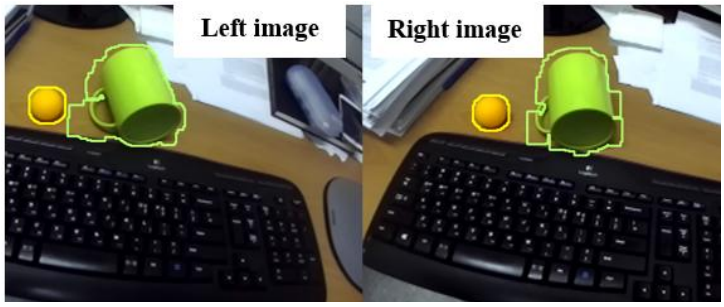

Fig. 5Segmentation of simple objects

The size of an object can affect the segmentation process. Therefore, two different sizes in the scene are evaluated. The scene contains the small size object and the large size object. Fig. 6 shows the result of the segmentation of the small object and the large object. The results of the segmentations produce for both the large object and the small object are very good. Especially, the segmentation of very small objects outperforms the expectations.

Fig. 6Segmentation of the small object and the large object

The shape of the object influences the segmentation process and can also have an impact on the results of segmentation. Fig. 7 shows the segmentation results of several objects with different shape of objects. The shape includes objects with sharp-edge, elongated object, and thin object. The experimental results in the segmentation of the edge detection for the large size are more difficult than those for the small size. When the large size objects are transformed to log-polar space, the displayed objects in the scene are deformed to long horizontal objects as they are bent around the fixation point. Fixation points are the view focuses on various salient locations. However, the graph cut algorithm tries to find a short vertical cut to minimize energy [69]. Therefore, long objects and objects with sharp edges tend to get cut off. The graph cut algorithm suffers from the same problem, but it can counteract this behavior by using the color information.

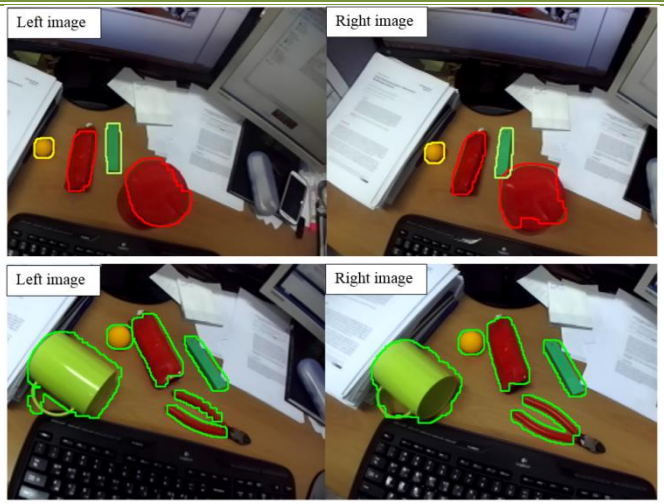

Fig. 7 Segmentation of the different shapes of objects

3.3 Real-time image segmentation results for fish object

This section describes the real-time segmentation process of the fish object. The surface structure of fish skin is categorized as a textured object. Object segmentation becomes a problem when the object surfaces have complex texture patterns and generate many local edges within the same region [70]. Textured objects are the most difficult objects to segment, where the computation of the texture gradient is neglected to increase the performance of the edge detection. The textured gradient is compensated by the large kernel filter in the edge detection and the color models in the graph cut algorithm.

In the segmentation process, if a segmentation result has low accuracy, the segmentation is considered as a failure. For segmentation of fish objects in real time, several trials were carried out to obtain satisfactory results. Fig. 8 shows the thresholded image that needed to be inverted from black color to white color after segmentation. As shown in Fig. 8(a), in this image segmentation process, the thresholded image is overlaid. The failure in segmentation causes failure in fish detection as shown in Fig. 8(b). In this trial, the segmentation of the fish object is done by choosing a high threshold (Th=255). Fig. 9 shows the image segmentation result of the fish object by choosing a medium threshold (Th=126). Fig. 9(a) shows that an image is not be segmented properly or an image get the wrong polarity after segmentation. The improper segmentation also causes failure in fish contour detection as shown in Fig. 9(b). Another case is needed to overcome this obstacle: i.e., image segmentation using the bounding box [12].

