A Dynamic Skin Detector Based on Face Skin Tone Color

Nada B. Ibrahim
Faculty of Computers and Informatics
Benha University, Benha
Nada.ahmed@fci.bu.edu.eg

Hala H. Zayed
Faculty of Computers and Informatics
Benha University, Benha
Hala.zayed@fci.bu.edu.eg

Mazen M. Selim
Faculty of Computers and Informatics
Benha University, Benha
SelimM@feng.bu.edu.eg

Abstract

In the last ten years, skin detection has been a milestone in most of the computer vision applications. But till now, there is no robust skin detector. The different degrees of the skin tone color are the obstacle that faces skin detection process. This paper proposes an adaptive skin modeling and detection technique which is based on face skin tone color. Face is a good indicator of different characteristics of skin tone color where it carries significant information about skin color. Skin modeling aims to develop adaptive margins of skin detector. These margins have been obtained after applying an online dynamic threshold to the pixels gathered around the major and minor axes of bounding rectangle of detected face. Experimental results show that the proposed method has promising results compared to state-of-the-art skin detection methods.

1. Introduction

Some computer vision applications rely on skin detection as a milestone. Face recognition [1, 2 and 3], image-content filtering [4], steganography [5] and gesture recognition [6] are some of the most popular computer vision applications that use skin detection as a fundamental step. Skin tone color pixels modeling and recognition is the objective of skin detection process. Challenge that faces skin detection can be confined in: skin tone color variation due to race, skin color-like confusing objects or backgrounds and illumination.

There are two ways to detect skin tone pixels in an image either a pixel-based way or a region-based way. The former classify solitary pixel as skin or non-skin pixel which is considered, from classification point of view, as a two-class problem. While the later, takes spatial arrangement aspects of pixels in consideration.

Skin detection has three main phases. Firstly, pixels are represented in a suitable color space which holds significant meaningful information that easily helps in the detection process. Modeling pixels using an appropriate distribution is the second phase. Finally, the modeled pixels are classified as skin and non-skin pixel clusters.

Although, color space choice is the primary phase for skin detection, there is no consensus on which color space is the best. RGB is the root color space in which any image captured from digital cameras is stored and represented in. But this color space cannot be used in skin detection due to its dependency on illumination and race. By either linear or non-linear transformation of RGB, other color spaces are derived that partially avoid the defects of RGB. A survey of different color spaces is given in [7, 8]. A good color space acts to decrease the overlap between skin and non-skin tone color. Chaves-Gonzalez et al. [9] perform a brief study on ten commonly used color spaces in skin detection to name the best one. In this study, HSV model was the winner and other color spaces that separate the luminance channel obtain high detection rates like YCbCr, YUV, YCgCb, YPbPr and YDbDr. From channel point of view, red channel of any color space carry most of the texture for skin tone color [9].

Skin modeling leads the separation between skin and non-skin pixels by building a decision rule. Explicit, non-parametric, parametric and neural network techniques are the general categories of skin modeling [7, 8]. The explicit techniques use single or multiple fixed boundaries for each color space. Skin pixels are the pixels that fall within these boundaries, others are non-skin pixels. In non-parametric techniques, a histogram for the given color space is built and then converted to a probability density function (PDF). If the PDF of a given pixel exceeds a predefined threshold it is considered as skin. On the other hand, parametric techniques use a modeled color space with a prescribed shape (e.g. Gaussian and elliptical boundary models). Skin pixels fall within a predefined slice of the shape. In neural network techniques, the network is trained by two datasets, one for skin images and the other for non-skin images to extract a decision rule. The most commonly used architectures are multi-layer perceptron (MLP) and self organized map (SOM) networks. A survey of different skin modeling methods is given in [7], where it is reported that the histogram techniques have high performance rate compared to Gaussian mixture method and MLP techniques.

