ISSN 2229-5518

Hand Segmentation from Depth Image using Anthropometric Approach in Natural Interface Development

Rayi Yanu Tara, Paulus Insap Santosa, Teguh Bharata Adji

Abstract—Hand gestures are often used as natural interface between human and robot. To acquire hand gesture from a captured image, hand segmentation procedure is performed. In this manuscript, a method for hand image segmentation from depth image is proposed. This method uses image thresholding technique to obtain human image part from a depth image. The threshold level is obtained by analyzing human posture dimension (anthropometry). By finding the centroid of the human image, left and right regions of human body can be separated. Assuming that each region has a hand image and the hand is positioned in front of its body, both hand images can be located. Hand segmentation is started by using an anthropometric data of hand pose. The data is used to compute the color values that represent the depth of hand in each image region. Thus, the acquired values are used as threshold for each image region. The thresholding operation resulted in completely segmented hand images. This proposed method has low computation time and works well when the basic assumptions are fulfilled.

Index Terms—anthropometric, depth image, gesture recognition, hand segmentation, natural interface, threshold.

1 Introduction

THE application of gesture recognition for human robot interaction has grown rapidly, especially for hand gesture. Human gesture interface will give human, as an operator, a natural way to interact with a robot [1]. The use of sophisticated joystick and control panel can be replaced with a gesture based interface system. Employing gesture interface will significantly reduce the needs of specific skill and training for the human operator to familiarize with the controller (i.e. joystick). Many studies have been conducted to build a natural interface using hand gestures between human and robot that operate in tele-operation mode. Hand gestures can be acquired using special made sensor gloves [2] or a camera [3], [4]. Lately, the usage of depth imager in building natural interface is more popular since the presence of Microsoft Kinect [5], a low-cost sensor packages that contains a depth camera, color camera, and microphone array.

A method for fingertip detection and center of palm detection distinctly for both hands using depth image was presented in [6]. The hands were segmented using depth vector and center of palms were detected using distance transformation of inverse image. This method was used as an input to a robot hand that emulates human hands operation. Using this method, a user needs to manually define the hand location before the segmentation is started which becomes the disadvantage of this system. Another research was conducted to do hand pose recognition using low-resolution depth images in real time [7]. This method take the user right hand into account since the right hand is the closest object in the depth image. An initialization must be carried out when this method is used for different users. Hand image is segmented by applying threshold to the depth image. Both of these methods are not automatically locating the hand locations in the image before performing the segmentation process.

In this paper, instead of manually choosing the hand location in the depth image before segmentation procedure is performed, an automatic detection of hand location is proposed. Through segmentation of human body from an image using anthropometric analysis, the centroid of human body can be calculated. Thus, left and right body regions can be separated. By assuming that each region has a hand image, the exact hand location for each region can be inferred by employing hand posture analysis. Therefore, the hand images are segmented from the original depth image by applying threshold operation with pre-computed values, which is related to each hand regions. Using this method, hand segmentation from depth image can be performed without any information from the user about the exact location of their hands. All of the illustrations and experiments used in this manuscript are built using ImageJ [8], Harpia [9], and OpenCV library [10]. This proposed method will be clearly explained in Section 2.

2 Proposed Hand Segmentation Method

An overview of the proposed method is shown in Figure 1. The proposed method works with some assumption while performing a gestural interaction: a) the human body in a depth image is the closest object to the depth imager, b) the human hands is positioned in front of the body, and c) there is no any object around the human that might occlude the human body. All of those procedures are executed sequentially for each captured frame from a depth imager.

- *Rayi Yanu Tara is currently pursuing master degree in EE & IT Department at Gadjah Mada University, Indonesia. E-mail: rayi.yanu.tara@ieee.org*
- *Paulus Insap Santosa, Ph.D is an Assistant Professor in EE & IT Department at Gadjah Mada University, Indonesia. E-mail: insap@mti.ugm.ac.id*
- *Teguh Bharata Adji, Ph.D is a Assistant Professor in EE & IT Department at Gadjah Mada University, Indonesia. E-mail: adji@mti.ugm.ac.id*

