Suppressing False Nagatives in Skin Segmentation

Roziati Zainuddin[1], Sinan Naji[1], and Jubair Al-Jaafar[2]

[1] Faculty of Computer Science and Information Technology, University of Malaya, 50603, Kuala Lumpur, Malaysia
roziati@um.edu.my, sinan@siswa.um.edu.my
[2] School of Computer Technology, Sunway University College, 64150, Kuala Lumpur, Malaysia
aljaafar@sunway.edu.my

Abstract. Human skin segmentation in colored images is closely related to face detection and recognition systems as preliminary required step. False negative errors degrade segmentation accuracy and therefore considered as critical problem in image segmentation. A general innovative approach for human skin segmentation that substantially suppresses false negative errors has been developed. This approach employed multi-skin models using HSV color space. Four skin color clustering models were used, namely: standard-skin model, shadow-skin model, light-skin model, and redness-skin model. The color information was used to segment skin-like regions by transforming the 3-D color space to 2-D subspace. A rule-based classifier produces four skin-maps layers. Each layer reflects its skin model. Pixel-based segmentation and region-based segmentation approaches has been combined to enhance the accuracy. The inspiring results obtained show that the suppression of false negatives is substantial and leads to better detection and recognition.

Keywords: Human skin segmentation, skin color modelling, face detection, HSV.

1 Introduction

Human skin segmentation in colored images is becoming an important task in many vision-based systems such as access control systems, face tracking, robot interaction, banking and financial transaction, video surveillance, videophone and teleconferencing, etc. The first task of such systems is to locate the face (or faces) within the image. It is not easy task since human face is a dynamic object and has a high degree of variability in its appearance (non-rigid object), which makes face detection a difficult problem in computer vision. Segmentation techniques based on color information as a cue has gained much attention motivated by four principle factors. First, color in general is a powerful descriptor that often simplifies object detection and extraction from a scene. Second, the processing of color information had proven to be much faster than processing of other facial features. Third, color information is robust against rotations, scaling and partial occlusions. Forth, Skin color information can be used as complimentary information to other features to improve detection. The challenges

T.-h. Kim et al. (Eds.): FGIT 2010, LNCS 6485, pp. 136–144, 2010.

addressed with skin color segmentation can be attributed to the following factors: illumination, race, complex background, number of persons, imaging conditions, image montage, individual characteristics, aging, makeup, etc. Apparently these variations complicate skin segmentation and the larger the variations are, the more difficult the problem is.

Different skin color appearance caused by unconstrained scene conditions degrades segmentation accuracy. Segmentation may cause two kinds of errors: False Negative errors in which a skin pixel classified as a non-skin pixel, and False Positive errors in which an image pixel is declared to be skin pixel, but it is not. The most critical problem of the two errors is false negatives, attributed to the fact that, image segmentation is the first step in image analysis. Subsequently, when a face region is missed, the following stages of the system cannot retrieve the missed face. Therefore, false negatives determine the eventual success or failure of the subsequent stages.

The research aims to suppress false negative errors to achieve precise skin segmentation in fast, robust, and reliable approach. A novel approach has been introduced, so that, the approach is shifted from mono-skin model to multi-skin models using HSV color space. The detail description of building multi-skin color clustering models and the classification boundaries is presented in Section 3. The proposed approach for skin color segmentation is presented in section 4. Experimental results of the proposed approach and conclusion are given in Section 5 and 6 respectively.

2 Background

Numerous methods for skin segmentation and detection have been proposed so far. Each has its advantages and limitations. Some is superior to others whilst some yields the same result when compared to other technique.

