A DYNAMIC THRESHOLD APPROACH FOR SKIN SEGMENTATION IN COLOR IMAGES

Pratheepan Yogarajah, Joan Condell, Kevin Curran, Abbas Cheddad and Paul McKevitt

Intelligent Systems Research Centre (ISRC), School of Computing and Intelligent Systems,
University of Ulster, United Kingdom.

ABSTRACT

This paper presents a novel dynamic threshold approach to discriminate skin pixels and non-skin pixels in color images. Fixed decision boundaries (or fixed threshold) classification approaches are successfully applied to segment human skin. These fixed thresholds mostly failed in two situations as they only search for a certain skin color range: 1) any non-skin object may be classified as skin if non-skin objects's color values belong to fixed threshold range. 2) any true skin may be mistakenly classified as non-skin if that skin color values do not belong to fixed threshold range. Therefore in this paper, instead of predefined fixed thresholds, novel online learned dynamic thresholds are used to overcome the above drawbacks. The experimental results show that our method is robust in overcoming these drawbacks.

***Index Terms*—** Skin segmentation, eye detection, skin tone, dynamic threshold, color images.

1. INTRODUCTION

Skin segmentation means differentiating skin regions from non-skin regions in an image. In the past, many techniques have been developed and successfully applied for skin segmentation using color information. The image pixels representation in a suitable color space is the primary step in skin segmentation in color images. A survey of different color spaces (e.g. RGB, YCbCr, HSV, CIE Lab, CIE Luv and normalised RGB) for skin-color representation and skin-pixel segmentation methods is given by Kakumanu et al. [1]. In [2], the conversion, which normalizes RGB color space, reduces brightness dependence. In [3], a fast and simple face detection is proposed to dynamically define ROI (Region-Of-Interest) according to Cr and RGB variance.

On top of the suitable color space selection a satisfactory classifier is needed to perform a best skin region segmentation. A satisfactory skin classifier must be able to discriminate between skin and non skin pixels for a wide range of people with different skin types such as white, yellow, brown and dark, etc. Skin segmentation approaches that rely on pixel-level classification often used a fixed decision boundaries technique. In the decision boundaries (also called fixed threshold) classification technique, the decision boundaries are often fixed by the designer to obtain best skin segmentation results. However, many of the decision boundaries techniques are limited in performance due to real-world conditions such as illumination and ethnicity.

Recently, Cheddad et al. [4, 5] developed a new skin segmentation method in color images using YCbCr color space and a fixed threshold classification technique. They compared their experimental results with other existing methods [6, 7] and showed very promising results. Also they argue that their method is insensitive with ethnicity and robust to illumination as well. In their approach, fixed threshold values are learned off-line using a certain amount of manually identified skin pixels and each image pixel is checked to see if its color value satisfies the fixed threshold values. They also mentioned that, like all existing algorithms, their method is not yet intelligent enough to discriminate whether a scene contains a skin colour or something that looks similar to it [5].

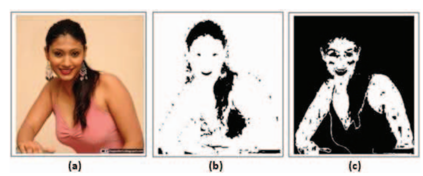

Fig. 1. (a) represents the sample image[1], (b) represents skin segmentation result using [4] and (c) represents skin segmentation result using our method. White pixels represent skin and black non skin in (b) and (c).

Skin color varies greatly between different human races and very often a fixed threshold for skin boundary is learned from a certain amount of different skin colors. Therefore skin segmentation that uses fixed threshold values may fail in unconstrained imaging conditions. Through experimentation we find that, as in Cheddad et al. [5], their fixed threshold classification technique is not robust to separate skin region from non skin region which have colors more similar to skin, see Fig 1.

Therefore we propose a new skin segmentation technique for color images that makes [4] intelligent and robust to separate skin from non skin which have colors more similar to

This work was part funded by the Invest NI Proof of Concept (PoC) fund.

[1]http://lankaactress.blogspot.com, last visit: 5th July 2010.

skin. The proposed technique uses the same color space used in [4, 5] as it is proved as robust with illumination. But we find the fixed threshold used in [4, 5] is limited in separating skin and non skin in color images. So instead of fixed threshold values we calculate dynamic threshold values via on-line learning by taking the color information of human face regions. Zheng et. al [8] argue that using the skin color property of the detected face region, the remainder of skin pixels in the image can be detected. Our skin segmentation method also takes advantage of the fact that the face and body of a person always shares the same colors. Using the detected face region color values, we calculate a dynamic threshold and our experimental results show that dynamic threshold is more robust in skin segmentation performance than previous experimental results [4, 5], see Fig 1.