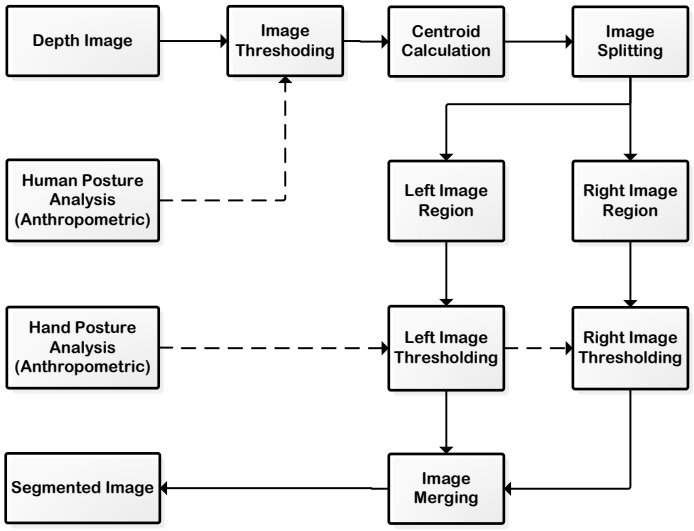

Fig. 1. Overview of the proposed method.

Mostly, those procedures can be grouped into this following step.

2.1 Depth Image Acquisiton

Microsoft Kinect [5] is used as depth imager in this research. The imager has 800 - 3500 mm operating range, producing depth image in VGA (640 x 480) resolution, and capable to capture a depth image up to 30 fps. Each pixel in the depth image is a measured distance in millimeter scale. The depth image needs to be converted into a grayscale image to visualize the data as a visible image. Conversion of the depth image into an 8-bit grayscale image is done using this following equation.

$$P_N = 255 \times \left(\frac{D_N - 800}{3500 - 800} \right), \qquad (1)$$

where D is the depth data, P is the new pixel data, and N is the pixel number.

Some pixels in the depth image captured by Kinect are not in the operating range value (classified as unknown pixel) because of the occlusion effect. To overcome this problem, some approach has been performed to fill the unknown pixel value by applying median filter [11], [12]. In this research, since human body is the closest object to the imager, any occlusion effect can be ignored. Therefore, the entire unknown pixel is set into the furthest distance (3500 mm). Figure 2 shows a captured depth image in 8-bit grayscale representation.

2.2 Anthropometric Analysis

Anthropometric is a measurement of the human individual, especially in physical posture. In this proposed method, the anthropometric data is used based on the characteristic of human pose as shown in the basic assumptions. The analysis of anthropometric data of human pose will resulted in three parameters, which are body depth level when performing gestural interaction pose (symbolized as D_A), hand depth level (symbolized as D_B), and body depth level when standing (symbolized as D_C). The human pose when performing a gestural interaction according to the premade assumptions is shown in Figure 3.

Fig. 2. Depth data visualization in 8-bit grayscale image.

Fig. 3. Human pose during gestural interaction. Upper body pose (*top*), palm opened pose (*bottom-left*), and grab pose (*bottom-right*).

Refer to the pose shown in Figure 3, the human arm length plays an important role in segmenting body image from a depth image because its equality with the distance between the palm and the human back. The hand dimension is also the key of the hand segmentation used in this proposed method. Since this research is conducted in Indonesia, the anthropometric data of Indonesian people is used.

Research in [13] shows that for 18-20 years old and 157 – 181 cm height, Indonesian people has: a) average arm length of 547 mm, b) average hand length of 188 mm, c) average palm

length of 110 mm, and d) average foot size of 250 mm. The illustration of the anthropometric data is shown in Figure 4.

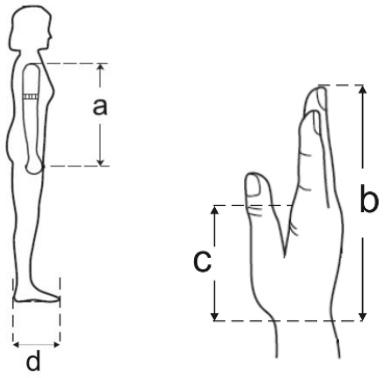

Fig. 4. Anthroprometric data of arm and hand.

The body depth level when performing gestural interaction poses (D_A) defined as a summation of arm length and hand length, resulted in 583 mm. According to the illustration shown in Figure 3, the horizontal distance of hand (started from wrist) in fully opened palm and grabbing postures are almost equal. Thus, the average palm length is used as the hand depth level (D_B). To calculate the body depth level when standing (D_C), the average foot length is used since the dimension of body width (from side view) is almost equal with the foot length.