A number of approaches have been proposed using different color spaces: RGB [1] [2], HSV or HSI [3] [4], YCbCr [5], YIQ [6], YES [7], CIE [8], YUV [9]. To build a skin color clustering model in the color space, many methods have been proposed. McKenna [10] has proposed Gaussian mixture models for the task which outperform single Gaussian model. Ruiz-del-Solar [2] has compensated for their color segmentation methods with additional features to obtain more valuable results robust to brightness variations. Gomez [11] has listed top components and made a hybrid color space from those. Jayaram [12] have proposed a method called an adaptive skin color filter that detects skin color regions in a color image by adaptively adjusting its threshold values. The Bayesian classifier with histogram technique has been used for skin detection by Jones [13]. Kim [4] proposed a skin color modeling approach in HSI color space while considering intensity information by adopting the B-spline curve fitting to make a mathematical model for statistical characteristics of a color with respect to intensity. Li [9] proposed an algorithm based on facial saliency map. Chen [1] proposed a hybrid-boost learning algorithm for multi-pose face detection and facial expression recognition. Juang [3] have used self-organizing Takagi–Sugeno-type fuzzy network with support vector is applied to skin color segmentation. Vezhnevets [14] wrote a survey on pixel-based skin color detection techniques.

A comprehensive survey on skin-color modeling and detection methods was written by Kakumanu [15].

3 Building Multi-skin Models

Generally, building skin model based on color information involves three main issues: First, what color space to choose. Second, how exactly the skin color distribution should be modeled, and finally, what will be the way of processing [14].

3.1 Choosing the Color Space

Algorithms based on skin color information need to deal with the sensitivity to the illumination conditions under which the input image is captured. HSV color space tends to be more realistic than other color systems. Its representation strongly relates to human perception of color, and it also eliminates the influence of illumination when describing the color of an object, Fig. 1(a). Hence, the HSV (and HSI) model is an ideal tool for developing image processing algorithms based on color descriptions [16]. The hue (H) is a measure of the spectral composition of a color and is represented as an angle from 0˚ to 360˚, while saturation (S) refers to the purity of a color and its value ranges from 0 to 1. Value component refers to the darkness of a color, which ranges also from 0 to 1. HSV color space had been chosen in our approach to build skin color models.

Generally, colored image contains millions of colors. The research also aims to reduce the number of colors in the source image. This will decrease the computational cost which is an essential to all systems. Reducing the number of colors process known as quantization. Each set of points of similar color is represented by a single color. Experimental results show that, clustering models of skin color are in the range of (0˚ ≤ Hue ≤ 49˚) and (339˚ ≤ Hue < 360˚) on the Hue wheel. Therefore, the range of colors at Hue wheel has been divided into equal intervals. Each interval step is (7˚). Hence, skin colors were reduced at Hue component to only eleven colors (Hue = 0, 7, 14, 21…, 49, and 339, 346, …, 360) which cover various types of skin; i.e. white, black, yellow, and redness skin colors under different lighting conditions.

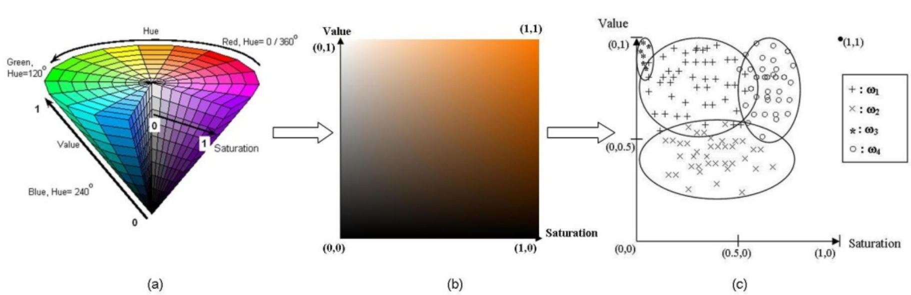

Fig. 1. Converting 3-D color space to 2-D subspace. (a) HSV color space. (b) 2-D S-V subspace, where Hue=28. (c) Four skin clusters of the pattern classes ω_1, ω_2, ω_3, and ω_4.

