Facial Expression Recognition Using Adaptive Template

A. Attipoe¹, J. Yan²
¹Department of Computer Science, Bowie State University, Bowie, Maryland, USA
²Department of Computer Science, Bowie State University, Bowie, Maryland, USA

Abstract - In this paper, we presented a classification method of facial expressions displayed in images or video sequences using adaptive template. In our approach, we trained the seven basic expressions templates known as “Happiness”, “Anger”, “Disgust”, “Fear”, “Surprise”, “Sadness” and “Neutral”. Then we take an input face image with unknown facial expressions and calculate the nearest distance between the feature points on that input face image and the feature points on each of the seven facial expressions templates, therefore to find the closest match. Experimental result shows that adaptive template is an effective method to identify facial expressions in an image that contains human face with unknown emotions.

Keywords: Facial expression recognition, feature points

1 Introduction

Facial expressions form a vital part of our interaction with people. Without facial expressions, it would be difficult to capture the whole meaning of a message being conveyed through conversation. Psychologist Mehrabian [8] noted that, facial expressions contributed to 55 percent to a message while 38 percent contributed to the vocal part such as voice intonation. Only 7 percent contributed to the verbal part, which is the spoken word. Facial expression recognition can be defined as a process performed by humans or computers, which consists of locating faces in a scene, extracting facial features from the detected face region (facial feature extraction), and analyzing the motion of the facial feature and/or the changes in the appearance of the facial features and classifying this information into some facial expression–interpretative categories such as facial muscle activations like smile, emotion categories like anger or attitude categories like (dis)liking [9]. Adaptive templates help us capture facial expressions variations or deformations. In our approach, we classify facial expression using adaptive templates.

2 Related Works

Lajevardi and Hussain [5] stated that an automatic classification of facial expressions consists of two stages: feature extraction and feature classification. And of the two stages, feature extraction is of key importance to the whole classification process. Lajevardi and Hussain [5] explained that, if inadequate features are used, even the best classifier could fail to achieve accurate recognition. Usually, extracted facial features are either geometric features such as the shapes of the facial components (eyes, mouth, etc) and the locations of facial fiducial points (corners of the eyes, mouth, etc), or appearance features representing the texture of the facial skin in specific areas including wrinkles, bulges, and furrows [9]. Brunelli and Poggio [1] performed a comparative analysis of geometric, feature-based matching and template matching. In the former, computation of a set of geometrical features from the picture of a face is performed. In the latter, the image, which is represented as a bidimensional array of intensity values, is compared with a suitable metric (typically the euclidean distance) with a single template representing the whole face. After performing face recognition using the two approaches, Brunelli and Poggio [1] concluded that the use of template matching is superior in recognition performance on their database. Zhang, Lyons, Schuster and Akamatsu [14] noted that, though geometric feature-based techniques are usually computationally more expensive than template-based techniques, they are more robust to variation in scale, size, orientation, and location of the face in the image.

3 Methodology

Our technique for facial expression recognition involves extracting facial features geometrically from the face region of an image and recognizing the face by classifying the extracted facial features. The processes covered in our experiment apply to images in both the training and testing set.
3.1 Data Acquisition

The images used for training and testing were taken from the Japanese Female Facial Expression database (JAFFE) and the Taiwanese Facial Expression Image Database (TFEID). The Taiwanese Facial Expression Image database consists of 7200 stimuli captured from 40 models (20 males and 20 females), each with eight facial expressions: neutral, anger, contempt, disgust, fear, happiness, sadness and surprise. Models were asked to gaze at two different angles (0° and 45°). Each expression included two kinds of intensities (high and slight), and was captured by two CCD-cameras simultaneously with different viewing angles (0° and 45°). A combination of grayscale and color images was used. The images taken from the Japanese Female Facial Expression database consists of 219 image variations of the six basic facial expressions and the neutral facial expressions of ten Japanese female models.

We took 175 images of the different facial expressions in total from the Taiwanese Facial Expression Image Database. 140 images were used for the testing set; each of the seven facial expressions had 20 input/test images. Then 35 images were used for the training set. Each of the seven facial expressions had five, sample images. We ensured that the facial expressions chosen for the training set were posed by the same male and female models. 77 images were chosen from the Japanese Female Facial Expression database. 70 images for the testing set (ten images for each facial expression) and seven images for the training set, each image representing one of the seven facial expressions. All the images had a frontal view.

3.2 Image Pre-processing

During image pre-processing, we scaled the images to 256x256 pixels and ensured that they were all of the same size and shape. We also normalized all the images to ensure that they had uniform intensity values.

Figure 1: Above is the normalized training set for the Japanese Female Facial Expression (JAFFE) database.

3.3 Feature Extraction

Feature points were chosen on the face region for all the images in our training and testing set. Feature points were chosen on the eye brows, eyes and mouth on the face region of each image retrieved from the Taiwanese facial expression database. And for the images retrieved from the Japanese female facial expression database, feature points were chosen on the eye brows, eyes, mouth, chin and around the contour of the face. The goal was to use areas of the face that showed the most deformation after a facial expression has been made. For example, a surprise facial expression will show that the eye brows raised up and eyes wide open whereas a neutral facial expression will show the reverse. Each feature point on the face of an image is represented by an x-axis and a y-axis. To extract some geometric measurements from the feature points, we matched each image in our testing set against the seven facial expressions in our training set. Then we calculated the distance between the feature points located at the same location on the two images. The distance for all the feature points was calculated and summed. For each input image from our testing set, seven different distances were generated.

