A Comparative Study of Skin-Color Models

Juwei Lu, Qian Gu*, K.N. Plataniotis, and Jie Wang

Bell Canada Multimedia Laboratory, The Edward S. Rogers Sr.,
Department of Electrical and Computer Engineering,
University of Toronto, Canada M5S 3G4

Abstract. In this paper, we report the results of a comparative study on skin-color models generally used for facial region location. These include two 2D Gaussian models developed in normalized RGB and HSV color spaces respectively, a 1D lookup table model of hue histogram, and an adaptive 3D threshold box model. Also, we present a new model - called "adaptive hue lookup table". The model is developed by introducing the so-called "Continuously Adaptive Mean Shift" (Camshift) technique into a traditional hue lookup table method. With the introduction of Camshift, the lookup table is able to adaptively adjust its parameters to fit the illumination conditions of different test images. In the experiments reported here, we compare the proposed method with the four typical skin-color filters in the scenarios of different human races and illuminations. The obtained results indicate that the proposed method reaches the best balance between false detection and detect rate.

1 Introduction

Automatic location of facial region is an important first step in face recognition/tracking systems. Its reliability has a major influence on the performance and usability of entire face recognition/tracking systems. Numerous solutions to the problem have been presented. Generally, they can be roughly classified into two classes [1]: (1) gray-level based methods and (2) color-based methods. Among the gray-level based methods, most are based on template matching techniques [2]. The input image is windowed (with varying window sizes) from location to location, and the sub-image in the window is classified as face or non-face. Although accurate in terms of detect rate, most of them are highly complicated and time-costing, which are major reasons why the kind of methods are not typically used in real-time tasks. Also, they are sensitive to facial variations due to view-points, scales, rotations and illuminations.

The color-based methods are normally based on various skin-color filters and region segmentation techniques [3,4]. This kind of methods have many advantages compared to the gray-level based methods. First, processing color is much faster than doing template matching. Second, color models are scale/orientation/rotion invariant. These properties are particularly important for a real-time

* Ms Qian Gu performed the work while at School of Electrical and Electronic Engineering, Nanyang Technological University, Singapore, 639798.

M. Kamel and A. Campilho (Eds.): ICIAR 2005, LNCS 3656, pp. 729–736, 2005.

face/human tracking system. Successful applications of color-based algorithms include some state-of-the-art face/human tracking systems, such as Pfinder [5] and Yang's face tracker [6]. In this work, we first study the properties of various skin-color filters, which are typically used for face detection tasks. Then, a new model - called "adaptive hue lookup table" (AH-LT) is developed. By introducing the so-called Camshift technique [7], the AH-LT model is able to take advantages of the available color information, and adaptively update its thresholds for different input images to identify skin-color. In the experiment reported here, we compare the AH-LT model with four typical skin-color filters in the scenarios of different human races and illuminations. The AH-LT model shows promising results.

2 Skin-Color Models

2.1 2D Gaussian Model in RGB Color Space (RG-GM)

In the RGB color space, a triple of $[R, G, B]$ represents not only color but also brightness. A typical way to separate chromatic colors or *pure* colors (r, g) from brightness is to apply a normalization process [8],

$$r = R/(R+G+B), \quad g = G/(R+G+B) \tag{1}$$

Eq.1 defines a mapping from $\mathbb{R}^3$ to $\mathbb{R}^2$, and the color blue is redundant after the normalization due to $r+g+b=1$. It is found by Yang *et al.* [6] that the human skin-color distribution tends to cluster in a small region in the (r, g) space, although in reality skin-colors of different people appear to vary over a wide range. These variations are generally believed to be mainly caused by brightness or intensity. Thus, the skin-color distribution can be represented exactly by a 2D Gaussian model $N\left(u, \Sigma^2\right)$ where $u = (\bar{r}, \bar{g})^T$,

$$\bar{r} = \frac{1}{N}\sum_{i=1}^{N} r_i, \quad \bar{g} = \frac{1}{N}\sum_{i=1}^{N} g_i, \quad \Sigma = \begin{bmatrix} \sigma_{rr} & \sigma_{rg} \\ \sigma_{gr} & \sigma_{gg} \end{bmatrix}, \tag{2}$$

N is the number of the training pixels. Fig.1:Left shows the skin-color distribution of a set of training samples, which are collected from different racial people. For simplicity, we call the model "RG-GM" hereafter. A successful application of RG-GM is Yang's face tracker [6].