3.2 False Negative Errors Survey

The false negatives degrade segmentation performance dramatically with the diversity of image types and sources. To diagnose this problem, a survey on false negatives has been done. The survey showed that skin pixels classified as non-skin pixels can be categorized into four pattern classes:

1) Shadow regions (70%): shadows, blackish skin, and poor lighting.
2) Light regions (15%): because of strong light reflection, skin color information may be lost in some areas of the face.
3) Redness regions (8%): usually because of makeup, montage process, and flushing.
4) Others (7%)

The survey also reveals two main reasons behind the false negatives: first, the limitations of mono-skin model to cover many skin color appearance (dark, light, redness, etc). The second reason is due to the fact that each colored pixel is treated individually in relation to the color space (skin or non-skin pixel) without any consideration to the content of neighboring pixels.

3.3 Shifting from Mono-skin Model to Multi-skin Models

The avoidance of false negatives is clearly one of the functions of segmentation algorithm. One solution is to handle different skin color appearance caused by illumination variations and other factors in a proper way. This results into a novel approach that shifts from mono-skin model to multi-skin models. There is a good reason to use multi-skin models: pixels that are indistinguishable in one model may be fully distinguishable in the other model that will suppress the false negatives.

More than 3,500 face skin patches are used. These patches contain about (300,000) pixels of skin color data acquired from several regions of human skin including forehead, cheeks, nose, neck, and so forth. According to the skin color appearance, these patches have been divided into four pattern classes ω_1, ω_2, ω_3, and ω_4 namely:

ω_1: Standard skin color with uniform lighting
ω_2: Shadow skin, dark and blackish skin
ω_3: Light skin regions
ω_4: Redness skin

The next step involves the determination of optimum decision boundaries which are needed in the classification process.

3.4 Classification Rules

Determination of decision boundaries in three-dimensional color space is more difficult than in the two-dimensional. There is no easy or obvious way to enclose arbitrary clusters in three-dimensional space to see which pixels are selected for specific cluster. In most papers, low-dimensional color space is chosen instead of high-dimensional color space, to ease the determination process, (the R-G space replaces the R-G-B color space, the H-S space replaces the H-S-V color space, and so on). When the lighting is uniform, the segmentation performance is acceptable; but the performance is bad when dealing with unconstrained lighting conditions. This is due to loss of some information when an image is expressed in a low-dimensional space instead of a high-dimensional space. The research uses the full color information (three components: Hue, Saturation, and Value) in a novel way by transforming the 3-D color space to 2-D subspace without any color information losses. The idea is to use 2-D subspace for each constant quantized Hue value instead of 3-D color space. The corresponding transformed subspace would be S-V plane as shown in Fig. 1(b). The pattern vectors therefore, are of the form $\mathbf{x}=(s,v)'$ where s represents the Saturation component ($0 \leq s \leq 1$) and v represents the Value component ($0 \leq v \leq 1$) of the color features. The training samples of each pattern class tend to cluster about a typical representation within some region of the S-V subspace. Fig. 1(c) illustrate the four clusters in the S-V subspace where (Hue=28)[1]. The clusters are overlapped (non-separable situation) and not equally distributed.

A set of classification boundaries have been found using training algorithm. The algorithm is deduced from k-means algorithm and the reward-punishment concept. Here is an example of the classification boundaries that was obtained for Hue=28:

$$P(x,y) \in \omega_1 \text{ if } (0.12 < s(x,y) \leq 0.65) \text{ and } (0.6 \leq v(x,y) \leq 1)$$
$$P(x,y) \in \omega_2 \text{ if } (0 \leq s(x,y) \leq 0.7) \text{ and } (0.33 \leq v(x,y) < 0.6)$$
$$P(x,y) \in \omega_3 \text{ if } (0 \leq s(x,y) \leq 0.12) \text{ and } (0.75 \leq v(x,y) \leq 1)$$
$$P(x,y) \in \omega_4 \text{ if } (0.65 < s(x,y) \leq 0.85) \text{ and } (0.6 \leq v(x,y) \leq 1)$$

Where $s(x,y)$ and $v(x,y)$ represent the Saturation and Value components of the pixel at location (x,y). The pixel is defined as part of a specific pattern class when its two components s and v lie within the selected ranges. For different quantized Hue values, there will be different classification rules. These multiple rules are combined with Boolean OR operation. This is logically equivalent to segmenting each hue channel individually, creating separate binary images, and then combine them with Boolean OR operation afterward. It is clear that classification boundaries are more easily adjusted using 2-D subspace because of direct access to color space.

