A Facial Expression Classification System based on Active Shape Model and Support Vector Machine

Chen-Chiung Hsieh and Meng-Kai Jiang
Department of Computer Science and Engineering
Tatung University
Taipei, Taiwan
cchsieh@ttu.edu.tw

Abstract—Most traditional expression classification systems track facial component regions such as eyes, eyebrows, and mouth for feature extraction. This paper utilized facial components to locate dynamic facial textures such as frown lines, nose wrinkle patterns, and nasolabial folds to classify facial expressions. Adaboost using Haar-like feature and Active Shape Model (ASM) are adopted to accurately detect face and acquire important facial feature regions. Gabor filter and Laplacian of Gaussian are used to extract texture information in the acquired feature regions. These texture feature vectors represent the changes of facial texture from one expression to another expression. Support Vector Machine (SVM) [3], the obtained information is classified into six types of expressions, namely neutral, happiness, surprise, anger, disgust, and fear. Cohn-Kanade database was used to test the feasibility of proposed method and the average recognition rate reached 91.7%.

Keywords—facial expression recognition; face detection; Gabor filter; Support Vector Machine; Active Shape Model

I. INTRODUCTION

Interpersonal communication is conducted simultaneously on auditory, visual, and tactile levels. Auditory communication involves communicating through speaking and visual communication involves communicating through the written word, gestures, and facial expressions. People communicate using these different channels, all of which complement one another, to express their feelings. However, the most important form of communication is facial expression.

As artificial intelligence (AI) focuses on interaction with users, the ability of computers to detect changes in emotions and generate appropriate feedback can benefit the application and development of AI. For instance, robots that can communicate ideas to humans will become more than a tool and can also be used in medical monitoring to detect changes in facial expressions of patients for use in diagnosis and to supplement patient care procedures.

This study proposes a facial expression recognition system that captures images through a camera. Adaboost [1] and active shape model (ASM) [2] are used to identify human faces and Gabor Filter and Laplacian of Gaussian (LoG) edge detection are likewise adopted to extract facial components and texture features. Afterwards, using the support vector machine (SVM) [3], the obtained information is classified into six types of expressions, namely neutral, happiness, surprise, anger, disgust, and fear.

This paper is organized as follows. The first section discusses the objectives of the study; the second section introduces related studies on facial expression recognition; the third section describes the system architecture and implementation of the study; the fourth section presents the results and analyses of the study; and the fifth section presents the conclusions.

II. LITERATURE REVIEW

Expression recognition has been the main focus of studies on human-machine interfaces in recent years. Fasel and Luettin [4] indicated that deriving facial feature deformation and facial motion from facial images are important stages in the analysis of facial expressions. Deformation extraction can be divided into two categories: image-based approaches [5] and model-based approaches [6].

Image-based approaches process facial images and delineated images to extract characteristics and do not require additional external information. Methods include neural network [7] and Gabor wavelets [8]. Model-based approaches, on the other hand, mainly use facial models to represent facial structures and therefore can effectively avoid environmental factors and can accurately determine the facial motion and deformation of facial features. The drawbacks of a model-based approach, however, is that most feature points require the manual setting of facial models and the setting procedures are fairly complex, such as those seen in the active appearance model (AAM) and point distribution model (PDM). Facial motion extraction usually refers to optical flow [9], motion model [10], and feature tracking [11].

Ou et al. [12] proposed the use of Gabor filter and expression analysis to implement automatic facial expression identification, the main approach being the use of the shape model of the 28 feature points to position the human face and the use of 40 Gabor filters comprising five different frequencies and eight different directions to identify facial features. Due to the high values of the derived eigenvector, principal component analysis (PCA) was used to reduce data dimensions while k-nearest neighbor (KNN) was used to categorize the results into six expressions. The drawbacks of this approach, however, are that the extraction of facial features needs to be set manually and normalized and that improper settings influence the results of processing and identification.
III. SYSTEM ARCHITECTURE

Figure 1 presents the system flowchart of the study. For each input image, Adaboost is used to detect the presence of a human face. If a human face is detected by Adaboost then ASM was used to process the detected face to determine the facial shape and facial features points. The ASM feature points were then used to delineate ROIs. Setting ROIs eliminates many unnecessary regions and acts as a noise filter. Afterwards, the multidirectional Gabor filter and LoG are used to identify facial features and lineal information to extract effective feature points and calculate the feature points of facial motions from a neutral expression to a given expression. This generated eigenvector values are used to do facial expression recognition using SVM.

A. Detection of facial textures

Facial feature positions and shapes, especially those of the eyebrows and mouth, differ when people make different facial expressions. In terms of lines, the nasolabial fold, nose wrinkles, and frown lines between eyebrows are the most prominent. This study employs various wave filters to detect facial and textural features. The Gabor filter is first used to detect facial features, such as the nasolabial folds on both cheeks and LoG edge detection is used to detect the mouth, nose wrinkles, and frown lines.

