Detection of Faces of Various Directions in Complex Backgrounds

Yuichi ARAKI Nobutaka SHIMADA Yoshiaki SHIRAI
Dept. of Computer-Controlled Mechanical Systems, Osaka University
2-1, Yamada-oka, Suita-shi, Osaka, 565-0871, Japan
E-mail:{araki, shimada, shirai}@cv.mech.eng.osaka-u.ac.jp

Abstract

This paper describes the detection of faces in complex backgrounds where their sizes, positions and directions are arbitrary. We detect the faces by extracting face components such as eyes, a mouth and so on. We first extract face features and then calculate their likelihoods as each face component. Second we detect the face features which satisfy geometrical relations of the face. In order to reduce the number of combinations of all labels for all features, we determine face candidate regions using generalized Hough transform and then apply a relaxation method to each candidate region.

1. Introduction

The face detection is useful not only for face recognition but also for person tracking. Many approaches of the face detection have been proposed, but most of them assume frontal faces and simple backgrounds. In the surveillance systems and pet-robots, it is important to detect faces whose position, size and direction are arbitrary in complex backgrounds.

One of the methods to detect faces tries to extract face features and then detect faces using geometrical relations[3]. Because this method can obtain distances between the face components, the precise face size is obtained and it is important for the face recognition. However, this method works only if all face components are extracted.

Another method handles the side-view faces and the partially occluded faces by computing the probabilities of face models of various component sets using Bayesian network[5]. A relaxation method[4] is also able to handle those kinds of faces because it uses geometric relations between two face components. However, in both of those methods above, incorrect frontal faces are detected instead of correct side-view faces because the number of components of the frontal face model is more than that of the side-view face model, and the frontal face is supported more than the side-view face.

In this paper, we propose a new method to detect faces whose position, size and direction are arbitrary. We first extract face features as candidates of six face components: both brows, both eyes, nostrils and a mouth. Then we calculate their likelihoods as those face components. Second we verify whether the face features satisfy geometrical relations of those face components. We consider all combinations of all labels for two face features. In order to reduce the number of the combinations, we determine face candidate regions using generalized Hough transform[2] and then apply a relaxation method to each candidate region. Only if we obtain the combination of the face features satisfying the geometrical relations, we detect it as a face.

2. Extraction of the face features

2.1 Detection of face search regions

In order to extract the face features, we first detect face search regions (Fig.1). The skin color regions are detected as the face search region if their areas are large enough.

As the skin color pixels, we extract the pixels satisfying all of the following conditions:

$$\begin{aligned} 0.333 &< r < 0.664, & r &> g, \\ 0.246 &< g < 0.398, & g &\geq 0.5 - 0.5r \end{aligned} \quad (1)$$

where, r=R/(R+G+B), g=G/(R+G+B). The skin color regions are obtained by connecting the skin color pixels.

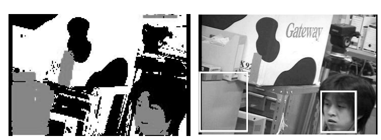
2.2. Calculation of the likelihoods as the face components

We search the face search region for the face components. All of them are not necessarily found in each region. We assume the face size(Fw) and the degree of the rotation angles (θ_y and θ_z) in Fig.2 as follows:

$$\mathrm{Ms} \leq \mathrm{Fw} \leq \mathrm{Cs} \quad (2)$$

$$-90 \leq \theta_y \leq 90, \quad -20 \leq \theta_z \leq 20, \quad (3)$$

where Ms and Cs denote the minimum face size(15 pixels in the following experiments) and the size of the face search region respectively. The face components are darker than the skin color regions in a face and their shapes are like a horizontal line or an ellipse. In order to extract them, we



(a) (b)

Figure 1. Detection of face candidate regions: (a)The skin color regions(gray color), (b)Face candidate regions

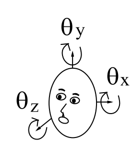


Figure 2. Definition of rotation angles

(a) (b) (c)

Figure 3. Extraction of face features (a)Face search region, (b)Skin color regions(white region), (c)Face features

extract locally darkest pixels and obtain face features by connecting the pixels. As shown in Fig.3, some features in the skin color background and in a face region are none of face components. We consider a feature may be others. Therefore, each feature has seven labels in all.