Classification rules in generally considered to be the high-end task of building skin color models. As shown in the experimental results section, the skin color models in this research can be used to detect the skin under different lighting conditions and under various types of skin color appearance.

4 Skin Color Segmentation

In general, finite level of accuracy can be achieved with pixel-based segmentation since each colored pixel is treated individually in relation to the color space (skin or non-skin pixel) without any consideration to the content of neighboring pixels. However, the skin of each human face has certain homogeneity between its pixels that could differentiate it from other objects. The approach in this research combines pixel-based segmentation and region-based segmentation to take in consideration the neighboring pixels that will improve segmentation accuracy. The input RGB image is

[1] For illustration, Hue=28 is found to be the most ideal case to graphically show the four skin clusters in Figure 1(c).

converted to the equivalent HSV color space. The number of colors in the input image is reduced by quantization. The input image is segmented using pixel-based segmentation approach which is based on classification rules. The classification rules produce four binary images called skin-maps in different separate layers (layer1, layer2, ..., layer4) as shown in Fig. 2(a) and Fig. 2(b). Fig. 2(b) shows that each layer reflects its relevant skin model. The merging process (region-based grow) between four skin-map layers is done to gather these pixels that satisfy the homogeneity between pixels of that region. The Region-based segmentation methods require the input of a number of seeds, might be manually, either individual seed pixels or regions. The approach in this research starts with the pixels in the first layer (ω1) as seed points. From these seed points, region grows by appending to the seed the neighboring pixels from the other layers (ω2, ω3, and ω4). Neighboring pixels are examined one at a time and added to the growing region if they are sufficiently similar. The conditions used:

- A pixel is adjacent to some seed pixel of the growing region,
- It belongs to some skin model at layers (ω_2, ω_3, and ω_4),
- Its color satisfies the similarity color of the growing region,
- It satisfies the texture analysis condition of the growing region and
- It is not an edge pixel.

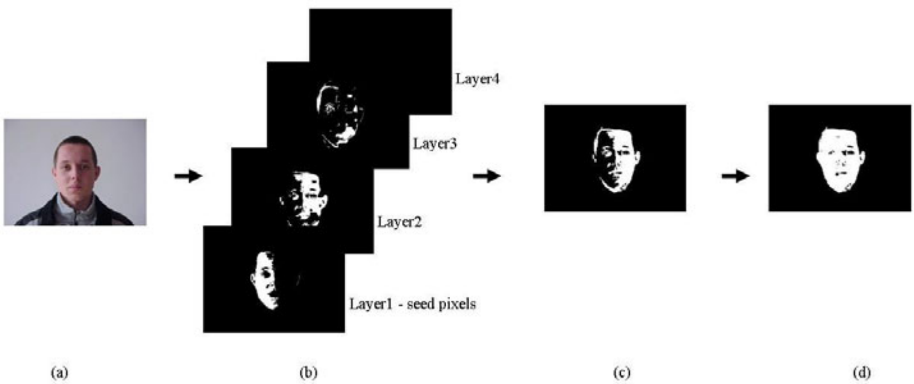

Fig. 2. Human skin segmentation. (a) Input image. (b) Pixel-based segmentation. (c) Region grows. (d) Skin segmentation output.

The color similarity condition used for growing region is not fixed; it is updated during merging process to reflect the actual mean intensity of the growing region to cover smooth changing in skin color. The Euclidian distance is used to measure color similarity:

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

where H_p, S_p, V_p are the intensity values of HSV components of the candidate pixel, and H_{av}, S_{av}, V_{av} are the average intensity value of HSV components of the growing region. The threshold of color intensity used is T, which is calculated locally.

