FEATURE-BASED DETECTION OF FACES IN COLOR IMAGES

Chi-Jaung Huang, Ming-Che Ho and Cheng-Chin Chiang

National Dong Hwa University
Department of Computer Science and Information Engineering
Shoufeng, Hualien, Taiwan, R.O.C., 974
ccchiang@mail.ndhu.edu.tw

ABSTRACT

This paper proposes a novel method to detect faces in color images with a wide range of variations in head poses, face tilting, facial expressions and body interferences. The proposed algorithm can locates not only the face regions, but also some important facial components such as mouth corners and eyes. The experimental results show that the proposed method exhibits satisfied performance in both well-controlled and real-world environments. The performance is also compared with that of a face detector developed by CMU to demonstrate the major strength of the proposed algorithm.

1. INTRODUCTION

Face detection plays a crucial role for many applications of face recognition, expression analyses and intelligent man-machine interfaces. Many face detection algorithms had been proposed in the past ten years [1, 2, 3, 4]. Generally, these algorithms can be categorized into image-based approach and feature-based approach. One of the typical methods in image-based approach is the neural-based face detector [3]. In the neural face detector, the faces are searched by using a sliding sub-window to examine every subimage on the whole input image. To detect faces of different sizes, several sliding sub-windows of different sizes must be used. Although the neural network approach is straightforward, it usually takes much more processing time and thus may be not suitable for real-time applications. Another weak point of this approach is that the detection performance depends significantly on the training samples which usually require much effort to collect for neural network training.

In contrast to the image-based approach, the feature-based approach focused on the analysis of facial components, including lips, mouth, eyes, etc. The feature-based approach usually achieves faster detection speed since it requires not much exhaustive searching over the whole image. Many useful cues and constraints can be used to reduce the searching space. These cues and constraints may include skin color of faces [5], spatial and geometrical relationship [6] between facial components. However, the major weakness of the feature-based approach is that the features of faces can be severely corrupted due to illumination, noise, and occlusion, etc. The accurate detection of facial components is somewhat difficult and unstable. Therefore, the goal in our study is to develop a robust feature-based face detection algorithm capable of dealing with a wide range of variations in head poses, face tilting, facial expressions and body interferences.

This work is supported by the project (grant no. NSC-90-2218-E-259-001-) funded by National Science Council, Taiwan, R. O. C.

2. POLYNOMIAL DISCRIMINANT MODELS FOR SKIN AND LIPS EXTRACTION

2.1. Light Regulation

The appearance of skin color may vary with the illuminating conditions. Therefore, a lighting regulation process on the image is necessary. For a given image, we define an intensity value W_{ref} called "reference white" for the image as $0.95b$, where b is the brightest luminance value in the gamma-corrected gray-level image. The pixel values of the given image I in R, G, and B channels are regulated by the following equation:

$$(x, y) = 255 * \phi\left(\frac{p_c(x,y)}{W_{ref}}\right)$$

where $p_c(x, y)$ denotes the intensity value in channel c ($c \in \{R, G, B\}$) of the pixel at position (x, y); the function $\phi(x)$ is defined as $\phi(x) = x$ for $0 \leq x \leq 1$, and $\phi(x) = 1$ for $x > 1$.

2.2. Polynomial Discriminant Functions for Skin and Lips

Skin color is an important cue for speeding up the detection of face. Many researchers proposed statistical approaches to model the distributions of skin colors over some color spaces [5, 7]. However, the calculation costs of the related probabilities in these statistical models are not low. To increase the computation efficiency, Soriano et al. [7] proposed a new formulation for the skin locus distributed in the 2D r-g color space. The transformation between RGB color space and r-g space is defined by:

$$r = \frac{p_R(x,y)}{p_R(x,y)+p_G(x,y)+p_B(x,y)} \text{ and } g = \frac{p_G(x,y)}{p_R(x,y)+p_G(x,y)+p_B(x,y)}.$$

