Combining Color, Contour and Region for Face Detection

Yuchun Fang, Yunhong Wang, Tieniu Tan
National Laboratory of Pattern Recognition, Institute of Automation,
Chinese Academy of Sciences, P. O. Box 2728, Beijing, P. R. China, 100080
E-mails: {ycfang, wangyh, tnt }@nlpr.ia.ac.cn

Abstract

In this paper, we propose a novel contour-region face detector based on the modified committee method. The detector fuses results from a face-contour-based classifier and a facial-area-based classifier, both of which are a SVM (Support Vector Machine). The former is used to identify face contour pattern, and the latter is aimed at discriminating facial area and non-facial area. A skin-color filter is adopted to accelerate the detection. Through serial and parallel fusion of multiple cues, a robust face detection algorithm is implemented through the complementarity of color, contour and region.

1. Introduction

To implement a successful facial analysis system, the first essential step is to detect face automatically, accurately and rapidly. Moreover, face detection can provide useful information for indexing video and image databases.

Although face is one of the most common patterns in the cognitive world of human beings, it is not an easy task for computers to automatically detect faces in static images or video sequences. Such difficulties lie mainly in the following aspects: the unknown size and number of faces in images, the diversity of facial poses and expression, the variety of lighting conditions and the complexity of background.

In the last decade, much effort has been devoted to solving these difficulties [1-16]. Generally, a face detection algorithm can be considered in a general hierarchical framework that detects object areas and decreases redundancy in images step by step as illustrated in Figure 1. The first step is to select candidate face areas in an image (Step ①). Then a face detector is adopted to determine whether a candidate face area really contains a face (Step ②). These first two steps are often repeated in order to facilitate multi-scale detection as a solution to the problem of uncertainty of face size (Step ③). The final detection result is derived by fusing the output of the above steps (Step ④).

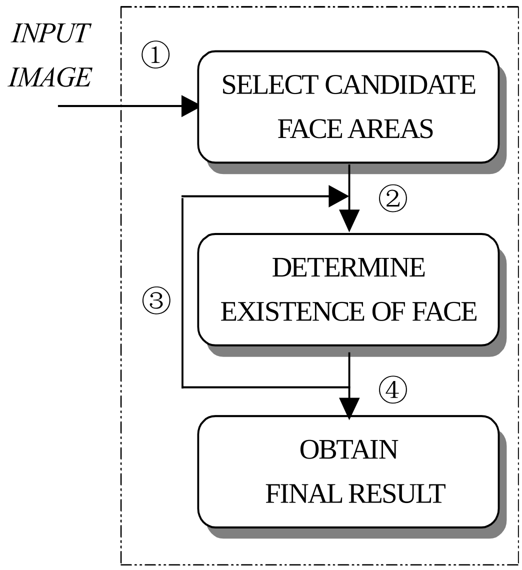


Figure 1. General face detection algorithm

As the location of faces is unknown, exhaustive search over the whole images is frequently performed. To fulfill a real-time algorithm, multi-resolution scheme [1], contour [2], position of facial feature points [3-5] and color filter [6,12,16] are adopted in order to speed up search. Step ②, in which a sub-image is judged as face or non-face, plays a most important role. Some researchers adopted rule-based judgment [1,2,4]. Various face templates were constructed based on color [6], shape [7] or region [8] information. Most researchers took facial area as a special pattern and trained a two-class classifier to make decision in Step ②[9-16]. Texture analysis [9], wavelet analysis [10,11], clustering projection [12] and moment [13] were frequently adopted as feature extraction schemes. The widely adopted classifiers are SVM [13,14] and neural network [12,15,16].

Although pattern classification based methods are more robust to the diversity of faces compared with rules or templates based methods, it is hard to find a sufficient number of non-face samples to train a classifier, and there are patches of natural scene that looks like facial area (see Figure 3 in [11] for an example). It is even impossible for human to decide whether such patch is facial area or not. In reality, one seldom performs face detection just according to facial area. They rely on multiple cues that result in the final highly accurate detection. The contour of head or face is one important cue as well as the facial area. Based on such observation, a contour-region face detector is established in this paper by fusing the outputs of a face-contour-based classifier and a facial-area-based classifier. A skin color filter is adopted to speed up search.