The background objects at layer2, layer3, and layer4 that do not merge will be rejected at early stage of computation. Figure 2(c) shows the region grows output and Figure 2(d) shows skin segmentation output.

5 Experimental Results

A series of experiments were performed to evaluate the proposed approach, comparing the performance of mono-skin model approach to multi-skin models approach. For these experiments, three different databases were used. “The CVL Database" [17], which contains a total 114 people × 7 images with different views and expressions taken with uniform background. The “LFW database - Labeled Faces in the Wild” [18], is a database contains more than 13,000 images of faces collected from the Web. However, from the point of view of our objective, a dataset that contains 450 images were collected from the Web. In the experimentation part, 175 real images chosen carefully have been employed.

Table 1. System performance

Lighting Conditions	Using Mono-Skin Model		Using Multi-Skin Models	
	False Negatives Pixels	False Positives Pixels	False Negatives Pixels	False Positives Pixels
Uniform lighting	11%	7%	4%	9%
Un-uniform lighting (sidelight)	35%	6%	14%	10.5%
Gloomy scene	28%	8%	15%	13%

The images contain single face or multiple faces with various sizes and different lighting conditions. These images have been divided into three sets. The first set, represents the uniform lighting, the second set denotes lighting under sidelight, and the third set comprises images in gloomy scene. It is not a simple task to evaluate segmentation performance and compare algorithms. However, one of the most widely used criteria for performance evaluation is whether the system can outline the desired regions in the image. The criteria used for our system performance evaluation is as follows:

- The skin-maps for the test images have been manually annotated as a ground truth.
- Two approaches were applied; the conventional static mono-skin model compared to multi-skin models.
- Accordingly, two parameters for each system output have been measured relative to the ground-truth:
  (1) False Negative ratio (Rfn) is the number of falsely missed pixels to the actual number of skin pixels.
  (2) False Positive ratio (Rfp) is the number of falsely detected pixels to the actual number of skin pixels.

Table 1, shows a performance comparison between mono-skin model and multi-skin models. The segmentation performance of multi-skin models showed better performance. The important achievement is the suppression of false negatives that reveals its ability to segment skin regions with robustness to varying illumination conditions and different racial groups leading to better recognition in subsequent stages.

Fig. 3 clearly shows that the experiment results are promising. Although in Fig. 3 Row 1, the light was sidelight, the right part of the face has been correctly detected. A drawback of the approach is the slight increase in computational cost, which is mainly caused by the complexity of the improved multi-skin models. The second drawback of the approach is that false positives were slightly increased. However, this is not an issue since image segmentation, in general, is a preprocessing stage in face detection and recognition systems. It is obvious that color information on its own is not sufficient and that another stage based on other facial features is required to classify the skin segment to be a face or non-face. The experiments were carried out using PC Pentium 4 with a 3.00 GHz CPU and 1GB of RAM memory. The system was implemented in MATLAB (Ver. 7.9).

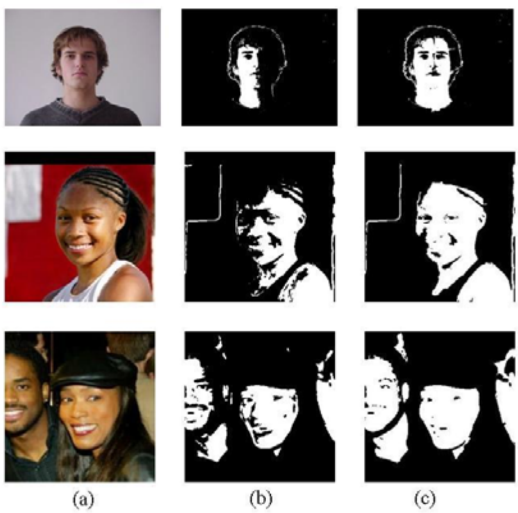

Fig. 3. Skin color segmentation. (a) Input image. (b) Using mono-skin model. (c) Using multi-skin models.

