ISSN 1009 – 3095 *Journal of Zhejiang University SCIENCE* V.4, No.2, P.162 – 165, Mar. – Apr., 2003
http://www.periodicals.com.cn; http://www.zju.edu.cn/English
http://www.zjupress.com; http://lib.zju.edu.cn/eindex.htm; jzus@zju.edu.cn

A color based face detection system using multiple templates*

WANG Tao (王涛)†, BU Jia-Jun (卜佳俊), CHEN Chun (陈纯)
(*College of Computer Science and Engineering, Zhejiang University, Hangzhou 310027, China*)
†E-mail: wt@cs.zju.edu.cn

Received Feb.26, 2002; revision accepted Oct.20, 2002

Abstract: A color based system using multiple templates was developed and implemented for detecting human faces in color images. The algorithm consists of three image processing steps. The first step is human skin color statistics. Then it separates skin regions from non-skin regions. After that, it locates the frontal human face(s) within the skin regions. In the first step, 250 skin samples from persons of different ethnicities are used to determine the color distribution of human skin in chromatic color space in order to get a chroma chart showing likelihoods of skin colors. This chroma chart is used to generate, from the original color image, a gray scale image whose gray value at a pixel shows its likelihood of representing the skin. The algorithm uses an adaptive thresholding process to achieve the optimal threshold value for dividing the gray scale image into separate skin regions from non skin regions. Finally, multiple face templates matching is used to determine if a given skin region represents a frontal human face or not. Test of the system with more than 400 color images showed that the resulting detection rate was 83%, which is better than most color-based face detection systems. The average speed for face detection is 0.8 second/image (400 × 300 pixels) on a Pentium 3 (800MHz) PC.

Key words: Color-based, Multiple templates matching, Face detection
Document code: A **CLC number:** TP391.41

INTRODUCTION

Images containing faces are essential for intelligent vision-based human computer interaction. Face processing methods include those for face recognition, face tracking, pose estimation, and expression recognition. However, many reported methods assume that the faces in an image or an image sequence have been identified and localized. To build fully automated systems that analyze the information contained in face images, robust and efficient face detection algorithms are required. Numerous techniques have been developed to detect faces in a single image. Four basic techniques are commonly used for dealing with the face detection problem: knowledge-based methods (Yang and Huang, 1994), feature invariant approaches (Yang and Waibel, 1996; Kjeldsem and Kender, 1996), template matching methods (Craw et al., 1992; Lanitis et al., 1995), and appearance-based methods (Turk and Pentland, 1991; Rowley et al., 1998; Osuna et al., 1997). Our system is a mixture of feature invariant approach and template matching, which takes skin color as feature invariant and uses multiple human faces as templates which include varied possible poses of a face so that the system can detect faces with almost all the possible poses.

SKIN COLOR STATISTICS

Segmenting human skin regions from non-skin regions based on color requires a reliable skin color model that is adaptable to people of different skin colors and to different lighting conditions. The common RGB representation of color images is not suitable for characterizing skin-color. In the RGB space, the three components (r, g, b) represent not only color but also luminance, which may vary across a person's face

* Project (No. 60203013) supported by National Natural Science Foundation of China

due to the ambient lighting and is not a reliable measure in separating skin from non-skin region. Luminance can be removed from the color representation in the chromatic color space. Chromatic colors, also known as "pure" colors in the absence of luminance, are defined by a normalization process shown below:

$$r = R/(R+G+B) \quad (1)$$

$$b = B/(R+G+B) \quad (2)$$

In fact, Eq.(1) and Eq.(2) define a $R3 \rightarrow R2$ mapping. Color blue is redundant after the normalization because $r + g + b = 1$.

Chromatic colors have been effectively used to segment color images in many applications. It is also well suited in this effort to segment skin regions from non-skin regions. The color distribution of skin colors of different people was found to be clustered in a small area of the chromatic color space. Although skin colors of different people appear to vary over a wide range, they differ much less in color than in brightness. In other words, skin colors of different people are very close, but they differ mainly in intensities. So we could proceed to develop a skin-color model in the chromatic color space.

Two-hundred and fifty skin samples were used to determine the color distribution of human skin in chromatic color space. Our samples were taken from persons of different ethnicities: Asian, Caucasian and African. Fig.1 shows the color distribution of these skin samples in the chromatic color space.