Further more, Fig 2 shows the performance of different skin-color detection approaches for the $Suzie$ image. Obviously, our proposed method can obtain more precise skin regions and result in less erroneous pixels than methods of [2] and [3]. Figures 2 (b) and (c) are obtained from [9].

Fig. 2. (a), (b), (c) and (d) represent $Suzie$ image and skin segmentation results using [2], [3] and our proposed method respectively. White pixels represent skin and black non skin in (b), (c) and (d).

This paper is organised as follows. In section 2, a classification point of view skin segmentation is analysed. Our proposed method and results and discussion are explained in sections 3 and 4 respectively. The conclusion is given in section 5.

2. SKIN SEGMENTATION A CLASSIFICATION POINT OF VIEW

From a classification point of view, skin segmentation can be viewed as a two class problem: skin-pixel vs. non skin-pixel classification. Different researchers have used different techniques to approach this problem. One of the easiest and often used methods is to define fixed decision boundaries for different color space components. Single or multiple ranges of threshold values for each color space component are defined and the image pixel values that fall within these predefined range(s) for all the chosen color components are defined as skin pixels.

Dai and Nakano [10] used a fixed range on I component in YIQ space for detecting skin pixels from images containing mostly people with yellow skin. All the pixel values in the range, $R_I = [0, 50]$ are described as skin pixels in this approach. Sobottka and Pitas [11] used fixed range values on the HS color space. The pixel values in the range $R_H = [0, 50]$ and $R_S = [0.23, 0.68]$ are defined as skin pixels. Chai and Ngan [12] proposed a face segmentation algorithm in which they used a fixed range skin-color map in the CbCr plane. The pixel values in the range $R_{Cb} = [77, 127]$, and $R_{Cr} = [133, 173]$ are defined as skin pixels. Wang and Yuan [13] have used threshold values in rg space and HSV space. The threshold values in the range $Rr = [0.36, 0.465]$, $Rg = [0.28, 0.363]$, $RH = [0, 50]$, $RS = [0.20, 0.68]$ and $RV = [0.35, 1.0]$ are used for discriminating skin and non-skin pixels.

Based on the above observations it is noticeable that most of the fixed threshold based skin classifiers were successfully applied in controlled imaging conditions, e.g. segment white and yellow skins. In order to handle uncontrolled imaging conditions a dynamic threshold classification technique is needed and so a new dynamic threshold approach is developed in our proposed method. The next section explains our proposed method.

3. OUR PROPOSED METHOD

As discussed previously, the color space discussed below is based on previous work [4]. The utilized transformation matrix is defined in Eq.1.

$$\hat{\alpha} = [0.298936021293775390 \;\; 0.587043074451121360 \;\; 0.140209042551032500]^T \quad (1)$$

where the superscript T denotes the transpose operator to allow for matrix multiplication. Let ϕ denote the 3D matrix containing the RGB vectors of the host image and let $x \epsilon [1, 2, ..., n]$, where n = length(R) = length(G) = length(B). Note that this method acts here on the RGB colours stored in double precision, i.e., linearly scaled to the interval [0 1]. The initial colour transformation is given in Eq. 2.

$$I(x) = (\phi(r(x), g(x), b(x)) \otimes \hat{\alpha}) \quad (2)$$

where $\otimes$ represents matrix multiplication. This reduces the RGB colour representation from 3D to 1D space. The vector I(x) eliminates the hue and saturation information whilst retaining the luminance. It is therefore regarded formally as a grayscale colour. Next, the algorithm tries to obtain another version of the luminance but this time without taking the R vector into account. Most skin colour tends to cluster in the red channel. The discarding of red colour is deliberate, as in the final stage it will help to calculate the error signal. Therefore, the new vector will have the largest elements taken from G or B as

$$\hat{I}(x) = arg \;\; max_{x \epsilon 1,...,n}(G(x), B(x)) \quad (3)$$

Eq. 3 is actually a modification of the way HSV (Hue, Saturation and Value) computes the V values. The only difference is that the method does not include in this case the red component in the calculation. Then for any value of x, the error signal is derived from the calculation of element-wise subtraction of the matrices generated by Eq. 2 and Eq. 3 which can be defined as $e(x) = I(x) - \hat{I}(x)$.

