Color Hand Gesture Segmentation for Images with Complex Background

Avinash Babu.D*, Dipak Kumar Ghosh and Samit Ari
Department of Electronics and Communication Engineering
National Institute of Technology Rourkela, Rourkela, India
Email:*avi.pandu6@gmail.com, dipakkumar05.ghosh@gmail.com, samit.ari@gmail.com
Telephone: +91-661-22462464/4464

***Abstract*—The aim of this paper is to automatically segment the hand gesture from a given image under different luminance conditions and complex backgrounds. The luminance value affects the color component of an image which leads to increase the noise level in the segmented image. This paper proposes a combined model of two color spaces i.e., HSI, YCbCr and morphological operations with labeling to improve the segmentation performance of color hand gesture from complex backgrounds in terms of completeness and correctness. The proposed color model separates the chrominance and luminance components of the image. The performance of the proposed method is demonstrated through simulation and the experimental results reveal that proposed method provides better performance accuracy compared to the HSI and YCbCr methods individually in terms of correctness and completeness.**

***Index Terms* — Segmentation, Color space, Labeling, YCbCr, HSI, Chrominance, Luminance.**

I. INTRODUCTION

Hand gesture recognition is an essential task of today's research due to the increasing demands for human–computer interactions (HCIs) in recent years. Image segmentation is the process of partitioning an image into its constituent components i.e. homogeneous and meaningful regions acceding to their identical set of properties or attributes. Segmentation algorithms are based on different parameters of an image like gray-level, color, texture, depth or motion [1]. The image segmentation process can be considered as one of the basic and important steps in digital image processing and computer vision applications such as tracking, pattern recognition and object identification. It is easy to distinguish the objects from the simple back ground but extraction of objects from the complex background of a digital image has been a challenging task in the field of digital image processing. With the increasing demand for complex image analysis and interpretation, the demand for accurate segmentation of images has also grown stronger and as a result many image segmentation methods and algorithms have been developed over the past few decades. The most popular method to perform image segmentation is gray level segmentation method which is based on thresholding because it is simple and having a high speed of operation and ease of implementation. However the disadvantage of thresholding method is performance limited and suitable for only simple background images.

All the color spaces are mathematical representation of a set of colors. All the color spaces are derived from the RGB information supplied by devices such as cameras and scanners. The most common are YCbCr, HSV, HSI, color spaces. The HSV (hue, saturation, value) color space is developed to be more intuitive in manipulating color and designed to approximate the way humans perceive and interpret color. The HSV color space is preferred for manipulation of hue and saturation i.e. to shift color or adjust the amount of color since it yields a greater dynamic range of saturation [1].

The HSI (hue, saturation and intensity) is similar to HSV model. The main difference between these two models is the computing of the brightness component (I and V), which determines the distribution and dynamic range of both the brightness and saturation. The HSI method is best color space for the traditional image processing function like Convolution, Equalization, and Histogram [2]. The YCbCr is another color space unlike the RGB color space, here the luminance or brightness or intensity is separated from the chrominance or pure color value [3]. The value of Y represents the luminance value and Cb and Cr represents the color or chrominance value, these are also known as color difference of the image.

In this paper, a combinational method of HSI and YCbCr is proposed to improve the segmentation performance. In the proposed model, brightness and luminance components are decoupled from the color information of the image which is not possible in the RGB color space. In this proposed method, the segmentation process is carried out by taking the Cb, Cr and H, S values into consideration and morphological operations with labeling is done to improve the performance of segmentation. This method uses unsupervised segmentation algorithm, hence no manual adjustment of any design parameter is needed in order to suit any particular input image. Moreover, the algorithm can be implemented in real time, and its underlying assumptions are minimal. In fact, the only principal assumption is that the person's hand must be present in the given image. In this paper, the proposed segmentation model is tested on subset of complex background color hand gesture images of standard database. The experimental results show that proposed model gives better performance in terms of completeness and correctness compared to HSI and YCbCr methods.

The remainder of the paper is organized as follows. Section 2 describes proposed methodology. Section 3 presents our results and discussion. Finally, the conclusion is given in Section 4.

