TOWARDS FACIAL FEATURE EXTRACTION AND VERIFICATION FOR OMNI-FACE DETECTION IN VIDEO/IMAGES

Xingquan Zhu[1], Jianping Fan[2], Ahmed K. Elmagarmid[1]

[1]*Dept. of Computer Science, Purdue University, West Lafayette, IN 47907, USA*
[2]*Dept. of Computer Science, University of North Carolina at Charlotte, NC 28223, USA*
Email:{zhuxq,ake}cs.purdue.edu; jfan@uncc.edu

ABSTRACT

Face detection is important in video/image content analysis and organization since the most important object in those medias is often human being. We propose a facial feature based omni-face detection algorithm in this paper. While utilizing the skin color model for face cue detection, a pairwise skin region refinement strategy is applied to eliminate the errors incurred by skin model. Then, a region based adaptive threshold selection scheme is employed for facial feature segmentation. After the facial feature filtering, an orientation, pose and scale invariant face verification strategy is utilized to verify the detected face candidate regions. Experimental results demonstrate successful detection over a wide variety of facial variation in background, scale, view and orientation from different types of video collections.

1. INTRODUCTION

The occurrences of different faces provide rich information for browsing, navigating and retrieval of huge amount of video/image data. Various approaches on face detection in images or video sequences have been discussed in [1]. In general, the face detection methods could be assigned into one of the two categories [2]: (1) feature-based method; and (2) classification–based method. The feature-based methods search for different facial features and use their spatial relationship to locate faces; the classification-based methods detect faces by classifying all possible sub-images of a given image as face or non-face sub-images. Due to the diversity of the faces, the classification-based approaches require a large number of training examples [8,9]. Feature-base strategies often meet the problem of facial feature extraction and verification [3,5,7,10].

The use of color information can simplify the task of face localization in complex environments [3], however, while using the skin color modal for face detection, a contradiction is usually implied: For sake of the flexibility and robustness, the skin model should address various kinds of races and illuminations; However, given any specific image or video sequence, the skin tone contained would occupy only a small portion of the skin model, thus, the rest part of the skin model would incur some pixels to be falsely segmented as skin pixels. To address this problem, a pairwse skin region refinement strategy is introduced in this paper; Moreover, a region based adaptive threshold selection scheme for facial feature segmentation and an orientation, pose and scale invariant face verification strategy are proposed to construct an efficient omni-face detection scheme.

2. SYSTEM DESCRIPTION

An overview of our face detection algorithm is illustrated in Fig. 1, which contains four major steps: (1) Skin color segmentation; (2) Pairwise skin region refinement; (3) Facial feature extraction; and (4) Face verification. Initially, a general skin color model is used to partition pixels into non-skin pixels *vs* skin pixels categories. Then, the pairwise skin region refinement scheme is adopted to classify all detected skin pixels into two layers (L_1 and L_2); hereby, three maps are gotten from the original image. Afterward, the facial feature extraction is applied on each face candidate region of each map to detect potential facial features. The region that passes the face verification successfully is reserved as the face area. After the algorithm is executed on all three maps, those successfully detected face regions will be jointed together as the result.

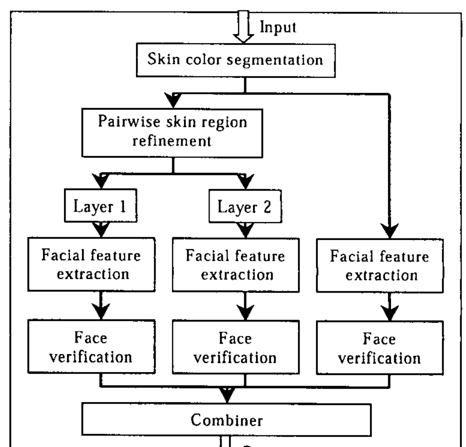

Figure 1. System architecture

2.1 Skin color segmentation

The statistical skin-color model is generated by means of supervised training skin-color regions. And the skin-color illumination brightness dependence is reduced trough intensity normalization in Eq. (1).

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

The colors (r,g) are known as chromatic colors. According to [4], the skin-color distribution in chromatic space could be approximated by the Gaussian model $N(m,\Sigma^2)$, where $m=(\bar{r},\bar{g})$ and

$$\bar{r}=\frac{1}{N}\sum_{i=1}^{N} r_i;\quad \bar{g}=\frac{1}{N}\sum_{i=1}^{N} g_i;\quad \Sigma=\begin{bmatrix}\sigma_{rr} & \sigma_{rg}\\ \sigma_{gr} & \sigma_{gg}\end{bmatrix}$$

With this Gaussian model, we can obtain the likelihood of skin for any pixel x of an image with Eq. (2).

$$L=\exp[-0.5(x-m)^T\Sigma^{-1}(x-m)],\quad where: x=(r,g)^T \tag{2}$$

With an appropriate threshold, the image can then be further transformed to a map showing skin and non-skin regions.

