Detecting Faces in Color Images

Wen-Hsiang Lai and *Chang-Tsun Li*
Department of Computer Science, University of Warwick,
Coventry CV4 7AL, UK
jj6066@dcs.warwick.ac.uk, ctli@dcs.warwick.ac.uk

ABSTRACT

We propose in this work a method for detecting faces in color images with complex backgrounds. The approach starts with the transformation of the image pixels from the RGB color space to the chrominance space (YCbCr). Secondly, a Gaussian model is fitted on the transformed image in order to calculate the likelihood of skin for each pixel and to create a likelihood image. Thirdly, by thresholding the likelihood image, skin pixels are segmented to form a binary *skin map*, which contains the candidate face regions. Finally, a verification process is carried out to determine whether these candidate face regions are real faces or not.

Keywords: Face detection, face recognition, skin Gaussian model, biometrics, security

1. INTRODUCTION

Face recognition is one of the main biometric techniques for security applications [1] and is the first step toward the success of face recognition. To design an automatic face recognition system, the key problem is to detect the face location quickly and efficiently. In addition, face detection also plays an important role in many applications such as video conferencing, human computer interface (HCI), and video surveillance. Face detection is a complex and challenging task because of the variation of the face in scale, location, orientation, and pose [2]. The challenges associated with face detection can be ascribed to 1) Facial features such as beards, mustaches, and glasses may or may not be present. 2) Faces may appear anywhere in images in different orientations and may be partially obstructed by other objects. 3) The imaging condition may affect the detection result. In this paper, we present a novel face detection approach based on a skin Gaussian model in an attempt to alleviate the influences of the afore mentioned factors.

2. FACE DETECTION

The framework of the proposed method comprises two main components: *skin region segmentation* and *face region verification* as shown in Figure 1, which is discussed as follows.

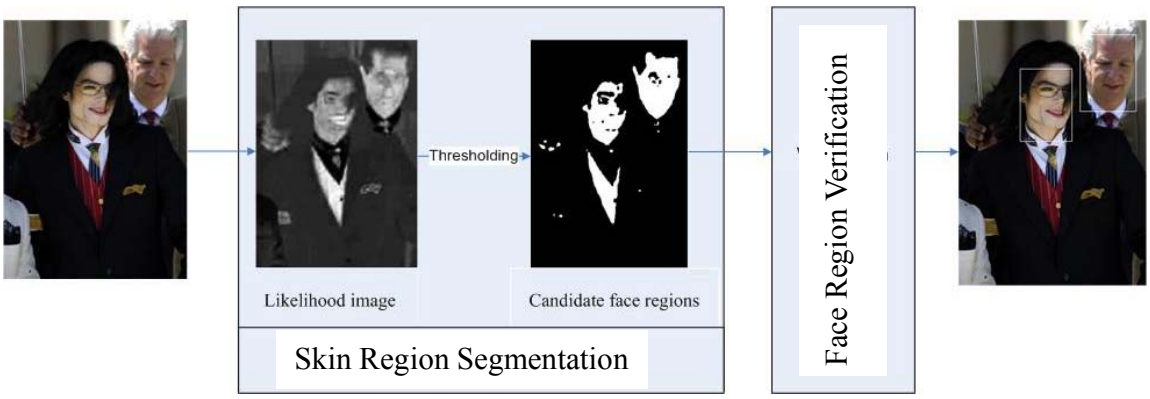


Figure 1. Framework of the approach face detection method.

