ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

Hand Gesture Segmentation Using Skin Color Detection in YCBCR Color Space

[1] Sarang Suresh Kakkoth, [2] Saylee Gharge
[1] ME Scholar, [2] Associate professor

Abstract: **- This paper discusses the segmentation of hand gestures using skin detection method based on the YCbCr color space. Here a face detection algorithm based on viola jones has been used to detect and delete the face. Thus the only area to detect based on skin color is that of the hand. After skin color detection and thresholding using rectangular selection over the input image the binary image received is further processed using image processing techniques like median filtering and morphological operation. Then robust tracking of the segmented hand contour using image moment concept further helps in localizing position of hand contour in the two dimensional space. These segmentation of hand gestures can be used in creating datasets of binary hand gestures and in recognition of hand gestures based on the machine learning, neural network, geometrical analysis of hand and so on in future.**

Index Terms **- Contour, Image Moment, Segmentation, Tracking.**

I. INTRODUCTION

This paper aims to provide automatic segmentation of hand gestures in real time without providing a specific rectangular box for hand to be placed within the live video frame, facilitating in ease of hand gesture segmentation based on skin color detection. According to recent trends in image processing and computer vision domain the technological trends is on improving camera technology and identifying features and recognizing these features using the machine learning technologies. Hand gestures is a body language that has specific meaning which is mainly conveyed through center of the palm and finger position and shape. For these feature extraction and ease of recognition in real time, it is necessary to form a robust segmentation of hand gestures in real time. Based on the various studies on skin detection based on different color space it is found that YCbCr has been effective enough in segmentation of hand gestures.

II. LITERATURE REVIEW

A robust segmentation of hand gestures in complex environment is the first step in any hand gesture recognition system. According to various segmentation schemes [1] researched, all methods depends upon the image acquisition, image processing and image understanding techniques of computer vision in digital image processing. Chung-Ju Liao et al. [2] improved the concept of image segmentation using with skin detection by dividing the image into two parts called frame division.

Stergios Poularakis and Ioannis Katsavounidis [3] proposed a complete gesture recognition framework based on maximum cosine similarity and fast nearest neighbor (NN) techniques. Various color space used in color based segmentation includes YCbCr, HSV, CIE LUV, CIE LAB, HLS color space models. These color space can be used in skin color detection by setting the minimum and maximum thresholds in their respective color space. By setting these thresholds it is possible to get a binary segmentation based on the set thresholds. Comparative studies of skin color detection and segmentation in HSV and YCbCr color space by Khamar Basha Shaik et al. [4] show that YCbCr is more effective and efficient in segmentation in terms of color in color images even under uneven lighting conditions. On-line skin classifier with multiple cues is proposed by Ying Zhao and Jiayong Yan [5]. A novel method called MACS for 2D hand tracking using motion and color information for headworn monocular color cameras is presented by JH Hammer et al. [6]. Yuan Yao et al. [7] introduces a framework that allows tracking hand gestures in 3-D space and matching gestures with simple contour model, supporting complex real–time interactions. Zhang Qiu-yu et al. [11] presents a method based on YCbCr color space and K-means clustering algorithm for segmentation hand gesture. There are various methods to detect edges in binary and grayscale images. These methods mainly aims at identifying discontinuities in digital images with brightness varying at different points. Edge detection is a fundamental tool in image processing, machine vision and computer

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

vision, particularly in the areas of feature detection and feature extraction. Gradient and derivative based methods [8, 9] like Sobel derivative operator, Scharr operator, Laplace filter, Canny edge detector [12] serves as a function to detect edges from grayscale images based on varying intensity levels in the image. Canny edge detector is the refined version of Laplace filter proposed by J. Canny in 1986. Similarly Sobel operator named after Irwin Sobel and Gary Feldman, colleagues at SAIL in 1968 aims at emphasizing the edges of the image. Border following algorithm proposed by Suzuki & Abe [10] for topological analysis of digital binary images helps in extracting some features without reconstructing the whole image. This algorithm finds the boundary edges in the binary images and sorts those borders accordingly. Box filter, median filter, Gaussian filter, bilateral filter are some of the smoothing filters [8] often used in reducing noise and camera artifacts. Smoothing filters are also used when there is need to reduce the resolution of the image for reducing the computation power needed in processing of images.