2.2 Pairwise skin region refinement

As shown in Fig.2(*b*), the segmented skin regions often contain some non-skin pixels, due to the contradiction between the flexibility of the skin model and the specialties of certain images. However, given any specific image or video frame, although those falsely segmented skin pixels and skin pixels do belong to the same skin model, they usually have some variances with their locations or colors. Hence, a pairwise skin region refinement scheme is adopted to separate them from each other.

Our pairwise skin region refinement is executed by a binary color space separation. We firstly calculate the color histogram of all detected skin-color pixels in chromatic space, the color which occupied the maximal number of pixels are selected out, we denote it as (r_{max}, g_{max}). Accordingly, a color region K is determined with Eq.(3). After that, the region K is used to separate the detected color pixels into two layers (L_1 and L_2), with any segmented pixel (r,g), if $(r,g)\subset K$ then $(r,g)\subset L_1$, otherwise, $(r,g)\subset L_2$, as shown in Fig.2(*c*) and Fig2(*d*). Then all these three maps (Fig.2(*b*), Fig.2(*c*) and Fig.2(*d*)) will be processed with strategies below to verify whether there is any face contained in.

$$K=\{r_{\max}-T_a\le r\le r_{\max}+T_a\,;\, g_{\max}-T_a\le g\le g_{\max}+T_a\}_{r,g} \tag{3}$$

T_a indicates the distinguishable color difference range in chromatic space with human perception, we set T_a=0.02 in our system.

Figure 2. Pairwise skin region refinement. ((*a*),(*b*),(*c*),(*d*) represent the original image, skin color segmentation result, layer 1 and layer 2 respectively).

2.3 Facial feature extraction

Given any image map (Fig.2(*b*), Fig.2(*c*) or Fig.2(*d*)), some general shape analyzing algorithms are firstly employed to detect the skin regions which may contain faces. Then, a threshold is determined to extract facial features (eyebrows, eyes, noise, mouth) contained in each face candidate region. Some other strategies have addressed the facial feature segmentation problem by constructing facial feature map [3], using template matching [5], morphological operator [7] or edge analyzing [10], etc. However, most of them use the predefined threshold for facial feature detection. As we know, different regions in the same image usually have large variances in gray and color distribution. Hence, any predefined threshold would barely work well in general situation. In this section, a region based adaptive threshold selection strategy for facial feature region segmentation is presented. And some filters are also developed to eliminate falsely detected facial features.

2.3.1 Facial feature segmentation with adaptive threshold selection

The intuition of our adaptive threshold selection strategy is derived from the fact that the facial features usually have low gray intensity on comparing with other facial areas: (1) The facial feature often belongs to the edge areas of the image; (2) Our experimental result shows that if there is a face contained in the candidate region, the facial features are always flooded among 60% of low gray intensity pixels of the region;

Hence, given any image I, we firstly use the Prewitt edge detector to detect all edge pixels (E_I) in I, as shown in Fig.3(*b*). For any face candidate region A (Fig.3(*c*)), we assume N_A denote the number of pixels in it, and $P_A^i=(A_Y^i, A_r^i, A_g^i)$, i=0,...,N_A denote the gray (A_Y^i) and color (A_r^i, A_g^i) information of pixel i in A. We then calculate the gray histogram of region A, and a threshold T_A is determined such that about 60% of pixels have intensity below it. For any pixel P_A^i in region A, if its gray intensity (A_Y^i) is lower than T_A, moreover, it belongs to the edge pixel (E_I) of image I, it will be taken as the candidate facial feature pixel of the region A. Assume the operator $[P_A^i|\{A_Y^i\le T_A, P_A^i\subset E_I, i=1,..,N_A\}]$ denotes the number of candidate facial feature pixels in region A, the adaptive threshold ($\hat{T}_A$) for region A is determined with Eq.(4).

$$\hat{T}_A=\frac{\sum_{\{A_Y^i\le T_A,\, P_A^i\subset E_I,\, i=1,\dots,N_A\}} A_Y^i}{[P_A^i|\{A_Y^i\le T_A,\, P_A^i\subset E_I,\, i=1,\dots,N_A\}]} \tag{4}$$

As we know, the edge pixels may cover both the high and low intensity area, however, the facial features only occupy the low intensity area. Hence, we use T_A to eliminate those high intensity pixels at first. Then, the edge pixels in region A, with their gray intensity less than T_A are taken as the candidate facial feature pixels. Their average intensity value is used as the threshold ($\hat{T}_A$), any pixel in A with its gray intensity less than this threshold will be segmented as the facial feature pixel.