The overall scheme of the proposed algorithm is shown in Section 2. In Section 3, a detailed introduction to the proposed contour-region face detector is given. Experimental results are presented in Section 4.

2. Framework of the proposed algorithm

Our face detection algorithm is also based on the general face detection framework shown in Figure 1. Search over the whole image in multiple resolutions is performed to cope with the unknown size and number of faces. Sub-images around or inside skin regions are candidate faces. Each candidate is judged as face or non-face by the contour-region face detector. The proposed algorithm is illustrated in Figure 2.

First, we adopt a skin filter to limit the search area in images. The adopted skin filter is based on our previously published skin detection algorithm [17]. This skin-filter is based on a skin color distribution model combined by two fuzzy membership functions on **rg** (the normalized **RGB** color space).

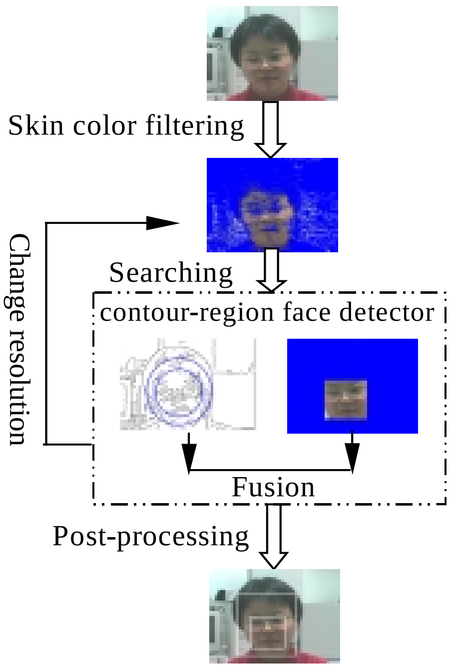


Figure 2. Framework of the algorithm

After pre-filtering, the search is performed either around or inside skin areas in both the input image and its edge image. In the edge image, feature vectors of elliptical rings around the searched sub-images are extracted based on edge points inside the rings. These feature vectors are the inputs to the facial-contour-based classifier. The feature vector of a pre-processed sub-image is the input to the facial-area-based classifier. The results of both classifiers are fused by the modified committee method [18] to decide the final detection result. The proposed contour-region face detector works through combining the above three processing modules. The above process is repeated at several levels of resolution in order to search faces of different sizes.

Finally, post-processing is performed to output the final detection result. Post-processing serves to arbitrate the detection results obtained from all search iterations. When we search in multiple scales around the images, the case of multiple results frequently happens around the same region. Post-processing evaluates the detection results and makes final decisions.

3. Contour-region face detector

Face detector serves to judge whether a candidate region is face or non-face. It is the most essential part in a face detection algorithm.

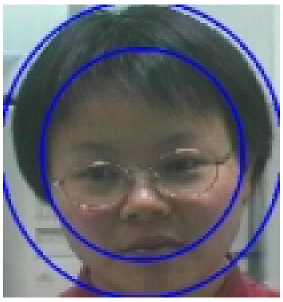
The proposed contour-region face detector is based on the modified committee method [18]. The inputs of the contour-region face detector are outputs of a face-contour-based classifier and a facial-area-based classifier. The feature vector of the former is extracted from edges around facial area, and that of the latter is obtained from the facial area. Because it is hard to obtain enough negative samples for either the facial-area-based classifier or the face-contour-based classifier, SVM, which can minimize the structure risk under small training sets, is selected as classifier for both.

