Skin segmentation based on multi pixel color clustering models

Sinan A. Naji, Roziati Zainuddin, Hamid A. Jalab *

Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, P.O. Box 50603, Malaysia

Article history:
Available online xxxx

Keywords:
Skin segmentation
Pixel classification
Skin color
Color space

Abstract
This paper presents a reliable color pixel clustering model for skin segmentation under unconstrained scene conditions. The proposed model can overcome sensitivity to variations in lighting conditions and complex backgrounds. Our approach is based on building multi-skin color clustering models using the Hue, Saturation, and Value color space and multi-level segmentation. Skin regions are extracted using four skin color clustering models, namely, the standard-skin, shadow-skin, light-skin, and high-red-skin models. Moreover, skin color correction (skin lighting) at the shadow-skin layer is used to improve the detection rate. The experimental results from a large image data set demonstrate that the proposed clustering models could achieve a true positive rate of 96.5% and a false positive rate of approximately 0.765%. The experimental results show that the color pixel clustering model is more efficient than other approaches.

© 2012 Elsevier Inc. All rights reserved.

1. Introduction

Skin detection is commonly used in determining pixels related to human skin. Skin detection is an important technique in image processing and the most distinctive and widely used key technology in a large number of applications [1,2], such as face detection [3], face tracking, human motion analysis [4], and naked image filters [5]. Skin detection could be defined as the process of identifying skin-colored pixels and regions in an image or a video. Skin detection has two main approaches, namely, pixel-based and region-based. In pixel-based methods, the features (e.g., color) are extracted from information coming from a pixel, whereas in region-based methods, the features (e.g., texture) are extracted from information about a pixel and its neighbors. Color pixel clustering refers to the process of grouping similar color pixels, whereas skin segmentation aims to locate skin regions in an unconstrained input image [6–8]. Color pixel clustering plays an important role in numerous computer vision applications, such as face detection, face tracking, human motion tracking, hand segmentation for gesture analysis, filtering of objectionable Web images, and other human-related image processing applications. In these tasks, the results of skin segmentation enable subsequent object detection to focus on reduced skin regions, instead of the entire input image.

Skin detection methodologies using skin color information as a cue have recently gained considerable attention. Skin color provides computationally effective, yet robust information against variations, scaling, and partial occlusions. Skin detection using color information can be a challenging task because the skin appearance in images is affected by various factors, such as illumination, background, camera characteristics, and ethnicity. Numerous techniques have been presented in the literature for skin detection using color [9]. Most existing skin segmentation approaches are based on skin color. Skin regions are detected by identifying pixels that have skin colors [6]. Only in the last 15 years has the value of skin color segmentation been fully realized in many computer vision applications. Three broad categories of methods exist for skin classification and segmentation, namely, parametric, nonparametric, and explicit threshold-based skin cluster classifiers [7]. The first category uses parametric models for skin color distributions. This model usually comprises a Gaussian or a mixture of Gaussian models used to model the skin color distribution in different color spaces [10]. The second category uses a nonparametric model. This type of method estimates the skin color distribution from the histogram of the training data without deriving an explicit model of the skin color [11]. The third category uses explicit rules on color values [12]. This type of method is generally very simple to implement and is computationally inexpensive. All mentioned approaches explicitly define the boundaries of the skin cluster in a color space.

A number of skin color segmentation approaches have been proposed using different color spaces [9]. Numerous methods have been proposed to build a skin color clustering model in a color space. Ruiz-del-Solar et al. [10] included additional features in their color segmentation methods to obtain more valuable results showing robust brightness variations. Gomez et al. [11] listed the top components from which they developed a hybrid color space. Kim et al. [12] proposed a skin color modeling approach in the
Hue, Saturation, and Intensity color space while considering intensity information. In their approach, they adopted B-spline curve fitting in creating a mathematical model for the statistical characteristics of a color with respect to intensity. Li et al. [13] proposed an algorithm based on the facial saliency map. Chen et al. [14] proposed a hybrid-boost learning algorithm for multi-pose face detection and facial expression recognition. A study of 10 commonly used color spaces for human skin detection and comparisons among them can be found in [15].