We define a likelihood $L_i(\lambda)$ of the i th feature for each face component. If the i th feature has image feature of the component of the label λ, we set the $L_i(\lambda)$ to a high value. We verify whether each feature has image feature of a brow and a closed eye, an open eye or a mouth.

- Brow and closed eye

  Let the maximal Y value of pixels in the i th feature be max_i and let the minimum Y value be min_i. The condition of the brow and the closed eye is the following:

  $$max_i - min_i < t_{bc}. \tag{4}$$

  In the experiments, we set t_{bc} to 20. We define L_i(Left Brow) and L_i(Right Brow) as follows:

  $$\begin{aligned} L_i(\text{Left Brow}) &= L_i(\text{Right Brow}) \\ &= \begin{cases} 0.4 & \text{if Eq.(4) holds} \\ 0.1 & \text{otherwise.} \end{cases} \end{aligned} \tag{5}$$

- Open eye

  Let the minimum Y value in the i th feature be min_i. Let the maximal Y value at the left side of the darkest pixel and the one at the right side be $lmax_i$ and $rmax_i$ respectively. The conditions of the open eye are the following:

  $$lmax_i - min_i > t_o \tag{6}$$
  $$rmax_i - min_i > t_o. \tag{7}$$

  In the experiments, we set t_o to 35. We define L_i(Left Eye) and L_i(Right Eye) as follows:

  $$\begin{aligned} &L_i(\text{Left Eye}) = L_i(\text{Right Eye}) \\ &= \begin{cases} 0.8 & \text{if Eqs.(6) and (7) hold} \\ 0.4 & \text{else if Eq.(4) holds} \\ 0.1 & \text{otherwise.} \end{cases} \end{aligned} \tag{8}$$

- Mouth

  We calculate θ value for all pixels in all features using the following equation:

  $$\theta = \arccos\left(\frac{0.5\times(2R-G-B)}{\sqrt{(R-G)(R-G)+(R-B)(G-B)}} \right). \tag{9}$$

  Let an average θ value of pixels in the ith feature be ave_i. The condition of the mouth is the following:

  $$ave_i < t_m. \tag{10}$$

  In the experiments, we set t_m to 90. We define L_i(Mouth) as follows:

  $$L_i(\text{Mouth}) = \begin{cases} 0.7 & \text{if Eq.(10) holds} \\ 0.1 & \text{otherwise.} \end{cases} \tag{11}$$

In cases of the labels of "Nostril" and "Others", we set L_i(Nostril) and L_i(Others) as follows:

$$L_i(\text{Nostril}) = L_i(\text{Others}) = 0.1. \tag{12}$$

For each label, the probability of that the i th feature is label λ is denoted as $p_i(\lambda)$ and the initial probability "$p_i^0(\lambda)$" is defined as follows:

$$p_i^0(\lambda) = \frac{L_i(\lambda)}{\sum_{\lambda'} L_i(\lambda')}. \tag{13}$$

3. Face detection using geometrical relations

We detect the face features which satisfy geometrical relations of the face by using a modified relaxation method. We describe a modification of the update manner in section 3.1 and that of the influences between the features in section 3.2. We describe the use of face candidate regions in section 3.3.

3.1. Use of a relaxation method with a modified update manner

In the normal relaxation method[1], the probability of each feature is influenced by the probabilities of all labels of all other features. The influence at the k+1 th iteration is given by the following equation:

$$q_i^{k+1}(\lambda) = \sum_j \sum_{\lambda'} [d_{i,j}(\lambda,\lambda')\gamma_{i,j}(\lambda,\lambda')p_j^k(\lambda')] \tag{14}$$

where $d_{i,j}(\lambda,\lambda')$ denotes the geometrical compatibility and $\gamma_{i,j}(\lambda,\lambda')$ denotes the influence coefficient. If the geometrical relation between the i th feature being the label λ and the j th feature being the label λ' are compatible, we set the $d_{i,j}(\lambda,\lambda')$ to 1. In the case of the incompatibility, we set the $d_{i,j}(\lambda,\lambda')$ to -1, otherwise 0. We set the $\gamma_{i,j}(\lambda,\lambda')$ values as shown in Tab.1.

