Comparative Study of Segmentation Technique for Hand Gesture Recognition System

Farah Farhana binti Mod Ma'asum, Suhana binti Sulaiman, Azilah binti Saparon
Centre for Electronics Engineering Studies (CEES), Faculty of Electrical Engineering, Universiti Teknologi Mara, 40450 Shah Alam, Selangor, Malaysia.
suhana832@salam.uitm.edu.my

***Abstract*—Hand gesture recognition system brings a lot of benefits to people in various industries. It consists of three main modules, which are the detection module, segmentation module and feature extraction module. The selection of appropriate segmentation method depends on the application and system environments. This paper concentrates on the following points: 1) The input images are converted to the HSV and YCbCr model by using threshold values, which satisfy the skin color segmentation based on skin color region; 2) Different edges operator are then applied on both color model images, namely Canny with fixed threshold, Canny with Otsu threshold and Sobel with fixed threshold. While there is no conclusion, which color space is best fit for skin color detection, Canny edge detector using Otsu threshold proves the best in creating contour of hand image since it is able to detect true weak edges compared to the other types of edge segmentation technique.**

***Index Terms*—HSV; YCbCr; Canny Edge Detector; Otsu Method Threshold; Sobel Edge Detector.**

I. INTRODUCTION

Numerous technologies nowadays comprise a lot of different interaction involving human and machine. Human Computer Interaction (HCI) is one of the examples of interaction, which uses the concept of emerging between computer (machine) and human [1]. Example of a technology that occupies HCI is Wii and Kinect application. The use of these two technologies shows that HCI is basically based on the ability to track a person's actions while understanding its gestures. There are two types of hand gestures recognition: glove-based gestural interfaces and vision-based gestural interfaces.

Kinect is one of the computer visions that provides a major potential in real time vision using infrared sensitive camera. However, it has the disadvantage of requiring many built-in sensors to capture the gestures and interpret the motion.

In this paper, we implemented a vision-based gesture recognition using web camera. We proposed three modules: detection module, segmentation module and feature extraction module, although there have been many methods used by other researchers. Segmentation technique can be defined as the process of dividing an image into non-overlapping regions, which have significant meaningful region representation that simplifies the analysis of the image [2]. The threshold between image processing will determine which pixel belongs to which regions [3]. Finding suitable local features by distinguishing it from background will produce a better image processing operations [3].

Hand segmentation of the tracking hand focuses on two different methods: Each frame of hand pose is processed separately, and the second method is conducted using hand-tracking information [4]. In the segmentation technique, we compared which three techniques, namely Canny, Sobel and Otsu, produces better image segmentation. This paper is organized as follow: Section 2 describes the skin-color representation, Section 3 presents the segmentation techniques, Section 4 presents the experimental results and finally conclusion is concluded in Section 5.

II. COLOR SPACE PRESENTATION

Hand image captured by camera needs to undergo skin color extraction to produce clear differentiation between the background and the hand itself. There are various types of skin color segmentation with different characteristics, but the main three commonly used color segmentations are RGB, HSV and YCbCr [5].

RGB color space, which is represented by red, green and blue has high correlation between channels, leading to luminance sensitivity. Mixtures of the light primary colors cover a large part of human color space, which enables skin to be recognized correctly [4]. Each of the three beams has its arbitrary intensity, which represents zero intensity for the darkest color (considered as black), while full intensity of each component gives a white color.

HSV, on the other hand is an acronym for hue-saturation-value. Hue describes pure color, while saturation defines the amount of white light mixed with a hue in which the value refers to the brightness of the image [6]. HSV and HSL are the improvisation of color space of RGB since the color arrangement of each hue is in radial slice, ranging from black at the bottom and white at the top of the radial slice.

YCbCr is a color space, which has a separation between its luminance and the chrominance data. The Y value represents the brightness component, while the Cb and Cr value each represents the chrominance component of the image [6]. The transformation used to convert from RGB to YCbCr color space is shown in the Equation (1):

$$\begin{bmatrix} Y \\ Cb \\ Cr \end{bmatrix} = \begin{bmatrix} 16 \\ 128 \\ 128 \end{bmatrix} + \begin{bmatrix} 65.481 & 128.553 & 24.966 \\ -37.797 & -74.203 & 112 \\ 112 & -93.786 & -18.214 \end{bmatrix} \begin{bmatrix} R \\ G \\ B \end{bmatrix} \quad (1)$$

Contrary to RGB, the YCbCr color space produced better performance since it is luma-independent [7].