Fig. 1 shows the skin locus on the r-g plane. Apparently, the upper and lower boundaries of this locus form two parabolic boundaries which can be modeled by two quadratic polynomials $f_u(r) = a_u r^2 + b_u r + c_u$ and $f_l(r) = a_l r^2 + b_l r + c_l$. By sampling a number of points on the two boundaries, the coefficients can be easily estimated using least-mean-square (LMS) error minimization method. One set of the derived coefficients $(a_u, b_u, c_u, a_l, b_l, c_l)$ is (-1.376, 1.074, 0.145, -0.776, 0.560, 0.176). To extract the skin-color pixels from a given image, the designed discriminating rule is:

$$S(r, g) = \begin{cases} 1, & \text{if } f_l(r) \leq g \leq f_u(r) \text{ and } W(r, g) > 0.0004, \\ 0, & \text{otherwise,} \end{cases}$$

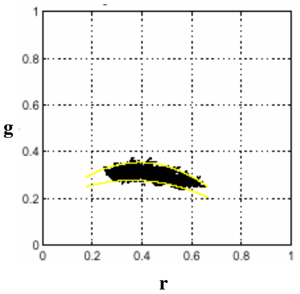

Fig. 1. The skin locus plotted by Soriano [7].

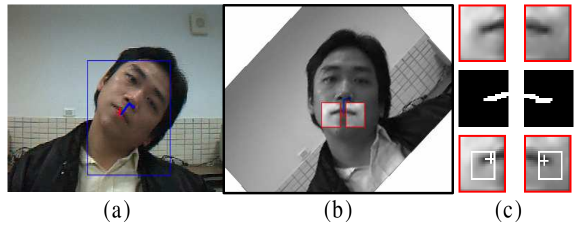

Fig. 2. Lip and mouth corners positioning: (a) a tilted face with the estimated tilt angle and the lip centroid; (b) the aligned face and the two searching window for mouth corners; (c) located mouth corners.

where $W(r,g) = (r-0.33)^2 + (g-0.33)^2$ and $0.2 \le r \le 0.6$. That is, a pixel is recognized as a skin pixel only when its transformed r-g value satisfies $S(r,g) = 1$. Notice that the term $W(r,g) < 0.0004$ is exploited to exclude the gray and near-gray pixels from the desired skin pixels.

Based on the novel modeling of skin colors given above, we also extend it to extract lip pixels because the pair of lips is one of the salient facial features for identifying and verifying faces. According to the general visual perception, lip colors can be regarded as some subsets of darker skin colors. We find that the colors of lip pixels distribute around the lower boundary of the crescent-like skin locus shown in Fig. 1. Hence, all we need is to define a new quadratic polynomial for the upper boundary of the lip pixel distribution. The new quadratic polynomial for the upper boundary is estimated to be

$$f_u^{'} = -0.776r^2 + 0.5601r + 0.21$$

For the lower boundary, the same polynomial for the lower boundary of the skin color distribution can be used. The above two polynomial discriminating models for skin and lips extraction are characterized by their simplicity in computations. In addition, these two models can efficiently extract the two kinds of desired pixels in only one scan of the input image. To identify the regions of interests for further analyses and processing, the extracted skin pixels form the desired regions, which are called facial masks, for this purpose. The extracted lips are helpful to locate the eyes for the proposed algorithm.

3. EXTRACTION OF FACIAL COMPONENTS AND VALID FACE REGION DETERMINATION

With the extracted facial masks, the next step of the proposed method is to identify the key facial components including mouth corners and eyes to determine the valid face regions. Before locating the facial components, we propose a method to align the facial masks to improve the robustness in detecting tilted faces. In what follows, we present the methods in details.