A collection of 147852 pixel samples was gathered from different skin regions exhibiting a range of races with extreme variations in lighting effects. Lower and upper decisions boundaries were calculated as 0.02511 and 0.1177 respectively. Finally $e(x)$ values in the range $[0.02511, 0.1177]$ are classified as skin.

Our experimental results showed that their fixed threshold $[0.02511, 0.1177]$ can not perform efficiently when background and cloths colors are more similar to skin color. Here our motivation is to get a dynamic threshold value via online learning for a particular image using human face skin color information. To get the face skin color information, human eyes are detected using the Machine Perception Toolbox [14]. When the eyes are detected, see Fig 4(a), the elliptical face region is generated, see Fig 4(b), using an elliptical mask model, see Fig 3.

Here (x_0, y_0) is the centre of the ellipse as well as the eyes symmetric point. Minor and major axes of ellipse are represented by $1.6D$ and $1.8D$ respectively, where D is distance between two eyes.

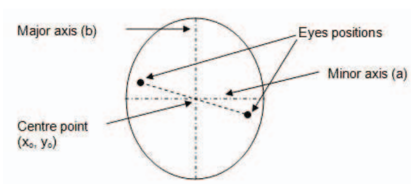

Fig. 3. Elliptical mask model generated using eyes coordinates.

We can see that the detected face region contains the smooth (i.e. skin) and non-smooth (i.e. eyes and mouth etc.) textures. As we only want to keep the smooth regions, the non-smooth regions are detected and removed. It is well known that edge pixels can be generated in non-smooth regions [15]. We applied the Sobel edge detector as it has computational simplicity [16], see Fig 4(c). The detected edge pixels are further dilated using dilation operation [15] to get the optimal non-smooth regions, see Fig 4(d). Finally the calculated non-smooth region is subtracted from the face region and smooth region is obtained, see Fig 4(e). Finally, the values from $e(x)$ which correspond to the detected smooth region are considered for the dynamic threshold calculation. For simplicity we assume those correspondence $e(x)$ values stored in array SR. Fig 5(a) shows the frequency distribution of SR.

Even though dilation operation is applied, there is no guaranty that 100% of the non-smooth regions are removed properly. Therefore a two-sided 95% confidence interval of a normal distribution, $\mathcal{N}(\mu, \sigma^2)$, is fitted on SR to generate

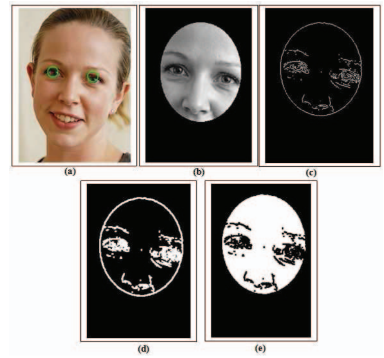

Fig. 4. (a), (b), (c), (d) and (e) represent eye detection, face region, edge detection, dilated image and smooth region respectively.

the dynamic threshold, see Fig 5(b). Here μ and σ are mean and standard deviation of SR and the lower and upper boundaries of the dynamic threshold values calculated based on the confidence interval of normal distribution. Based on these calculated dynamic threshold values, each and every pixel in the color image is classified as skin and non-skin. When more than one face region is detected in the image, each region is used to construct a dynamic threshold and each threshold is used to perform skin segmentation on the whole image. The final result is a logical OR combination of each of the segmented regions obtained respectively with each dynamic threshold. The next section shows some experimental results using our dynamic threshold approach.

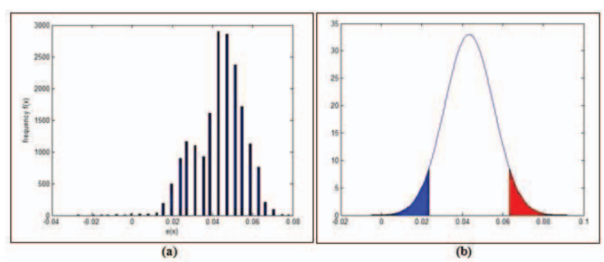

Fig. 5. (a) and (b) represent the frequency and Normal distribution with 95% confidence interval respectively.