3.1. Face-contour-based classifier

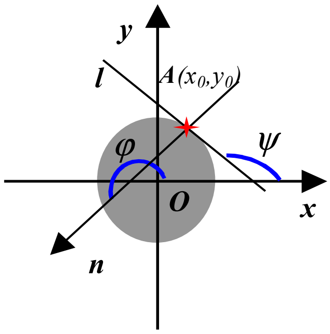
In general, the shape of face or head can be viewed as an ellipse, so the contour of face or head forms a semi-ellipse in a well-detected edge image. Such contour information has been adopted in much previous work for face detection [2,8].

In our work, the feature of edges inside an elliptical ring is explored as feature of face or head contour. As shown in Figure 3.(a), the contour of face or head can be nearly encircled by two ellipses of suitable aspect ratio. The elliptical ring is viewed as a pattern, and the face-contour-based classifier serves to identify whether it is face contour or not. A 9-dimensional feature vector is extracted based on edge points inside the ring.

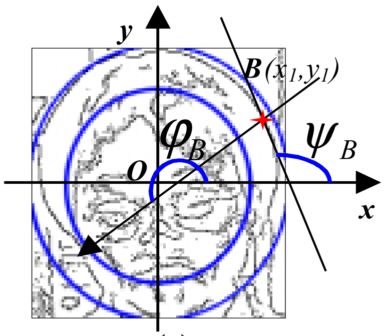
We divide a ring into eight equal bins according to the angles. The number of edge points inside each bin is accumulated. Through normalizing these eight numbers with the perimeter of the ring and total number of edge pixels inside the ring, we obtain eight elements of the feature vector. They reflect density of the edge points distributing along the elliptical ring.



(a)



(b)



(c)

Figure 3. Feature extraction of elliptical ring

On a given ellipse as shown in Figure 3.(b), select an arbitrary edge point $A(x_0,y_0)$, the direction of its gradient is $\boldsymbol{n}$, and the angle between $\boldsymbol{n}$ and $\boldsymbol{x}$ axis is φ. Line $\boldsymbol{l}$ is the tangent of A to the ellipse, ψ is the obliquity of $\boldsymbol{l}$. It is ease to show that φ and ψ are fit to Eqn.(1). If the ellipse has $\boldsymbol{N}$ edge points, the value of Eqn.(2) is 0.

$$|\varphi-\psi|=\frac{\pi}{2} \tag{1}$$

$$z=\frac{1}{N}\sum_{i=1}^{N}(|\varphi_i-\psi_i|-\frac{\pi}{2}) \tag{2}$$

Let a and b be the length of long axis and short axis of the outer ellipse of the elliptical ring respectively. For an arbitrary edge points $B(x_1,y_1)$ inside an elliptical ring (see Figure 3.(c)), we could obtain φ_B through edge detection. If we assume that B locates on an ellipse with the same aspect ratio as the outer ellipse of the ring, ψ_B can be calculated according to Eqn.(3).

$$\psi_B = a\tan\left(-\frac{b^2x_1}{a^2y_1}\right) \tag{3}$$

Thus, we could obtain the 9th elemant of the feature vector according to Eqn.(2). It can be regarded as a kind of average accumulated error of edge points. Such error reflects the likeness of an edge point as a point on an ellipse.

With the obtained feature vectors on training sets, a SVM is trained as the classifier.

The contour-based classifier is simple and less time-consuming. But the correct detection heavily relies on the edge detection algorithm, which often has several parameters to adjust. In addition, crowded background and elliptical shape like objects in background will deteriorate the algorithm. Hence, the contour-based detection has relatively poor adaptation ability.

3.2. Facial-area-based classifier

The facial area possesses great homogeneity in the distribution of facial features. As shown in Figure 4.(a), the facial area is regarded as one spatially well-defined pattern [15]. It is natural to design a classifier to identify face and non-face. The adopted facial-area-based classifier serves to classify any square patch of image as face or non-face. It is a verification problem.

Preprocessing of the patch is performed before feature extraction. First, a given square patch is normalized to 20*20, and then it is masked to avoid possible influence of non-facial area. Finally, the brightness is adjusted with histogram equalization in the remaining part. An example is shown in Figure 4.(b).

