中華民國八十六年全國計算機會議

臉部特徵變化之自動偵測
Automatic Detection of Facial Feature Variations

高民晟
Ming Chung Kao

莊仁輝
Jen-Hui Chuang

國立交通大學資訊科學所
Institute of Computer and Information Science
National Chiao Tung University, Hsinchu, Taiwan, R.O.C.
Tel: (03)5712121-59253 Fax: (03)5736797
Email: kaoming@act.nctu.edu.tw

摘要

此篇論文提出了一個新的方法可以自動地在一張任意表情的臉部影像中擷取出眼睛和嘴巴的形狀與長寬值. 整個方法分爲三個階段. 第一個階段先由一張沒有表情的臉部影像產生此張臉的參考數據, 第二個階段則利用這一些參考數據在任意表情的臉部影像中擷取出所要的特徵. 而第三個階段則是將所擷取出來的特徵數量化. 由實驗結果證實此一方法在適當的光線和臉部沒有太大傾斜的情況下可以得到良好的擷取結果. 此一方法也已被製作成一個即時的臉部特徵擷取系統.

Abstract

This paper proposed a new approach to automatic extraction of shape-variable features, such as eyes and mouth, in a facial image of arbitrary expression. The overall approach consists of three separated but interrelated stages. The purpose of first stage is to locate the bounding boxes of the features and to extract the features in an expressionless image. In the second stage, similar procedures are carried out for an arbitrary expression image. The purposes of the two stages are not the same, because the changes in the shapes of the features are not known a priori. Quantitative descriptions of these changes are then generated in the last stage of the proposed approach. According to the examples considered, when the lighting condition is normal and the tilt of the face is small, the approach can produce reasonably good results. A real-time system based on the proposed approach has been implemented. [1]

1 Introduction

Analyzing facial expression is an easy task for human beings, but not for computers. One of the difficulties lies in the feature extraction process. In this paper, we will focus our attention on extracting features with variable shapes in a facial image with an arbitrary expression. The features include the eyes and mouth, and a person may wear glasses. Environmental effects such as lighting confition and the movement of face will also be considered.

1.1 Previous Works

Feature extraction is usually accomplished by thresholding and model matching. The thresholding methods use gray values to separate feature points from whole facial image, since gray values of image pixels of eyes and mouth features are smaller than other pixels. In [1], the eyes are located by looking for a pair of minima (pupils) in image gray values located below the eyebrows.

For model (template) matching, a model is established for each facial feature and a set of parameters can be associated with each model to handle changes in the size, orientation and scale of a facial feature [2] [3]. For example, in [4] the deformable template for an eye consists of two parabolic sections (the bounding contour), one circle (iris) and two points (centers of the whites of the eye) as shown in Fig. 1. Similar templates consisting of a circle and a tailored exponential function is presented in [5].

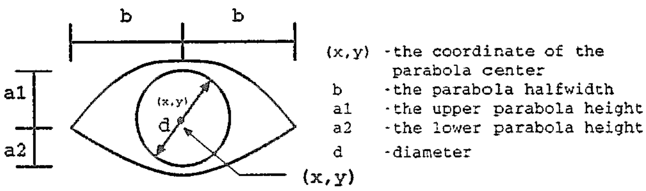

Figure 1: Eye deformable template.

Using the deformable template for feature extraction is time consuming, because the approach requires searches in a 2-D image and each additional parameter will increase the time complexity. Therefore, one may

[1] This research is supported by R.O.C. National Science Council under NSC-86-2213-E009-105.

need to reduce the resolution of a face image, or to segment the image [6].

1.2 A Brief Description of Proposed Approach

The proposed approach extracts eyes and mouth features in an arbitrary expression image. The approach can be divided into three stages as shown in Fig. 2. The purpose of the first stage is to find the reference position of each feature, from an expressionless image. The goal of the second stage is to extract features from an arbitrary expression image, where the facial organs can have any shapes. To achieve such a goal, instead of using the model-based methods directly, the information obtained in the first stage is used to generate the bounding boxes of different features to minimize the search space. The purpose of the third stage is to provide quantitative descriptions of the extracted features. The descriptions include the height and width of each extracted feature. More detailed description of these three stages will be given in the following.

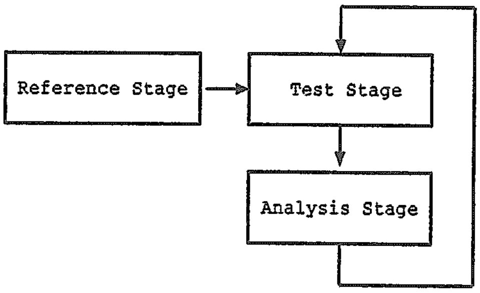

Figure 2: The feature extraction procedure.