In the method[1], the probability "$p_i^{k+1}(\lambda)$" at the k+1 th iteration is updated by following equation:

$$p_i^{k+1}(\lambda) = p_i^k(\lambda)\{1+q_i^k(\lambda)\} \Big/ [\sum_{\lambda'} p_i^k(\lambda')\{1+q_i^k(\lambda')\}]. \tag{15}$$

The probability is guaranteed to be nonnegative by assuming the influence "$q_i^k(\lambda)$" is in the range [-1,1]. However,

Table 1. $\gamma_{i,j}(\lambda,\lambda')$ values

λ'	λ						
	LB	RB	LE	RE	NO	MO	OT
Left Brow(LB)	-0.5	0.3	1.0	0.3	0.3	0.3	0.1
Right Brow(RB)	0.3	-0.5	0.3	1.0	0.3	0.3	0.1
Left Eye(LE)	1.0	0.3	-0.5	0.3	0.35	0.35	0.1
Right Eye(RE)	0.3	1.0	0.3	-0.5	0.35	0.35	0.1
Nose(NO)	0.3	0.3	0.35	0.35	-0.5	1.0	0.1
Mouth(MO)	0.3	0.3	0.35	0.35	1.0	-0.1	0.1
Others(OT)	0.0	0.0	0.0	0.0	0.0	0.0	0.0

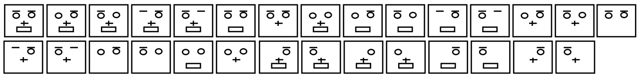


Figure 4. 29 combinations of the face components

the negative value of "$q_i^k(\lambda)$" contributes to the probability "$p_i^k(\lambda)$" more than the positive value.

In order to equalize both contributions of negative values and positive values, we use the following update equations instead of Eq.(15):

$$s_i^{k+1}(\lambda) = \begin{cases} p_i^k(\lambda)[1 + q_i^{k+1}(\lambda)] & \text{if } q_i^{k+1}(\lambda) \geq 0 \\ p_i^k(\lambda) \;/\; [1 - q_i^{k+1}(\lambda)] & \text{if } q_i^{k+1}(\lambda) < 0 \end{cases} \quad (16)$$

$$p_i^{k+1}(\lambda) = \frac{s_i^{k+1}(\lambda)}{\sum_{\lambda'} s_i^{k+1}(\lambda')}. \quad (17)$$

The assumption that the $q_i^k(\lambda)$ is in the range [-1,1] is not necessary.

We stop the iteration when either of the following two conditions is satisfied for all features.

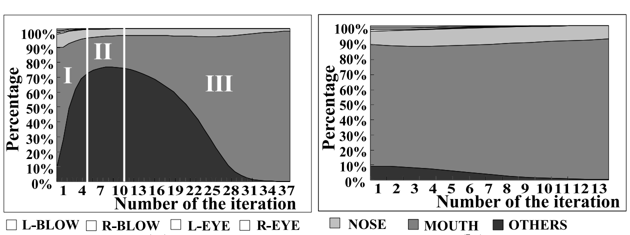
1. The probability of a certain label is more than 0.9.
2. The probabilities of all labels converge.

If the second condition is satisfied for a feature, we consider that the label of the feature is one of labels with probability more than 0.1.

In order to handle arbitrary-posed faces and partially occluded faces, we consider 29 combinations of the face components as shown in Fig.4, where the lines, the circles, the cross and the rectangle show the brows, the eyes, the nostril and the mouth respectively. If the combination of the face features satisfying the geometrical relations is one of them, we detect it as the face.