II. METHODOLOGY

In this paper, a hybrid method is proposed by combining HSI and YCbCr color models to extract the hand gesture from the images with complex background. The outline of the proposed method is shown in Fig 1.

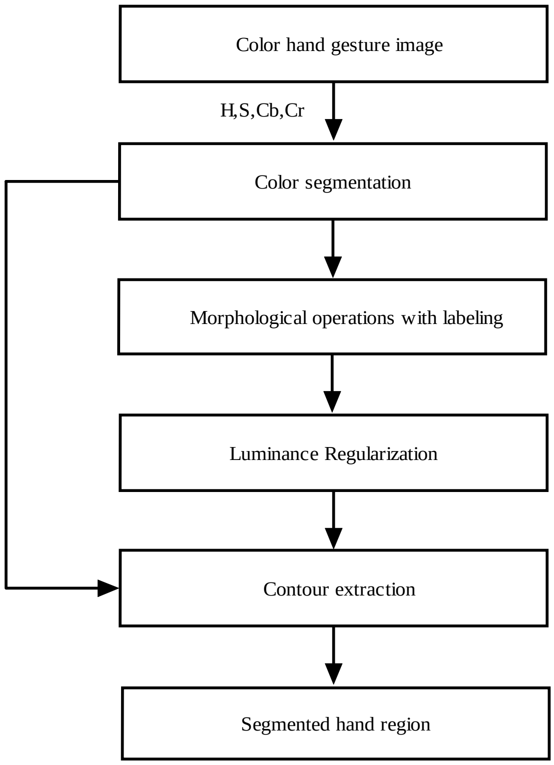

Fig 1: Proposed hybrid model.

The steps of proposed hybrid model are described as follows:

Step 1: *Color Segmentation*

The aim of this step is to divide the image into two classes one is skin color and another is non skin color. In this step, the color segmentation process is carried out by taking Cb, Cr and H, S values in to consideration. For this purpose, the skin color distribution in the YCbCr and HSI color space is derived. By using the histogram method we have found that a skin-color region can be identified by the presence of a certain set of chrominance (i.e, Cr and Cb) values narrowly and consistently distributed in the YCbCr color space. The location of these chrominance values have been found and can be illustrated using the histogram of the skin color diagram which is shown in Fig 2.

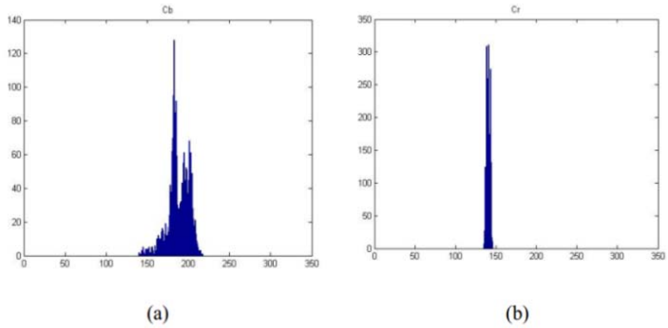

Fig 2: Histogram of Cb and Cr of skin color.

The threshold value for the HSI method is calculated using the histogram method. By using histogram method we have been found that range of H and S for skin color is 0.01 to 0.43 and 0.001 to 0.278. The histogram of H, S for different skin colors is shown in Fig 3.

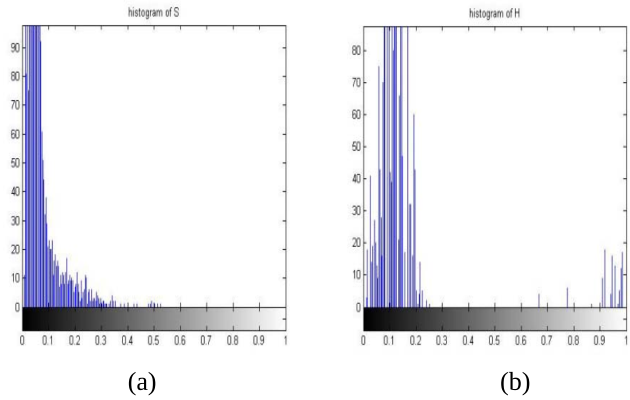

Fig 3: Histogram of H and S of skin color.