(a)

(b)

Figure 4. An example of facial area pattern

After preprocessing, the gray values of pixels inside the unmasked area are extended to a feature vector of 360 dimensions. A SVM is trained as a classifier. The adopted facial-area-based classifier is with the similar idea as the detection algorithm of Osune et al [13]. The difference lies in the selection of the kernel function of

SVM. While Osune et al. took the polynomial function, we adopted the radial basis function.

There are also obstacles with the facial-area-based classifier. As mentioned in Section 1, the facial area is easy to be confused with natural scenes. Another difficulty is the representation ability of negative samples.

3.3. Fusion algorithm

As we all know that one obtains knowledge not just through edge, color or region, fusion of multiple cues is often involved in visual perception. We try to adopt the modified committee method to fuse the face-contour-based classifier and the facial-area-based classifier. The modified committee method is as follows [18]:

Assume there are M classes (C_1, C_2,…,C_M) in the sample spaces Λ. And there are K classifiers $e_k(x)$ (k=1,…,K), the output of each classifier is shown in Eqn.(4)

$$T_k(x \in C_i) = \begin{cases} 1 & (e_k(x) = i \text{ and } i \in \Lambda) \\ 0 & \text{others} \end{cases} \tag{4}$$

To classifier k, a confidence R_k is assigned according to Eqn.(5)

$$R_k = \frac{P_k}{1 - N_k} \tag{5}$$

where P_k and N_k denote the recognition rate and rejection rate of classifier k respectively. Let

$$T_E(x \in C_i) = \sum_{k=1}^{K} R_k \cdot T_k(x \in C_i) \tag{6}$$

The final decision is given with Eqn.(7)

$$E(x) = \begin{cases} j, & T_E(x \in C_j) = \max_i T_E(x \in C_i) > \alpha \\ M+1 & \text{others} \end{cases} \tag{7}$$

where $\boldsymbol{\alpha}$ is the threshold.

In practice, we directly take the outputs of the two SVM classifiers as inputs to the fusion module. With the adoption of the empirical confidence, the fusion between the two classifiers can be performed with the modified committee method.

In the proposed algorithm, fusion takes place between the facial-area-based classifier for a candidate facial area and the face-contour-based classifier for the best matching elliptical ring. The best matching elliptical ring is determined through comparison among 4 elliptical rings with the same aspect ratio as that of the average head size around facial area. The relative position between the facial area and the 4 candidate elliptical rings is sufficient to adapt to the general change of head pose.

4. Experiments and Analysis

4.1. Training of classifiers

Three classifiers are trained in our system. To train the face-contour-based classifier, 924 positive samples and 3,194 negative samples are adopted. 2,304 positive samples and 3,000 negative samples are obtained to train the facial-area-based classifier. We select the radial basis function as the kernel function of both SVM classifiers. The confidence in the fusion classifier is adopted through performing the face-contour-based classifier and facial-area-classifier on two manually obtained test sets.

4.2. Experimental results

We tested the proposed algorithm on three different sets of color images A, B and C collected by ourselves. The images are either scanned, or captured with CCD and digital cameras. The test images are taken under a wide variety of conditions.

Set A contains 924 manually well-cut head images and 3,198 randomly cropped non-head images of animal and scene with aspect ratios similar to that of head. We test the performance of the three classifiers (FCBC: face-contour-based classifier; FABC: facial-area-based classifier; FCRC: integrated contour-region classifier) on Set A respectively. The test results are listed in Table 1, which shows that the FABC is more reliable than the FCBC, while the fusion of the two classifiers further improves the correct recognition rate.