III. PROPOSED METHOD

Based on the research works by various researchers hand gesture enabling technologies based on vision can be divided as hand gesture acquisition, segmentation, tracking and recognition. As the name suggests vision based technologies use RGB cameras to acquire the real time camera input which is further subjected to various pre-processing steps for foreground segmentation in this case which is the hand from the background.

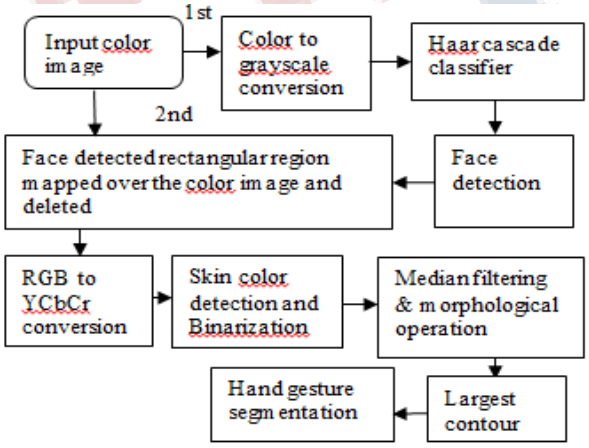


Fig. 1. Block diagram for robust hand gesture segmentation

For proper real time recognition of hand gestures, seamless segmentation of hand from the background without any lag is necessary. For this purpose, skin detection method using YCbCr color space serves as a medium for segmentation of hand as well as the face. Since only hand gestures are to be segmented, initially face deletion is to be performed. A frontal face Haar cascade classifier based on Viola jones algorithm proves to be efficient technique for detection and deletion of face in real time. The next step after face deletion is skin detection where we use RGB to YCbCr color space conversion. Then to remove the noise in the background we perform some image processing techniques to get a clear segmentation of hand gestures. The proposed method for real time hand gesture segmentation can be divided into as follows:

A. Image acquisition

In this paper a hand gesture segmentation using a simple web camera is proposed. For this purpose a camera capturing images at a resolution of 640 × 480 is sufficient. These images are continuously captured in loop at a frame rate of 6 fps having a capture time of 0.15 to 0.2 seconds until a stop or pause command is given. The images captured are 8 bit images in RGB format which can processed using image processing and video processing modules which are present in OpenCV [8] , a computer vision library. Computer vision is mainly transformation of data from still or video camera into some decision or new representation.

Fig. 2. Input image using RGB camera

B. Face Detection and Deletion

After acquiring continuous RGB camera feed of images the next step is face detection and deletion for which we use a pre-trained frontal face haar cascade classifier present in OpenCV 3 library.

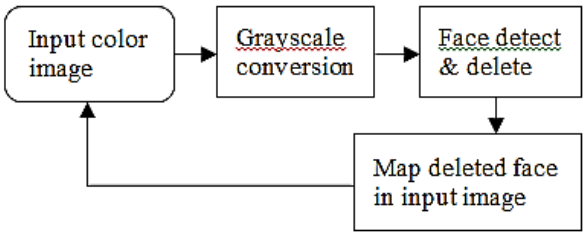


Fig. 3. RGB to Grayscale conversion for face detection

The cascade classifier implemented in OpenCV is a technique for face-detection first developed by Paul Viola and Michael Jones, commonly known as the Viola-Jones detector. Since haar cascade classifier takes grayscale images as input we first perform RGB to grayscale conversion as follows

$$Y \leftarrow 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \quad (1)$$

Then we load the pre-trained Haar cascade classifier and detects the number of face detected and marks them with rectangular box of respective scale. For face deletion we mark the face detected with a filled rectangle box of black

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

pixels and maps it into the color input image as shown in Fig. 4.