The range of H and S for segmentation is 0.01 to 0.43 and 0.001 to 0.278 and for Cb and Cr is 118 to 255 and 128 to 165 which are constant for database system. The output pixel at point (x, y) is classified as skin color and set to one if H, S, Cr and Cb value at that point falls inside their respective ranges R_H, R_S, R_{cb}, and R_{cr}. Other-wise, the pixel is classified as non-skin color and set to zero. The segmentation equation can be written as

$$o_1(x,y) = 1, if\,[H(x,y) \in R_H] \cap [S(x,y) \in R_S] \cap [C_r(x,y) \in R_{cr}] \cap [C_b(x,y) \in R_{cb}]$$

$$0, \text{otherwise.} \qquad (1)$$

where $x = 0,1,2 \ldots, \left(\frac{M}{4} - 1\right)$ and $y = 0,\ 1,2 \ldots, (\frac{N}{4} - 1)$. R_H, R_S, R_{cb}, and R_{cr} represents the respective histogram skin color ranges of H, S, Cb, and Cr respectively. The output resolution of the segmented bitmap image in this stage is same as input image i.e.$M \times N$.

Step 2: *Morphological operations with labeling*

The main purpose of this step is to reduce the noise in the image using erosion and dilation process. The density map for the proposed model can be written as

$$D(x,y) = \sum_{i=0}^{3}\sum_{j=0}^{3} o_1(4x+i, 4y+j) \qquad (2)$$

where $x = 0, 1, 2 \ldots, \left(\frac{M}{4} - 1\right)$ and $y = 0, 1, 2 \ldots, (\frac{N}{4} - 1)$. The equation (2) gives the density value of a 4×4 window of the image and its value lies within the range of 0 to 16. This density map is divided in to 3 parts, full density, intermediate density and zero density points. A group of points with zero density value will represent a non skin region, while a group of full density points will signify a cluster of skin-color pixels and a high probability of belonging to a skin region. Any point of intermediate density value will indicate the presence of noise. Once the density map is derived, the process termed as density regularization starts.

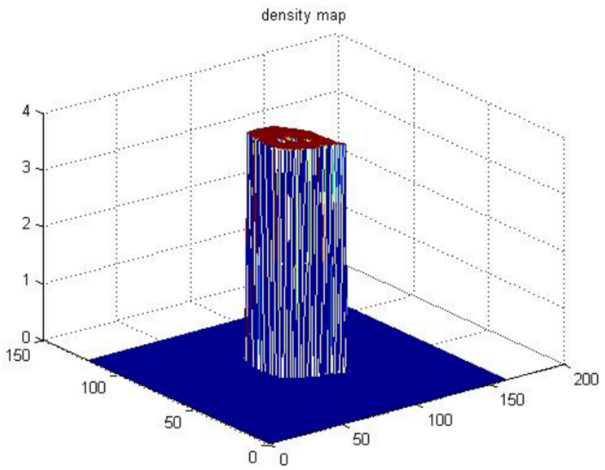

Fig 4: Density map of a hand image.

After finding the density map of the image the four steps of operation for density regularization is performed. Density regularization involves the following four steps:

• Discard all points at the boundary or edge of the density map, i.e., set D(0, y)=D($\frac{M}{4} - 1$, y) =D(x,0)= D(x,$\frac{N}{4} - 1$)=0. for all $x = 0, 1, 2 \ldots, \left(\frac{M}{4} - 1\right)$ and $y = 0,1,2 \ldots, (\frac{N}{4} - 1)$.
• Erode any full-density point (i.e., the points having value 16 set to zero) if it is surrounded by less than five other full-density points in its local 3×3 neighborhood.
• Labeling works by scanning an image pixel by pixel (from top to bottom and left to right) in order to identify connected pixels. It removes all the minimum connected parts in image (i.e., noise) which is unwanted and maximum connected part will present in the image.
• Dilate any point of either zero or intermediate density (i.e., set to 16) if there are more than two full-density points in its local 3×3 neighborhood.
Then the density map is converted to bitmap using the equation (3).

$$o_2(x,y) = \begin{cases} 1, if D(x,y) = 16 \\ 0, \text{otherwise} \end{cases} \qquad (3)$$

for all $x = 0, 1, 2 \ldots, \left(\frac{M}{4} - 1\right)$ and $y = 0, 1, 2 \ldots, (\frac{N}{4} - 1)$.