The objective of this paper is to segment the human skin area in a given color image using the Hue, Saturation, and Value (HSV) color space, based on the fact that the color distribution of a colored object is invariant, with respect to brightness and saturation variations. In this approach, skin detection using color pixel clustering models is utilized to detect skin areas. A number of experiments have verified that the proposed algorithm shows robust results in terms of irregular illumination variations and abrupt changes in brightness by its application to different image databases. Fig. 1 shows the general system design.

The remainder of this paper is organized as follows: Section 2 presents the proposed skin color segmentation. Sections 3 and 4 present the analysis of the multi-skin color clustering models and the classification boundaries. Sections 5 and 6 show the experimental results of the proposed approach and the conclusion, respectively.

2. Skin color segmentation

Growing research interest has recently been directed toward the problem of skin color segmentation. Most skin color segmentation techniques involve the classification of individual image pixels into skin and non-skin categories based on pixel color.

Algorithms based on skin color segmentation have to deal with sensitivity to the illumination conditions under which the input image is captured. One of the aspects from which we can observe illumination influence is the choice of proper color space. Color spaces, such as RGB, CVM, CIE, and so on, work very well for technical reproduction purposes. However, human vision tends to look at the world in a different fashion. We do not see the color of an object as quantities of primary mixed colors (red, green, and blue). When humans view a colored object, we tend to describe it by its hue (color). The HSV color space tends to be more realistic in that its representation strongly relates to the human perception of color, thus eliminating the influence of illumination when describing the color of an object. As a result, the HSV model is an ideal tool for developing image processing algorithms based on color descriptions. The transformation between HSV and RGB is non-linear [8].

However, skin color segmentation is inadequately robust for dealing with complex backgrounds and different image collections. Complex backgrounds usually increase false positive (FP) errors. An FP refers to the proportion of non-skin pixels classified incorrectly as skin. FP can be solved by the system’s subsequent steps. In this paper, we deal with this problem using reliable, fast, and robust solutions.

Initially, we built a skin-clustering model using the HSV color space and conducted extensive experiments by applying different skin color appearances to identify false negative (FN) errors. We found two main reasons behind these cases. The first error lies on the limitations of the single skin clustering model in covering all skin color appearances, such as dark shadow regions and blackish skin. Strong light reflection may cause skin color information to be lost. In addition, makeup, montage, and image reproduction influence the skin color appearance to a red concentrated appearance.

The second reason lies in the fact that each colored pixel is treated individually in relation to the color space (skin or non-skin pixel), without considering the content of neighboring pixels. Thus, if the skin color clustering model is too general, it may yield a large number of FP errors, that is, a non-skin pixel classified as a skin pixel, such as furniture, clothes, buildings, rocks, and so on. On the other hand, if the skin model is tight, then it may yield numerous FN errors, in which the skin pixels are missed.

The skin color clustering problem leads us to create a novel approach for human skin color segmentation that combines pixel-based segmentation and region-based segmentation in a way that would cover different skin color appearances. Our approach suggests building multi-skin color clustering models, instead of a single skin model, wherein each skin color clustering model represents the clustering of a pattern class. The skin color models are as follows:

- $m_1$: standard skin,
- $m_2$: shadow skin and blackish skin,
- $m_3$: light skin,
- $m_4$: red concentrated skin and lips.

Once we built these clustering models, we segmented the input image into four layers of skin maps, with each layer reflecting its skin-clustering model. Fig. 2 shows segmentation output using multi-skin color clustering models for some test images. Fig. 2(a) shows the output of skin color segmentation at Layer1 for a white-skinned woman with no makeup and under uniform lighting. Fig. 2(b) shows the effectiveness of applying multi-skin color models on images containing people of different races. In this figure, the woman with dark skin is completely recognized in Layer2, although she is misclassified in Layer1. Fig. 2(c) shows an example of various skin tones for different racial groups. The dark reddish hands are correctly classified at Layer3 only. Although the hands that tend to be yellow are partially classified at Layer1, the convex-hull algorithm can greatly recover the missing regions in such cases. Fig. 2(d) shows that the skin of the same individual varies under non-uniform illumination. The right half of the face with low lighting is correctly classified at Layer2.