3.2. More efficient face detection

For calculation of the $d_{i,j}(\lambda, \lambda')$ in Eq.(14), we consider all combinations of all labels(λ, λ') for a pair of two features(i, j). If one of a pair of labels is "Others", it is geometrically compatible. There are many such pairs. However, most other pairs are geometrically incompatible. Moreover, in the beginning of the iteration, there are many features whose probabilities "$p_i(\lambda)$" of all labels are high. Therefore, the influence of "Others" is larger than that of the others. As a result, the probability "p_i(Others)" increases while the probabilities of the face components decrease(Fig.5(a)-I). Features whose labels are "Others" don't influence the probabilities of other labels of features. When the probabilities "p_i(Others)" of all features become large, the changes of all probabilities of all features become small (Fig.5(a)-II). If a feature satisfies the geometrical relation of a component, its probability of the component "$p_i(\lambda)$" increases(Fig.5(a)-III). This is not efficient even if features are labeled correctly.



(a) (b)

Figure 5. Example of convergence: (a)Result before modification, (b)Result by the modified method

In order to solve this problem, we modify the relaxation method as follows.

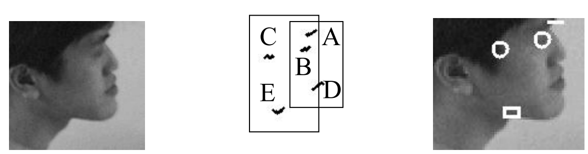
1. The geometrical incompatibilities are not considered. For incompatible labels of λ and λ', the $d_{i,j}(\lambda, \lambda')$ is set to 0 instead of -1.
2. Negative influences between the same labels are considered only if two features are close. However, a mouth may be divided into upper lip and lower lip when it is open. Therefore, negative influences between the labels "Mouth" are not considered if two features are close vertically.

Fig.5(b) shows the convergence result of the same feature by the modified method.

3.3. Determination of face candidate regions by generalized Hough transform

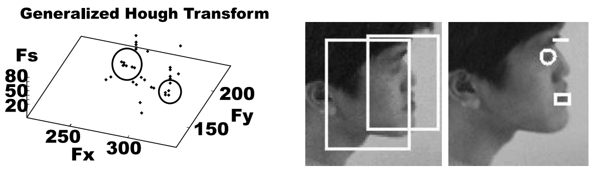
Suppose a face feature satisfying both geometrical relations of frontal faces and side-view faces. Frontal faces have a maximum of six face components instead of four in the case of side-view faces. Therefore, the influence of the frontal face relations is stronger than that of the side-view face relations. As a result, frontal faces are detected instead of side-view faces. For example, in Fig.6(b), the labeling(B⇐Left Eye) satisfies the geometrical relations with three labelings: (A⇐Left Brow), (C⇐Right Eye), (E⇐Mouth). On the other hand, the labeling(B⇐Right Eye) satisfies the geometrical relations with only two labelings: (A⇐Right Brow), (D⇐Mouth).

In order to solve this problem, we determine face candidate regions using generalized Hough transform and then apply the relaxation method to each candidate region. We regard the 2-D position of the nose as the face center(Fx, Fy) and the distance between both eyes as the face size(Fs). We consider combinations of all labels of three features. If the labeling of a combination of the features satisfies the geometrical relations, we estimate the Fx, Fy and Fs based on the positions of the features. The Fx, Fy and Fs are voted



(a) (b) (c)

Figure 6. The false detection of the side-view face:(a)Face search region, (b)Face features, (c)Result of face detection

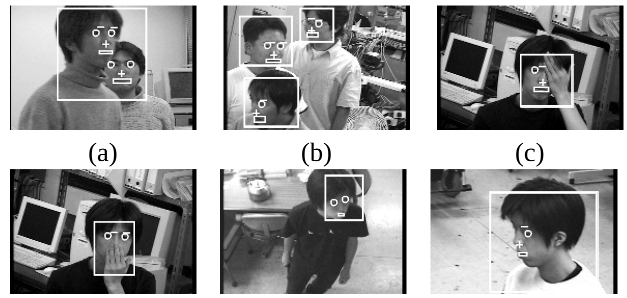


(a) (b) (c)