While image I has clutter background or the illumination of the face area has large variances, other methods [3,5,7,10] may partially fail to locate the facial feature, since the edge information might be too complicated to distinguish the facial feature and the variance of the color may also induce lots of errors. However, our method can address these problems by adjusting the segmentation threshold adaptively.

2.3.2 Facial feature region filtering

Fig.3(*d*) presents the facial feature segmentation result, it shows that most facial feature regions (eyes, mouth, nose, eyebrow) are successfully detected out, however, some regions (such as the hair, and shadow of the chin, etc.) are also falsely segmented as the facial features. In this section, some facial feature filters are developed to eliminate those regions. Given any face candidate region A, we denote one of its facial feature regions as B_i, N_{Bi} means the number of pixels in B_i. We define R_{Bi} as the Minimal Rectangle Box (*MRB*) of region B_i, and $W(R_{Bi})$, $H(R_{Bi})$ indicate the width and height of R_{Bi} respectively (As a general hypothesis, we assume there is only one face contained in each candidate region A). Then, the rules below are utilized to construct our facial feature filters:

1. Derived from the usual width and height ratio of facial features, any facial feature region with $W(R_{Bi})/H(R_{Bi})>T_1$ or $W(R_{Bi})/H(R_{Bi})<1/T_1$ is eliminated.
2. Since the facial feature occupies only a portion of the face candidate region (A), any facial feature region with $W(R_{Bi})/W(R_A)>T_2$ or $H(R_{Bi})/H(R_A)>T_2$ is eliminated. Moreover, the facial feature region with $N_{Bi}/N_A>T_3$ is also eliminated.
3. In general, the facial feature regions are embedded in the skin region, and the skin usually has higher gray intensity than that of facial features. Hence, the *MRB* of the facial feature region should contain both high and low intensity pixels, its intensity variance ($V(R_{Bi})$) would be larger than a threshold T_4.
4. After the filtering with strategies above, all reserved facial feature regions' directions are calculated out. We then map all these directions into two categories {(0° ~ 45°, 135° ~ 180°), (45° ~ 135°)}. The category with the maximal number of members is reserved, and the region with its direction is not in this category is eliminated.

To get the direction of the facial feature region, we firstly find out the center of the region, then, the line across the center with the sum of the squared distances between region points and this line is minimum would be selected as the axis of the region. Its direction is taken as the orientation of the region.

The face candidate region contains less than three facial features would be relatively hard for verification; accordingly, it is eliminated as the candidate. Fig.3(e) presents the facial feature filtering result, it can be seen that most non-facial feature regions could be successfully eliminated with proposed strategies.

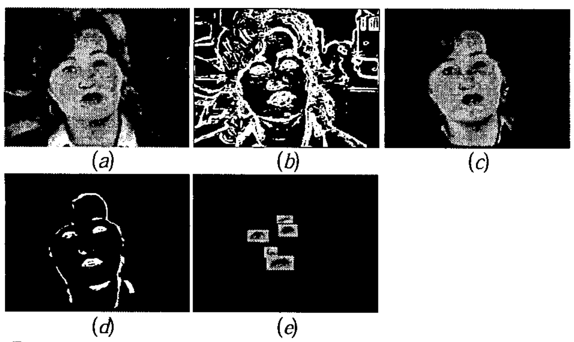

Figure 3. Facial feature extraction ((a),(b),(c),(d),(e) represent the original image (I), Prewitt edge of the image (E_I), segmented face candidate region (A), facial feature segmentation result and filtered facial feature regions (*MRB*) respectively)

2.4 Face verification

Due to different pose, orientation and scale of the face, the template image based face verification methods usually fail in finding the best-matched point to verify face candidate regions which contain side-view faces or faces with different angles [8,9]. In section above, we have detected a set of face candidates (such as A, *etc.*) and their facial feature regions ($B_1,...,B_M$, $M>2$). In this section, a view, scale and orientation invariant face verification strategy is proposed.

In general, the facial features of human being have two distinct characteristics below:

- Almost all facial feature regions in the face are parallel with each other, i.e. they have the same direction. The projections of the gray intensity of the face region along this direction are invariant with the scale, orientation and pose of the face.
- Since the face of human being is symmetrical, in theory, the gray intensity projection of the profile face along facial feature regions' direction would be the same with that of the whole face.