The next step in this direction was to use region-based segmentation, in which the analysis of the pixel neighborhood was conducted to perform the iterative merging process (region grow) with the other layers. Iterative merge is a procedure that groups
pixels or sub-regions into a larger region based on a predefined criterion for merging.

Our merging step started with the objects in the skin layer (Layer1) as the seed points. From these seed points, the region was grown by appending to the seed those neighboring pixels in the other layers that had predefined properties similar to those of the seed. The candidate pixel was merged if it satisfied the following conditions:

- it was adjacent to some pixels of the growing region;
- it belonged to one of the other layers' regions;
- it satisfied the similarity of the mean color intensity of the growing region;
- it satisfied the texture analysis condition of the growing region; and
- it was not an edge pixel.

The similarity condition of the mean color intensity of the growing region was not fixed and was updated during the growing process to reflect the actual mean intensity of the growing region to cover the smooth change in skin color. We used the Euclidian distance to measure the skin color similarity. The condition used is

$$\sqrt{(H_p - H_{av})^2 + (S_p - S_{av})^2 + (V_p - V_{av})^2} < T$$  \hspace{1cm} (1)

where $H_p$, $S_p$, and $V_p$ are the intensity values of the HSV components of the candidate pixel, respectively; and $H_{av}$, $S_{av}$, and $V_{av}$ are the growing region's average intensity values of HSV components, respectively. The threshold of color intensity used is $T = 0.25$ (identified experimentally). The unmerged background regions in Layer2, Layer3, and Layer4 are rejected at the early stage of computation.

3. Building multi-skin color clustering models

Manually, we prepared 2485 patches extracted from still images from different databases for different people and different lighting conditions covering a large range of skin appearances. These images were divided into four pattern classes ($m_1$, $m_2$, $m_3$, and $m_4$). These patches contain 15,333,922 pixels of skin and non-skin samples. The skin samples acquired from several regions of human skin, including the forehead, cheeks, nose, neck, and so forth. Skin pixels that were blackish and those containing strong highlights and dark shadows were included. To ensure that skin samples were taken by different cameras, samples were collected from images that came from different sources and types. We found that the skin color clustering for the four models was in the range of $(0^\circ \leq \text{Hue} \leq 49^\circ)$ and $(339^\circ \leq \text{Hue} < 360^\circ)$ on the HSV wheel, [see Fig. 3(b)]. Our purpose is to transform the three-dimensional (3D) color space into a two-dimensional (2D) space without losing color information. We divided the total range of skin color at the Hue channel component into equal intervals on the Hue wheel with $(+7^\circ)$. For each quantized constant Hue, the correspondence transformed subspace in the 2D color space shall be as shown in Fig. 3(c). Thus, we have only 10 slides of SV-planes ($Hue = 0, 7, 14, \ldots, 49, 339, 346, 353$). Each colored pixel may now be viewed as a point in a 2D space characterized by the two measurements, $S$ and $V$, where $(0 \leq S \leq 1)$ and $(0 \leq V \leq 1)$. Our experiments show that the distributions of skin color are clustered around a typical representation. Fig. 4 shows the skin color distribution for the skin color classes. We report here an example of the classification boundaries for a known Hue where ($\text{Hue} = 14$):

$m_1(S, V)$: if $(0.12 < S \leq 0.55)$ and $(0.6 \leq V \leq 1)$,

$m_2(S, V)$: if $(0.12 \leq S \leq 0.7)$ and $(0.30 \leq V < 0.6)$.
the average intensity of the growing region, and adjustment gain determined experimentally by trial and error.