2.1 Skin region segmentation

The task of this component is to create a binary *skin map* by segmenting the skin regions using a Gaussian model. We extracted 45,000 skin pixels from 90 face images, transformed their values from the *RGB* color space to the chrominance space ($Y\ C_bC_r$), and

observed that their C_b-C_r components cluster in one small area even though these images are from different ethnicities. Since Y component is not independent of hue information, like most color method, we do not involve this component in our approach. The C_b-C_r distribution is shown in Figure 2. With $x = [C_b\ C_r]$ denoting a two-component color vector, we employ the ITU.BT-601 transformation formula as follows to do the transformation.

$$x^T = \begin{bmatrix} C_b \\ C_r \end{bmatrix} = \begin{bmatrix} -0.169 \cdot r - 0.331 \cdot g + \cdot 0.500 \cdot b \\ 0.500 \cdot r - 0.418 \cdot g - 0.082 \cdot b \end{bmatrix} \tag{1}$$

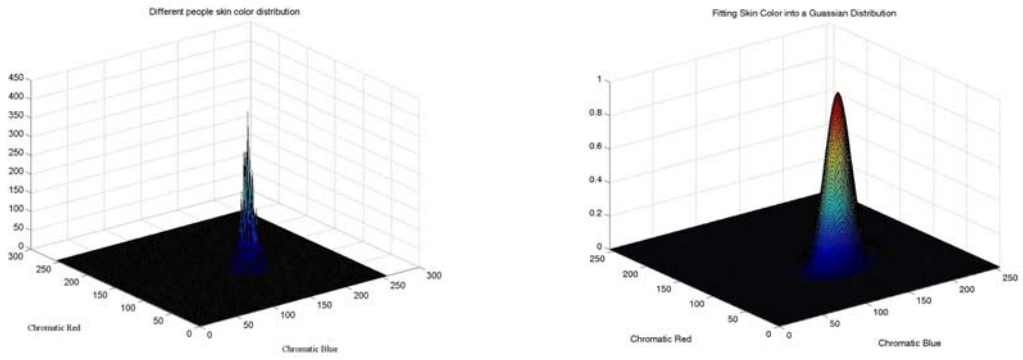
where $r = R / (R + G + B)$, $g = G / (R + G + B)$, and $b = b / (R + G + B)$. By fitting a Gaussian model (Figure 2(b)) on the transformed data (Figure 2(a)), we can denote the likelihood of the skin color x given its class ω $(\omega \in \{skin, non-skin\})$ as

$$P(x \mid \omega) = e^{-0.5(x-\mu)^T \cdot \upsilon^{-1} \cdot (x-\mu)} \tag{2}$$

where μ is the mean and $\nu = E[(x-\mu)\cdot(x-\mu)^T]$ is the variance. The values of μ and ν extracted directly from the Gaussian model are as follows.

$$\mu = [156.56. \quad 117.43] \tag{3}$$

$$\upsilon = \begin{bmatrix} 160.13 & 12.143 \\ 12.143 & 299.46 \end{bmatrix} \tag{4}$$



(a) (b)

Figure 2. (a) Skin color distributions of our data (45000 pixels) in C_b-C_r space. (b) A Gaussian model fitted on the data shown in Figure 2(a).

Having obtained the likelihood of skin for each pixel and created a likelihood image, we generate a binary *skin map* with value 1 representing *skin* if the likelihood is greater than a threshold *T* and 0 representing *non-skin* otherwise. The regions of *skin* pixels in the skin map are taken as *candidate face regions* for further verification. Since there is no theoretical backing for determining the threshold *T*, after experimenting with several values (see Table 2 and Figure 5), we found 0.5 an empirically optimal threshold.

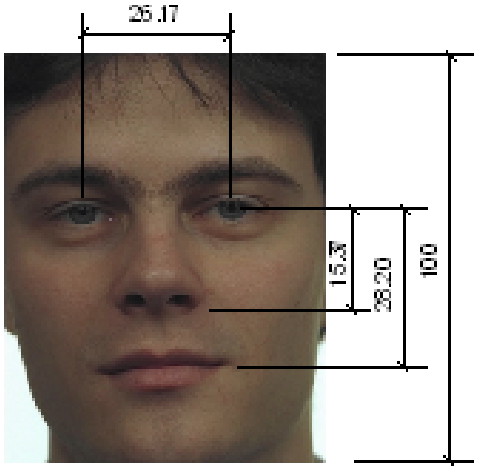


Figure 3. A position map of facial features.

2.2 Face Region Verification

Since the identified *candidate face regions* may be actually parts of the human body other than faces, further verification should be performed to locate the real face region. Three criteria as explained below are used in our verification process.