3.1. Alignment of Facial Masks

In general cases, faces in the input image may be not always in upright direction. As the subsequent analyses are based on the normal upright faces, the proposed algorithm must align the tilted faces to their upright directions before the analyses. For each facial mask generated by preceding steps, we collect the non-skin pixels inside the facial mask. These non-skin pixels usually contain the salient facial components including eyes, eyebrows, nostrils, and mouth cavity. The orientation of the distribution of these non-skin pixels can be used to estimate the tilt angle for the face. An alternative option for estimating the tilt angle is to find the orientation of the distribution of skin pixels in the facial masks. However, this estimate is more sensitive to the body interference among faces and other skin-color bodies. Therefore, we choose the non-skin pixels as our cue for tilt angle estimation.

The orientation of the non-skin pixels can be estimated by the second-order moments analysis. For a distribution of non-skin pixels, $R = \{p(x_i, y_i) | 1 \le i \le N\}$, let $(\bar{x}, \bar{y})$ be the sample mean of R, i.e., $(\bar{x}, \bar{y}) = \left(\sum_{i=1}^{N} x_i/N, \sum_{i=1}^{N} y_i/N\right)$. The orientation θ of R can be estimated according to the following equation:

$$\theta = \frac{1}{2}\tan^{-1}\left(\frac{2u_{xy}}{u_{xx} - u_{yy}}\right)$$

where $u_{xx} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})^2$, $u_{xy} = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x})(y_i - \overline{y})$, and $u_{yy} = \frac{1}{N}\sum_{i=1}^{N}(y_i - \overline{y})^2$. With the determined tilt angle θ, each facial mask can be aligned accordingly to get an upright or near-upright facial mask.

3.2. Positioning of Lips and Mouth Corners

Previously, we have presented the quadratic discriminating model for extracting lip pixels. Therefore, the centroid of the lip pixels on an aligned facial mask can be easily determined. According to the located lip centroid, two searching windows ($\gamma \times \gamma$ pixels) at both sides of the lip centroid are defined for detecting the possible mouth corners. Within each searching window, we apply the Sobel edge detector and find the largest connected component of edge pixels. The leftmost point for the largest connected component in the left-side window and the rightmost point for the largest connected component in the right-side window are then considered as two seed points for finding the left and right mouth corners, respectively. Starting from each seed point, we identify the darkest point within a rectangle area ($\lambda \times \lambda$ pixels) around the seed point. This found darkest point is the desired mouth corner. Fig. 2 shows the example for mouth corner positioning.

3.3. Locating of Eyes

In normal cases, the two eyes are right above the two mouth corners on an upright face. Therefore, the eyes can be easily

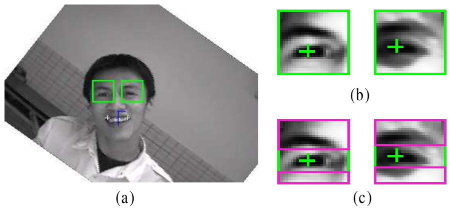

Fig. 3. Locating the eyes:(a) two searching windows for eyes; (b) darkest points within the searching window; (c) eyebrow and eyes differentiation.

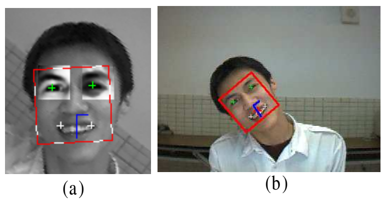

Fig. 4. Determination of valid face region:(a) determined valid region on the aligned facial mask; (b) determined valid face region on the original image.

found as long as the two mouth corners can be positioned on aligned facial masks. For this reason, given a facial mask of $W \times H$ pixels and the position of a found mouth corner, a rectangle searching area of $W/3 \times H/3$ pixels right above the mouth corner is defined for locating eyes. The distance between the center of the searching area and the mouth corner is $1.5L$, where L is the distance between the two mouth corners.