Step 3: *Luminance Regularization*

In complex background images, the brightness is non uniform throughout the skin region, while the background region tends to have a more even distribution of brightness. Hence, in this stage based on these characteristics, background region that was previously detected due to its skin-color appearance can be further eliminated. The analysis employed in this stage involves the spatial distribution characteristic of the luminance values since they define the brightness of the image. In this proposed method, it uses standard deviation as the statistical measure of the distribution. The size of the previously obtained bitmap of $o_2(x,y)$ is (M/4 × N/4). Hence the each pixel of the image will have information of 4 ×4 window. 'W' is the intensity value represented by the each pixel of the image and its standard deviation of its corresponding group of luminance values can be written as

$$\sigma(x,y) = \sqrt{E[W^2] - (E[W])^2)} \qquad (4)$$

The bitmap equation for stage 3 can be written as

$$o_3(x,y) = \begin{cases} 1, if\, o_2(x,y) = 1 \; and \; \sigma(x,y) \geq 2 \\ 0, \text{otherwise} \end{cases} \qquad (5)$$

for all $x = 0, 1, 2 \ldots, \left(\frac{M}{4} - 1\right)$ and $y = 0, 1, 2 \ldots, (\frac{N}{4} - 1)$.

Step 4: *Contour Extraction*

In this final step, the (M/4 × N/4) output bitmap of step four is converted back to the dimension of (M × N). To achieve the increase in spatial resolution, this work utilizes the edge information that is already made available by the color segmentation in step one. Therefore, all the boundary points in the previous bitmap is being mapped into the corresponding group of 4×4 pixels with the value of each pixel as defined in the output bitmap of step one.

III. EXPERIMENTAL RESULTS AND DISCUSSION

In this paper, the proposed segmentation model is tested on subset of complex background color hand gesture images of standard database [4]. This data base consists of 60 complex background images of 15 hand postures performed by 4 persons under different luminance conditions.

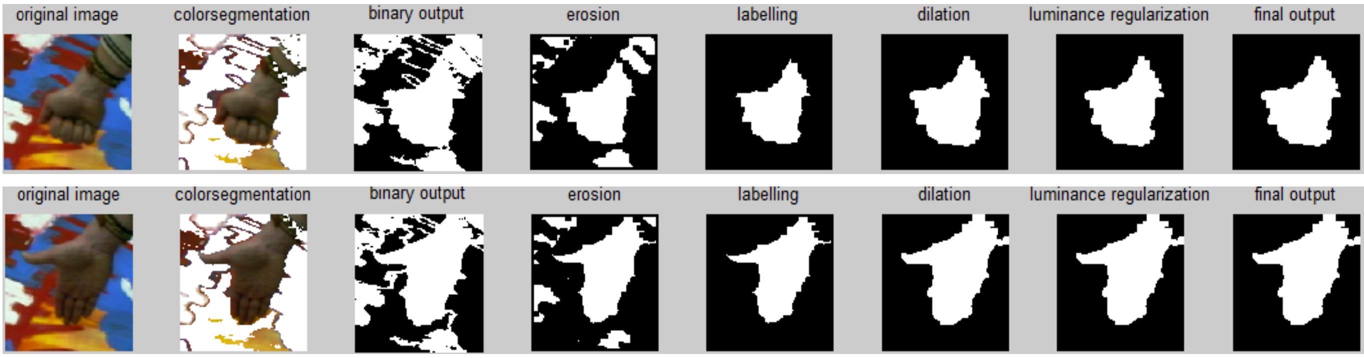

Fig 5: Step-wise output of Proposed method for two hand gesture images.

The step wise output of the proposed method is shown in Fig 5. The performance of the proposed method is compared with HSI [5] and YCbCr [6] methods in terms of completeness and correctness [7]. The definition of completeness and correctness are described below.

• Completeness is the percentage of the ground truth region extracted by the segmentation algorithm.