2.3 Body Segmentation

Body segmentation is performed by applying image threshold operation to the depth image. This threshold operation simply removes any background component in the image and leaves the human body part. In order to determine the threshold level (symbolized as TH_{body}), the closest object distance in the depth image must be searched by finding the lowest pixel value in the image (symbolized as P_{min}). Using parameter obtained from the previous section (i.e. anthropometric analysis), the threshold level can be calculated using Equation (2), which is derived from Equation (1).

$$TH_{body} = P_{\min} + \left(255 \times \left(\frac{D_A - 800}{3500 - 800} \right) \right), \qquad (2)$$

where D_A is the parameter obtained from the anthropometric analysis. After the threshold operation is completed, the dimension of the human body is then calculated to determine a region of interest (ROI) of the image. The ROI is used as a boundary of the specific image region that needs to be processed later. Therefore, the rest-processing load will be reduced. Since abs, chest and legs have the smallest color deviation, the highest distributed color level in the ROI (symbolized as P_{body}) is also obtained in order to deduce the threshold limit of hand segmentation step (symbolized as TH_{lim}). The TH_{lim} value is obtained from TH_{val} in Equation (2) by replacing D_A with D_C, and P_{min} with P_{body}.The human body ROI in the segmented image is shown in Figure 5.

The binary version of the segmented image is used to calculate the centroid of ROI by employing Equation (3). Thus, the grayscale image of the human body is divided into two ROIs (i.e. left body region and right body region) by drawing a vertical line across the centroid coordinate.

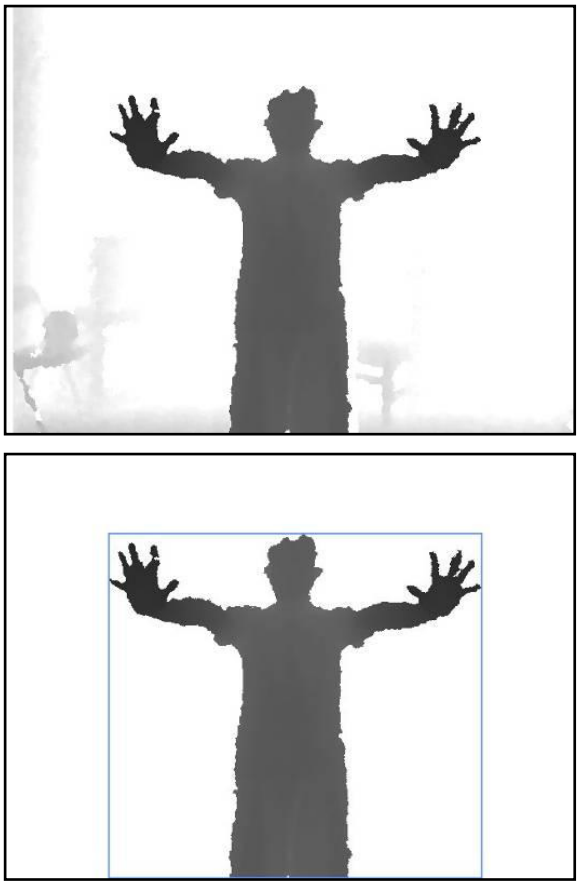

Fig. 5. Original image (*top*), segmented image (*bottom*), and body ROI (*blue rectangle*).

$$\bar{x} = \frac{\sum_{i=0}^{n} x_i}{n}; \bar{y} = \frac{\sum_{i=0}^{n} y_i}{n}, \qquad (3)$$

where x_i and y_i are the x and y coordinates of i^{th} pixel in the image, and n refer to the number of pixels in the image.

Each region of the separated grayscale image contains a single hand image that refers to human hand position in the body. Figure 6 shows the centroid coordinates and the separated image ROIs.