Cheddad et al. [5, 10] introduced a new 1D color space for skin detection. Using this color space, a skin probability map (SPM) was built. An explicit modeling method that used a fixed threshold learned from offline samples. The experiments show significant improvement compared with other methods. But this method confuses with skin-like objects or backgrounds. Yogarajah et al. [11] refine the performance of [5] by using a dynamic threshold that was calculated from color of pixels collected from face detected after removing eyes and mouth as they are the non-smooth texture in face . Finally, 95% confidence interval of dynamic threshold normal distribution is applied on the whole image to identify skin pixels. In this method, if eyes and mouth are not recognized, this method is not applicable. In addition, its false positives (FP) are high where a lot of non-skin pixels are detected as skin.

In this paper, a dynamic skin detector based on face skin tone color is proposed. YCbCr color space is used after discarding the luminance channel. Face detector is applied to a static frontal face image. The values of PDF histogram bins are calculated and trimmed at 0.005. To avoid eyes and mouth regions to be recognized as skin, a threshold is applied on remaining PDF values after trimming. The pixels along the major and minor axes of the bounding rectangle of detected face are used to calculate a dynamic threshold. This threshold is applied to the face image to identify skin pixels. Then the threshold is updated by increasing the pixels around the axes until 95% from face pixels are recognized as skin. Finally, this threshold is applied to the entire image. The experimental results show that the proposed method overcomes the defects of [11].

The upcoming sections are arranged as follows: the proposed approach including a novel modeling and detection method is described in section 2. Results and evaluation are outlined in section 3. Finally, conclusions are given in section 4.

2. The proposed approach

As mentioned previously, skin detection has three main phases: choosing a suitable color space, skin modeling and classification of pixels. For the proposed skin detector, a pre-processing phase is inserted after choosing color space. Each phase is illustrated as in the following sub-sections.

2.1. Color space

In this approach, YCbCr color space is used. This color space is an orthogonal color space that discriminates luminance channel from chrominance channels. As stated in [9], "*luminance channel is a negative point for skin detection*". So, by neglecting luminance channel the dependency of the color space on illumination will fade and the dimensions of this color space will be marked down.

YCbCr is obtained from RGB by applying the following transformation [8]:

$$\text{YCbCr:}\begin{cases} Y = 0.299R + 0.587G + 0.114B \\ Cb = (B - Y) \\ Cr = (R - Y) \end{cases} \quad (1)$$

2.2. Pre-processing

This phase includes image resizing. Experiments show that resizing images to 150 x 150 preserves the important features. Then, filter the image by removing the white and black pixels as they never represent skin tone color. The goal of inserting this phase is to facilitate the upcoming calculations by decreasing the number of pixels to deal with.

2.3. Skin Modeling

The main target of this phase is to get the margins that narrow down the color space to skin tone color sub-space. Margins are calculated with the help of color vector of the face pixels. The face skin tone color is a good indicator for different characteristics of the whole body skin tone color. Viola and Jones face detector [12] is used to extract region of interest (ROI) around the face. ROI is represented by the bounding rectangle around detected face, as shown in Figure1(c). This phase depends on the efficiency of the used face detector. As the bounding rectangle is bounded tightly to the face excluding background, hair and accessories this leads to better results.

(a) (b) (c)

Figure 1. (a) RGB image, (b) YCbCr image and (c) ROI of image.

Let p be the color vector of the pixel. This vector has two components Cb and Cr which represent red and blue chrominance components, respectively. The minimum and maximum margins that control Cb and Cr components for skin cluster are Cbmin, Cbmax, Crmin and Crmax. The pixel is classified as skin if the components of its color vector fall within the ranges:

$$\begin{aligned} Cbmin \leq Cb \leq Cbmax \\ Crmin \leq Cr \leq Crmax \end{aligned} \quad (2)$$

The face by its nature composes of smooth and non smooth regions. Non-smooth regions represented by the eyes and mouth must be discarded from face. Detecting the eyes and the mouth can be achieved by cascade boosting and then removed from face [12]. But this method is a time consuming method. It is observed that the region around the major and minor axes of the bounding rectangle of the detected face includes a little non-smooth region than the entire face. Therefore, the p vectors for these axes are quantized into histogram bins. The value of PDF for each bin is calculated, as shown in Figure 2(a). The PDF is corresponding to the likelihood that the given p belongs to skin cluster. Trim the p vectors bins that has PDF<0.005 from the upper and the lower terminals of p vectors which are marked by an ovals in Figure 2(b). This trimmed p vectors has the worst likelihood to be skin. It may be the earrings or some hair.