Each and every color space has its own effectiveness. The intended purpose of segmentation will determine which color space is suitable to use. For instance, D.Chai et. al [6] prefer to apply YCbCr as segmentation technique in his research due to its effective use of the chrominance information for modeling human skin color in video coding format. Chai and Ngan [6] have developed algorithm that exploits the spatial

characteristics of human skin color. They also have found that the range of Cb and Cr which represents best in skin colour reference map are:

$$77 \leq Cb \leq 127 \text{ and } 133 \leq Cr \leq 173$$

M.M. Hassan et. al [8] on the other hand focused on HSV color space segmentation for splitting the hand area from the image gesture, which then will be normalized in order to obtain gesture database that are invariant against position, scale and rotation. While Yishen Xu et. al [9] converts hand image format which has been captured using RGB color space into HSL in order to reduce the impact of changes in lighting intensity. This is because hue is a good color-invariant model.

A. Limitation of Skin Color Segmentation

Segmentation based on skin color map requires good differentiation between the skin color and the background objects to produce an accurate result. The background whose color is similar to the skin will interrupt the result of the output since it could not locate which one is the skin of hand. For the purpose of this research, the hand is considered as the main image, while other parts of the body are considered as the background.

III. Classification Of Segmentation Technique

Image segmentation can be defined as a division of an image into regions or categories, in which each and every pixel in an image is allocated to one another [10]. There are three types of segmentation technique, which are the thresholding method, edge-based method and region-based method. In the following subsections, the types of segmentation-based techniques are briefly discussed.

A. Thresholding Technique

Thresholding technique is the simplest approach for segmenting images where the pixels are allocated to categories according to the range of values in which a pixel lies [10]. Thresholding creates binary images from grey-level one by turning all pixels below threshold to zero, and all pixels above threshold to one. A fixed threshold can be defined as in Equation (2) below:

$$g(x,y) = \begin{cases} 0, & f(x,y) < T; \\ 1, & f(x,y) \geq T; \end{cases} \quad (2)$$

where T is the threshold value of the image pixels, f(x,y).

a. Otsu Method

Otsu method is the type of global thresholding that depends on the gray value of the image only. It is widely used because it is a simple and effective method [11]. Otsu method is deduced by least square method based on grey histogram which suits best in automatically selecting an optimal threshold value. Otsu algorithms calculate the optimum threshold value by separating the two classes so that intra-class variance is minimal while inter-class variance is maximal.

B. Edge-based Segmentation

In edge-based segmentation, edge filter is applied on the image in order to determine whether the pixels can be classified either as edge or non-edge. Segmentation can be achieved by allocating all non-pixels, which are not separated by an edge to a single category [10]. Image contours can be constructed by connecting the edge pixels gain from the edge filter itself. Examples of some edge detection filters are canny detector, Prewitt's filter and Laplacian of Gaussian filter.

a. Canny Edge Detector Technique

Canny edge detector algorithm is simple and easy to implement and it has been defined as the standard benchmark for comparison with other edge detection methods. To determine an optimal edge, three criteria must be met which are: Minimizing error rate, Marking edges as closely as possible to the actual edges to maximise localization, and marking edges only when a single edge exists for minimal response [12]. Canny detector algorithm consists of four steps as mentioned in [5]: (1) Gaussian filter is applied to smooth the images by reducing noises, (2) Intensity gradients of the image is calculated by using Sobel operator, (3) Edge direction is used to search along the edges and suppress any non-edge pixels, and (4) Broken edges are eliminated using high and low thresholds. H.S. Yeo et. al [13] applied Canny detector to find edges around contour of both hand and face during a face removal process. Hand contour can be separated effectively from face contour by drawing thick lines along contour perimeter.

b. Sobel Edge Detector Technique

Sobel operator consists of a pair of 3×3 convolution kernels, where one of the kernels is rotated by 90˚ (see Figure 1).

-1	0	+1
-2	0	+2
-1	0	+1

Gx

+1	+2	+1
0	0	0
-1	-2	-1

Gy

Figure 1: Mask used by Sobel Operator

These kernels respond vertically and horizontally relative to the pixel grid, where one kernel is meant for each of the two perpendicular orientations. Initially, the kernels are applied separately (Gx and Gy) in order to produce separate measurements of the gradient component. These can then be combined to find the absolute magnitude of the gradient of each point [14]. The gradient of the magnitude is given by Equation (3):

$$|G| = \sqrt{{G_x}^2 + {G_y}^2} \quad (3)$$

IV. Hand Gesture Detection System

Using skin color as a primitive feature for detecting hand gestures has several advantages in term of robustness. However, people all around the world have different skin color tone depending on ethnics and races. Figure 2 shows the image transformation from the RGB color space to HSV and YCbCr color space respectively. Next, thresholding technique is applied to determine the best results that could represent hand gestures. The image then undergoes morphology technique to reduce noises by dilating and

eroding technique. The next step is that Canny, Sobel and Canny edge detector with Otsu Threshold are compared to determine which detector produces the best edge segmentation on these binary images. After the comparison, the best edge segmentation technique is applied to mark the edges as closely as possible to the actual edges of hand. Next, contouring technique with bounding box is applied on the detected hand gesture for the next analysis.