The color histogram revealed that the skin-color distribution of different people are clustered

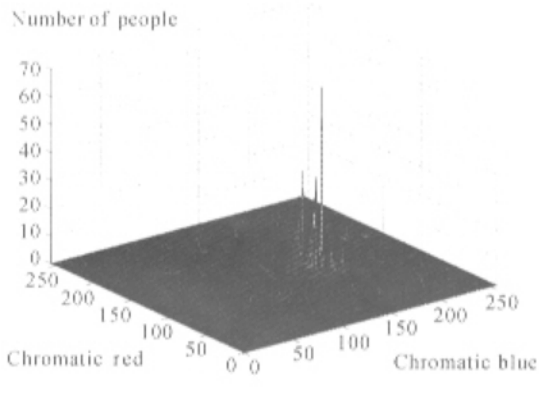

Fig.1 Color distribution for skin-color of different people

in the chromatic color space and that skin color distribution can be represented by a Gaussian model $N(m, C)$, where:

$$\text{Mean:} \quad m = E\{x\} \text{ where } x = (r, b)^{T} \quad (3)$$

$$\text{Covariance:} \quad C = E\{(x-m)(x-m)^{T}\} \quad (4)$$

Fig. 2 shows the Gaussian Distribution $N(m, C)$ fitted by our data.

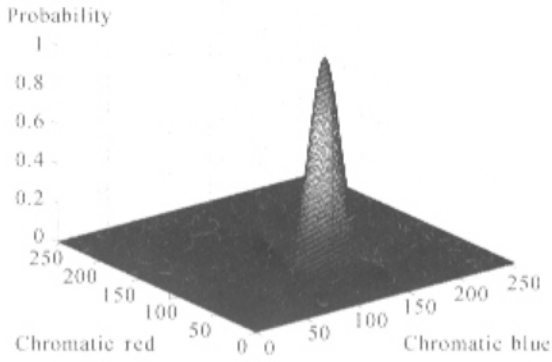

Fig.2 Fitting skin color into a Gaussian distribution

With this Gaussian fitted skin color model, we can now obtain the likelihood of it representing the skin for any pixel of an image. Therefore, if a pixel, having transformed from RGB color space to chromatic color space, has a chromatic pair value of (r, b), the likelihood of it representing the skin for this pixel can then be computed as follows:

$$Likelihood = P(r, b) = \exp[-0.5 \cdot (x - m)^{T} C^{-1} (x - m)] \quad (5)$$

where: $x = (r, b)^{T}$

Hence, this skin color model can transform a color image into a gray scale image such that the gray value at each pixel shows the likelihood of the pixel belonging to the skin. Fig.3 shows the original color image and skin-likelihood image.

Fig.3 (a) the original color image; (b) the skin-likelihood image

SKIN SEGMENTATION

With this Gaussian fitted skin color model, we get a chroma chart that shows likelihoods of skin colors. This chroma chart is used to generate a gray scale image from the original color image. This image has the property that the gray value at a pixel shows the likelihood of that pixel representing the skin.

However, it is important to note that the detected regions may not necessarily correspond to skin. It is only reliable to conclude that the detected regions have the same color as that of the skin. The important point here is that this process can reliably point out regions that do not have the color of the skin and such regions would not need to be considered anymore in the face finding process.

Since the skin regions are brighter than the other parts of the images, the skin regions can be segmented from the rest of the image through a thresholding process. To process different images of different people with different skin, a fixed threshold value is not possible to be found. Since people with different skins have different likelihood, an adaptive thresholding process is required to achieve the optimal threshold value for each run.

The adaptive thresholding is based on the observation that stepping the threshold value down may intuitively increase the segmented region. However, the increase in segmented region will gradually decrease (as percentage of skin regions detected approaches 100%), but will increase sharply when the threshold value is considerably so small that other non-skin regions get included. The threshold value at which the minimum increase in region size is observed while stepping down the threshold value will be the optimal threshold. In our program, the threshold value is decremented from 0.65 to 0.05 in steps of 0.1. If the minimum increase occurs when the threshold value was changed from 0.45 to 0.35, then the optimal threshold will be taken as 0.4. Using this technique of adaptive thresholding, many images yield good results; the skin-colored regions are effectively segmented from the non-skin colored regions.

Fig.4 Skin-segmented image

TEMPLATE MATCHING

This section shows how to do the matching between the part of the image corresponding to the skin region and the template face. The templates include 11 most frequent poses of a face could be in an image (Fig.5).

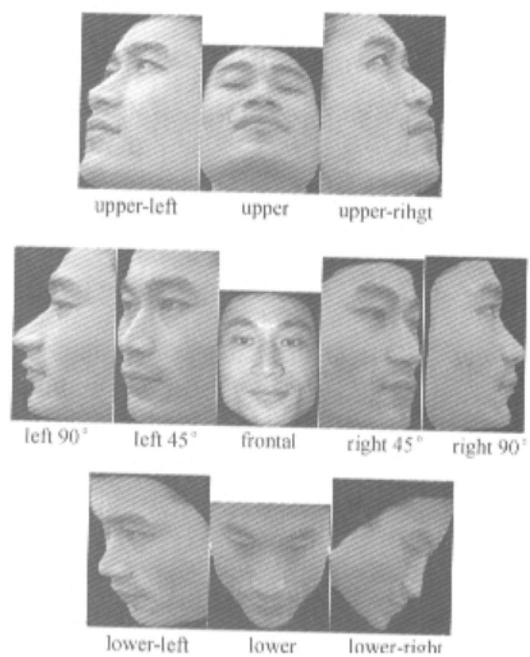

Fig.5 Human faces used in template matching

The template face has to be positioned and rotated in the same coordinate as the skin region image. This is done as follows:

For (every template)

Begin

1. Resize the template according to the height and width of the region.

2. Rotate the resized template so the template is aligned in the same direction the skin region is.

3. Compute the cross-correlation value be-

tween the template and the skin region.