2 Reference Stage

The purpose of this stage is to generate facial feature information from an expressionless image, which can then be used to simplify the feature extraction process for various, more complex, expressions later in the test stage. Fig. 3 shows the processes of the reference and test stages.

2.1 Preprocessing

A gray-level transfer function is used to reduce the noise and to enhance facial features. Fig. 4 shows an expressionless image before and after the transformation, respectively. Subsequently, the Prewitt mask is applied to the whole image for edge detection. Fig. 5 shows the result of Prewitt operation.

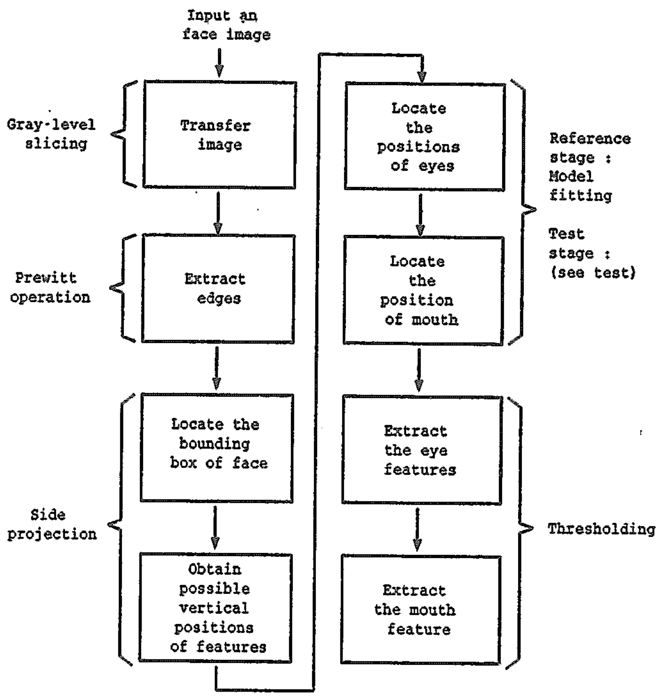

Figure 3: Processes of the reference and test stages.

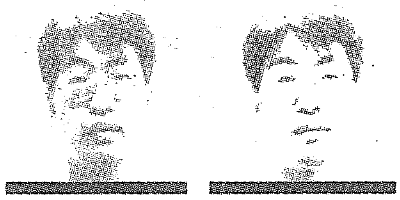

Figure 4: Expressionless images before (left) and after (right) gray-level slicing.

2.2 Locating Eyes and Mouth

2.2.1 Location of Face

We assume that there is hair in the face image. For example, the accumulated gray values in rows of the image near the hair area will have smaller values than the other rows in Fig. 6 where each valley in the *x-axis* represents a possible vertical location of the hair area. Because the vertical position of hair is higher than the other features in a facial image, the position of the first valley will give the vertical position of hair. Similarly, two peaks in the *y-axis* can represent the two sides of the face. These

Figure 5: A result of Prewitt operation.

valleys also give possible vertical positions of eyes and mouth. The advantage of using the side projection is to reduce a 2-D search problem to a 1-D one that makes the search more efficient. Fig. 7 shows an example of the bounding box of a face and possible vertical positions of eyes and mouth.

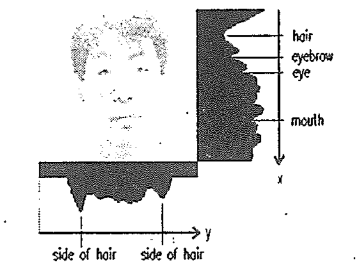

Figure 6: Side projections of face.

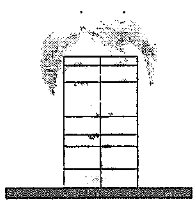

Figure 7: Bounding box of a face and possible vertical positions of eyes and mouth.

2.2.2 Eyes Fitting

We define a model to depict iris, as shown in Fig. 8. The following eye fitting algorithm calculates the numbers of pixels which have gray value larger than a threshold value of 60 along the edges associated with the model of eye. For each *(x,y)* which gives the center of model, the radius of iris γ is tested from four to ten, and *(k,l)'s* represent image pixels located on the edges of the model. The variable *count* gives the number of pixels whose gray values are larger than the threshold.