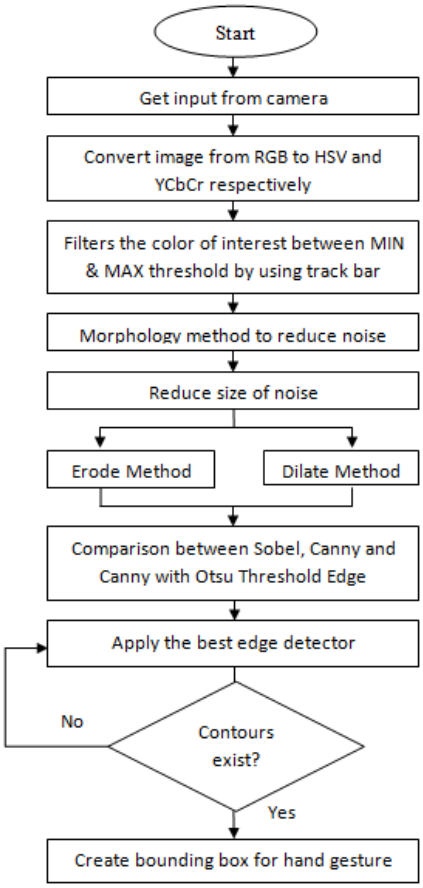
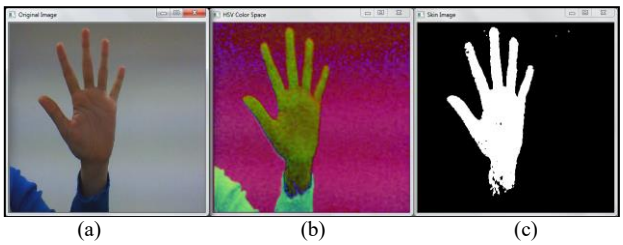


Figure 2: Flowchart of system

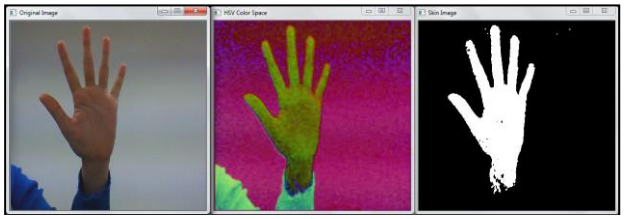
V. EXPERIMENTAL RESULTS

The hand recognition is done in a static background in order to reduce noise constraints. A test is done using Microsoft Visual Studio 2012 and OpenCV 2.4.10 using C++ language. The results of real-time image of hand which has been converted into binary images using HSV and YCbCr colour space representation respectively are shown below. Figure 3 shows the boundaries, which were obtained by thresholding the hand image using HSV color segmentation, while Figure 4 shows the image after being segmented by thresholding of YCbCr color representation.



(a) (b) (c)

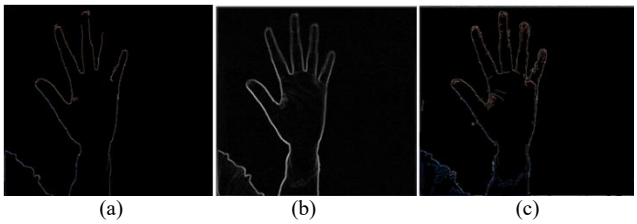
Figure 3(a): Original image, 3(b): Image of RGB-HSV, 3(c): Image after threshold



(a) (b) (c)

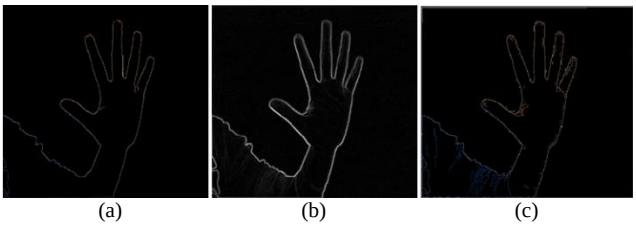
Figure 4(a): Original image, 4(b): Image of RGB-YCbCr, 4(c): Image after threshold

This binary image is then compared by applying different types of edge detector, namely the Canny, Sobel and Canny using Otsu threshold to find edges around the contour of hand: This is shown in Figure 5 and 6.