Classification boundaries of normal white skin will be in the range (0 ≤ S ≤ 1), becoming smaller in the range (0 ≤ S ≤ 1), becoming smaller in the range (0 ≤ S ≤ 1), becoming smaller as we move through the Hue wheel. At Hue = 0 degree, the classification boundaries will be in the range (0.10 ≤ S ≤ 0.45 and 0.65 ≤ V ≤ 1), becoming smaller in the range (0.8 ≤ S ≤ 0.28 and 0.70 ≤ V ≤ 1) at Hue = 346 degrees.

Classification boundaries are more easily adjusted using 2D space because of the direct access to the color space. Another advantage in our approach of using multi-skin clustering models is that this approach may be used to perform skin color correction (lightening) on regions in the shadow layer without losing color information. Given that the darkness of the color in the HSV color space is represented in the value component \( V \), we can adjust the darkness of the pixel color by adjusting \( V \) with the constant bias \( C \) computed locally for each skin candidate region, as follows:

\[
C = \beta \times \text{abs}(V_{sh} - V_{av})
\]

where \( V_{sh} \) is the average intensity of the shadow region, \( V_{av} \) is the average intensity of the growing region, and \( \beta = 0.25 \) is the adjustment gain determined experimentally by trial and error.

We found some cases in which the saturation component \( S \) also had a good effect on the skin color correction of images with highly concentrated red color. Thus, we used the same idea to adjust the saturation component. Fig. 5 shows the idea of skin color correction using multi layers. Figs. 5(a) and (c) illustrate pixel-based skin segmentation. Fig. 5(d) shows the skin color correction of skin regions at Layer2 by adjusting the darkness of each pixel. The iterative merge used to create a candidate face region is shown in Fig. 5(e). The newly generated image (i.e., lighting-corrected image) is shown in Fig. 5(f).

4. Boundary classification algorithm

The goal of skin color detection is to build decision functions or classification boundaries to discriminate between different classes. We sought to use an algorithm that would take the measured features of an unknown pixel (i.e., three-color components \( H, S, \) and \( V \)) as input and then predict the true class membership as output. The term “predict” indicates that this process may not always be possible without error. The rate of false predictions depends on numerous factors, such as the degree of overlapping among the classes in the feature space, the amount of invalid data, noise, and the generality of training samples. In an ideal case, classification boundaries could completely separate different classes. In real-world implementation, different methods usually produce different classification boundaries. Simpler classification boundaries are favored over those that are needlessly complicated. Classification boundaries based on overly complex models often result in lower accuracy of the classifier [16].

Based on the collected training samples (pixels), the different methods yielded different classification boundaries using Matlab Statistics Toolbox, the classification boundaries of Bayes classifier are shown in Fig. 6(a), the Linear Discriminate Analysis (LDA) are shown in Fig. 6(b), and the Quadratic Discriminate Analysis (QDA) are shown Fig. 6(c). Although such classification boundaries may facilitate very good classification of the current training samples, they also result in poor performance on the novel pixels, that is, the pixels not yet seen. With such solutions though, our satisfaction would be premature because the central aim of designing a classifier is to suggest actions when presented with novel pixels. These classification boundaries are not likely to provide a good generalization because they seem to be “tuned” to the particular training samples, rather than some underlying characteristics or true models of the classes that have to be separated. For example, classification boundaries with thin gulls or holes are not practical for the task of skin segmentation because the assumption that some regions in a small neighborhood belong to the skin, whereas the other regions that fall in between do not is improbable. The adjacent entries in a color space show very similar colors. Clearly, the classification method is supposed to satisfy the following:

\[
m_3(S, V): \quad \text{if } (0 \leq S \leq 0.12) \text{ and } (0.75 \leq V \leq 1),
\]

\[
m_4(S, V): \quad \text{if } (0.55 \leq S \leq 0.75) \text{ and } (0.6 \leq V \leq 1).
\]
Fig. 5. Human skin segmentation and skin color correction; (a) RGB source image; (b) HSV image; (c) skin color segmentation; (d) skin color correction; (e) iterative merge; and (f) newly generated image (i.e., lighting-corrected image).