Within each searching area, the color image is converted to gray-level one and the contrast is enhanced by histogram equalization. Afterwards, the darkest point is searched on the enhanced gray-level image. Normally, the identified darkest pixel belongs to either eyebrow or eyes. These two ambiguous cases can be easily differentiated by comparing the average gray-level intensities of the regions above and below the darkest pixel in the searching area. Let the average intensities for the regions above and below the darkest pixel be G_u and G_l, respectively. If $G_u \geq G_l$, then it means that the current darkest point is very likely to belong to eyebrow. In this case, we switch to search the darkest point in the lower region as our desired eye point. On the other hand, if $G_u < G_l$, then the darkest point is considered as the desired eye point. Fig. 3 illustrates an example of the eye locating. As soon as the exact mouth corners and eyes are detected, the valid face region can be determined by drawing a rectangle according to the geometry depicted in Fig. 4. Note that the four corner points must be rotated according to the estimated tilt angle.

4. EXPERIMENTAL RESULTS

4.1. Testing Data Sets

Two sets of testing images are collected for evaluating the proposed method. The first testing set contains 450 color images (15 persons x 30 samples) of near-frontal views. These images are acquired under several well-controlled lighting conditions and simple backgrounds. This testing set is used mainly for examining the capability of the proposed algorithm in handling variations in head posed, face tilting, facial expressions, and body interference.

The second testing set contains 185 color images. A lot of challenging cases appear in this set because we collect the images from both unconstrained natural environments and Internet websites. This image set is mainly for evaluating the performance of the proposed algorithm for real-world applications.

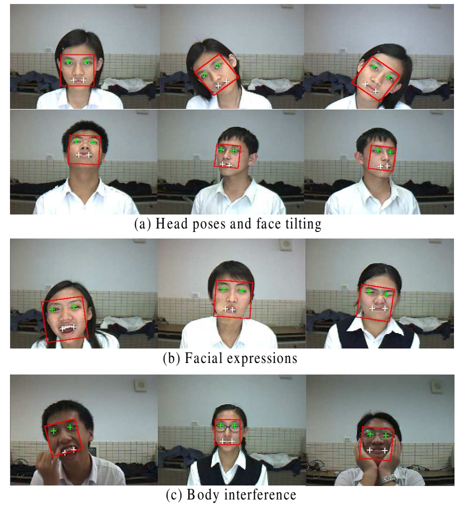

Fig. 5. Some face detection results for variations in head poses, face tilting, facial expressions and body interference.

4.2. Face Tilting and Head Poses

For face tilting, the proposed method can detect any angle of tilting. This is a critical improvement over many other existing face detection methods. For head poses, the allowable range of angles for lifting/bending/rotating head is between +22.5^0 and -22.5^0. Fig. 5(a) shows some testing results.

4.3. Expression Variations and Body Interference

The variations of facial expressions and body interference may significantly affect the face detection results. In our method, as long as the facial expressions do not drastically cause the indistinguishableness of eyes and mouth corners, the facial expressions have very little influence to our algorithm. Fig. 5(b) demonstrates the effectiveness of our algorithm in handling different facial expressions.

For the problem of body interference, we get the same conclusion. That is, as long as the major facial components such as eyes and mouth corners are not occluded, our algorithm detects the faces very well. Fig. 5(c) illustrates some cases of the detection results.

4.4. Performance Evaluation on Detection Accuracy

Table 1 shows the overall system performance for our evaluations on the two testing sets. All images in the first set have the same size of 320×240 pixels, while the second image set contains images of variational sizes. The algorithm

Table 1. Detection Accuracy for Image Set I and II.

	No. of Images	No. of Faces	No. of Misses	No. of False Positives	Detection Rate
Set I	450	450	33	12	92.6%
Set II	185	396	49	19	87.6%

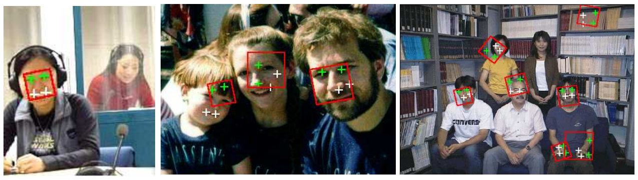

Fig. 6. Some incorrect detection results.

achieves about 0.6 second per face on a PC with Pentium 1.4 GHz CPU and 256 MB memory. We have found that most of the incorrect detections result from the extreme situations of light illuminations and color biasing. Some incorrect detections are shown in Fig. 6.