- Since a real face region contains some unique features that differentiate the face from other skin regions, we exploit this characteristic by assuming that a real face, whether it is a frontal face or not, should contain at least one eye. Therefore, a region containing no other region(s) should be ruled out as a face region. A position map of face features is shown in Figure 3. However, the faces are not always oriented vertically in the images and the degree they deviate from vertical orientation is unknown. So, before detecting the eye regions, each candidate face region is rotated incrementally by 15 degree and then verify whether there is at least one smaller region (i.e. an eye) contained in the rotated region located at one of the two eyes' positions as shown in Figure 3. It is regarded as a face if the result is positive.

- In addition, there may be regions covering only a small number of *skin* pixels, which are too small to be treated as faces. So we exclude those regions whose size is smaller than 5% of the size of the largest region.
- According to [3], the aspect ratio of an "ideal frontal face" is close to $\frac{width}{height} = \frac{2}{1+\sqrt{5}}$. However, in a complex scene, faces do not always face the camera and may be partially obstructed by other objects. Therefore, in our work, if the aspect ratio of a region is less than 0.5, we regard it as a non-face region.

3. EXPERIMENTS

Although there are several well-known face image databases publicly available but most of them are meant for face recognition applications. There is no acknowledged face image database for face detection algorithm so far. Due to this reason, we use three face image databases to validate the proposed method. We use Visual C++ 6.0 as our system development tool. The computer is equipped with a Pentium III 1.2 GHz CPU and 512 MB memory. There are four factors used to evaluate our algorithm:

1. Detection Rate (D.T.R %): The number of correctly detected faces divided by the total number of faces.
2. False Reject Rate (F.R.R %): The number of falsely-rejected faces divided by the total number of faces. A falsely-rejected face is a face region that is not detected.
3. False Accept Rate (F.A.R %): The number of falsely-accepted faces divided by the total number of faces. A falsely-accepted face is an image region that is mistaken as a face.
4. Frames per sec (fps): The total detection time divided by the total number of images (frames).

The first experiment is conducted on the FERET face image database [4] containing 11128 images of 994 different faces each at three different resolutions. The results are shown in Table 1. It has no challenge in achieving high detection rate on the FERET face image database because the faces appear in uncluttered backgrounds. However, there are still few faces undetected due to the dark illumination. Furthermore, to give a clear view on the performance of the proposed algorithm when applied to the images with different head poses, the same results shown in Table 1 are displayed in Figure 4(a), (b), and (c) according to the head poses and resolutions. In Figure 4, the two-letter codes *fa, fb, pl, pr, hl, hr, ql, qr, ra, rb, rc, rd,* and *re* stands for the head poses of 'frontal face with regular expression (*fa*)', 'frontal face with alternative expression (*fb*)', 'profile left (*pl*)', 'profile right (*pr*)', 'half left (*hl*), 'half right (*hr*)', 'quarter left (*ql*)', 'quarter right (*qr*)', 'random pose *a* (*ra*)', 'random pose *b* (*rb*)', 'random pose *c* (*rc*)', 'random pose *d* (*rd*)', 'random pose *e* (*re*)', respectively. According to Figure 4, the detection rates for two poses (*pr* and *pl*) are relatively lower due to the difficulty in extracting the feature of eyes. Overall, our method has reached a good detection rate with some degree of sensitivity to dark lighting condition. The detection speed is close to 12 fps at 128 × 192 resolution, making the proposed

algorithm suitable for real-time application. However the false accept rate is relatively high. The reason is that a face is a 3D surface, when projected on a 2D image, illumination may create shadowy regions that could have ill impact on the skin segmentation result.

Table 2 presents the experimental results based on the Oulu face video database [5], which contains 403 images with 22 humans. The images in this database have only one frontal face under various illumination in complex background, which make them suitable for evaluating the performance of the algorithm when different values of threshold T is used for creating the binary skin map. Figure 5(a) illustrates the same results as in Table 2 and suggests that the value of 0.5 for the threshold T is a reasonable choice.