Fig. 6. Classification boundaries yielded by different methods: (a) using Bayes classifier; (b) using LDA classifier; and (c) using QDA classifier; and (d) using the proposed algorithm.

1. The classification boundaries should match the corresponding real distribution of training data.
2. The classification boundaries should form compact regions. By compact region, we mean no broken regions, no holes, and no bridge gaps.
3. The contour of a region should be smooth, with no long thin gulfs and no protrusions.
4. The graphical representation of the region's shapes should not overlap (e.g., snaky, complex, zigzag, and so on).
5. The shape of regions among adjacent slides should be changed gradually in the 3D space that forms a sold 3D body. An abrupt change in the shape through successive adjacent slides is undesirable.
The Bayes classifier, as shown in Fig. 6(a), yields regions in such a way that a small neighborhood belongs to the skin, whereas other regions that fall in between do not. The LDA classifier, as shown in Fig. 6(b), yields regions that do not match the actual distribution of training samples. The QDA classifier, as shown in Fig. 6(c), has the same problem as that of the Bayes classifier. The visual inspection of the remaining set of classification boundaries for other SV-plane slides shows that the abovementioned conditions did not hold true for many cases.

The classification boundaries of the proposed algorithm are shown in Fig. 6(d). Clearly, these boundaries are simple and smooth. In addition, the boundaries match the actual distribution shown in Fig. 6(d). Clearly, these boundaries are simple and smooth. The group of connected pixels (i.e., pixels have the same class label) form a region (object) in that image, and the label of the pixels refers to the region’s number. For example, the pixels labeled 1 make up region one; the pixels labeled 2 make up a second region; and so on. Generally, the scattered plot of raw data forms regions with thin gulfs, protrusions, and holes (i.e., non-compact regions). The goal of this stage is to fill holes and remove small protrusions and thin gulfs. We proposed the use of erosion and dilation operations inspired from morphological operations in binary images. Morphological operations can be described simply in terms of adding or removing pixels from a region according to certain rules, which depend on the pattern of neighboring pixels. Erosion removes pixels from a region in an image or, equivalently, turns OFF pixels that were originally ON. The purpose is to remove pixels that should not be there (i.e., in our case, noise data). Dilation can be used to add pixels to a region. For this purpose, a structuring element with a size of 3 × 3 pixels is used to smooth the border of each region. This step will remove thin gulfs and protrusions to create compact regions. Fig. 7(d) shows examples of the removal of thin gulfs and smoothing of a region’s border.

5. Experimental results

The experimental results are presented to show the effectiveness of the proposed color pixel clustering models. Our system was conducted on a 3.00 GHz Intel (R) Core TM 2Duo processor with 8 GB RAM running the Windows Vista platform using MATLAB R2010a. Three different databases were used to evaluate the proposed system. “The CVL Database” contains a total of 114 people × 7 images with different views and expressions taken with uniform background [17].

The “Labeled Faces in the Wild (LFW) database,” contains over 13,000 images of faces collected from the Internet. A total of 1680 people pictured have two or more distinct photos in the data set [18]. Finally, we created the third database by collecting face images from the Internet. The third database contains 75 images with a different number of faces and a complex background. Our experimental dataset comprised 125 images taken from the abovementioned databases as samples. The segmentation performance evaluation was measured in terms of True Positive (TP), FP, True Negative (TN), and FN and were computed for all pixels in the “skin classifier testing set” through skin segmentation testing. TP is the proportion of skin pixels classified correctly as skin, whereas FP is the proportion of non-skin pixels classified incorrectly as skin; FN and TN serve as the complements of TP and FP, respectively [1]. Table 1 shows the system performance when using multi-skin color segmentation models. The best performance of TP was 96.5% of skin pixels with a very low FP of 0.76%. The comparison of experimental results of the proposed approach with other standard skin classifiers is presented in Table 2. The proposed approach shows a good increase in the detection rate and a decrease in the false detection error compared with the other methods.
Fig. 7. The proposed algorithm; (a) projected 2D SV-space of class-labeled summary table; (b) sub-sampling SV-plane of HSV color space; (c) output of sub-sampling the class labeled SV-plane based on dominant-class filter; and (d) smoothing region's border by removing thin gulfs, correcting invalid data, and filling holes.