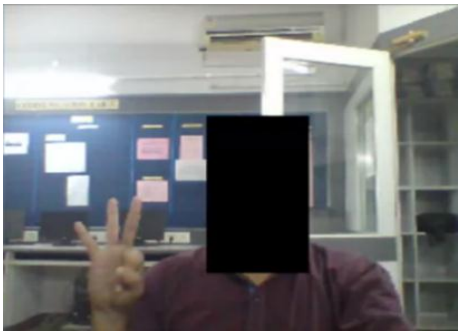


Fig. 4. Face detection and deletion

C. Skin color detection

Skin color detection is not possible in RGB color format due to varying chrominance values as compared to other color space models. Also YCbCr color space proved most efficient in skin color detection as compared to other color space. Hence we implement an YCbCr based skin detection technique in this paper. In YCbCr color space, Y is luminance. Cb and Cr are chromaticity of blue and red colors. Cb and Cr are two-dimensional independent. Conversion formula is as follows:

$$\begin{aligned} Y &\leftarrow 0.299 \cdot R + 0.587 \cdot G + 0.114 \cdot B \\ Cr &\leftarrow (R-Y) \cdot 0.713 + delta \\ Cb &\leftarrow (B-Y) \cdot 0.564 + delta \end{aligned} \quad (2)$$

Where

$$delta = \begin{cases} 128 & \text{for } 8-\text{bit images} \\ 32768 & \text{for } 16-\text{bit images} \\ 0.5 & \text{for floating}-\text{point images} \end{cases}$$

Fig. 5. RGB to YCbCr color space conversion

The YCbCr converted image of the input is obtained as shown in Fig. 5. For skin color detection, threshold the YCbCr values to minimum and maximum values depending upon the skin color so that only skin colored pixels in the input frame get thresholded to binary image as in Fig. 6.

Fig. 6. Binary threshold after face deletion

D. Image processing

An image may be defined as a two-dimensional function, f(x, y), where x and y are spatial (plane) coordinates, and the amplitude of 'f' are any pair of coordinates (x, y) is called the intensity or gray level of the image at that point. When x, y, and the amplitude values of f are all finite, discrete quantities, we call the image a digital image [9]. The field of digital image processing refers to processing digital images by means of a digital computer. Image processing functions performed here to remove the noise present in the thresholded binary image as obtained in Fig. 6 include median filtering and morphological operations. Median filter is an order-statistic spatial non-linear filter. The median filter replaces each pixel by the median or "middle-valued" pixel (as opposed to the mean pixel) in a rectangular neighborhood around the center pixel as shown in Fig. 7. Consider 3 x 3 neighborhood image values: (10, 20, 20, 20, 15, 20, 20, 25, 100). Sorted values in ascending order: (10, 15, 20, 20, 20, 20, 20, 25, 100). Median value: 20

10	20	20
20	15	20
20	25	100

(a)

10	20	20
20	20	20
20	25	100

(b)

Fig. 7. Median filtering operation (a) Original neighborhood image (b) Median filtered image

The Gaussian blur is a type of image-blurring filter that uses a Gaussian function. Gaussian filtering is done by convolving each point in the input array with a Gaussian kernel and then summing them all to produce the output array. A 2D Gaussian can be represented as:

$$G_0(x,y) = Ae^{\frac{-(x-\mu x)^2}{2\sigma x^2} + \frac{-(y-\mu y)}{2\sigma y^2}} \quad (3)$$

where, μ is the mean (the peak) and σ represents the variance.

Gaussian filtering of images works in the way that pixels of the image should vary slowly correlating with neighborhood pixels in that image. Thus Gaussian filtering reduces noise

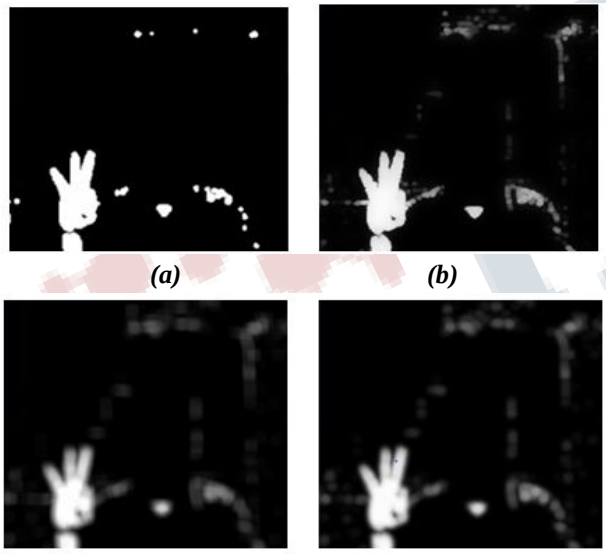
ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018