Finally, we test the algorithm on a database containing 40 images with 128 faces taken from the Internet. Faces appear in these images in different size, location and orientation with complex background, making the face detection a challenging task. The results are shown in Table 3 and some detection examples are shown in Figure 6. Our algorithm can detect multiple faces in different sizes with a wide range of facial variations in the images. Furthermore, the algorithm can detect non-frontal faces in different orientation (e.g. slanted face) and the faces with various facial expressions (e.g. closed eyes or opened mouth) and facial details (e.g. wearing glasses, having a beard).

4. CONCLUSIONS

It is confirmed in our experiments that the distribution of human skin is relatively clustered in a small area in the chrominance space and the influence of illumination can be alleviated to some extent by transforming the image from the RGB color space to the chrominance space. By fitting a Gaussian model on the transformed data, our algorithm is capable of detecting faces in different sizes, orientations, and poses. However, there are still some possible improvements for further works. For example, applying a lighting compensation model to correct the color bias can be an effective way of increasing the algorithm's robustness against dark lighting condition. In addition, combining multiple facial features, such as mouth and nose, can also be desirable in detecting faces in poor illumination and dark lighting condition if the computational complexity can be balanced.

ACKNOWLEDGEMENTS

Portions of the research in this paper use the Color FERET database of facial images collected under the FERET program. Portions of the research in this paper use the Oulu face video database of facial images.

REFERENCES

[1] C.-T. Li, "The Role of Biometrics in Virtual Communities and Digital Governments," in *Encyclopedia of Virtual Communities and Technologies,* ed. by S. Dasgupta, Idea Group Publishing, July 2005.

[2] M. H. Yang, D. J. Kriegman, and N. Ahuja, “Detecting faces in images: A Survey,” *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34-58, 2002.
[3] L. G. Frakas and I. R. Munro, *Anthropometric facial proportions in medicine*. Charles C. Thomas, Publisher Ltd, 1987.
[4] P. J. Phillips, H. Moon, P. J. Rauss, S. Rizvi, “The FERET evaluation methodology for face recognition algorithms,” *IEEE Transactions on Pattern Analysis and Machine Intelligence,* vol. 22, no. 20, pp. 1090-1104, 2000.
[5] B. Martinkauppi, N. Soriano, S. Huovinen, M. Laaksonen, “Face video database,” in *Proceedings of the First European Conference on Color in Graphics, Imaging and Vision*, pp380-383, 2002.

Table 1 Detection results based on the FERET face image database at different resolutions.

Image resolution	No. of images	Duration(sec)	No. of D.T	No. of F.R	No. of F.A
		Frames/sec	D.T.R %	F.R.R %	F.A.R %
512 x 768	11128	87612.61	10202	926	4048
		0.13	91.68%	8.32%	36.38%
256 x 384	11128	5848.22	10035	1093	2847
		1.90	90.18%	9.82%	25.58%
128 x 192	11128	942.95	9722	1406	918
		11.80	87.37%	12.63%	8.25%

Table 2 Test results on the Oulu face video database with different thresholds.

Threshold *T*	Total images	No. of D.T	No.of F.R	No. of F.A	D.T.R%	F.R.R %	F.A.R %	Duration (sec)	fps
0.10	403	20	383	8	4.96%	95.04%	1.99%	300.62	1.34
0.20	403	131	272	80	32.51%	67.49%	19.85%	575.2	0.70
0.30	403	245	158	109	60.79%	39.21%	27.05%	347.75	1.16
0.40	403	366	37	96	90.82%	9.18%	23.82%	224.64	1.79
0.45	403	377	26	96	93.55%	6.45%	23.82%	206.96	1.95
0.50	403	370	33	121	91.81%	8.19%	30.02%	174.57	2.31
0.55	403	359	44	108	89.08%	10.92%	26.80%	162.85	2.47
0.60	403	342	61	101	84.86%	15.14%	25.06%	131.57	3.06
0.70	403	312	91	89	77.42%	22.58%	22.08%	103.58	3.89
0.80	403	239	164	67	59.31%	40.69%	16.63%	77.39	5.21
0.90	403	150	253	45	37.22%	62.78%	11.17%	61.1	6.60

Table 3 Detection results based on randomly images collected.

