I-101

Gaussian Mixture Modeling of Skin Color and Adaptive Thresholding for Face Detection

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

1. Introduction

Processing of facial information has been an active research area for purposes such as intelligent vision-based human-computer interaction, face tracking, face recognition and facial expression analysis. 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. 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. While many of the earlier applications of face detection processed images captured under controlled lighting condition and with a rather simple background, new applications have emerged in recent years, which require faces to be detected in presence of complex background and under varying lighting condition.

In this paper, we propose a method to extract efficiently skin-color regions to locate human faces in digital images. Our approach is based on a Gaussian mixture model of skin color and a skin-color segmentation using an automatic adaptive thresholding technique to segment the image into skin and non-skin regions. Finally, the number of candidate faces is reduced by using morphological operations.

2. Processing of color information

HSV (Hue, Saturation, Value) color space is used to process color information because it has shown to be one of the most adapted to skin-color detection. It is also compatible with the human color perception. We use the cone representation of the HSV color space, where H, S and V are all normalized in the range [0,1]. H and S represent the chromatic information, while V represents the luminance information. 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 [1]. Therefore, V component is discarded to reduce dependence to lighting conditions and a 2D model of the skin-color distribution is built in the chromatic space (H,S).

3. Skin-color modeling

Previous research works have modeled the skin-color distribution with a multivariate normal (Gaussian) distribution in the chromatic space. 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. Figure 1(a) shows the distribution of skin-color samples taken from 253 persons of different ethnicities and in different illumination conditions. We see that different modes co-exist and that one single Gaussian is not effective enough to model properly this distribution. Furthermore, one main problem arising when using skin-color segmentation results for face detection is partial extraction of the face due for example to the presence of facial highlights. Besides, 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 these problems, we model the skin-color distribution with a Gaussian mixture model (GMM):

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

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

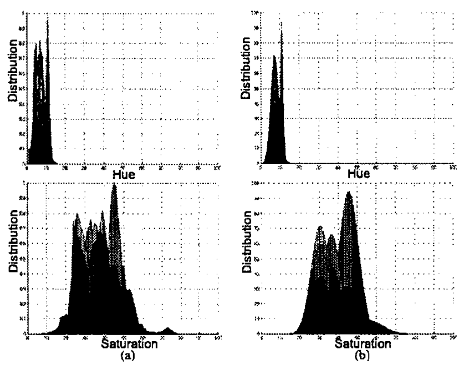

Figure 1: (a) distribution of skin-color samples, (b) GMM.

are estimated using the Expectation-Maximization algorithm. Figure 1(b) shows the mixture model built with 4 Gaussian components.

4. Skin-color segmentation

We call the *likelihood image* a grayscale image where the gray value of each pixel shows the probability of the pixel to represent skin color. Using the skin-color model, we can compute such a likelihood image from the original color image. For each entry pixel of chromatic pair (H,S), a corresponding skin-likelihood value for this pixel can be computed using (1). A label $l \in [1, k]$ is assigned to the pixel, according to the Gaussian component that has the highest probability. With appropriate thresholding, the pixel is then decided to be skin or non-skin pixel based on the following rule:

$$label = \begin{cases} l & \text{if likelihood>threshold} \\ 0 & \text{else} \end{cases} \quad (2)$$

For a likelihood greater than a certain threshold value, the pixel is labeled with the selected Gaussian l. Otherwise, it is labeled as a non-face pixel, i.e. $l = 0$. By doing so, we can divide the image into $k + 1$ decision regions, i.e. k skin-color regions and the non skin-color region. Next, we obtain a binary image showing skin and non-skin regions by assigning level 0 to non-skin region and level 1 to all the detected skin regions. Finally, morphological operations, such as opening and closing, are performed to refine the extracted skin regions and build a face mask. Applying the mask on the original color image, we get the final candidate face regions.

5. Adaptive thresholding

While previous segmentation techniques were based on an a priori fixed threshold value [2, 3], we compute automatically the optimal threshold value using an adaptive method. 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 background 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. T_i and T_f are chosen to be different according to each Gaussian component G_l. Furthermore, one optimal threshold is estimated for each of the k Gaussians, resulting in k *optimal threshold values.* The final segmentation result is a logical *OR* combination of each of the segmentated region obtained respectively with each threshold value.

6. Experimental results

Experimental results of Figure 2 show that a GMM can provide a robust representation of the human skin color to accommodate with large color variations. Using this model, the entire facial regions could be extracted efficiently in images with complex background, varying illumination condition and persons of different ethnic types, regardless of face orientation, size, occlusion and the number of persons in the image. However, using only color information, it is not possible to separate completely the background from the true face regions. In Figure 2, we see that non-face objects, such as hands, may also be falsely extracted as face regions because of color similarity with face color.

(a) (b)

Figure 2: Face extraction examples: (a) original image, (b) face region extraction.

7. Conclusion

We have proposed a method to extract efficiently the entire facial regions in images with varying lighting condition and in presence of complex background, with people of different ethnicities and with several persons within the image. Experimental results show that a mixture of Gaussians can provide a robust representation of the human skin color to accommodate with large color variations. While previous skin-color segmentation have used a single threshold value, defined arbitrarily, we compute automatically multiple optimal threshold values to perform efficient skin-color segmentation. By using some additional information about facial structure and facial features, a complete face detection system can be implemented.

References

[1] J. Yang and A. Waibel, "A Real-Time Face Tracker", *Proceedings of the 3rd Workshop on Applications of Computer Vision*, pp. 142–147, 1996.

[2] M-H. Yang and N. Ahuja, "Gaussian Mixture Model for Human-Skin Color and its Applications in Image and Video Database", *Conference on Storage and Retrieval for Image and Video Databases, SPIE '99*, vol. 3656, pp. 458–466, 1999.

[3] H. Greenspan, J. Goldberger, and I. Esthet, "Mixture Model for Face-Color Modeling and Segmentation", *Pattern Recognition Letters*, vol. 22, pp. 1525–2536, 2001.