• Correctness can be defined as the percentage of correctly extracted region (ground truth) by the segmentation algorithm.

The correctness and the completeness can be calculated using the following formulas.

$$Correctness = \frac{TP}{TP + FP} \times 100\% \quad (6)$$

$$Completeness = \frac{TP}{TP + FN} \times 100\% \quad (7)$$

• TP (True Positive image) is the common part of Ground truth image and segmented image.

• FN (False Negative image) difference between ground truth image and true positive image.

• FP (False Positive image) difference between segmented image and true positive image.

Here the ground truth image represents the actual output. In this work, the ground truth image is manually created by using *Gimp software* [8].

Fig 6: Segmented images and the corresponding Ground truth images of the Proposed model.

The ground truth images and corresponding output of the proposed hybrid model is shown in Fig 6. Table 1 shows the comparative performance of the proposed segmentation algorithm in terms of correctness and completeness with respect to HSI [5] and YCbCr [6] methods. The proposed method yields 90.47% of correctness and 98.77% of completeness where as HSI method provides 84.54% and 93.41% and YCbCr method provides 86.30% and 96.55% respectively. Therefore, the proposed method provides better performance for the color segmentation as compared to the earlier reported HSI [5] and YCbCr [6] methods.

TABLE I.COMPARISON RESULT

Method	**Correctness in%**	**Completeness in%**
HSI [5]	84.54	93.41
YCbCr [6]	86.30	96.55
Proposed Model	90.47	98.77

IV. CONCLUSION

In this paper, a combinational method of HSI and YCbCr along with the morphological operations with labeling is proposed for background noise reduction and better performance of segmentation. In this work the hand gesture image is segmented by performing region segmentation using skin color map. This is feasible because the skin color distribution is different from background color distribution. By using skin color map, this work classifies pixel of the input image into skin color and non-skin color. Experimental results reveal that the proposed method provides better performance accuracy compared to HSI and YCbCr methods in terms of correctness and completeness. This technique can be used in the field like bio-medical imaging, satellite imaging and robotics, face detection, gesture recognition where accuracy has the higher priority over complexity.

REFERENCES

[1] R. C. Gonzalez, R. E. Woods., Digital Image Processing. Pearson Education Third Edition, 2009.

[2] P. L. Correia and F. Pereira, "Objective Evaluation of Video Segmentation Quality," *IEEE transactions On Image Processing*, vol. 12, no. 2, Feb. 2003.

[3] D. Chai and K. M. Nagan , "Face segmentation using skin color mapping video phone applications," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 9, no. 4, June 1999.

[4] Database is available online (http://www.idiap.ch/marcel/Databases/ main.html).

[5] G. M. N. R Gajanayake and R. D. Yapa, "Comparison of Standard Image Segmentation Methods for Segmentation of Brain Tumors from 2D MR Images," *International conference on Industrial and Information Systems*, Dec. 2009.

[6] M. Silveir, "Comparison for Segmentation Method for Melanoma Diagnosis in Dermoscopy Images," *IEEE Journal of Selected Topics in Signal Processing*, vol. 3, no. 1, Feb. 2009.

[7] Y. Zhu, H. Ren, G. Xu, X. Lin, "Toward real-time human – computer interaction with continuous dynamic hand gestures," in *Proc. Fourth Intl.Conf. Automatic. Face Gesture Recognition, Grenoble, France, IEEE Comput. Soc.*, pp. 544 – 549, 2000.

[8] N. Otsu, "A Threshold Selection Method from Gray-Level Histogram," *IEEE Trans. Systems Man, and Cybernetics*, vol.9, pp.62-66, 1979.

[9] C. K. P. Clarke, Color Encoding and Decoding Techniques for Line-Locked Sampled PAL and NTSC Television Signals, BBC Research Department Report BBC RD 1986/2.

[10] M. T. Chan, "HMM-based audio-visual speech recognition integrating geometric and appearance-based visual features," in *IEEE Fourth Workshop on Multimedia Signal Processing, Cannes, France*, October 2001, pp.9–14.

[11] M. Kolsch and M. Turk, "Robust Hand Detection," in *Proc. IEEE Intl. Conference on Automatic Face and Gesture Recognition*, May 2004.