D.T.R %	F.R.R %	F.A.R %
86.73 %	13.27 %	21.75%

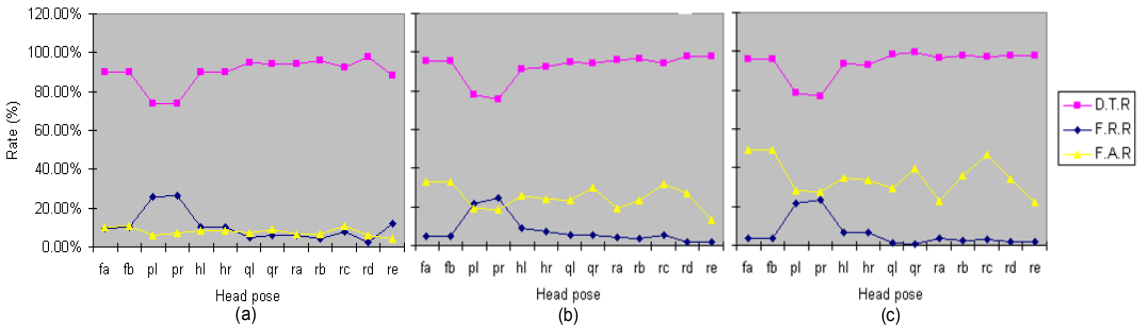


Figure 4 Detection results in different poses in FERET face database at three resolutions: (a) 128 × 192. (b) 256 × 384. (c) 512 × 768.

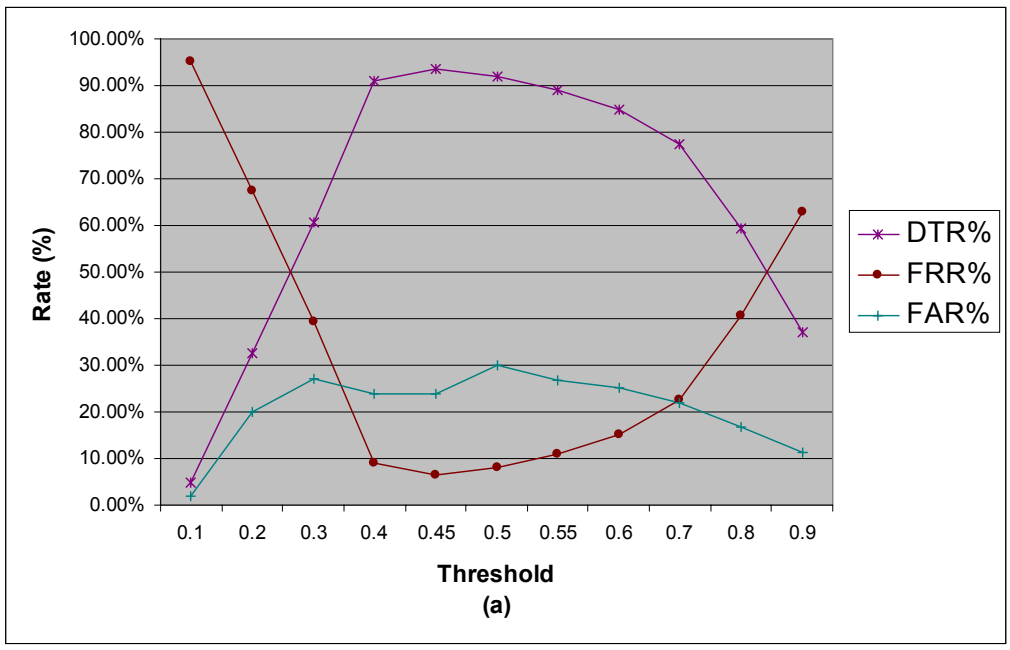


Figure 5. Detection results on the Oulu face video database with different threshold *T* mentioned in Section 2.1.