preserving the image information. However Gaussian filtering operation blurs the edges and is not much useful in reduction of background noise in the form of white pixels. Box linear filter is the simplest filter where each output pixel is the mean of its kernel neighbors. Its kernel can be represented as

$$K = \frac{1}{K_{width} \cdot K_{height}} \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \\ . & . & . & \dots & 1 \\ . & . & . & \dots & 1 \\ 1 & 1 & 1 & \dots & 1 \end{bmatrix} \quad (4)$$

Bilateral filtering does the same function of Gaussian filtering without blurring the edges at the cost of substantially more processing time. Median filtering and bilateral filtering preserves the edges of the image and gives more accurate image. As shown in Fig. 8, median filtering operation performs better noise reduction as compared to other smoothing operations like bilateral filtering, box linear filtering and Gaussian filtering. Thus in this paper median filtering have been chosen for noise reduction over other smoothing filters.



(a) *(b)* *(c)* *(d)*

Fig. 8. Binary image after using smoothing filters and morphological operations (a) Median filtering (b) Bilateral filtering (c) Box linear filtering (d) Gaussian filtering

E. Largest contour detection

There are various method to find the boundaries or edges of the digital image based on intensity variation within the image. The contour found using border following algorithm is shown in fig 10.a. This algorithm being used in topological analysis of binary images further arranges the contour returned in hierarchy, helping in contour analysis. This algorithm can be used to find contours within images and arrange them accordingly with reference to parent border. Canny edge detection of a binary image shown in Fig. 10.b is a complex process as it also performs the function of removing speckle noise adding to the computation power required. Scharr filter is a continuation of Sobel filter but is more fast and accurate compared to Sobel filter. The operator uses two 3×3 kernels which are convolved with the original image to calculate approximations of the derivatives – one for horizontal changes, and one for vertical. Both these operators gives foggy edges as compared to border following and canny edge detector. Based on the various edge detection methods and its properties it can be shown that border following algorithm is computationally effective and much faster as compared to other algorithms.

Here we apply border following algorithm proposed by Suzuki & Abe [10] to find contours as implemented in OpenCV. These contours are then indexed accordingly and the number of contours formed is detected. These contours are then filtered out based on two largest contour area. The contour area is found out using the image moments concept. Contour moments represent certain high-level characteristics of a contour, an image, or a set of points. Numerically, the moments are defined by the following formula:

$$m_{p,q} = \sum_{i=1}^{N} I(x_i, y_i) x^p y^q \quad (5)$$

The moment mp,q is defined as a sum over all of the pixels in the object, in which the value of the pixel at point x, y is multiplied by the factor xp yq. We compute central moments since those values are invariant of position of contour.

$$\mu_{p,q} = \sum_{i=0}^{N} I(x_i, y_i)(x - \bar{x})^p (y - \bar{y})^q \quad (6)$$

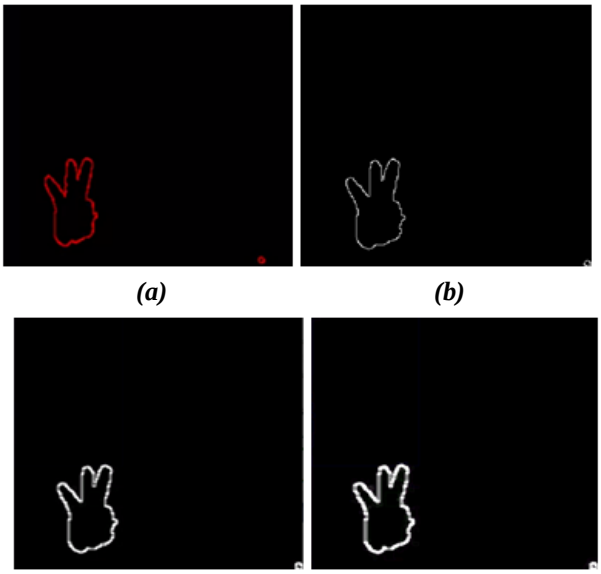
Where $\bar{x} = \frac{m_{10}}{m_{00}}$ and $\bar{y} = \frac{m_{01}}{m_{00}}$

The resultant output obtained after largest contour detection is as shown in Fig. 9.