Algorithm Eye-fitting

Step 1. Initialize the values of *(x,y)*, and let $\gamma = 4$.
Step 2. Use *(x,y)* and γ to determine the positions of edge pixels, *(k,l)'s*.
Step 3. Calculate $count(x, y, \gamma)$ which is equal to the number of *(k,l)'s* having gray values larger than the threshold.
Step 4. Increase γ.
Step 5. If $\gamma \leq 10$, goto *Step 2*.
Step 6. Determine the next position of *(x,y)*. If all possible *(x,y)'s* are tested, goto *Step 8*.
Step 7. Let $\gamma = 4$, goto *Step 2*.
Step 8. Record the maximum value of *count* and the corresponding *(x,y)*, and γ.

2.2.3 Mouth Fitting

The mouth is supposed to be closed in the reference image. It corresponds to an obvious edge between the upper and the lower lips after the Prewitt operation. Thus, the model is chosen to consist of a line segment with the width of four pixels as shown Fig. 9. The final width of mouth is determined by side projection. For the side projection result along the vertical direction, as shown in Fig. 10, one can see that there are two peaks at two sides of the mouth. The distance between the two peaks gives the width of the mouth. The white box indicates the area where the side projection is performed.

2.3 Extracting Feature Points

We use thresholding to extract the eyes and mouth features. The threshold values of eye and mouth are set empirically to be 190 and 180, respectively. Fig. 11 shows bounding boxes of eyes, mouth and face, and the extracted features of eyes and mouth obtained in the reference stage.

3 Test Stage

For the test image considered in this stage, which can be an arbitrary expression image, we don't use simple model fitting to determine the bounding boxes of eyes and mouth. Instead, we use information obtained from the reference stage to generate the bounding boxes of eye and mouth. Fig. 12 shows the test image and the feature extraction result. For locating the eyes and mouth, we must locate the face first and then find the possible vertical positions of eyes and mouth. In general, the widths of the bounding box of the face obtained in the reference image and test image, respectively, are about the same. Therefore, only the shift of the bounding box need to be determined (see Fig. 13).

3.0.1 Location of Mouth

For vertical position of bounding box of mouth, consider the bounding box of face and possible vertical positions of eyes and mouth shown in Fig. 12, The most probable position of the mouth in test stage is chosen to be the one

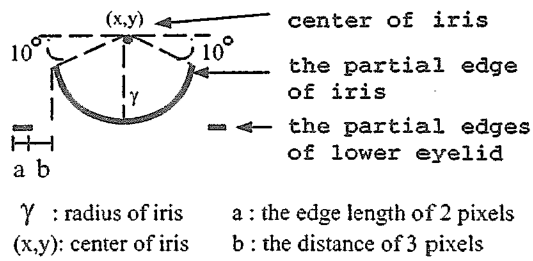

Figure 8: The model of eye.

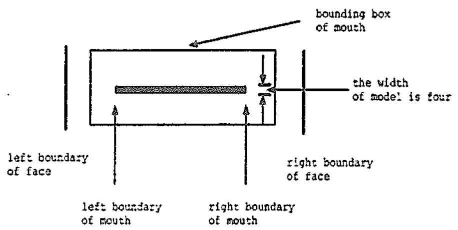

Figure 9: The model of mouth.

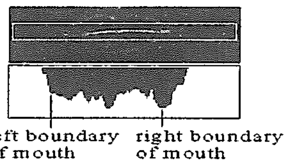

Figure 10: Side projection of mouth bounding box.

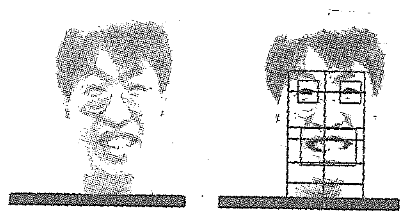

Figure 12: The test image and result.

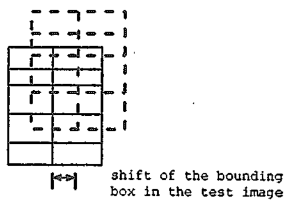

Figure 13: Horizontal shift of the bounding box of face.

which is closest to that obtained in the reference stage as shown in Fig. 14. This simple method can be used to determine the vertical position of mouth because there is little noise near the mouth area.

3.0.2 Location of Eye

The above method dose not yield good results in locating the eyes because there are more edges in the neighborhood of eyes. The three quantities used to determine the most probable vertical position of an eye include :
(1) the vertical distance between the mouth and the eye, and
(2) the vertical distance between upper boundary of the face and the eye.
(1) has higher priority, since the upper boundaries of the faces may be not equal in test and reference images due to possible change in the hair style (see Fig. 15).