4.5. Comparing with CMU Face Detector

We compare the proposed detector with the detector developed by CMU [8], which is designed based on a statistical decision model. We submit the second image set to their WWW demonstration system at http://vasc.ri.cmu.edu/cgi-bin/demos/findface.cgi. Table 2 shows the comparative results. Although the CMU face detector is slightly better than our detector, it generates more false positives. The major strength of CMU detector is its insensitivity to the variations in illuminations and colors because it processes gray-level images. However, when dealing with tilted faces, our detector is much better than the CMU detector. According to our experiments, the CMU detector fails to detect almost all the faces tilted more than 30 degrees. Some of the compared detection results are sown in Fig. 7.

5. CONCLUDING REMARKS

In this paper, we present a feature-based face detection algorithm to cope with a wide range of variations in head poses, face tilting, facial expressions and body interferences. The key ingredients in the proposed algorithm include an efficient polynomial model for skin and lips extraction, the method to estimate the tilt angles for faces and the method to detect mouth corners and eyes. We have conducted many experiments to demonstrate the effectiveness of the proposed algorithm.

Table 2. Comparison between the CMU detector and ours.

	No. of Misses	No. of False Positives	Detection Rate
CMU Detector	46	24	88.4%
Our Detector	49	19	87.6%

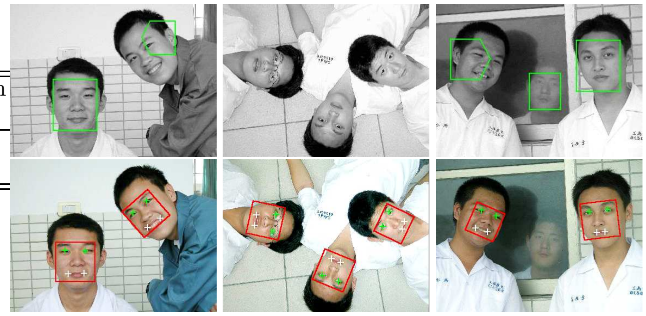

Fig. 7. Some compared detection results for CMU face detector (top row) and the proposed algorithm (bottom row).

In the future, the research directions of this study will be focused on

- the development of adaptive color model for skin extraction to improve the robustness in handling different lighting conditions, and
- the method of detecting faces with partially of fully occluded facial components.

6. REFERENCES

[1] M. H. Yang, D. J. Kriegman, and Narendra Ahuja, "Detecting faces in images: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 1, pp. 34–58, 2002.

[2] E. Hjelmas and B. K. Low, "Face detection: A survey," *Computer Vision and Image Understanding*, vol. 83, no. 3, pp. 236–274, 2001.

[3] H. A. Rowley, S. Baluja, and T. Kanade, "Rotation invariant neural network-based face detection," in *Proceeding of the IEEE International Conference on Computer Vision and Pattern Recognitoin*, 1998, pp. 38–44.

[4] K. K. Sung and T. Poggio, "Example-based learning for view-based human face detection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 1, pp. 39–51, 1998.

[5] C. Garcia and G. Tziritas, "Face detection using quantized skin color regions merging and wavelet packet analysis," *IEEE Transactions on Multimedia*, vol. 1, no. 3, pp. 264–277, 1999.

[6] C.H. Lin and K.C. Fan, "Human face detection using triangle relationship," in *International Conference on Pattern Recognition*, 2000, vol. 2, pp. 941–945.

[7] M. Soriano, S. Huovinen, B. Martinkauppi, and M. Laaksonen, "Using the skin locus to cope with changing illumination conditions in color-based face tracking," in *IEEE Nordic Signal Processing Symposium*, Kolmarden, Sweden, 2000, pp. 383–386.

[8] H. Schneiderman and T. Kanade, "A statistical method for 3d object detection applied to faces and cars," in *IEEE Conference on Computer Vision and Pattern Recognition*, 2000, vol. 1, pp. 746–751.