(a) (b) (c)

Figure 5: Comparison between edge segmentation on HSV image (a) Canny operator using fixed threshold value 50 (b) Sobel operator (c) Canny using Otsu threshold



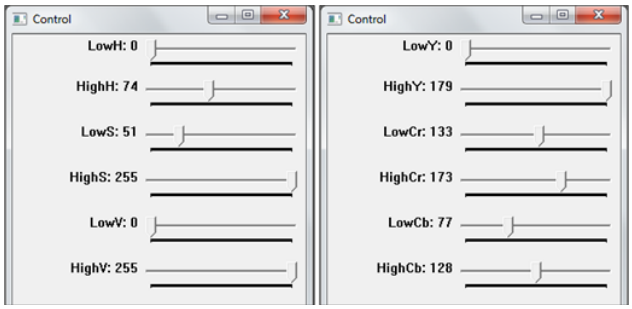
(a) (b) (c)

Figure 6: Comparison between edge segmentation on YCbCr image (a) Canny operator using fixed threshold value 50 (b) Sobel operator (c) Canny using Otsu threshold

a. Discussion of the Results

As we can see above, both color spaces produced a similar result in terms of skin detection. This is because a camera in RGB color space was used to capture the image of hand. The RGB color space has luminance or brightness information embedded into each layer of its color space [15]. However, RGB has instability issue due to varying levels of brightness in an image [16]. This issue can be solved by the application of the YCbCr and HSV color spaces in image processing. In the YCbCr color space, the information luminance of color was at Y, while the chrominance was distinguished in the Cb and Cr layers. In the HSV color space, the luminance information was placed in the V (Value) layer and the chromaticity information was placed in the H (Hue) and S (Saturation) [16].

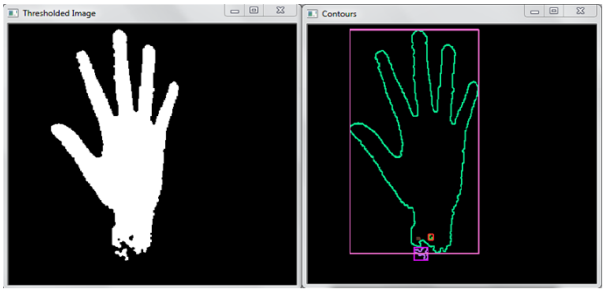
During thresholding segmentation of color space, both HSV and YCbCr play important role in determining the best skin segmentation. Figure 8(a) and 8(b) show the tracking bar, which can be adjusted by user. This bar was used to determine the best skin color map representing each user's color skin depending on his/her ethnic and race. After being tested with various values of thresholding, it has proven that Chai and Ngan [6] algorithm of YCbCr skin color range (Figure 7 (b)) is the best representation of human skin color (Refer Figure 4(c)).



(a) (b)

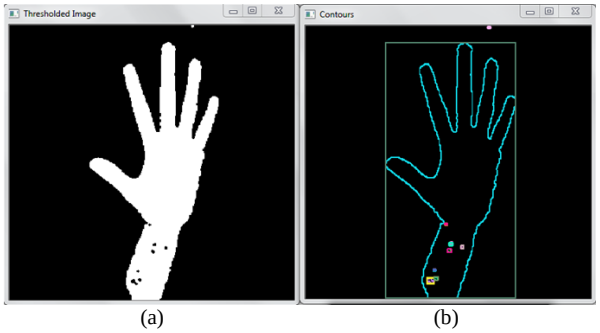
Figure 7: (a) Tracking bar for HSV thresholding, (b) Tracking bar for YCbCr thresholding

Canny edge detector was used as a filter method that produced hand image contours. Canny was chosen as the method of edge-based segmentation because it has the advantage of fast execution time compared to other edge-based segmentation technique [17]. Figure 5 and 6, show that Canny edge detection with Otsu thresholding technique yielded the most accurate edge detection. This is because Otsu thresholding technique was able to minimize the intra-class variance which managed to produce thin lines. Applying the Canny algorithm using Otsu was more efficient in recognizing the main edges in the hand image compared to the manually Canny thresholding method.