Hence, instead of using template images, we employ a face projection curve for verification, as shown in Fig.4. Given any three facial feature regions (B_1, B_2, B_3) in face candidate region A, we firstly calculate the average direction of B_1, B_2 and B_3, and denote this direction as $\overline{D}_B$, the vertical direction of $\overline{D}_B$ is denoted as $\hat{\overline{D}}_B$. Then, the line parallel to $\hat{\overline{D}}_B$ is taken as the horizontal projection line. The projection curve of the facial feature regions (B_1, B_2, B_3) is obtained with strategies below:

1. Given directions $\overline{D}_B$ and $\hat{\overline{D}}_B$ which are vertical with each other, a coordinate system ($\overrightarrow{\overline{D}_B\hat{\overline{D}}_B}$) is set up. The *MRB* of the facial feature regions (B_1, B_2, B_3) in this coordinate system is also detected out. As shown in Fig.4, the P_U, P_D, P_L and P_R form the *MRB* of B_1, B_2 and B_3.
2. Among the *MRB* of B_1, B_2 and B_3, for any line i with its direction parallel with $\overline{D}_B$, we calculate the average gray intensity of all pixels on this line and denote it as GA_i. We then use PC_i ($PC_i=255-GA_i$) as the projection value of this line in the projection curve.
3. After all projection values ($PC_1,...,PC_W$) of region A have been worked out, they will be used to compare with the face template curve ($PT_1,...,PT_S$) to calculate the likelihood of current region belongs to the face. For that, the projection curve is normalized at first, then, the Hausdorff [6] distance is used to calculate the similarity between it and template curves. A simple threshold selection strategy could determine whether there is a face contained or not.

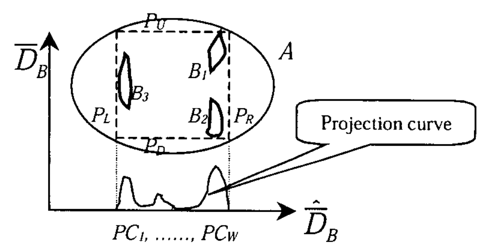

Figure 4. Facial feature projection

Given any face candidate region A, it may not always contain eyes and mouth area for verification, some facial features might be missed while feature extraction. Hence, two template curves are used in our system: one curve is gotten from the projection of the eyes and mouth triangle area; the other one is gotten from the eyes and nose triangle area. That is, if some facial features are missed while feature extraction, or faces are partially occluded, the face candidate region can still be verified successfully.

The analysis of the chrominance components indicated that high C_b and low C_r values are found around the eyes, and high C_r values are found in the mouth areas [3]. Hence, for facial feature regions which pass the verification successfully, the chrominance color contained in them are used to classify them into eyes, mouth, etc. Meanwhile, the triangle which connect these facial features are used to indicate the central location of the face, as shown in Fig.5(*c*).

The predominance of using this projection curve matching strategy is threefold: (1) it's orientation invariant; (2) it's scale invariant; (3) it's view invariant, As shown in Fig.5, even side-view faces or faces with various orientation are verified successfully.

3. EXPERIMENTAL RESULTS

In this section, a set of experimental results are demonstrated to verify the effectiveness and efficiency of the proposed strategy. We randomly select about 200 frames from two types of videos (medical and News) with about 160 frames contain the faces; the other 40 frames contain no face. In these test frames, human faces are presented in various environment, pose, orientation and scale. While facial feature filtering, we set T_1, T_2, T_3 and T_4 equal to 4, 0.5, 0.05 and 15 respectively.

In average, we need about 2 second with a *PIII 900MHz PC* to process each of the test frames. The experimental result is shown in Table 1. We use N_c, N_m and N_f to indicate the number of face which are correctly detected, missed and falsely detected. Then, the *detection rate* (*DR*) and *false rate* (*FR*) which are defined with Eq.(5) are employed to evaluate the system efficiency. Experimental results show that our strategy can locate the face, even profile faces, efficiently. The average *detection rate* and *false rate* are about 84.2% and 11.9% respectively.

$$DR=N_c/(N_c + N_m); \qquad FR=N_f/(N_c+ N_f) \qquad (5)$$

Table 1. Face detection results

Head pose	Frontal	Half-Profile	Profile	With-Angle	All
Number	91	44	11	26	172
Image Size	320*240, 352*288				
DR	87.3%	82.5%	62.0%	78.5%	84.2%
FR	8.8%	11.5%	19.6%	9.5%	11.9%