Fig. 10 shows the segmentation results of the fish object with the lowest threshold (Th = 0) and the bounding box. Fig. 10(a) shows that the best result is gained in this trial. Since it has high accuracy, the feature segmentation results obtained in this trial will be used on the next step. Next trial is using feature segmentation results obtained in Fig. 10. By applying a graph cut algorithm, the results in Fig. 11 are obtained. Finally, image segmentation in the *bw* image of the object is obtained as shown in Fig. 11(a).

The segmentation results of the proposed method were compared with the segmentation methods in [13] and [14] as shown in Fig. 12. Fig. 12(c) showed that the proposed method obtains good segmentation result compared to the method by the watershed method and the method by the threshold value of Fig. 12(a) and Fig. 12(b). As can be seen in Fig. 12(a) and Fig. 12(b), the results of segmentation are not perfectly segmented.

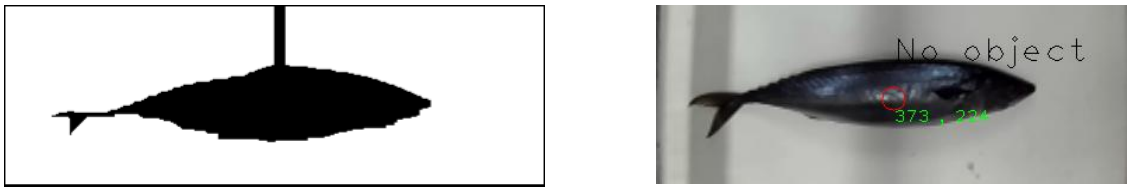

Fig. 8 Segmentation of the fish object with *Th* = 255

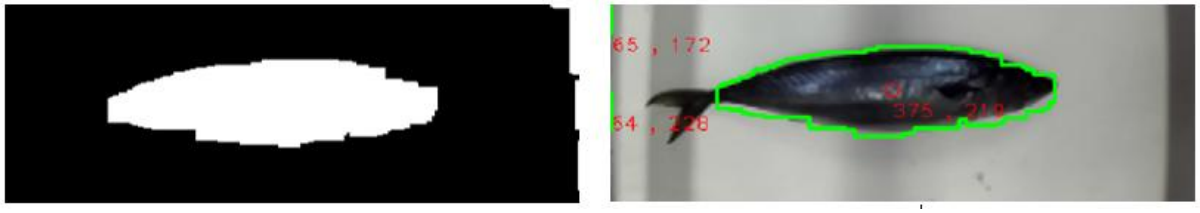

Fig. 9Segmentation of the fish object with *Th* = 126

Fig. 10Segmentation of the fish object with *Th* = 0 and bounding box

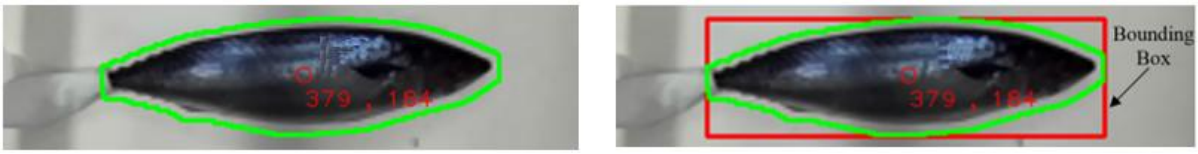

Fig. 11Segmentation of the fish object with *Th* = 0, bounding box, and graph cut