(a) (b)

Figure 8: HSV: (a) Thresholded image (b) Contouring with bounding box



(a) (b)

Figure 9: YCbCr: (a) Thresholded image (b) Contouring with bounding box

The results in Figure 8 and Figure 9 show similarity in terms of tracking skin color map. It shows that both HSV and YCbCr can be used to extract human skin color from the background, which then will produce quite similar results. The use of a 5x5 structuring element in morphological filter results in minimizing and reducing noises in the image.

VI. Conclusion

In this paper, it can be concluded that both HSV and YCbCr color representation can be used as skin color segmentation technique in hand gesture recognition. Both color space representation required thresholding technique to ensure that the hand gesture is properly extracted, in which other colors are extracted besides the skin color. The best threshold for skin-color map in HSV is $H \leq 74$ and $51 \leq S \leq 255$, while YCbCr is $77 \leq Cb \leq 127$ and $133 \leq Cr \leq 173$. Canny edge detector with Otsu threshold also works best in creating contour of hand image considering that it has the ability to detect true weak edges since it is not susceptible to noise interference.

Acknowledgment

The authors would like to give a special thank for the support given to this research by the Ministry of Higher Education (MOHE), Malaysia and Universiti Teknologi Mara (UiTM) for providing the facilities required in this research.

References

[1] Gupta, R. 2012. Human Computer Interaction – A Modern Overview. *International Journal Computer Technology Application*. 3(5): 1736–1740.

[2] Raju P. D. R., Neelima G., and Rao K. P. 2011. Image segmentation-MR Images Segmentation with A Modified Gaussian Mixture Model, *International Journal Computer Science Information Technology*. 2(6): 2573–2578.

[3] Singh K. K. and Singh A. 2010. A Study Of Image Segmentation Algorithms For Different Types Of Images. *International Journal Computer Scence Issues*. 7(5): 5–8.

[4] Hasan M. M. and Mishra P. K. 2012. Hand Gesture Modeling and Recognition using Geometric Features : A Review. *Canadian Journal Image Processing Computer Vision*. 3(1): 12–26.

[5] Ibraheem N., Khan R., and Hasan M. 2013. Comparative Study of Skin Color based Segmentation Techniques. *International Journal Application Information System*. 5(10): 24–38.

[6] Chai D. and Ngan K. N. 1999. Face Segmentation Using Skin-Color Map in Videophone Applications. *IEEE Transactions Circuits and System Video Technology*. 9(4): 551–564.

[7] Basilio J. A. M., Torres G. A., Pérez G. S., Medina L. K. T., and Meana H. M. P. 2011. Explicit Image Detection using YCbCr Space Color Model as Skin Detection. *Application Mathematics Computer Engineering*. 123–128.

[8] Hasan M. M. and Pramod Misra K. 2011. HSV Brightness Factor Matching for Gesture Recognition System. *International Journal Image Processing*. 4(5): 456–467.

[9] Xu Y., Gu J., Tao Z., and Wu D. 2009. Bare Hand Gesture Recognition with a Single Color Camera. *2nd International Congress Image Signal Processing*. 1–4.

[10] Glasbey C. and Horgan G. W. 1995. *Chapter 4: Image Analysis for the Biological Sciences*.

[11] Vala H. J. and Baxi A. 2013. A Review on Otsu Image Segmentation Algorithm, *International Journal Advanced Research Computer Engineering Technology*. 2(2): 387–389.

[12] Han D. D., Zhang T. C., and Zhang J. 2014. Research and Implementation of an Improved Canny Edge Detection Algorithm. *Key Engineering Materials*. 572: 566–569.

[13] Yeo H.-S., Lee B.-G., and Lim H. 2013. Hand Tracking And Gesture Recognition System For Human-Computer Interaction Using Low-Cost Hardware. *Multimedia Tools Application*.

[14] Maini R. and Aggarwal H. 2009. Study and comparison of various image edge detection techniques. *International Journal Image Processing*. 147002(3): 1–12.

[15] Kim W., Cho D., and Kim H. M. 2004. Inter-plane Prediction for RGB. (1): 785–788.

[16] Sebastian P., Voon Y. V., and Comley R. 2010. Colour Space Effect On Tracking In Video Surveillance. *International Journal Electrical Engineering Informatics*. 2(4): 298–312.

[17] Almadhoun M. D. 2013. Improving and Measuring Color Edge Detection Algorithm in RGB Color Space. *International Journal Digital Information Wireless Communication*. 3(1): 19–24.W. D. Doyle, “Magnetization reversal in films with biaxial anisotropy,” in *1987 Proc. INTERMAG Conf.*, pp. 2.2-1–2.2-6.