2.4 Hand Segmentation

Hand segmentation step uses a similar procedure with the body segmentation step. The differences of those steps are the uses of hand posture analysis and body width dimension to define the parameter in calculating the threshold level (TH_{val}) in each ROI. Refer to the assumption that human hand always positioned ahead of the body, the hand location in each region can be located by finding the closest distance object. Since the

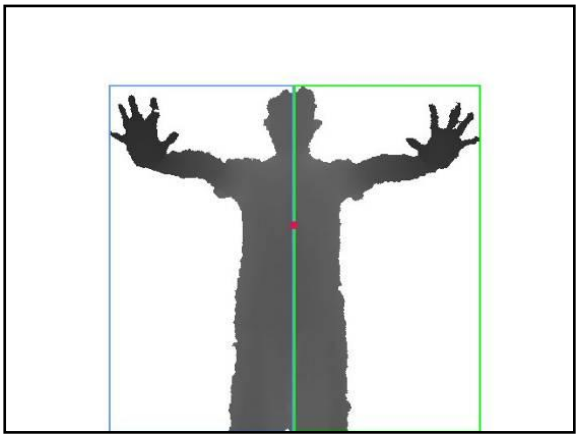

Fig. 6. Image centroid location (*red dot*), left ROI (*blue rectangle*), and right ROI (*green rectangle*).

closer the distance, the lower the intensity is, thus the lowest pixel value in each image region (ROI) is searched. Equation (4), which is derived from Equation (2), is used to find the threshold level of each image region.

$$TH_{val}^{i} = P_{\min}^{i} + \left(255 \times \left(\frac{D_B - 800}{3500 - 800} \right) \right) \quad (4)$$

where i is the image region (i.e. left ROI or right ROI). To avoid segmentation error that caused by partially segmented of another body parts, the value of TH_{val} of each region is re-evaluated using Equation (5).

$$TH_{val}^{i} = \min \left(TH_{val}^{i}, TH_{\lim} \right) \quad (5)$$

Once the calculations are completed, the image thresholding operation is applied to each ROI using each related threshold level. All the ROIs are then merged into a single grayscale image, which contains both hands. The image result conserves the hand grayscale value and removes other parts in the image, as shown in Figure 7.

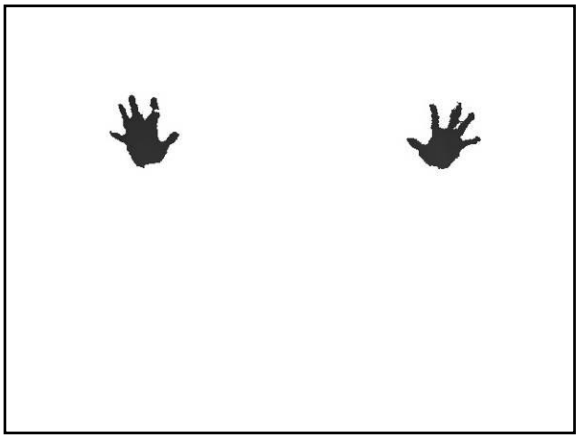

Fig. 7. Segmented hand image.

3 Experiments and Results

In order to measure the performance of this method, two depth images are used in this experiment. The first image contains a human image where the left hand is placed on a line with the human body and the right hand is positioned ahead. The

(a)

(b)

Fig. 8. Ilustrations of the experiment. First image (a) and second image (b). Original image, segmented body image, and segmented hand image (*left-to-right*).

second image contains a human image where both hands is placed on a line with the human body so the human body and hands have the same distance towards the imager. Illustrations of the experiment are shown in Figure 8.

According to the result shown in the illustrations, the proposed method is satisfactory enough in performing hand segmentation. The segmentation is failed when the distance between hand and camera is equal or further than the distance between body and camera, resulted in no segmented object in the image.

Computation latency is also evaluated in this section. The experiment is performed by measuring computation latency time in processing three images (i.e. one in Figure 5 and two in Figure 8). Another three computation time data is also obtained from live captured image of human pose. This experiment is conducted using a PC with Ubuntu OS with 2.0 GHz dual core processors and 3 GB RAM. The result of this test is shown in Table 1. Refer to the result, this proposed method is capable to process a VGA sized image up to 255 frames per second and still much higher than the imager (i.e. Kinect) frequency rate.