Fig. 9. Binary image filtered out using two largest area concept

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)
Vol 5, Issue 2, February 2018



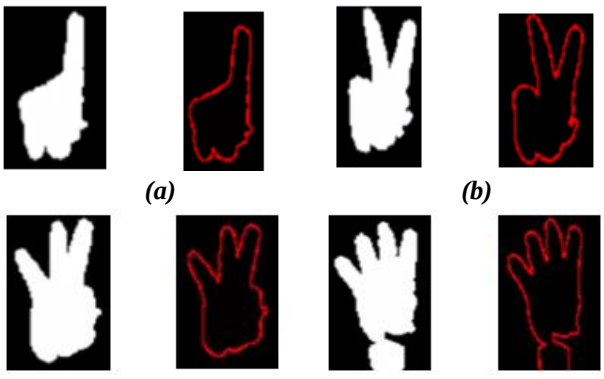
(a) *(b)*

(c) *(d)*

Fig. 10. Contour formation over 2 largest contours area (a) Border following algorithm (b) Canny edge detection (c) Sobel derivative operator (d) Scharr operator

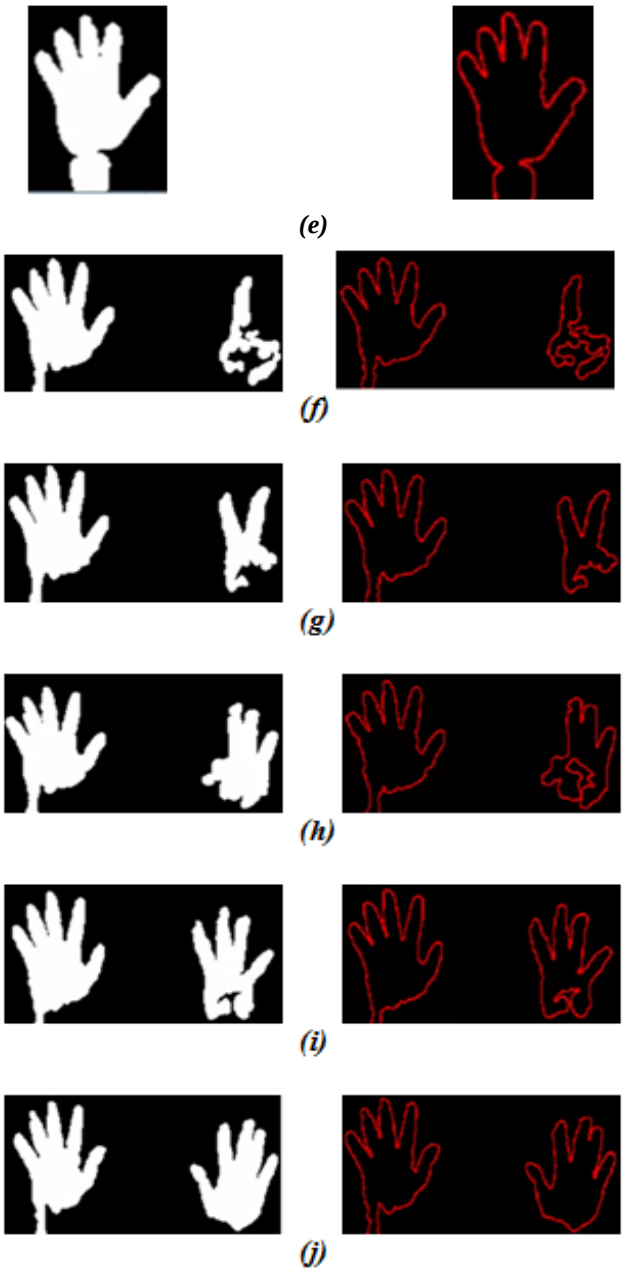
IV. EXPERIMENTAL RESULTS

Several experiments performed in different backgrounds and varying gestures suggests that RGB to YCbCr color space conversion performed better in thresholding the skin color with respect to other color space like HSV. We also try to auto threshold skin color by drawing a rectangle over the hand palm region and recording its corresponding YCbCr values. These thresholded values are used for binarization and hand gesture segmentation. Various sign gestures performed in this experiment include hand sign gesture segmentation results along with their contour information as shown in Fig. 11.



(a) *(b)*

(c) *(d)*



(e) *(f)* *(g)* *(h)* *(i)* *(j)*

Fig. 11. Hand gesture segmentation and contour formation

V. CONCLUSION

The segmentation of hand performed better in YCbCr color space as compared to HSV color space. Also using morphological operations like dilation under specific structuring element or kernel size the background noise in the form of small white pixels are removed. The contour information through image moments can be used for hand centroidtracking and hand gesture recognition. These hand

gesture recognition system can be used in applications like ASL, HCI, robotics, home utility sector, manual traffic signal control systems for traffic police and so on, in the future. Segmentation of hand in low light environment and overlapping of hands are other some of the challenges which should to taken into consideration. Combining these segmentation with hand motion segmentation using optical flow techniques also can probably help in overcoming these challenges. Better segmentation, tracking and recognition algorithm along with improvement in capturing device hardware will help in providing a continuous friendly hand gesture recognition system output possible. With depth cameras having infrared sensors, hand gesture segmentation in any lighting conditions is possible however they are expensive and not easily accessible as compared to simple color cameras.