Figure 2: These are ten images of the "fear" facial expression in the testing set, which were taken from the Taiwanese Facial Expression Image Database and used for experimental analysis. Each of the images is 256x256 pixels.
<table>
<thead>
<tr>
<th><strong>Image Acquisition</strong></th>
<th>An image is extracted from a video sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Normalization</strong></td>
<td>The image is normalized; it is scaled to the shape and size of the images in the training set</td>
</tr>
<tr>
<td><strong>Feature Point Extraction</strong></td>
<td>Feature points are extracted from the image</td>
</tr>
<tr>
<td><strong>Feature Template</strong></td>
<td>The feature template represents the final template after the feature points have been extracted.</td>
</tr>
<tr>
<td><strong>Training Set</strong></td>
<td>The training set contains normalized images of the seven facial expressions</td>
</tr>
<tr>
<td><strong>Classification</strong></td>
<td>The input image is classified by matching the feature template against the seven facial expressions in the training set</td>
</tr>
</tbody>
</table>

**Table 1**: These steps show the facial expression recognition approach used in our research

**Figure 3**: The facial expression recognition technique used in this research

The diagram shows the process involved in facial expression recognition. Once the input image is detected, it is normalized, feature points are extracted and the resulting template is matched against the training set. Classification is done to recognize the facial expression.

**Recognized Image**
### Experimental Results for Disgust Facial Expression
Grayscale Images – Male Models

<table>
<thead>
<tr>
<th>Input Image</th>
<th>Anger</th>
<th>Disgust</th>
<th>Fear</th>
<th>Happiness</th>
<th>Neutral</th>
<th>Sadness</th>
<th>Surprise</th>
<th>Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image1</td>
<td>588.274</td>
<td>282.826</td>
<td>736.591</td>
<td>641.45</td>
<td>570.307</td>
<td>610.487</td>
<td>953.741</td>
<td>Matches</td>
</tr>
<tr>
<td>Image2</td>
<td>671.77</td>
<td>391.888</td>
<td>840.107</td>
<td>702.337</td>
<td>643.916</td>
<td>715.799</td>
<td>1029.16</td>
<td>Matches</td>
</tr>
<tr>
<td>Image3</td>
<td>707.52</td>
<td>373.836</td>
<td>832.939</td>
<td>679.12</td>
<td>662.373</td>
<td>689.664</td>
<td>1033.89</td>
<td>Matches</td>
</tr>
<tr>
<td>Image4</td>
<td>565.241</td>
<td>279.052</td>
<td>685.508</td>
<td>551.4</td>
<td>516.064</td>
<td>536.249</td>
<td>896.51</td>
<td>Matches</td>
</tr>
<tr>
<td>Image5</td>
<td>626.419</td>
<td>438.834</td>
<td>605.351</td>
<td>424.115</td>
<td>536.194</td>
<td>447.965</td>
<td>810.888</td>
<td>False Alarm</td>
</tr>
<tr>
<td>Image6</td>
<td>485.41</td>
<td>300.429</td>
<td>625.11</td>
<td>449.845</td>
<td>442.746</td>
<td>533.819</td>
<td>841.982</td>
<td>Matches</td>
</tr>
<tr>
<td>Image7</td>
<td>458.051</td>
<td>293.513</td>
<td>563.366</td>
<td>391.773</td>
<td>399.355</td>
<td>486.053</td>
<td>779.307</td>
<td>Matches</td>
</tr>
<tr>
<td>Image8</td>
<td>787.42</td>
<td>439.974</td>
<td>923.111</td>
<td>845.738</td>
<td>755.025</td>
<td>712.19</td>
<td>1060.76</td>
<td>Matches</td>
</tr>
<tr>
<td>Image9</td>
<td>601.599</td>
<td>328.846</td>
<td>734.675</td>
<td>613.035</td>
<td>548.826</td>
<td>585.717</td>
<td>925.44</td>
<td>Matches</td>
</tr>
<tr>
<td>Image10</td>
<td>467.804</td>
<td>349.293</td>
<td>632.663</td>
<td>632.302</td>
<td>434.449</td>
<td>607.6</td>
<td>867.875</td>
<td>Matches</td>
</tr>
</tbody>
</table>

Table 2: Sample results of the calculated distance for all the seven facial expressions in the training set using ten disgust input images.

![Number of Matches vs. False Alarms](image-url)

Figure 4: Each facial expression has 20 input images. The chart shows the number of input images that match the facial expressions and the number that gave false alarms.
3.4 Classification

Assuming the input image under consideration was the disgust facial expression. Matching the anger facial expression against each facial expression template will yield seven distance values. We scan through the seven distance values and choose the smallest value or shortest distance and its corresponding facial expression. Since the input image under consideration is the disgust input image, our goal is to have the shortest distance correspond to the disgust facial expression template, signifying a match and hence positive recognition. Our aim in using a minimum of 20 input images for each facial expression is to detect how well our approach recognizes facial expressions.

4 Experimental Results

Table 2 is an example of part of the results gathered for the disgust facial expression. You will notice that, for each disgust input image, the disgust image in the training set (i.e. the disgust template) had the smallest distances, except for input image 5. Ten of the fear images in the testing set yielded 100% matches in the first experiment. In the second experiment, the next ten “fear” input images yielded five false alarms. The results indicated that the distance of another facial expression in the training set was smaller than the fear facial expression. In that case, our technique could not recognize the input image, which was fear. Figure 4 shows the number of matches’ versus false alarms. The “disgust” facial expression had the fewest false alarms, followed by the “surprise” facial expression. The anger facial expression had the highest number of false alarms. The recognition rate for the seven facial expressions is shown in Figure 5.

5 Conclusions

Our goal in this research was to recognize facial expressions displayed in images or video by using adaptive template. The recognition rate for the seven facial expressions ranged from 60% to 90%. This result is an indication that adaptive templates can be used to gain good recognition results.
6 References