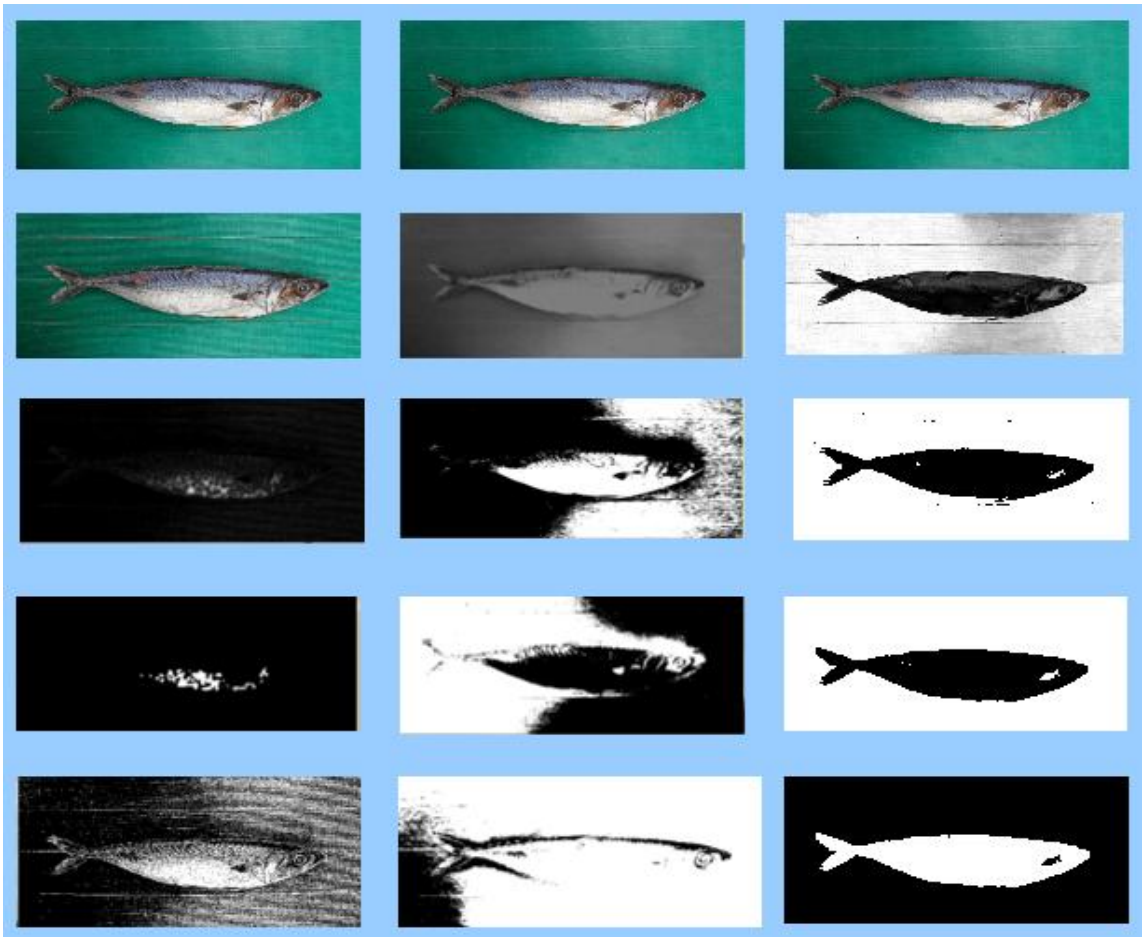

Fig. 12 Image segmentation results by watershed transformation, threshold value, and the proposed method

IV. CONCLUSION

The proposed real-time image segmentation method could solve the problems of an image captured by a stereo-vision such as excessive information, complex disparities, and the significant change of the shape and appearance of regions. The real-time image segmentation method could increase the capacity detection, accuracy and execution time of multi object. For real-time applications of image segmentation and feature-extraction, open source computer vision (OpenCV), C++, and stereo-vision were used.

