MVA2002 IAPR Workshop on Machine Vision Applications, Dec. 11-13, 2002, Nara-ken New Public Hall, Nara, Japan

3—3

Skin-Color-Based Image Segmentation and Its Application in Face Detection

Quan Huynh-Thu Mitsuhiko Meguro Masahide Kaneko
Department of Electronic Engineering, The University of Electro-Communications *

Abstract

In this paper, we propose a technique to efficiently detect human skin in color images with complex background and varying illumination conditions. The strategy consists of modeling the distribution of skin colors using a Gaussian mixture model (GMM) and segmenting the image into skin and non-skin parts, using threshold values that are computed automatically by an adaptive technique. Morphological operators are finally applied to refine the final skin regions and obtain candidate face regions. We show that a mixture model of Gaussians can provide a robust representation of the human skin color to accommodate with large color variations. Using images taken from the Internet and in-house pictures containing persons with various skin-color types in different illumination conditions, experimental results show that the proposed method can cope with a wide range of illumination situations and complex backgrounds.

1 Introduction

Within the last decade, research on processing and analysis of images containing human faces has been constantly increasing for purposes such as intelligent vision-based human-computer interaction, face tracking, face recognition, facial expression analysis and human emotion recognition systems. However, most of the facial image processing methods assume that faces have been previously localized and identified within the image. A robust face detection technique is therefore a requirement to build fully automated systems that analyze facial information. With sizes of digital image databases growing constantly, managing digital images using other search functions than name, date or image description will become more and more necessary. In the future, organization of personal pictures could be made using search functions such as a person's face. However, personal pictures often contain images taken under a wide range of uncontrolled conditions, such as complex backgrounds and varying illumination conditions.

The task of human face detection is to determine in an arbitrary image whether or not there are any human faces in the image, and if present, localize each face and its extent in the image, regardless of its three-dimensional position and orientation. Such a problem is a very challenging task because faces are non-rigid forms and have a high degree of variability in size, color, shape and texture. In this paper we propose a method to extract skin regions in color images in cases of a wide range of skin-color variations, varying lighting conditions, different face orientations and complex backgrounds. Based on a set of sample pixels extracted from a wide variety of persons of different skin colors and captured in different lighting conditions, a Gaussian mixture model of this distribution is built. Using this skin-color model, skin-color segmentation is performed to extract skin regions. Our skin-color segmentation uses threshold values that are computed automatically using an adaptive technique. Finally, skin regions are refined by using morphological operations in order to produce the candidate face regions.

*Address: 1-5-1 Chofugaoka, Chofu-shi, Tokyo 182-8585, Japan. E-mail: {quan, meguro, kaneko}@radish.ee.uec.ac.jp

2 Color information processing

Because color is a powerful fundamental feature and because it is, under constant illumination, almost invariant to scale, orientation and partial occlusion, we propose a method using color information to extract efficiently human skin in images captured in complex environmental conditions. Skin-color image segmentation is computationally inexpensive and is robust to cluttered background.

In this research, we use the HSV (hue, saturation, value) color representation because it is compatible with the human color perception and because previous work has shown that this color space is one of the most adapted for skin-color detection [12]. The HSV color space is obtained by a non-linear transformation of the fundamental RGB color space. We use the cone representation of the HSV color space, where H, S and V are all normalized in the range [0,1]. The H and S components represent the chromatic information, while V represents the luminance information.

3 Skin-color model

Although people from different ethnicities have different skin colors in appearance, experiments have shown that skin colors of individuals cluster closely in the color space, i.e. color appearances in human faces differ more in intensity than in chrominance [2, 7, 8, 11]. Therefore, the V component is discarded in order to reduce dependence to lighting conditions, while H and S are used to build a 2D model of the skin-color distribution.

Figure 1 shows the distribution in the chromatic space of samples of skin colors from 253 persons from different ethnicities, e.g. Caucasian, Asian, African and dark-skin types, and captured in different illumination conditions. Figure 1(b) shows the plot of H and S values of these samples. We see that even though samples are from persons of different types of skin-color, a cluster exists in the chromatic space. The samples were extracted from a set of face images randomly taken from the WWW and a set of in-house data. The illumination conditions for these samples ranged from poor lighting to bright lighting, with no control on the direction of the illumination source. These pictures were captured in both outdoor and indoor situations, and are thought to represent a large sample of the typical variations met in personal pictures captured with different digital cameras. We only considered images that were thought to be captured with a white illuminant, as a non-white illuminant changes completely the true skin color.