REFERENCES

1. Siddharth S. Rautaray, Anupam Agrawal, "Vision based hand gesture recognition for human computer interaction: a survey," Springer Science Business Media Dordrecht 2012.

2. Chung-Ju Liao, Shun-Feng Su, Ming-Chang Chen, "Vision-based Hand Gesture Recognition System for a Dynamic and Complicated Environment," IEEE Int. Conf. Syst, Man, & Cybernetics, 2015.

3. Stergios Poularakis & Ioannis Katsavounidis "Low-Complexity Hand Gesture Recognition System for Continuous Streams of Digits and Letters," IEEE Trans. Cybern., vol. 46, no. 9, Sep 2016, pp. 2094-2107.

4. Khamar basha shaik et al., "Comparative Study of Skin Color Detection and Segmentation in HSV and YCbCr Color Space," in Proc. Int. Conf. on Recent Trends in Computing, vol. 57, 2015, pp. 41-48

5. Ying Zhao, Jiayong Yan, "Online Dynamic Hand Gesture Recognition with Multiple Cues," Int. Congress Image and Sig Process, pp. 219-223.

6. Jan Hendrik Hammer, Michael Voit, Jurgen Beyerer. "Motion Segmentation and Appearance Change Detection Based 2D Hand Tracking," International Conference on Information Fusion Heidelberg, pp. 1743-1750, July, 2016.

7. Yuan Yao, and Yun Fu, "Contour Model-Based Hand-Gesture Recognition Using the Kinect Sensor," IEEE Trans. Circuits And Systems For Video Technology, vol. 24, no. 11, Nov 2014.

8. G. Bradski, A. Kaebler, "Learning OpenCV 3," 1st ed., United States, O'Reilly Media, 2016, pp 1037-1071.

9. Rafael C. Gonzalez and Richard Eugene Woods, "Digital image processing," pearson international edition, pearson education 3rd ed.

10. Suzuki, S., and K. Abe, "Topological structural analysis of digital binary images by border following," Computer Vision, Graphics and Image Processing 30 (1985): 32–46.

11. Zhang Qiu-yu, Lu Jun-chi, Zhang Mo-yi, Duan Hong-xiang and Lv Lu, "Hand Gesture Segmentation Method Based on YCbCr Color Space and K-Means Clustering," International Journal of Signal Processing, Image Processing and Pattern Recognition, vol. 8, No. 5 (2015), pp. 105-116.