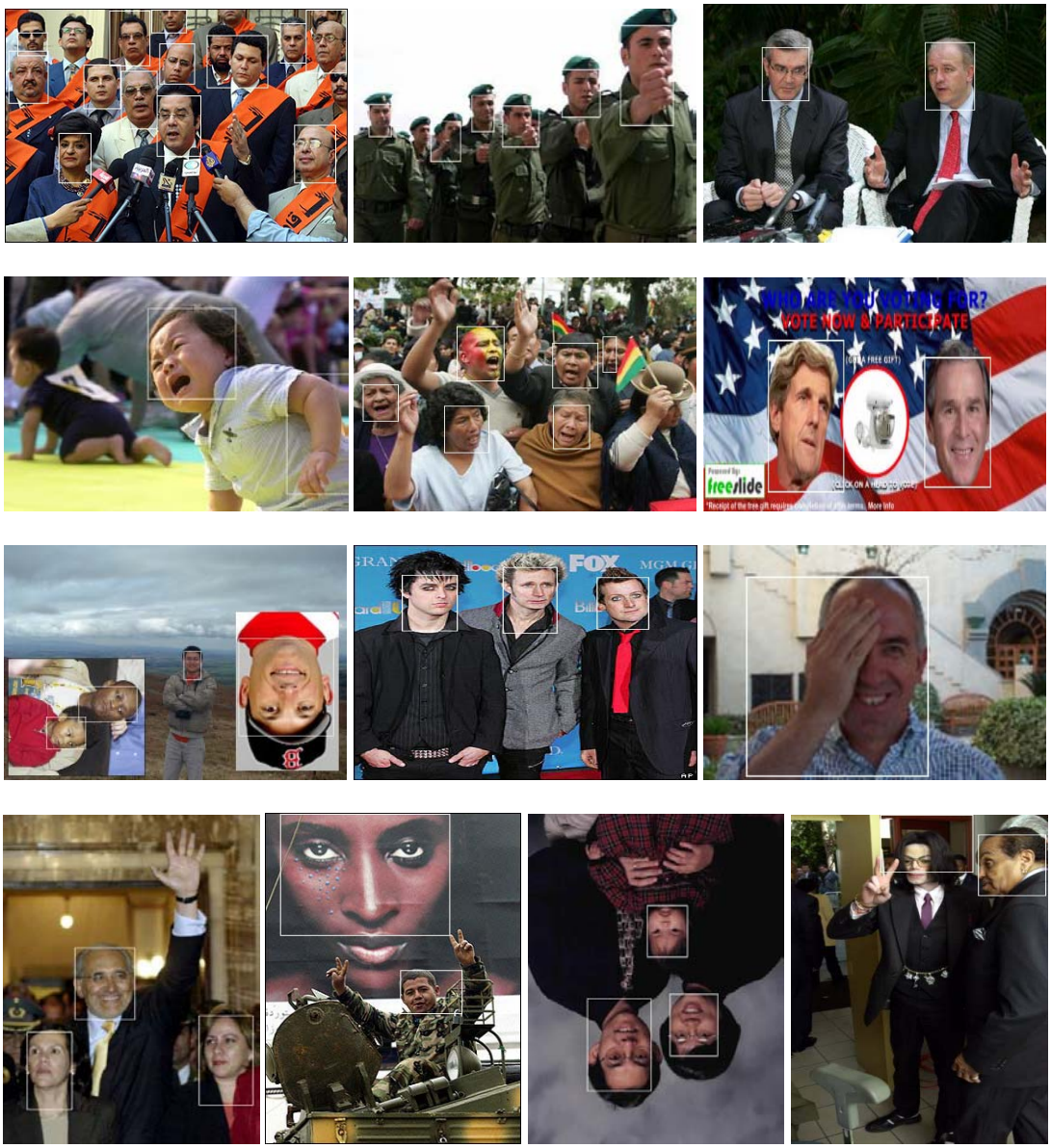


Figure 6. Face detection results on some photos.