As eyes and mouth represent a significant proportion of face, its p vector bin's PDF is greater than 0.005 and falls in the remaining range of p vectors that is marked by two solid lines in Figure 2(c). To avoid eyes and mouth regions to be recognized as skin, a mid-way threshold is applied to the remaining p vectors to overcome confusion and segment image into skin and non-skin clusters. The margins of the cluster with the higher average PDF values are considered as skin cluster margins which are labeled as S in Figure 2(d).

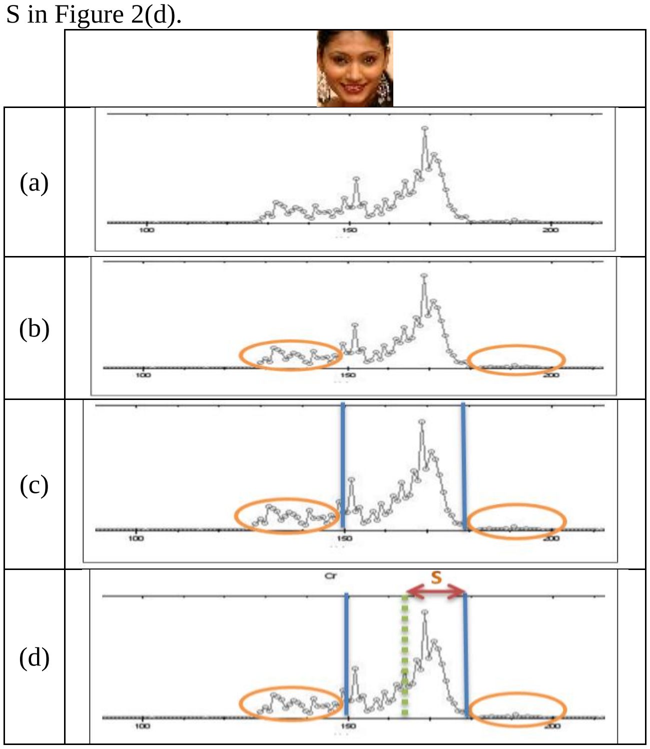

Figure 2. Cr-component of p vector: (a) PDF of histogram bins, (b) trimming bins at PDF<0.005, (c) trimmed PDFs and (d) skin margins after applying mid-way threshold.

Repeat the proposed approach by increasing the width of the region around the major and minor axes of the bounding rectangle of detected face until 95% of the face pixels are labeled as skin. The width is increased by step of two. Figure 3 shows the effect of increasing the width of the region around the major and minor axes of the bounding rectangle of detected face on the entire image not only the face. It is obviously seen that as the width increases the number of detected pixels of the face increases.

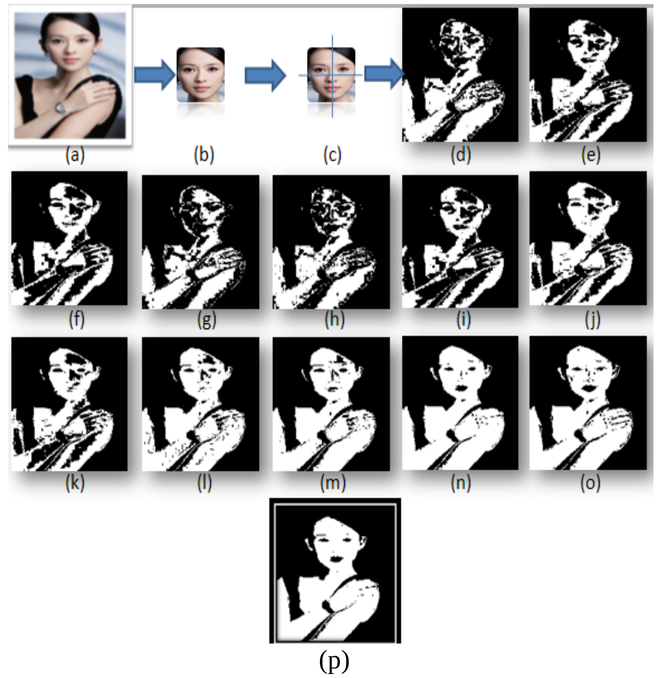

Figure 3. Applying the proposed approach: (a) original image, (b) detected face, (c) identified major and minor axes of the face and calculate mid-way threshold, (d) applying calculated threshold with region width=1, (e-o) increase region width with step 2 (e.g. 3,5,7,9,….. etc.) and (p) resulted image with region width of 25.