TABLE 1
COMPUTATION LATENCY TEST RESULT

Image Source	Computation Time (ms)
Figure 5	3,914489
Figure 8.a	3,884144
Figure 8.b	3,885941
Live process 1	3,921104
Live process 2	3,927903
Live process 3	3,907144
Average	3,906788

4 CONCLUSION

In this manuscript, a method to segment human hands from a depth image is proposed. The proposed method works with basic assumptions that human hands always positioned in front of human body and there is no obstruction in front of the human image. Anthropometric analysis is the core approach in defining the parameters that are used in this proposed method. From the experimental results, this proposed method works well as long as the assumptions are satisfied. The result shows that the segmentation is correct if the hands are not on a line with human body (hand and body does not have the same distance with the imager). Moreover, having average computation time of 3.906788 ms in processing a VGA sized image, this proposed method is considerably fast to process an image up to 255 frames per second. The image result conserves the grayscale level of the segmented hands. The conservation property is useful when the real pixel values of the segmented hands will be used in further processing mechanism.

Although the proposed method is successfull in segmenting hands, the use of static centroid has a limitation when is employed as a reference in defining exact location of human hands. A problem might occur when the hand exact location of one image region is overlapping another image region (e.g. right hand is positioned in left image region, which is the location of the left hand). Therefore, an adaptive centroid estimation and region growing method might be used in the future research to adressing this problem.

Based on the background of this research, which is aimed at the development of natural interface between human and robot using hand gestures, thus this proposed method is suitable to be used. Furthermore, since this proposed method is a kind of a pre-processing stage in hand gesture recognition, the performance of this method needs to be evaluated more in the real application.

REFERENCES

[1] Y. Xu, M. Guillemot, and T. Nishida, "An Experiment Study of Gesture-Based Human-Robot Interface," *Complex Medical Engineering, 2007. CME 2007. IEEE/ICME International Conference on* , vol., no., pp.457-463, 23-27 May 2007.

[2] R. Huertas, J.R.M. Marcelo, V. Romero, and A.M. Hector, "A Robotic Arm Telemanipulated through a Digital Glove", *Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007)*, pp.470-475, 2007.

[3] G. Ariyanto, "Hand Gesture Recognition Using Neural Networks for Robotic Arm Control", National *Conference on Computer Science & Information Technology*, 2007.

[4] J. Wachs, "Real-Time Hand Gesture Telerobotic System Using the Fuzzy C-Means Clustering", *Fifth Bianual World Automation Congress*. 2007.

[5] Microsoft Kinect. http://en.wikipedia.org/wiki/Kinect. 2011.

[6] J.L. Raheja, A. Chaudhary, and K. Singal, "Tracking of Fingertips and Centers of Palm Using KINECT," *Computational Intelligence, Modelling and Simulation (CIMSiM), 2011 Third International Conference on* , vol., no., pp.248-252, 20-22 Sept. 2011.

[7] Z. Mo, U. Neumann, "Real-time Hand Pose Recognition Using Low-Resolution Depth Images," *Computer Vision and Pattern Recognition, 2006 IEEE Computer Society Conference on* , vol.2, no., pp. 1499- 1505, 2006.

[8] M.D. Abramoff, P.J. Magalhaes, and S.J. Ram, "Image Processing with ImageJ", *Biophotonics International, volume 11, issue 7, pp. 36-42*, 2004.

[9] S2i - Industrial Intelligent Systems, "HARPIA", Departamento de Automação e Sistemas (DAS) da Universidade Federal de Santa Catarina (UFSC), Brazil, http://s2i.das.ufsc.br/harpia/en/home.html, 2007-2009.

[10] G. Bradski, A. Kaehler, *Learning OpenCV - Computer Vision with the OpenCV Library*, O'Reilly Media Inc. Sebastopol, 2008.

[11] L. Xia, Chia-Chih Chen, JK Aggarwal, "Human Detection Using Depth Information by Kinect",*Computer Vision and Pattern Recognition Workshops (CVPRW), 2011 IEEE Computer Society Conference on*. 2011: 15-22.

[12] K Lai, L. Bo, X. Ren, D. Fox, "A Large-Scale Hierarchical Multi-View RGB-D Object Dataset", *Robotics and Automation (ICRA), 2011 IEEE International Conference on*. May 2011: 1817-1824.

[13] Sutjana, I. D. Putu, M. Sutajaya, S. Purnawati, P. Adiatmika, K. Tunas, E. Suardana, and I.B.A. Swamardika, "Preliminary Anthropometric Data of Medical Students for Equipment Applications", *Journal of Human Ergology* 37, no. 1: 45-48, 2008.