6 Conclusions

1- This approach based on using multi-skin models that substantially suppress false negative errors that reveals its ability to segment skin regions with robustness to varying illumination conditions and different racial groups leading to better detection and recognition in subsequent stages.
2- Higher suppression ratio can be achieved. However, it is undesired because the false positives will be slightly increased accordingly.

3- Multi-skin models caused higher computational cost than mono-skin model; however that is a normal outcome.
4- The approach does not require any preconditions and assumptions.
5- The approach can be applied to region segmentation for arbitrary objects (e.g., vehicles, hand tracking, aero imaging, etc).

References

1. Chen, W.C., Wang, M.S.: Region-Based and Content Adaptive Skin Detection in Color Images. International journal of pattern recognition and artificial intelligence 21(5), 831–853 (2007)
2. Ruiz-del-Solar, J., Verschae, R.: Skin Detection Using Neighborhood Information. In: IEEE International Conference on Automatic Face and Gesture Recognition, pp. 463–468 (2004)
3. Juang, C.-F., Shiu, S.-J.: Using Self-Organizing Fuzzy Network With Support Vector Learning For Face Detection in Color Images. Neurocomputing 71(16-18) (2008)
4. Kim, C., You, B.-J., Jeong, M.-H., Kim, H.: Color Segmentation Robust to Brightness Variations by Using B-Spline Curve Modeling. Pattern Recognition 41(1), 22–37 (2008)
5. Chen, H.-Y., Huang, C.-L., Fu, C.-M.: Hybrid-Boost Learning for Multi-Pose Face Detection and Facial Expression Recognition. Pattern Recognition 41(3), 1173–1185 (2008)
6. Dai, Y., Nakano, Y.: Face-Texture Model Based on SGLD and Its Application in Face Detection in a Color Scene. Pattern Recognition 29(6), 1007–1017 (1996)
7. Saber, E.M., Tekalp, A.M.: Frontal-View Face Detection and Facial Feature Extraction Using Color, Shape and Symmetry Based Cost Functions. Pattern Recognition Letters 17(8), 669–680 (1998)
8. Kawato, S., Ohya, J.: Real-Time Detection of Nodding and Head-Shaking by Directly Detecting and Tracking The Between Eyes. In: IEEE International Conf. on Automatic Face and Gesture Recognition, Grenoble, pp. 40–45 (2000)
9. Li, B., Xue, X.Y., Fan, J.P.: A Robust Incremental Learning Framework for Accurate Skin Region Segmentation in Color Images. Pattern Recognition 40(12), 3621–3632 (2007)
10. McKenna, S., Gong, S., Raja, Y.: Modelling Facial Colour And Identity with Gaussian Mixtures. Pattern Recognition 31, 1883–1892 (1998)
11. Gomez, G.: On Selecting Colour Components For Skin Detection. In: International Conference on Pattern Recognition, vol. 2, pp. 961–964 (2002)
12. Jayaram, S., Schmugge, S., Shin, M.C., Tsap, L.V.: Effect of Color Space Transformation the Illuminance Component, and Color Modeling on Skin detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 813–818 (2004)
13. Jones, M.J., Rehg, J.M.: Statistical Color Models with Application to Skin Detection. Int. J. Computer Vision 46(1), 81–96 (2002)
14. Vezhnevets, V., Sazonov, V., Andreeva, A.: A Survey on Pixel-Based Skin Color Detection Techniques. Graphicon (2003)
15. Kakumanu, P., Makrogiannis, S., Bourbakis, N.: A Survey of Skin-Color Modeling and Detection Methods. Pattern Recognition 40, 1106–1122 (2007)
16. Gonzalez, R.C., Woods, R.E.: Digital Image Processing. Addison-Wesley, Reading (2002)
17. The CVL Database, `http://lrv.fri.uni-lj.si/facedb.html`
18. The LFW Database, `http://vis-www.cs.umass.edu/lfw/`