Table 1. Recognition results on Set A

Classifier	FCBC	FABC	FCRC
False reject #	154	55	29
False accept #	284	12	1
Rate of Error	10.6%	1.6%	0.7%

The performance of the proposed face detection algorithm is also evaluated on Set B and Set C that are composed of natural images containing faces. The total number of candidate face regions produced by skin color filter, the total number of faces genuine in the set, the correct detection rate and the number of false detections are recorded in Table 2, where the correct detection rate is the ratio between the number of correctly detected faces and the total number of faces, and a false detection means that a non-face candidate region judged as a face.

Because much face analysis research involves indoor images, we evaluate the performance of the proposed algorithm on Set B composed of such images. In Set B, the size of facial area ranges from 25*25 to 120*120, and for each image we perform search on 4 or 5 levels of resolution.

The 46 images in Set C are much more challenging for the face detection algorithm. They are images either taken under crowded outdoor background, or containing multiple persons with different sizes of heads and small sizes of faces (as small as 22*22). Some examples are shown in Figure 6. The detection results on both sets are summarized in Table 2.

Table 2. Detection results on Set B and Set C

	Set B	Set C
Total # of candidate face regions	**13,259**	**484,856**
Total # of faces	**78**	**145**
Correct detection rate	**96.2%**	**91.0%**
False detection #	**4**	**55**

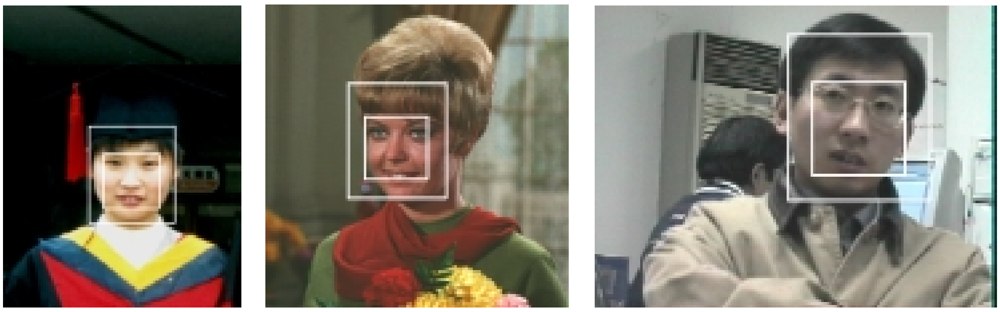
From Table 2, we see that the proposed algorithm obtains a high accuracy of 96.2% on Set B and 91.0% on Set C respectively. For Set B, a total number of 13,259 candidate face regions were produced by the skin color filter, 13,076 of which are rejected (i.e. they are considered as containing no face), and the other 183 are accepted. The number of false rejects and the false accepts are 50 and 25 respectively. So the overall error rate of the integrated contour-region classifier is as low as 0.6% (=75/13259). Without taking into account search on very small scale, the average detection time is no more than 1 second for images in Set B, so the proposed algorithm is fast enough for face detection. Although the proposed algorithm has 55 false detections in Set C, in comparison with the 484,856 sub-images being searched, it is still a small number.

Some sample detection results are shown in Figure 5 and Figure 6. The small white rectangle frame shows the position of facial area, and the bigger one shows the range of the face contours.

Figure 5 (a)-(c) shows several detection results of images in Set B. It can be found that the proposed algorithm has considerable tolerance to in-plane and out-plane rotations of head. Detection results of sample images of Set C are showed in Figure 6 (a)-(c). It can be seen that the proposed algorithm can detect faces of different size in the same image. The false detection can be eliminated in the followed process in face analysis. We have analyzed the missed faces, some of them are lost in arbitration phase, and few are due to the failure of the proposed detector. Improving the arbitration scheme and the searching scheme will obtain better result.

5. Conclusion and future work

In this work, a novel face detection algorithm has been presented which combines color, contour and region information. Performing detection through combining multiple-cues is in accordance with the mechanism of the biological perception system. The adopted color filter speeds up the algorithm through decreasing the search region. The proposed contour-region face detector improves the correct detection rate and results in more precise detection through complementarity between the face-contour based classifier and the facial-area-based classifier. The performance of the facial-area-based classifier can be improved further through bootstrapping training to the SVM classifier. Applying optimization search in the parameter space of the ellipse is another potential research direction that will improve the face-contour-based classifier.