2.2 2D Gaussian Model in HSV Color Space (HS-GM)

Compared to RG-GM, a better way to extract chromatic colors seems to transform color representation from the (R, G, B) space to the (H, S, V) space, where H denotes Hue distinguishing pure colors such as red, green, purple and yellow, S denotes Saturation referring to how far color is from a gray of equal intensity, and V denotes Value embodying the lightness or intensity. Compared with the (R, G, B) color space, the (H, S, V) space embodies the artist's ideas of

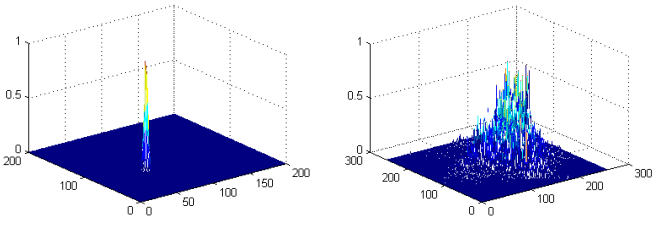

Fig. 1. The skin color distributions in the normalized (r, g) space (Left), and the (h_x, h_y) space (Right), respectively

tint, shade and tone. Unlike the normalization of (R, G, B), the mapping from (R, G, B) to (H, S, V) is nonlinear. Research studies indicate that HSV-based models outperform RGB-based models in skin pixel classification [3,7].

Similar to the normalized RGB space, observations show the human skin-color distribution also tends to cluster in a small region in the HSV space [9]. As such, we can model the skin-color distribution using a 2D Gaussian Model of (H, S) components. It is well-known that H represents angle and S represents distance. In order to combine the two variables with different units, we can derive a pair of new variables (h_x,h_y) from (H, S) to represent color pixels, where $h_x = S \cdot \cos(H)$, and $h_y = S \cdot \sin(H)$. Then, the skin-color distribution can be modeled by a 2D Gaussian distribution: $N(\mu, K)$ with

$$\mu = \frac{1}{N}\sum_{n=1}^{N} H_n, \quad K = \frac{1}{N}\sum_{n=1}^{N} (H_n - \mu)(H_n - \mu)^T \tag{3}$$

where $H_n = \left[h_x^{(n)}, h_y^{(n)}\right]^T$, and $\left(h_x^{(n)} h_y^{(n)}\right)$ denotes the nth pixel. The model has been applied into extraction of hand region [9]. For simplicity, we call it HS-GM hereafter. Fig.1:Right shows the distribution of a set of training skin-color pixels used in our experiments.

2.3 1D Lookup Table Based on Hue Histogram (H-LT)

Some researchers [7] found that saturation is also influenced by lightness. Thereby, a simple but efficient model is derived only from the histogram of the H (hue) component. In this method, a hue histogram of the training skin-color pixels is first built. Then, the histogram is smoothed by a Gaussian low-pass filer. The values in each bin are further normalized to the range $[0, 1]$. The obtained histogram is called hue lookup table (H-LT hereafter). The values in the H-LT cells reflect the likelihood that the corresponding color is classified to the skin color. Fig.2 shows the hue histogram used in our experiments.

2.4 Adaptive Hue Lookup Table Model (AH-LT)

Often it is found to be insufficient to use only hue information for skin-color classification in practical applications [4]. In order to improve the performance,