4 Analysis Stage

The purpose of this stage is to generate quantitative descriptions of facial features. In the previous two stages, we extract features in expressionless and arbitrary expression images. Some of the extracted features of eyes and mouths are shown in Figs. 16 and 17, respectively.

We define the size of the opening of an eye as the maximum number of feature points in a column of its bounding box. As for the mouth, similar definition is used. The main reason why we choose these quantitative descriptions of eyes and mouth is that the changes in these values are large and often have key effects on the variation of expression.

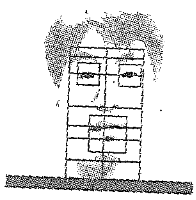

Figure 11: The results of feature extraction in the reference stage.

5 Experimental Results

In this section, the proposed approach is evaluated by examining the correctness of bounding boxes found and correctness of extracted features. We apply the proposed approach to an image database and analyze the result. There are ten persons' facial images in the database. Each person has eight different kinds of facial expressions.

5.1 Correctness of Bounding Boxes

We use four numbers to evaluate the correctness of the bounding boxes found, as shown in Table 1. Examples of some evaluation results are shown in Fig. 18. According to Table 2, the obtained bounding boxes of face, eyes and mouth are correct in most situations. Thus, extracting facial features by the proposed two-stage (reference stage and test stage) procedure seems to be a good approach. The drawback of the proposed two-stage proce-

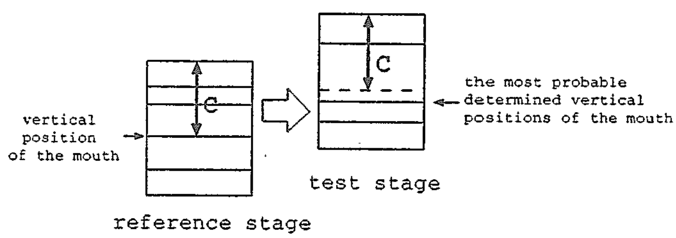

Figure 14: The most probable vertical position of mouth.

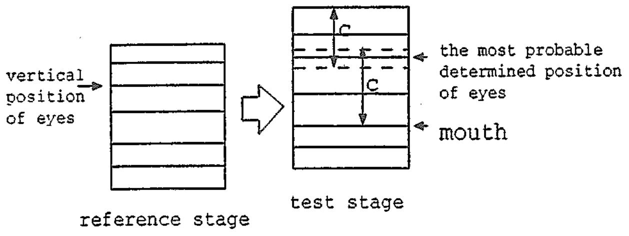

Figure 15: The most probable vertical position of an eye.

Figure 16: Some extracted eye features.

dure is that extra time is spent in the reference stage. The main reason of bad results in winked and tilted expression cases is that we do not consider the tilt of face in the test stage, and it is assumed that the two eyes are of similar shapes for both reference and test images.

5.2 Correctness of Extracted Features

Table 3 shows the sizes (height) of eyes opening in number of image pixels. The two numbers obtained for each image are the sizes of left eye and right eye, respectively. Table 4 shows the width and height of extracted mouth. The two numbers represent the width and height of mouth, respectively. A * mark indicates meaningless results due to a "3" evaluation in Table 2.

We have the following general observations from these two tables:

1. The sizes of left and right eyes are almost equal. It is reasonable. As another observation, we can see that the sizes of a closed-eye are smaller than an open eye, i.e., expressionless > sadness.

2. The height of a mouth varies according to the opening of the mouth. From Table 4, we find that laugh $\simeq$ surprise > smile > expressionless $\simeq$ sadness in general.

3. The width of a mouth varies according to different facial expressions. However, the relative width of a mouth for each facial expression is not clear. But the widths are almost equal when the mouth is closed, i.e., expressionless $\simeq$ sadness.

Figure 17: Some extracted mouth features.

Table 1: Evaluations of the bounding boxes found.

0 :	each bounding box contains the right feature
1 :	one bounding box misses the feature
2 :	two or more bounding boxes are incorrect
3 :	unsuccessful

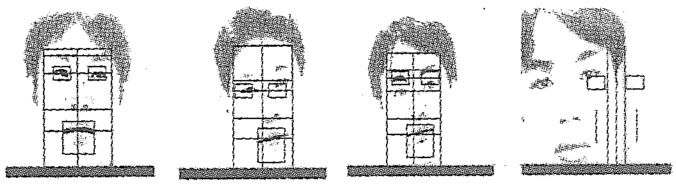

Figure 18: Examples of some evaluation results for the correctness of the bounding boxes obtained with the proposed approach.