(a) (b) (c)

Figure 5. Several detection results on Set B

(a) (b)

(c)

Figure 6. Several detection results on Set C

Acknowledgements

This work is funded by research grants from the NSFC (Grant No. 59825105), the 863 Program (Grant No. 863-306-ZT-06-06-5 and 863-317-01-10-99) and the Chinese Academy of Sciences.

References

[1] G. Yang, T. S. Huang, Human Face Detection in a Complex Background, *Pattern Recognition*, 1994, Vol.27, No.1, pp.53-63;

[2] V. Govindaraju, Locating Human Faces in Photograghs, *International Journal of Computer Vision*, 1996, Vol.19, No.2, pp.129-46;

[3] C. Y. Kin, R. Cipolla, Fearture-Based Human Face Detection, *Image and Vision Computing*, 1997, Vol.15. No.9, pp.713-35;

[4] J. Miao, B. Yin, et al. A Hierarchical Multiscale and Multiangle System for Human Face Detection in a Complex Background Using Gravity-Center Template, *Pattern Recognition*, 1999, Vol.32, No.5, pp.1237-48;

[5] C. Han, H. M. Liao, et al., Fast Face Detection Via Morphology-Based Pre-Processing, *Pattern Recognition*, 2000, Vol.33, pp.1701-12;

[6] J. Cai, A. Goshtasby, Detecting Human Faces in Color Images, *Image and Vision Computing*, 1999, Vol.18, pp.63-75;

[7] H. Wu, Q. Chen, M. Yachida, Face Detection From Color Images Using a Fuzzy Pattern Matching Method, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1999, Vol.21, No.6, pp.557-63;

[8] J. Wang, T. Tan, A New Face Detection Method Based on Shape Information, *Pattern Recognition Letters*, 2000, Vol.21, pp.463-471;

[9] Y. Dai, Y. Nakano, Face-Texture Model Based on SGLD and its Application in Face Detection in a Color Scene, *Pattern Recognition*, 1996, Vol.29, No.6, pp.1007-1017;

[10] C. P. Papageorgiou, M. Oren, T. Poggio, A General Framework for Object Detection, 1998, *Int. Conf. on Computer Vision*;

[11] C. Garcia, G. Tziritas, Face Detection Using Quantized Skin Color Regions Merging and Wavelet Packet Analysis, *IEEE Transaction on Multimedia*, 1999, Vol.1, No.3, pp.264-77;

[12] J. Terrillon, David, et al., Invariant Face Detection with support vector machine, *The 15th Int. Conf. on Pattern Recognition*, 2000, Vol.4, pp.210-17;

[13] E. Osuna, R. Freund, F. Girosi, Training Support Vector Machines: an Application to Face Detection, *Int. Conf. on Computer Vision and Pattern Recognition*, 1997;

[14] H. A. Rowley, S. Baluja, T. Kanade, Neural Network-Based Face Detection, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1998, Vol.20, No.1, pp.23-38;

[15] K. Sung, T. Poggio, Example-Based Learning for View-Based Human Face Detection, *IEEE Transaction on Pattern Analysis and Machine Intelligence*, 1998, Vol.20, No.1, pp.39-50;

[16] R. Feraud, O. J. Bernier, et al., A Fast and Accurate Face Detector Based on Neural Networks, *IEEE Transaction on Pattern Analysis and Machine Intelligence,* 2001, Vol.23, No.1, pp.42-53;

[17] Y. Fang, T. Tan, A Novel Adaptive Colour Segmentation Algorithm and Its Application to Skin Detection, The 11th British Machine Vision Conference, 2000, Vol.1, pp.23-31;

[18] Y. H. Wang, Application of Neural Networks in Radar Target Recognition, Ph.D. Dissertation, 1998, Nanjing University of Science and Technology.