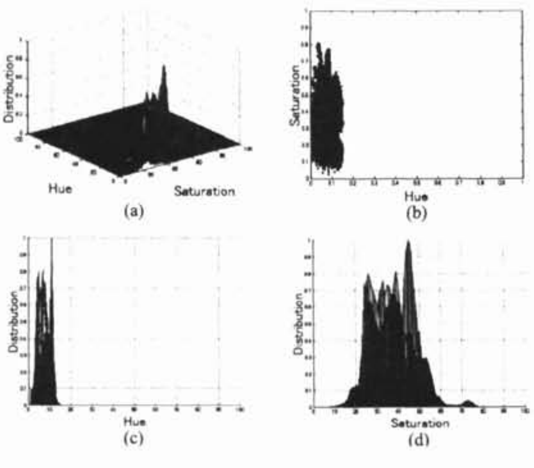

Figure 1: Distribution of samples of skin-pixels: (a) distribution, (b) chromatic chart, (c) projection in hue and (d) saturation axes.

Previous research works have modeled the skin-color distribution by a multivariate normal (Gaussian) distribution in the chromatic space [1, 5, 6, 9]. However, even if there exists a small cluster in the chromatic space, using a single Gaussian function is not effective enough to model human skin-color distribution. One main problem occurring when using a Gaussian model is partial extraction of the face due for example to the presence of facial highlights and complex lighting. Furthermore, color appearance is influenced by different lighting environments or object movement. Different cameras may produce different color values, even for the same person under the same constraints of pose and illumination. Finally, human skin color differs from person to person. To address all these problems, we propose to model the skin-color distribution using a Gaussian mixture model (GMM):

$$\begin{aligned} P(x,\theta) &= \sum_{i=1}^{k} \alpha_i \frac{1}{\sqrt{(2\pi)^d |\Sigma_i|}} \\ &\quad \times \exp\{-\frac{1}{2}(x-\mu_i)^T \Sigma_i^{-1} (x-\mu_i)\} \end{aligned} \quad (1)$$

where k is the number of Gaussians in the mixture model, the parameter set $\theta = \{\alpha_i, \mu_i, \Sigma_i\}_{i=1}^{k}$ is such that: $\sum_{i=1}^{k} \alpha_i = 1$, $\alpha_i > 0$, $\mu_i \in I\!R^d$ and Σ_i is a $d \times d$ positive definite matrix.

The parameters of the model are estimated using the EM (Expectation-Maximization) algorithm [4]. This algorithm is an iterative method to obtain the maximum likelihood estimation of the parameters:

$$\alpha_i^{new} = \frac{1}{N} \sum_{j=1}^{N} P(i|x_j, \theta_i) \quad (2)$$

$$\mu_i^{new} = \frac{\sum_{j=1}^{N} x_j P(i|x_j, \theta_i)}{\sum_{j=1}^{N} P(i|x_j, \theta_i)} \quad (3)$$

$$\Sigma_i^{new} = \frac{\sum_{j=1}^{N} P(i|x_j, \theta_i)(x_j - \mu_i^{new})(x_j - \mu_i^{new})^T}{\sum_{j=1}^{N} P(i|x_j, \theta_i)} \quad (4)$$

The initialization of the mixture parameters is done using the K-mean algorithm. Using Equations (2)-(4), the newly derived parameter values are then used as the guess for the next iteration.

Figure 2 shows the mixture model of the skin-color distribution of Figure 1, built with 4 Gaussian components using Equation (1).

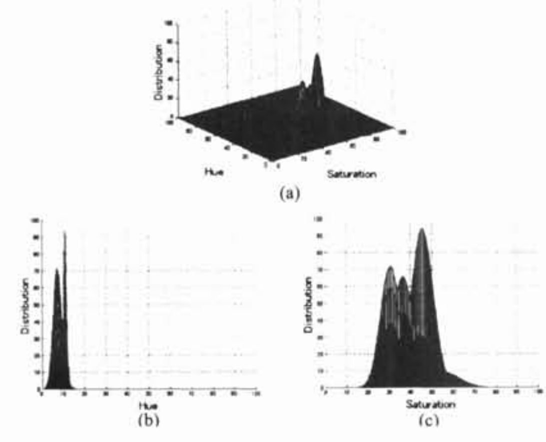

Figure 2: Gaussian mixture model of the skin-color distribution of Figure 1: (a) distribution, (b) projection in hue and (c) saturation axes.