For the sad expression in the fourth image, since the eyes are wide open, the heights of eyes are larger than that of eyes of the expressionless one.

6 Conclusion

We have proposed a feature extraction approach which can extract the shape of eyes and mouth in a facial image. The main feature of the approach is that it can extract changeable features in a facial image with an arbitrary expression. Such an ability is made possible by using the locations of facial features obtained from an expressionless image to help the search of similar features, but may be with different appearances, in other image. The approach is quite efficient and has been implemented in a real time system. Applications of such a system include man-machine interface, computer vision, etc.

Table 2: Correctness of bounding boxes found.

	1	2	3	4	5	6	7	8	9	10
no exp.	0	0	0	0	0	0	0	0	0	0
smile	0	0	0	0	0	0	0	1	0	0
laugh	0	0	0	0	0	0	0	0	0	0
sadness	0	0	0	0	0	0	0	1	0	0
surprise	0	0	0	0	0	0	0	0	0	0
teeth	0	0	0	0	0	0	0	0	0	0
wink	1	2	0	2	0	1	0	1	0	0
tilt	0	1	1	3	1	1	3	2	2	2

Table 3: Sizes of eyes opening.

	1	2	3	4	5
no exp.	7, 8	11, 8	8, 9	5, 4	6, 6
smile	6, 6	10, 7	7, 7	8, 8	5, 5
yell	7, 7	10, 7	8, 9	8, 6	6, 6
sad	2, 2	4, 4	6, 5	9, 9	2, 2
surprise	6, 6	9, 7	10,11	7, 7	7, 7
teeth	7, 7	4, 5	7, 9	10, 9	3, 5
wink	2, 7	6, 1	0, 0	8, 4	4, 6
tilt	7, 7	4, 6	0,10	*9, 3	5, 2
	6	7	8	9	10
no exp.	8, 9	10,11	15,13	6, 9	10, 9
smile	4, 6	10,11	5,14	6, 6	11, 9
yell	6, 5	12,12	7,13	5, 7	10, 9
sad	3, 3	5, 7	4,16	8, 8	5, 5
surprise	7, 8	10,12	10,13	10,10	11,10
teeth	8, 8	10, 8	6,14	10,10	10, 9
wink	5, 4	7, 6	2,15	9,10	8, 5
tilt	10, 1	*0, 3	14, 4	11, 2	7,11

References

[1] I. C. H. Ellis and J. R. Lishman, "Automatic Extraction of Face-Feature," *Pattern Recognition Letters*, pages 183–187, 1987.

[2] S. N. Srihari. V. Govindaraju and D. B. Sher, "A Computational Model for Face Location," *Proc. 3rd Int. Conf. Comput. Vision*, pages 718–721, 1990.

[3] R. K. Srihari. V. Govindraju, D. B. Sher and S. N. Srihari, "Locating Human Faces in Newspaper Photographs," *Proc. CVPR*, pages 545–554, 1989.

[4] D. S. Cohen. A. L. Yulie and P. W. Hallinan, "Feature Extraction from Faces Using Deformable Templates," *Proc. CVPR*, pages 104–109, 1989.

[5] Nixon, "Eye Spacing Measurements for Facial Recognition," *SPIE Proc. 575, Applications of Digital Image Processing VIII*, pages 279–285, 1985.

[6] G. Chow and X. Li, "Towards a System for Automatic Facial Feature Detection," *Pattern Recognition*, 26(12):1739–1755, 1993.

Table 4: Width and height of extracted mouth.

	1	2	3	4	5
no exp.	35, 1	50, 2	72, 2	39, 2	44, 3
smile	44, 8	55, 5	69,34	60, 6	71, 9
yell	40, 2	52,10	55, 6	48, 5	59, 4
sad	37, 0	52, 4	72, 4	39,16	44, 5
surprise	33,17	45,10	63,22	38,23	41,12
teeth	46, 5	54, 6	77,13	55, 5	63,12
wink	64, 5	49, 4	74, 1	50, 3	55, 3
tilt	37, 3	46, 4	55, 4	*70, 0	40, 4
	6	7	8	9	10
no exp.	52, 4	44, 5	53, 3	44, 4	44, 5
smile	62,12	69,20	61,14	54,23	40,18
yell	70, 7	64, 9	60, 4	46, 6	46, 4
sad	54, 3	61, 5	53, 4	46,13	44, 7
surprise	73,22	48,16	42, 8	40, 5	32,25
teeth	77,15	72,17	68, 9	52,12	46, 8
wink	56,15	56, 5	50, 3	47, 3	43, 8
tilt	44,13	*70, 0	67, 5	34, 7	57, 4