2.4. Pixels Classification

The margins that satisfy the condition that 95% of the face pixels are labeled as skin are applied to the entire image to get a binary skin image where the white pixels correspond to skin while black pixels labeled non- skin cluster. If the image contains more than one face, the proposed approach is applied for each face. Then, the results of each individual face are gathered together.

A summarized flow chart of the proposed approach is shown in Figure 4.

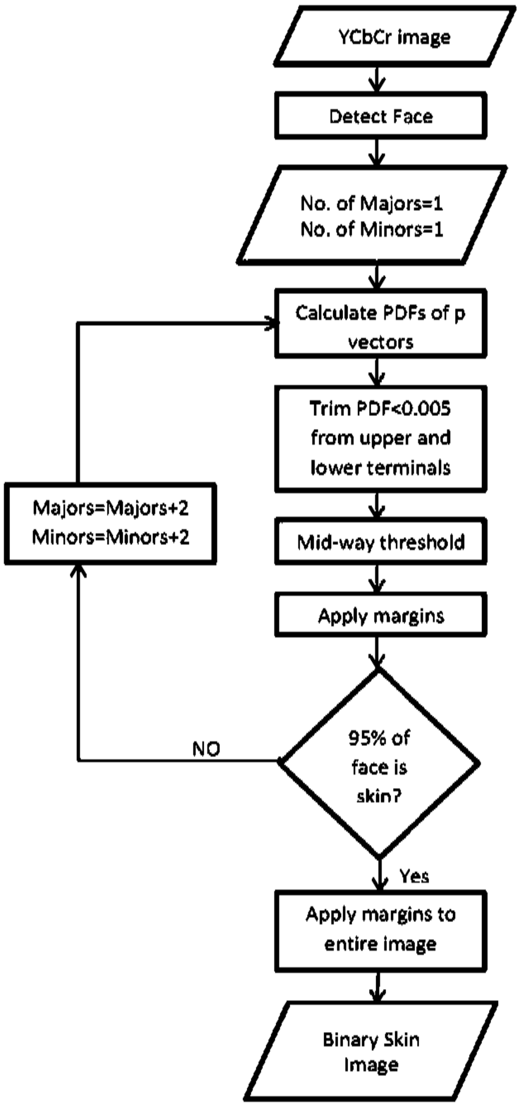

Figure 4. Summarized flow chart for the proposed approach.

3. Results and evaluation

A set of static images that include at least one frontal face were gathered from the web. The corresponding binary skin image is cropped by coloring skin pixels with white and non- skin pixels with black. These images are for people from different races. Some of the images in the set with their corresponding cropped image are shown in Figure 5(a) and (b).Then, the proposed approach is applied and the resulting images are shown in Figure 5(c). It is clear from Figure 5 that the proposed system could correctly detect the skin of people of different races.

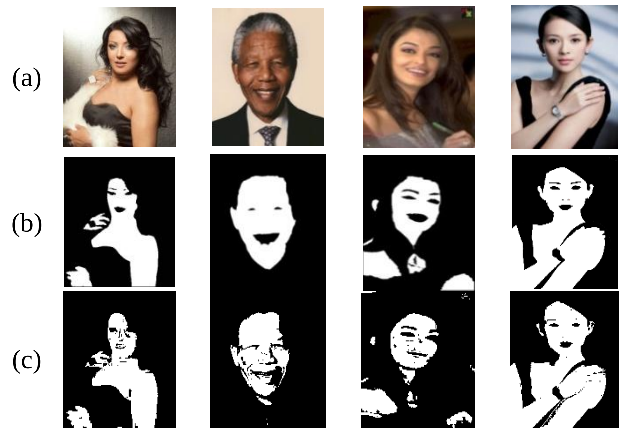

Figure 5. (a) original image, (b) cropped images and (c) results of the proposed approach.