4. RESULTS AND DISCUSSION

To carry out skin segmentation, a set of images was downloaded randomly from Google. These random images may have been captured with a range of different cameras using different color enhancement and under different illumination. The method of Cheddad et al. [4] and our proposed method are applied and Fig 6 shows a sample of results obtained for four images. From the results it can be firmly argued that our dynamic threshold method can optimally discriminate skin from non-skin, even though non-skin color is more similar to skin color, than Cheddad et al.'s [4]. Further more, Fig

Fig. 6. columns (a), (b) and (c) represent random images from google, skin segmentation using Cheddad et al. [4] and our methods respectively.

6 shows that Cheddad et al.'s method detected human skin and also detected much non-skin objects as skin. But if we further analyse the third row of Fig 6 then we can see that Cheddad et al.'s method failed to detect most of the skin area. Therefore as we discussed previously in this paper about the drawback of fixed threshold methods, Cheddad et al.'s method also searches for specific ranges of skin colors. We can conclude that our method is more robust for any skin color as our method learns dynamic threshold values instantly from the current image.

5. CONCLUSION

The proposed skin segmentation technique performs better than the fixed threshold skin segmentation proposed previously [4]. The proposed method is robust to imaging conditions and not biased by human ethnicity. Two main drawbacks with fixed threshold classification outlined in this paper are solved by our novel dynamic skin segmentation technique. However, it is noted that if there are no eyes detected in the image then our method can not be applied to skin segmentation problems.

6. REFERENCES

[1] P. Kakumanu, S. Makrogiannis, and N. Bourbakis, "A survey of skin-color modeling and detection methods," in *Pattern Recognition*, 2007, pp. 1106–1122.

[2] M. Soriano, B. Martinkauppi, S. Huovinen, and M. Laaksonen, "Skin detection in video under changing illumination conditions," in *ICPR*, 2000, pp. 839–842.

[3] M.J. Chen, M.C. Chi, C.T. Hsu, and J.W. Chen, "Roi video coding based on h.263+ with robust skin-color detection technique," in *IEEE Trans. Consum. Electron*, 2003, pp. 724–730.

[4] A. Cheddad, J.V. Condell, K.J. Curran, and P. McKevitt, "A new colour space for skin tone detection," in *In Proceedings of The IEEE ICIP*, 2009, pp. 7–11.

[5] A. Cheddad, J.V. Condell, K.J. Curran, and P. McKevitt, "A skin tone detection algorithm for an adaptive approach to steganography," in *Elsevier - Journal of Signal Processing*, 2009, pp. 2465–2478.

[6] R.L. Hsu, M. Abdel-Mottaleb, and A.K. Jain, "Face detection in color images," in *PAMI*, 2002, pp. 696–702.

[7] J. Kovac, P. Peer, and F. Solina, "Human skin colour clustering for face detection," in *in Proc. of Int. Conf. on Computer as a Tool*, 2003, pp. 144–148.

[8] Qing-Fang Zheng, Ming-Ji Zhang, and Wei-Qiang Wang, "A hybrid approach to detect adult web images," in *PCM (2)*, 2004, pp. 609–616.

[9] M.C. Chi, J.A. Jhu, M.J. Chen, and C.T. Hsu, "H.263+ region-of-interest video coding with efficient skin-color extraction," in *IEEE Int. Conf. on Cons. Elect.*, 2006, pp. 381–382.

[10] Y. Dai and Y. Nakano, "Face-texture model based on sgld and its application in face detection in a color scene," in *Pattern Recognition*, 1996, pp. 1007–1017.

[11] K. Sobottka and I. Pitas, "A novel method for automatic face segmentation, facial feature extraction and tracking," in *Signal Proc. Image Comm.*, 1998, pp. 263–281.

[12] D. Chai and K.N. Ngan, "Face segmentation using skin-color map in videophone applications," in *IEEE Trans. Circuits Syst. Video Technol.*, 1999, pp. 551–564.

[13] Y. Wang and B. Yuan, "A novel approach for human face detection from color images under complex background," in *Pattern Recognition*, 2001, pp. 1983–1992.

[14] I. Fasel, B. Fortenberry, and J. R. Movellan, "A generative framework for real time object detection and classification," in *CVIU*, 2005, pp. 182–210.

[15] R.C. Gonzalez and R.E. Woods, "Digital image processing 3rd ed," in *Prentice Hall*, 2008.

[16] S.L. Phung, D. Chai, and A Bouzerdoum, "Adaptive skin segmentation in color images," in *IEEE International Conference on ASSP*, 2005, pp. 182–210.