Figure 7. Face detection by determining face candidate regions: (a)Result of the generalized Hough transform, (b)Two face candidate regions, (c)Result of face detection

(a) (b) (c)

Figure 8. Some examples of face detection



(a) (b) (c)
(d) (e) (f)

Figure 9. Some results of face detection (The rectangle enclosing the face shows the face search region)

in a 3-D space as shown in Fig.7(a). Two circles in Fig.7(a) represent the locally dense positions. Fig.7(b) shows the face candidate regions with the white rectangles. In order to reject extra faces, we track the face over time sequence. While tracking, even if more than one face candidate region are found, we select only the one nearest to the previous position. Fig.7(c) shows the result of the face detection by tracking. When faces are not detected previously, we treat the face as follows:

- If we obtain more than one faces sharing the same face features with each other, we don't detect the faces.
- Otherwise, we detect the face and begin to track it.

4. Experimental results

We made three experiments of the face detection.

First, we verified the robustness to pose changes. We used image sequences of 15 people taken by a fixed camera in a corridor and obtained 4878 images. In these images, there were arbitrary-posed faces. The detection rate of the face was about 85%. Most of the rests 15% were detected as the face search regions, but the positions of some facial components were incorrect.

Secondly, we verified the robustness to pose changes using 2700 face images of 9 directions in a Japanese face database[6]. Tab.2 shows the face detection rate. Some results of the face detection are shown in Fig.8.

Thirdly, we verified the robustness to complex backgrounds. Some results of the face detection are shown in Fig.9. Fig.9(a) and (b) show the detection of multiple faces. Fig.9(c) and (d) show the detection of partially occluded faces. Fig.9(e) shows the detection of a face taken from higher position. In Fig.9(f), the size of the face search region is much larger than the face size because of the skin color clothes and floor. In spite of that, the face is detected.

Our system can detect one face in 0.15 seconds using 1.0 GHz CPU(Pentium III).

Table 2. Face detection rate(%)

vertical direction (degrees)	horizontal direction(degrees)		
	0	right:30	right:60
up:30	87.4	82.0	76.8
0	97.0	96.4	91.3
down:30	92.4	86.7	91.7

5. Conclusion and future work

We propose a new algorithm to detect faces in the complex backgrounds where face position, size and pose are arbitrary. We first extract face features and calculate their likelihoods as each face component. Then we detect the faces using the geometrical relations. We reduce the number of the combinations of all labels of all features using the generalized Hough transform. By experiments, we showed that this method is robust to the changes of face poses, partial occlusions, and complex backgrounds. For future work, we will extract face components more correctly and detect small faces.

The facial data in this paper are used by permission of Softopia Japan, Research and Development Division, HOIP Laboratory. It is strictly prohibited to copy, use or distribute the facial data without permission.

References

[1] A.Rosenfeld, A.Hummel, and S.W.Zucker. Scene labeling by relaxation operations. *IEEE Trans.*, SMC-6(6):420–433, 1976.

[2] D.H.Ballard. Generalizing the hough transform to detect arbitrary shapes. *Pattern Recognition*, 13(2):111–122, 1981.

[3] K. Fukui and O. Yamaguchi. Facial feature point extraction method based on combination of shape extraction and pattern matching. *Trans.IEICE(D-II)*, J80-D-II(8):2170–2177, 1997.

[4] H. J. Lee, D. G. Sim, and R. H. Park. Detection of facial features based on the relaxation algorithm. In *Visual, Proc. of Communications and Image Processing*, pages 1074–1081, 2000.

[5] K. C. Yow and R. Cipolla. Feature-based human face detection. *Image and Vision Computing*, 15:713–735, 1997.

[6] http://www.hoip.softopia.pref.gifu.jp/web_catalog/top.html.