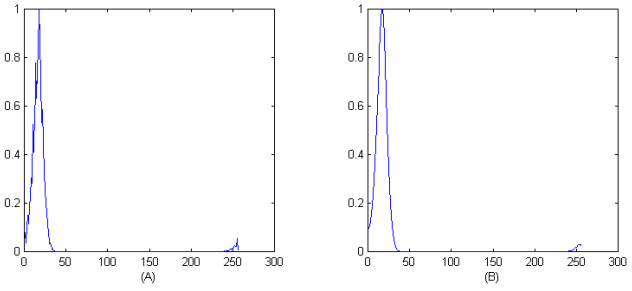

Fig. 2. (A): Original hue histogram, (B): Smoothed hue histogram

it is important to integrate saturation and value information into the skin-color models. However, since the saturation and value of skin-color often vary with illumination conditions, the models have to be able to adaptively update their decision boundaries on (S, V) according to the illumination conditions of the input images. To this end, Cho *et al.* [10] presented an adaptive 3D threshold box model (3D-TBox) in the HSV color space. The 3D box is constructed by 6 thresholds, *i.e.* upper and lower thresholds of H, S and V respectively. All the pixels whose (h, s, v) fall in the box are identified as skin-color pixels. In the model, the thresholds of hue is considered to be stable and fixed, while those of saturation and value are adaptively updated when a new image is tested. The update is implemented by a simple search procedure of gravity center. We found that the updating algorithm can be improved using an efficient alternative method - called "Continuously Adaptive Mean Shift Algorithm" (Camshift), which is based on the Meanshift technique [11]. Also, it seems to be not a good way to classify skin color using a pair of *hard* hue thresholds. An obviously better alternative is the *soft* H-LT model.

Based on the above two points, an improved model - called "Adaptive Hue Lookup Table Model" (AH-LT) is proposed here. The AH-LT model integrates Camshift, H-LT and the adaptive scheme of 3D-TBox together. The advantages of such a combination will be demonstrated in our experiments. The detail procedure of how the AH-LT model works can be divided into two steps: offline and online. In the offline learning step, we firstly build the H-LT model from a set of given skin-color pixels. Then, the initial upper and lower thresholds for S, V are chosen manually by observing the skin color distributions of training sample images that are obtained under various illumination conditions. In order to compare our method with the 3D-TBox as fair as possible, we use the same initial threshold values for S and V as [10].

In the online test step, we first go through the input image using H-LT to find all skin-color pixel candidates. The threshold for the likelihood is set to 0.3 in our experiments. Then, we find the distribution of the skin-color pixel candidates in the (S, V) space by constructing a 2D (S, V) histogram of these candidates. All the values in the 2D histogram is linearly normalized to the range $[0, 1]$. Thus, we can obtain a likelihood $p(s_i, v_i) \in [0, 1]$ for any candidate pixel

from the normalized histogram or lookup table. Since (S, V) of skin-color often shift with variation of lightness conditions, we have to adaptively update the thresholds of (S, V) for different test images. This can be done by finding the mode of the probability distribution $p(s_i, v_i)$ using the Camshift algorithm [7].

Since the Camshift algorithm is derived from the Meanshift [11], it is necessary to introduce the Meanshift prior to the Camshift. The Meanshift is a non-parametric technique that climbs the gradient of a probability distribution to find the nearest dominant mode (peak). The procedure to calculate the Meanshift algorithm is given as follows:

1. Set size and location of initial search window W_0.
2. Compute the mean location in the search window. Let M_{00}, M_{10} and M_{01} denote the zero-th and first moments for (s, v). These moments can be found by $M_{00} = \sum_{(s,v)\in W_i} p(s,v)$, $M_{10} = \sum_{(s,v)\in W_i} s \cdot p(s,v)$, and $M_{01} = \sum_{(s,v)\in W_i} v \cdot p(s,v)$, where W_i denotes current search window. Then we have the mean location (s_c, v_c), where $s_c = M_{10}/M_{00}$ and $v_c = M_{01}/M_{00}$.
3. Center the search window at the mean location computed in Step 2.
4. Repeat Step 2 and 3 until convergence..

A shortcoming of the Meanshift algorithm is that the size of the search window cannot be updated, but it is overcomed in the Camshift algorithm. The complete procedure to update the thresholds of (S, V) using Camshift is given as follows:

1. Set size and location of initial search window W_0 (as shown in [10]).
2. Do Meanshift as above.
3. Update the search window size. Let M_{20} and M_{02} denote the second moments, and we have
$$M_{20} = \sum_{(s,v)\in W_i} s^2 \cdot p(s,v),\ M_{02} = \sum_{(s,v)\in W_i} v^2 \cdot p(s,v). \tag{4}$$
Then the length and width of the probability distribution "blob" can be found as in [12]. Let $a = \frac{M_{20}}{M_{00}} - s_c^2$, $b = 2\left(\frac{M_{11}}{M_{00}} - s_c \cdot v_c\right)$, and $c = \frac{M_{02}}{M_{00}} - v_c^2$. We have the length and width of the new search window,
$$l = \sqrt{(a + c + \sqrt{b^2 + (a-c)^2})/2},\ w = \sqrt{(a + c - \sqrt{b^2 + (a-c)^2})/2}. \tag{5}$$
4. Repeat Steps 2 and 3 until convergence. The final search window gives new thresholds of (S, V).