4. CONCLUSION

In this paper, an omni-face detection system is presented to detect faces by facial feature extracting and verifying. The novel feature that distinguishes our framework from the existing strategies is threefold: (1) To overcome the contradiction between the flexibility of the skin model and the specialties of specific images, a pairwise skin region refinement strategy is firstly proposed to detach the falsely segmented non-skin pixels from the skin region. By doing this, the inputted frame is separated into three maps, the face detection results on all these three maps are combined together to enhance the detection accuracy; (2) A region based adaptive threshold selection strategy for facial feature segmentation is proposed. It can be used to extract the facial features from regions vary in intensity and color distribution; (3) Instead of using the traditional template image for face verification, an orientation, pose and scale invariant template, face projection curve, is employed in this paper. Detection results for frames from various types of videos have been presented. The experimental results reveal the efficiency of the proposed method in detecting omni-faces.

5. REFERENCES

[1] M.H. Yang, D. Kriegman, N. Ahuaja, "Detecting face in images: A survey", IEEE T. PAMI, Vol.24, No.1, p.34-58,2001.

[2] Gang Wei, Ishwar K. Sethi, "Omni-face detection for video/image content description", ACM international conference on MM, CA, 2001.

[3] R. L. Hsu, M.A. Mottaleb, A.K. Jain, "Face detection in color images", IEEE 8th ICIP, Greece, Oct. 2001.

[4] J. Yang, W. Lu and A. Waibel, "Skin-color modeling and adaptation", Technical Report, CMU-CS-97-146, May, 1997.

[5] W. Huang, Q. Sun, C.P. Lam and J.K. Wu, "A robust approach to face and eyes detection from images with cluttered background", ICPR, vol.1, p110-114, Aug. 1998.

[6] D.P. Huttenlocher, G.A. Klanderman and W.J. Rucklidge, "Comparing images using the Hausdorff distance", IEEE Trans. PAMI, Vol.15, p850-863,1993.

[7] C. Kotropoulos, A. Tefas and I.Pitas, "Frontal face authentication using morphological elastic graph matching", IEEE Trans. IP, vol.9, p. 555-560, Apr. 2000.

[8] H.A. Rowley, S. Baluja and T. Kanade, "Neural network-based face detection", IEEE T. PAMI, v.20, p.23-38, Jan. 1998.

[9] H. Schneiderman and T. Kanade, "A statistical method for 3D object detection applied to face and cars", IEEE CVPR, June, 2000.

[10] G.Z. Yang, T.S. Huang, "Human face detection in a complex background". Pattern Recognition, vol. 27, No.1, p.53-63, 1994.

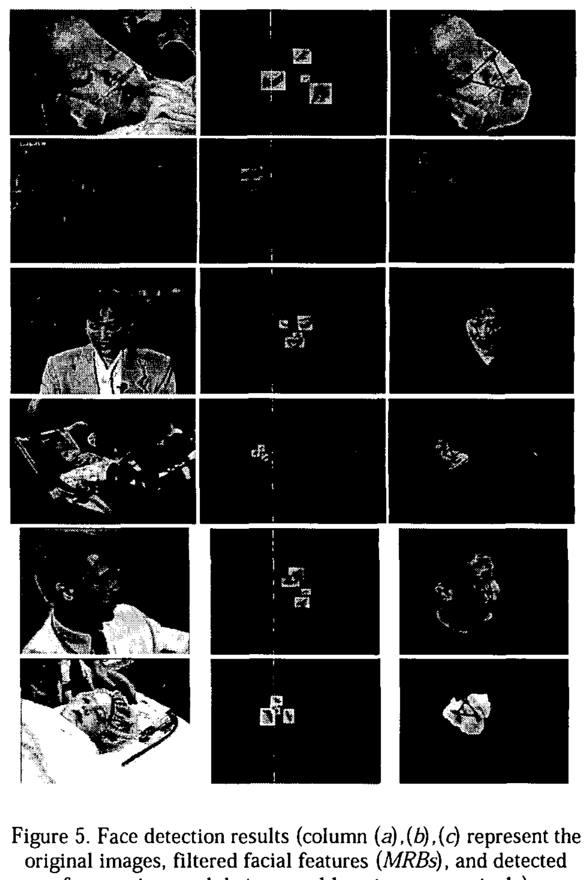

Figure 5. Face detection results (column (*a*),(*b*),(*c*) represent the original images, filtered facial features (*MRBs*), and detected face regions and their central locations respectively)