For evaluation purpose, the results of the proposed approach are compared with that of Yogarajah et al. [11] and Cheddad [5, 10] as state-of-art methods. As shown in Figure 6, when applying Yogarajah et al. method on the first image it can be seen that a lot of skin pixels are recognized as non-skin, on the other hand, the proposed method seems to have a much smooth image. For the rest of images except the last image, the proposed method has a little confusion with the skin-like background compared with Yogarajah et al. method. The proposed method failed with the last image because the detected ROI around the face include a lot of non-skin pixels from hair and background.

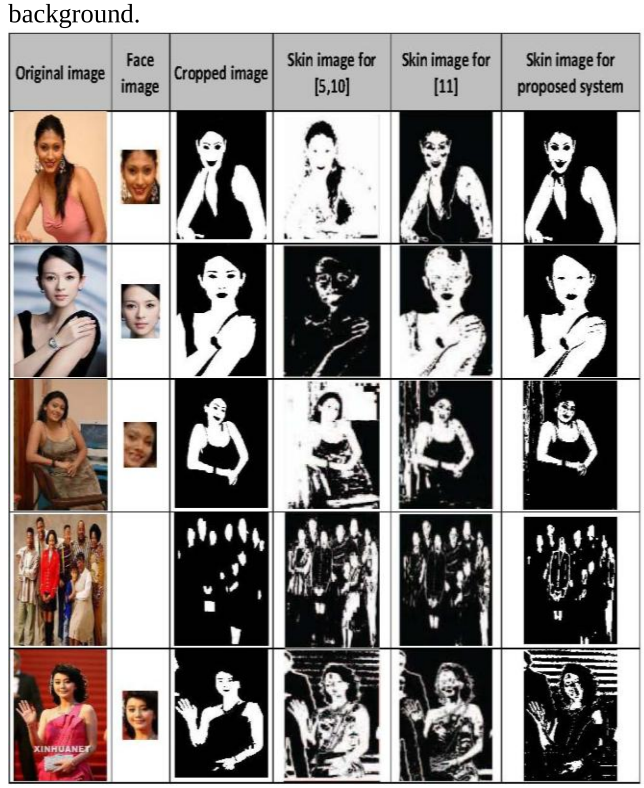

Figure 6. Experiments results.

Image No.	Face size	Region width	% of face	% of entire image
1	51 x 51	11	38%	4%
2	62 x 62	25	64%	11%
3	24 x 24	7	50%	1%
4	-----------	-----------	-----------	-----------
5	33 x 33	11	56%	3%
Average			**52%**	**5%**

Table 1. The effect of the proposed method on the percentage of pixels used for skin detection.

The effect of the proposed method on the percentage of the pixels used in the calculations for both the face and the entire image is illustrated in Table 1. The percentage of the face pixels to deal with when applying the proposed method has an average of 52%. And the average percentage of the pixels from the entire image is nearly 5%. This decreases dramatically the number of pixels intended for calculations. This leads to decreasing the time consumed compared with the other non-parametric histogram methods that use the entire pixels of the image to detect skin.

In the context of skin classification, false positive rates (FPR) and false negative rates (FNR) are used to evaluate skin classifiers. FPR refers to the errors due to recognizing non-skin pixels as skin while the inverse of the previous case is detected by FNR. These scores can be calculated from the following equations [13]:

$$FPR = FP / (FP+TN)$$
$$FNR = FN / (FN+TP) \quad (3)$$

Where: TN, FN, FP and TP are true negative, false negative, false positive and true positive scores; respectively.

A good skin classifier should provide low FNR and low FPR. But there is always a trade-off between these two scores. This means that if one of the two scores increases the other decrease. The results of evaluation are illustrated in Table 2 and Table 3.