4 Skin-color segmentation

4.1 Skin likelihood

Using the skin-color model, the original color image can be transformed into a likelihood image. This likelihood image is a grayscale image where the gray value of each pixel shows the probability of the pixel to represent skin color. Bright pixels have a high probability to represent skin color, while dark pixels have a low probability to represent skin color. Figure 3 shows an example of such a likelihood image. With an *appropriate threshold value*, the likelihood image can be segmented into a binary image showing skin and non-skin regions:

$$pixel = \begin{cases} \text{skin pixel} & \text{if likelihood>threshold} \\ \text{non-skin pixel} & \text{else} \end{cases}$$

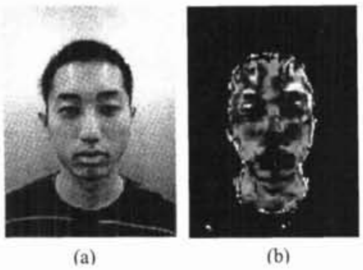

Figure 3: (a) original image, (b) likelihood image.

4.2 Adaptive threshold

GMM has been used by previous researchers for skin-color modeling [3, 10]. In these works, skin-color segmentation was based on a unique threshold value arbitrarily decided. However, when processing images of different people taken in different imaging conditions, the use of only one single threshold value is not adapted to deal with the wide range of variations.

We use a method to compute automatically the optimal threshold value using an adaptive technique. The algorithm is based on the observation that a high threshold value will give a small segmented area, while a low threshold value will give a larger one. Decreasing the threshold from an initial high value, the size of the detected skin region will likely remain stable under a certain range of threshold values until the threshold value becomes too small such that non-skin regions merge with skin-regions, resulting in a sharp increase of the total segmented area. If the increase in size of a region is plotted as a function of the decrease in threshold value, the obtained curve gradually decreases up to a point where it sharply increases. The optimal threshold is the one at which the minimum increase in region size is observed while stepping down the threshold value from an initial value T_i to a final value T_f. We use a decreasing step of 0.1, with different T_i and T_f according to each Gaussian component G_i:

	G_1	G_2	G_3	G_4
T_i	0.75	0.75	0.75	0.35
T_f	0.05	0.35	0.05	0.05

Furthermore, as we use a mixture model composed of k Gaussian components, one optimal threshold is estimated independently for each of the k components, i.e. we compute k *optimal threshold values*. These are estimated *a posteriori* according to the input image and are therefore not fixed a priori. Each Gaussian component with its corresponding threshold value will extract a different skin region on the face. On the original image in Figure 3, the top-down lighting produced clearly some highlights on the forehead, while most of the remaining part of the face is darker. Figure 4 shows the skin regions detected by each of the four Gaussian components G_i of our model. It can be seen that each Gaussian component has detected a different part of the facial area, corresponding to a different skin-color variation due to the illumination environment.

The final segmentation image is a logical *OR* combination of each of the segmented skin region obtained respectively with each Gaussian component and its corresponding threshold value. Figure 5(a) shows the skin segmentation result obtained by combining the partial results of Figure 4.

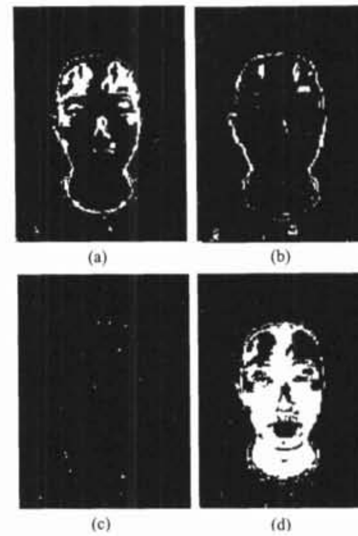

Figure 4: Partial segmentation results of Figure 3, for each Gaussian component G_i: (a) G_1 (T_1=0.7), (b) G_2 (T_2=0.7), (c) G_3 (T_3=0.6) and (d) G_4 (T_4=0.3).

5 Refinement of candidate regions

Due to the existence in the background of colors similar to skin tones, skin-color segmentation in images with complex background can yield small isolated groups of pixels in the segmented image. These regions have typically a size of a few pixels and may be assimilated to noise, which can be eliminated using a median filtering. Such small blobs can be seen in Figure 5(a). Morphological operators, such as opening and closing, are then applied to refine the remaining regions and construct a face mask. These morphological operators use a structural element, which is a matrix that defines a neighborhood shape and size. As human faces are likely to have round shapes, our structural element is a disk. The size of the structural element used for each morphological operation is set automatically according to the sizes of the image and the skin region. Applying the face mask on the original image, we get the final candidate face regions. Figure 5(b) shows the final candidate face region of the image in Figure 3(a), after morphological cleaning of the segmentation result. In this case, the size of the structural element was computed to include 13 neighbor pixels.