An assumption behind AH-LT is that the areas of real skin-color regions are comparable to (or larger than) the areas of those regions similar to skin-color. Otherwise, the Camshift may converge to the largest false skin-color region. Fig.3 show an example obtained by AH-LT. The initial detect result without threshold updating is shown in Fig.3(B), where some parts of the clothes were detected as well as the face region. However, one can see that most of false detects has been removed in Fig.3(C), after the threshold values of (S, V) are updated accordingly.

Fig. 3. An skin-color detect example using AH-LT. (a): Input image; (b): Result obtained with the initial threshold values recommended in [10]; (c): Result obtained with the updated threshold values.

3 Experimental Results

In the experiments, a set of human images were collected from the Internet. They are partitioned into two sets: training set and test set with no overlapping between the two. Skin-color pixels are cut manually from the training images to form a set of skin-color samples, which contain a total of 56279 pixels. The testing set contains 200 images, consisting of different racial people (Asian, African and Caucasian). Since the images were taken and digitized under various conditions, it can be said that no special illuminations or other constraints are imposed on the test images. Two experiments are implemented for the evaluation. One is designed to compare performance of the five skin color filters for different racial people. To this end, the testing set is manually partitioned into three groups according to races of people in the images. Another is designed to compare performance of the skin-color filters in different illuminations. In this experiment, according to the illumination variations of the skin regions in the images, we partition the test set to four groups: normal (161), reddish (13), bright (17) and dark (9). The five skin-color filters are applied in the two experiments, and the obtained results are shown in Table 1, where we define a detect if at least half of the true skin-color regions are found in a given test image.

In the first experiment, there are 88 images of Asian, 12 of African and 100 of Caucasian in the test set. From Table 1(Races), it can be seen that there are not significant difference in terms of detect rate among the three groups for each skin-color model. This demonstrates the observation by Yang *et al.* [6], that is, human skin colors tend to cluster into a small region in a color space. Also, not surprisingly, the HSV-based models (HS-GM, H-LT, 3D-TBox and AH-LT) are overall superior to the RGB-based model (RG-GM) in all the three groups. This result is consistent with our analysis in previous sections.

In the second experiment, it can be seen from Table 1(Illuminations) that RG-GM has the lowest detect rate among the five models. Specifically, it is difficult for RG-GM to detect the reddish or too dark skin-color regions. For example, only 2 out of 13 reddish images and 1 out of 9 dark images are detected by the

Table 1. Detect rates (%) of the five skin-color models in different experimental conditions of race and illumination

Algs.	Illuminations					Races		
	Normal	Reddish	Bright	Dark	Overall	Asian	African	Caucasian
RG-GM	80.1	18.2	100	11.1	74.5	78.4	83.3	73
HS-GM	82	92.3	64.7	88.9	81.5	79.5	91.7	82
H-LT	93.8	92.3	100	100	94.5	93.2	91.7	96
3D-TBox	85.7	76.9	82.4	100	85.5	84.1	91.7	89
AH-LT	90	76.9	94.1	100	90	87.5	91.7	92

method. This shows that the RGB-based models are rather sensitive to variations of illumination, because the saturation component influenced by illumination is not separated. Similar observations are also found by Bradski *et al.* [7].

In Table 1, the overall detect rate of HS-GM is 81.5%. This result is higher than that of RG-GM, but lower than those of the other three models. It failed when the skin-color regions are too bright. Although H-LT obtained the highest detect rate among the five models, it is found that its false alarm is much higher than AH-LT. One reason is that H-LT cannot adaptively update its parameters for the specific illumination conditions in different images. In contrast with H-LT, saturation and value can be appropriately adjusted in the AH-LT approach to fit the requirements of different inputs. As a result, some false detects can be removed and skin-color regions can be more accurately extracted as shown in Fig.3. Therefore, we have reasons to believe that the AH-LT method embodies a better trade-off between detect rate and false alarm.