Table 1
System performance.

<table>
<thead>
<tr>
<th>No. of images</th>
<th>Proposed multi-skin models</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>TP %</td>
</tr>
<tr>
<td>125</td>
<td>96.5</td>
</tr>
</tbody>
</table>

6. Conclusion

In this paper, we demonstrated a reliable color pixel clustering model for skin segmentation under unconstrained scene conditions based on building multi-skin color clustering models using HSV color space. A serious issue in skin color segmentation is the building of a skin model cluster that could cover different skin appearances over different kinds of image collections. If the skin model is too general, it may yield a large number of FPs. On the other hand, if the skin model is tight, then it may yield numerous FNs. The multi-skin color clustering model is a promising method that may be used for skin color segmentation in uncontrolled conditions. We gained two advantages. First, we overcame the limitations of pixel-based segmentation by combining it with region-based segmentation using an iterative merge with other layers that consider the neighborhood pixels. Second, we addressed the problem of shadow regions by lightening these regions (skin color correction), which improved the detection rate. A number of experiments have verified that the proposed algorithm is robust to irregular illumination variations and abrupt changes in brightness through its application on different image databases.

Table 2
Performance comparison in terms of detection rates.

<table>
<thead>
<tr>
<th>Method</th>
<th>TP %</th>
<th>FP %</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed approach</td>
<td>96.5</td>
<td>0.76</td>
</tr>
<tr>
<td>Color and texture using Neural Network [1]</td>
<td>95.6</td>
<td>0.875</td>
</tr>
<tr>
<td>Color and texture using Fuzzy Inference System [2]</td>
<td>90.0</td>
<td>0.22</td>
</tr>
<tr>
<td>Color Distance Map [7]</td>
<td>89.97</td>
<td>9.26</td>
</tr>
<tr>
<td>Bayesian classifier [7]</td>
<td>83.92</td>
<td>10.91</td>
</tr>
</tbody>
</table>

References


Please cite this article in press as: S.A. Naji et al., Skin segmentation based on multi pixel color clustering models, Digital Signal Process. (2012), http://dx.doi.org/10.1016/j.dsp.2012.05.004


Sinan A. Naji received his B.Sc. and M.Sc. degrees in computer science from University of Technology, Iraq in 1987 and 1990, respectively. Currently, he is a Ph.D. student at the department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya. His research interests include neural network, pattern recognition, digital image processing, and natural language processing.

Rozia Zainudin is a professor at the department of Artificial Intelligence, Faculty of Computer Science and Information Technology, University of Malaya. She obtained her B.Sc. from Monash University, Australia; M.Sc. (Reading University) and Ph.D. from University of East Anglia. Her areas of interest are intelligent multimedia, image and speech processing, computational fluid dynamics and bio-medical informatics. Her research work has been published in several international journals and conference publications. Several awards have been won for her research outcome at software exhibitions. Her professional duties include reviewing articles, editing journals, supervising research students, and she is appointed as an external examiner for postgraduate research.

Hamid A. Jalab received his Bachelor's degree of Electrical Engineering from University of Technology, Baghdad, Iraq, M.Sc. and Ph.D. degrees in Computer System from Odessa National Polytechnic University. He is currently a senior lecturer at the Faculty of Computer Science and Information Technology in University of Malaya, Malaysia. His areas of research interest are signal and digital image processing, wavelets, neural networks and image retrieval. His research work has been published in several international journals and conference publications.

Please cite this article in press as: S.A. Naji et al., Skin segmentation based on multi pixel color clustering models, Digital Signal Process. (2012). http://dx.doi.org/10.1016/j.dsp.2012.05.004