The stereo-vision was calibrated to collect its intrinsic parameters and distortion parameters. Real-time image segmentation and feature extraction were done to extract objects in a graph-based image. These were done by combining HSV color space, threshold, mathematical morphological transformation, and contour detection techniques. Finally, the object coordinate, surface area and volume of the fish were obtained.

Experiment results showed that real-time image segmentation of the proposed method had a good result. The experiments demonstrated that the calibration process could quickly detect the chessboard corners. After the calibration results, focal length of the left camera and right camera were about 2.8289 mm and 2.8311 mm, respectively. The focal length difference between two cameras was about 0.00221 mm.The execution times of image segmentation and feature extraction in the offline method and the real-time method were 23 milliseconds and 0.0019 millisecond, respectively.

The real-time image segmentation method proposed in this work could be useful for fish recognition and sorting applications. The multi object detection based stereo-vision could be applied to an automated fish processing system to handle fish on a conveyor belt.

Acknowledgements

This research was funded by PNBP BLU 2022, Sam Ratulangi University, Manado, Indonesia.

REFERENCES

[1] P. I. Corke and S. A. Hutchinson, Real-time Vision, Tracking and Control,*Proceedings of the International Conference on Robotics and Automation*, 2000, 622–629.

[2] S. Helmer and D. Lowe, Using Stereo for Object Recognition, *Proceedings of the International Conference on Robotics and Automation*, 2010, 3121–3127.

[3] Y. Sumi, Y. Kawai, T. Yoshimi, and Tomita S, 3D Object Recognition in Cluttered Environments by Segment-Based Stereo Camera,*International Journal of Computer Vision, Vol. 46(1)*,2002, 5–23.

[4] L. J. Belaid and W. Mouru, Image Segmentation: a Watershed Transformation Algorithm,*International Society for Stereology & Image Analysis*, Tunis, Tunisia, 2009, 93–102.

[5] S. Zhu, X. Xia, Q. Zhang, and K. Belloulata, An Image Segmentation Algorithm in Image Processing Based on Threshold Segmentation, *Proceedings of the Third International IEEE Conference on Signal-Image Technologies and Internet Based System*, Shanghai, China, 2007, 673–678.

[6] B. R. Lee, Q.B. Truong, V. H. Pham, and H. S Kim, Automatic Thresholding Selection for Image Segmentation Based on Genetic Algorithm,*Journal of Institute of Control, Robotics and Systems (in Korean), 17(6), 2011*, 587–595.

[7] Z. Zhang, A Flexible New Technique for Camera Calibration, *IEEE Transactions on Pattern Analysis and Machine Intelligence. 22(11)*, 2000, 1330–1334.

[8] https://sourishghosh.com/2016/stereo-calibration-cpp-opencv/

[9] R. Cucchiara, C. Grana, A. Prati, and R. Vezzani, A Hough Transform-Based Method for Radial Lens Distortion Correction, *Proc. Int. IEEE Conf. Image Analysis Processing*, 2003, 182–187.

[10] F. Yi and I. Moon, Image Segmentation: a Survey of Graph-Cut Methods, *International Conference on Systems and Informatics*, 2012,1936 –1941,2012.

[11] X. Huang, Z. Qian, R. Huang, and D. Metaxas, Deformable-Model Based Textured Object Segmentation,*Proceedings of International conference on Energy Minimization Methods in Computer Vision and Pattern Recognition*, 2005, 119–135.

[12] V. Lempitsky, P. Kohli, C. Rother, and T. Sharp, Image Segmentation with a Bounding Box Prior, *IEEE 12th International Conference on Computer Vision*, 2009, 277–284.

[13] R. J. Hughes, Estimation of Shell Surface Area from Measurements of Length, Breadth, and Weight of Hen Eggs, *Poultry Science,63(12)*, 1984, 2471–2474.

[14] N. N. Mohsenin, Chapter 3: Physical Characteristics in Physical Properties of Plant and Animal Materials, *Gordon and Breach Science Publishers*: New York, NY, 1970, 51–87.