4 Conclusion

In this paper, a new skin-color model is introduced by combining several commonly used techniques, such as the hue lookup table, the continuously adaptive mean shift, and the adaptive update of thresholds. Also, a comparative study between the proposed method and four traditional methods is carried out in various experimental settings such as races and illuminations. The obtained results indicate that the proposed AH-LT method is a promising solution to balance the tradeoff of detection rate and false alarm. Due to low computational costs and insensitivity to most facial variations such as view-points, scale, rotation and expressions, we expect that the AH-LT method can be used as an important pre-processing step in a real-time face location/tracking system.

Following the work presented here, there are several interesting topics to be conducted in the future. First, many of existing color-based methods use Gaussian models to approximate the skin-color distributions. However, it has been found that the practical distributions in the color spaces are much more complicated than Gaussian. Thus, it seems to be a better solution to map the color spaces to a feature space, where the assumption is closer to be true. Such a

mapping can be implemented by using a kernel technique such as [13]. Furthermore, more sophisticated pattern recognition techniques, such as discriminant analysis used in face recognition [14,15,16] can be applied in the feature spaces to enhance the separability between the two classes, skin and non-skin pixels.

References

1. Li, S.Z., Lu, J.: Face detection, alignment and recognition. In Medioni, G., Kang, S.B., eds.: EMERGING TOPICS IN COMPUTER VISION. Prentice-Hall, Upper Saddle River, New Jersey, ISBN: 0-13-101366-1 (2004)
2. Yang, M.H., Kriegman, D.J., Ahuja, N.: Detecting faces in images: A survey. IEEE Transactions on Pattern Analysis and Machine Intelligence **24** (2002) 34–58
3. B., Z., B., S., , F., Q.: Comparison of five color models in skin pixel classification. In: Proc. International Workshop on Recognition, Analysis, and Tracking of Faces and Gestures in Real-Time Systems, Corfu, Greece (1999) 58–63
4. Phung, S.L., Bouzerdoum, A., Chai, D.: Skin segmentation using color pixel classification: Analysis and comparison. IEEE Transactions on Pattern Analysis and Machine Intelligence **27** (2005) 148–154
5. Wren, C., Azarbayejani, A., Darrell, T., Pentland, A.: Pfinder: real-time tracking of the human body. In: Proc. SPIE. Volume 2615. (1996) 89–98
6. Yang, J., Waibel, A.: Tracking human faces in real-time. Technical report (CMU-CS-95-210), Carnegie Mellon University (1995)
7. Bradski, G.R.: Computer vision face tracking for use in a perceptual user interface. Intel Technology Journal (1998)
8. Wyszecki, G., Styles, W.: Color Science: Concepts and Methods, Quantitative Data and Formulae. 2 edn. John Wiley and Sons, Inc., New York (1982)
9. S., T., A., K., T., W., Y., M., M., I.: Extraction of hand region and specification of finger tips from color image. In: Proc. of International Conference on Virtual Systems and MultiMedia. (1997)
10. Cho, K.M., Jang, J.H., Hong, K.S.: Adaptive skin-color filter. Pattern Recognition **34** (2001) 1067–1073
11. Comaniciu, D., Meer, P.: Mean shift analysis and applications. In: Proc. IEEE Int'l Conf. Comp. Vis., Kerkyra, Greece (1999) 1197–1203
12. Freeman, W., Tanaka, K., J.Ohta, Kyuma, K.: Computer vision for computer games. In: Int. Conf On Automatic Face and Gesture Recognition. (1996) 100–105
13. Lu, J., Plataniotis, K., Venetsanopoulos, A., Wang, J.: An efficient kernel discriminant analysis method. to appear in Pattern Recognition (2005)
14. Lu, J., Plataniotis, K., Venetsanopoulos, A.: Face recognition using LDA based algorithms. IEEE Transactions on Neural Networks **14** (2003) 195–200
15. Lu, J., Plataniotis, K., Venetsanopoulos, A.: Face recognition using kernel direct discriminant analysis algorithms. IEEE Transactions on Neural Networks **14** (2003) 117–126
16. Lu, J., Plataniotis, K., Venetsanopoulos, A.: Regularization studies of linear discriminant analysis in small sample size scenarios with application to face recognition. Pattern Recognition Letter **26** (2005) 181–191