Table 2 shows that the proposed method has a less average FNR score than the other methods. On the other hand, the FPR of the proposed approach is nearly half its equivalent for Yogarajah method, as shown in Table 3.

Image No.	Proposed method	Yogarajah [6]	Cheddad [2,3]
1	0.042	0.069	0.011
2	0.044	0.051	0.744
3	0.030	0.023	0.029
4	0.029	0.059	0.018
5	0.176	0.138	0.107
Average	**0.064**	**0.068**	**0.182**

Table 2. FNR scores of the proposed method compared with other reported in literature.

Image No.	Proposed method	Yogarajah [6]	Cheddad [2,3]
1	0.010	0.020	1.926
2	0.021	0.249	0.039
3	0.169	0.477	3.246
4	0.217	0.900	1.450
5	0.593	0.489	1.224
Average	**0.202**	**0.427**	**1.577**

Table 3. FPR scores of the proposed method compared with other reported in literature.

4. Conclusions

The proposed dynamic mid-way threshold histogram skin detector uses the pixels of the region around the major and minor axes of the face to obtain the margins of the skin detector. This approach has better performance than the state-of-art methods. The proposed method succeeded in partially beating the two main defects of Yogarajah method as it does not rely on eye and mouth detection. It lowers the FPR by about half while keeping the FNR approximately the same. It also decreases the number of pixels to deal with to be about 52% of the entire face which dramatically decreases the detection time. So, it is recommended for real-time applications. Finally, it is applicable to face of different races due to its adaptive dynamic nature.

References

[1] M. Shemshaki and R. Amjadifard. Face Detection Based on Fuzzy Granulation and Skin Color Segmentation. *In Proc. of CICSyN*, 219-224, 2011.

[2] A. Tofighi and S.A. Monadjemi. Face Detection and Recognition Using Skin Color and AdaBoost Algorithm Combined with Gabor Features and SVM Classifier. *In Proc. of CMSP*, 1: 141-145, 2011.

[3] S. aya, A. Karaali and A.T. Erdem. Combining Haar Feature and skin color based classifiers for face detection. *In Proc. of ICASSP*, 1497 – 1500, 2011.

[4] S. Elfkihi Lrit, A. Jilbab and D. Aboutajdine. Recognition of adult video by combining skin detection features with motion information. *In Proc. of ICMCS*, 1–5, 2011.

[5] A. Cheddad, J.V. Condell, K.J. Curran, and P. McKevitt. A skin tone detection algorithm for an adaptive approach to steganography. *Elsevier - Journal of Signal Processing*, 89:2465–2478, 2009.

[6] M.M. Zaki and S.I. Shaheen. Sign language recognition using a combination of new vision based features. *Pattern Recognition Letters*, 32:572-577, 2011.

[7] P. Kakumanu, S. Makrogiannis, and N. Bourbakis. A survey of skin-color modeling and detection methods. *Pattern Recognition*, 40:1106–1122, 2007.

[8] V. Vezhnevets, V. Sazonov, A. Andreeva. A survey on pixel-based skin color detection techniques. *In Proc. of GRAPHICON03*, 85–92, 2003.

[9] J.M. Chaves-González, M.A. Vega-Rodríguez, J.A.G. Pulido, and J.M. Sánchez-Pérez. Detecting skin in face recognition

systems: A colour spaces study. *Elsevier - Journal of Digital Signal Processing*, 20:806-823, 2010.
[10] A. Cheddad, J.V. Condell, K.J. Curran, and P. McKevitt. A new colour space for skin tone detection. *In Proc. of IEEE ICIP*, 7–11, 2009.
[11] P. Yogarajah, J. Condell, K. Curran and et al. A dynamic threshold approach for skin segmentation in color images. *In Proc. of IEEE Int. Conf. on Image Processing*, 2225-2228, 2010.
[12] P. Viola, M. J. Jones. Robust real-time face detection. *Int. Journal of Computer Vision*, 57(2):137–154, 2004.
[13] J.P. Vink, T. Gritti, Y. Hu and G. De Haan. Robust skin detection using multi-spectral illumination. *In Proc. of FG*, 448-455, 2011.