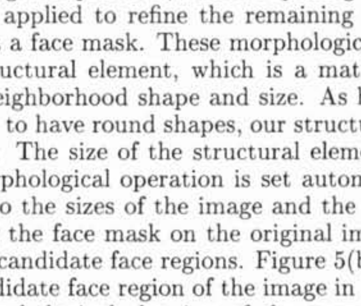

Figure 5: Skin-color extraction results: (a) skin segmentation result of Figure 3, (b) candidate face region after morphological cleaning of (a).

6 Experimental results

Figure 6 shows skin detection results on images with cluttered background, varying illumination conditions, with persons from different ethnic types and sometimes with several persons contained in the image. These results show that the entire facial regions are clearly extracted, regardless of face orientation. In some of these examples we see that non-face objects with colors similar to skin-color, such as hands and clothing, may also be falsely extracted. Using only color information, it is not possible to separate completely the true face regions from regions similar to skin tones. However, we have shown that our method can be a robust step in localizing candidate face regions in a face detection algorithm.

Figure 6: Skin-color extraction for face detection: (a) original image, (b) skin-color extraction result.

7 Conclusion

Face detection using color information usually consists of two main steps: (1) localization of candidate face regions and (2) validation of the face hypothesis using some additional information about face structure. We have proposed a method to extract efficiently candidate face regions in images with varying lighting condition and in presence of complex background, with people of different ethnicities and with several persons contained within the image. While Gaussian modeling and hard thresholding usually fail to extract completely skin regions in complex illumination conditions, our method manages to detect skin regions on the entire face. Furthermore, while other GMM-based methods use skin-color segmentation with a unique threshold value defined arbitrarily, we compute automatically and adaptively the threshold value for each of the component of our model. This makes our method more effective to process a large range of environmental variations in images. The next step in our research is the detection of facial features and the use of additional information about face structure to validate or reject the candidate regions and build a complete face detection system.

References

[1] J. Cai and A. Goshtasby. Detecting Human Faces in Color Images. *Image and Vision Computing*, 18:63–75, Feb 1999.

[2] H.P. Graf, E. Cosatto, D. Gibbon, M. Kocheisen, and E. Petajan. Multimodal System for Locating Heads and Faces. In *Proceedings of the Second IEEE International Conference on Automatic Face and Gesture Recognition*, pages 88–93, 1996.

[3] H. Greenspan, J. Goldberger, and I. Esthet. Mixture Model for Face-Color Modeling and Segmentation. *Pattern Recognition Letters*, 22:1525–1536, 2001.

[4] R. A. Redner and H. F. Walker. Mixture Densities, Maximum Likelihood and the EM Algorithm. *SIAM Review*, 26(2):195–239, 1984.

[5] E. Saber and A.M. Tekalp. Frontal-view Face Detection and Facial Feature Extraction using Color, Shape and Symmetry Based Cost Functions. *Pattern Recognition Letters*, 8:669–680, June 1998.

[6] J-C. Terrillon, M. David, and S. Akamatsu. Detection of Human Faces in Natural Scene Images by Use of a Skin Color Model and Invariant Moments. In *Proceedings of IEEE Third International Conference on Automatic Face and Gesture Recognition*, pages 112–117, 1998.

[7] J. Yang, W. Lu, and A. Waibel. Skin-Color Modeling and Adaptation. Technical Report CMU-CS-97-146, Carnegie Mellon University, 1997.

[8] J. Yang and A. Waibel. A Real-Time Face Tracker. In *Proceedings of the Third Workshop on Applications of Computer Vision*, pages 142–147, 1996.

[9] M-H Yang and N. Ahuja. Detecting Human Faces in Color Images. In *Proceedings of IEEE International Conference on Image Processing*, pages 127–139, 1998.

[10] M-H. Yang and N. Ahuja. Gaussian Mixture Model for Human Skin Color and Its Applications in Image and Video Database. In *Conference on storage and retrieval for image and video databases, SPIE '99*, volume 3656, pages 458–466, 1999.

[11] H. Yao and W. Gao. Face Detection and Location Based on Skin Chrominance and Lip Chrominance Transformation from Color Images. *Pattern Recognition*, 34:1555–1564, 2001.

[12] B.D. Zarit, B.J. Super, and F.K.H. Quek. Comparison of Five Color Models in Skin Pixel Classification. In *ICCV'99 International Workshop on Recognition, Analysis and Tracking of Faces and Gestures in Real-Time Systems*, pages 58–63, 1999.