4. Ifthe cross-correlation value is greater than a given threshold value, we take the region as a human face.

End

Cross-correlation value =

$$\frac{\sum_{m}\sum_{n}(A-\bar{A})\cdot(B-\bar{B})}{\sqrt{(\sum_{m}\sum_{n}(A-\bar{A})^2(\sum_{m}\sum_{n}(B-\bar{B})^2}} \quad (7)$$

$\boldsymbol{A}$ is the skin region $m \times n$ matrix and $\boldsymbol{B}$ is the template $m \times n$ matrix. $\bar{\boldsymbol{A}}$ is the mean of matrix elements of $\boldsymbol{A}$, $\bar{\boldsymbol{B}}$ is the mean of matrix elements of $\boldsymbol{B}$.

We empirically determined, from our experiments, that a good threshold value for classifying a region as a face is if the resulting cross-correlation value is greater that 0.6.

EXPERIMENTATION AND RESULT

The system was tested with more than 400 color images. The pictures were taken from persons of different ethnicities: Asian, Caucasian and African. The resulting detection rate was 83%, better than most color-based face detection systems. This is attributed to the multiple template faces. Although the poses of human faces vary greatly in images, at least one cross-correlation value between template and the face region is smaller than the threshold. And the color-based method determines the system can do its job in real-time. The average speed for face detection is 0.8 second/image (400 × 300 pixels) on a Pentium 3 (800MHz) PC. The following pictures in Fig.6 are some examples of the test images, and the rectangle in the picture is the face region marked by the system.

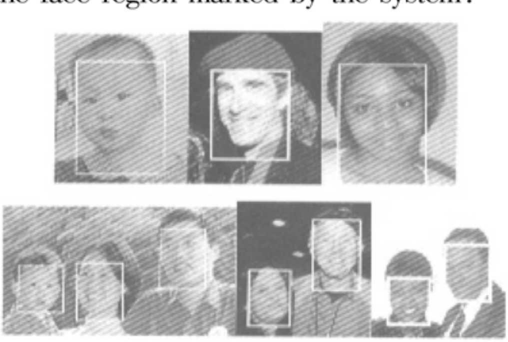

Fig.6 Examples of the result

Here is a comparison between our method and some other methods (Ming et al., 2001).

Table 1 Comparison between our method and some other methods

Method	Detection Rate
Distribution based	81.9%
Neural network	90.3%
Naïve Bayes classifier	91.2%
Kullback relative information	N/A
Support vector machine	74.2%
Mixture of factor analyzers	89.4%
Fisher linear discriminant	91.5%
SNoW with primitive features	93.6%
SNoW with multi-scale features	94.1%
Inductive learning	N/A
Our method	83%

Although the detection rate of our method seems not as good as some other algorithms, it runs in real time.

References

Craw, I., Tock, D. and Bennett, A., 1992. Finding face features. *In*: Proceedings of the Second European Conference on Computer Vision, Italy, p.92 – 96.

Kjeldsem, R. and Kender, J., 1996. Finding skin in color images. *In*: Proceedings of the Second International Conference on Automatic Face and Gesture Recognition, IEEE Computer Society Press, Vermont, USA, p.312 – 317.

Lanitis, A., Taylor, C. J. and Cootes, T. F., 1995. An automatic face identification system using flexible appearance models. *Image and Vision Computing*, **13**(5):393 – 401.

Ming, Y., David, K. and Narendra, A., 2001. Detecting faces in Images: A Survey. *IEEE Transactions on Pattern Analysis and Machine Intelligance*, **24**(1):34 – 58.

Osuna, E., Freund, R. and Girosi, F., 1997. Training support vector machines: An application to face detection. *In*: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Puerto Rico, p.130 – 136.

Rowley, H., Baluja, S. and Kanade, T., 1998. Neural network-based face detection. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **20**(1):23 – 38.

Turk, M. and Pentland, A., 1991. Eigenfaces for recognition. *Journal of Cognitive Neuroscience*, **3**(1):71 – 86.

Yang, G. and Huang, T. S., 1994. Human face detection in complex background. *Pattern Recognition*, **27**(1):53 – 63.

Yang, J. and Waibel, A., 1996. A real-time face tracker. *In*: Proceedings of Third Workshop on Applications of Computer Vision, Sarasoto, USA, p.142 – 147.