To reduce the effect of noise on edge detection and the accurate detection of textural features, this study combined the use of Gaussian smoothing and Laplace filter to form the Laplace of Gaussian (LoG) wave filter.

The Gabor filter is less effective in detecting upper and lower jaw movements, so this study employed LoG edge detection to extract mouth shape (Fig. 3(a)) and nose wrinkles and frown lines (Fig. 3(b) and 3(c)).

B. Facial feature points

The inner region of the eyebrows is designated as a feature point, as can be seen in Fig. 4. Our eyebrows usually move up and down when expressing surprise or fear. Euclidean distance is derived by calculating for the distance BtoE between the average of the y-coordinate axes of the inner feature points between eyebrows (EyebrowL (x,y) and EyebrowR (x,y)) to the average of the y-coordinate axes between the edges of the person’s eyes (EyeL (x,y) and EyeR (x,y)). The eyebrows contract when people express anger so the x-coordinate distance EB of the feature point between the eyebrows is used instead, as can be seen in Fig. 5.

Equation (1) presents the equation for calculating the ratio between the pixel count and total area of the texture features between eyebrows, nose wrinkles, and nasolabial folds on both cheeks:

$$\text{Equation (1)}$$
The ratio between mouth width and facial width, on the other hand, is used as the eigenvector, as can be seen in (2) and Fig. 6. Six features have been detected to form the expression eigenvectors of the human face. These are then entered into SVM to perform facial expression recognition.

\[
\text{Face}_{\text{eigenvector}} = \frac{\sum_{i=1}^{n} \text{ROI}_{\text{area}} \sum_{j=1}^{m} \text{Pixel}_{ij}}{\text{ROI}_{\text{area}}}, \text{Pixel} \in \text{Texture Area}
\]  

(1)

The detection results are shown in Table I. The results indicate that the recognition rate for neutral expressions was 100 % and the average expression recognition rate reached 93.08 %. This study also employed the Cohn-Kanade AU-coded facial expression database [13] to reach 93.08 %. This study also employed the Cohn-Kanade AU-coded facial expression database [13] to reach 93.08 %. This study also employed the Cohn-Kanade AU-coded facial expression database [13] to reach 93.08 %.

Table I. Classification rates for facial types of expressions.

<table>
<thead>
<tr>
<th>Feature</th>
<th>T</th>
<th>N</th>
<th>H</th>
<th>S</th>
<th>A</th>
<th>D</th>
<th>F</th>
<th>RR (%)</th>
<th>ARR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Mouth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>94.5</td>
<td></td>
</tr>
<tr>
<td>Eye</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96.6</td>
<td></td>
</tr>
<tr>
<td>Nose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90.4</td>
<td></td>
</tr>
<tr>
<td>Chin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>87.9</td>
<td></td>
</tr>
<tr>
<td>Jaw</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.1</td>
<td></td>
</tr>
</tbody>
</table>

IV. EXPERIMENTAL RESULTS AND ANALYSIS

The researchers studied the facial expressions of five people between the ages of 24 and 26. Each person was photographed making six expressions; each expression was photographed 50 times, 10 of which were used as training data. The total number of images for each person was 300. Each of the participants also produced 40 training images, totaling 1,200.

The detection results are shown in Table I. The results indicate that the recognition rate for neutral expressions was 100 % and the average expression recognition rate reached 93.08 %. This study also employed the Cohn-Kanade AU-coded facial expression database [13] to confirm the research findings. This database is commonly used in facial expression recognition research. The 486 image sequences showed neutral, happy, surprised, angry, disgusted, fearful, and hurt expressions. Each sequence started with neutral expressions and ended with the target facial action code. This study retained 341 sets for use as training and testing data. The training data included 20 images of each expression and the testing data included 221 images. The detection results are shown in Table II. The results indicate that the recognition rate for neutral expressions was 100 % and the average expression recognition rate reached 91.7 %.

Table II. Recognition rates using Cohn-Kanade database.

<table>
<thead>
<tr>
<th>Feature</th>
<th>T</th>
<th>N</th>
<th>H</th>
<th>S</th>
<th>A</th>
<th>D</th>
<th>F</th>
<th>RR (%)</th>
<th>ARR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Face</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>100</td>
<td></td>
</tr>
<tr>
<td>Mouth</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>94.5</td>
<td></td>
</tr>
<tr>
<td>Eye</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>96.6</td>
<td></td>
</tr>
<tr>
<td>Nose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>90.4</td>
<td></td>
</tr>
<tr>
<td>Chin</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>87.9</td>
<td></td>
</tr>
<tr>
<td>Jaw</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>89.1</td>
<td></td>
</tr>
</tbody>
</table>

V. CONCLUSIONS

This study proposed a facial expression recognition module that not only identifies facial features but also dynamic facial textures such as frown lines, nose wrinkles, and nasolabial folds to detect changes in expression. Past studies converted whole facial images for classification or manually set feature points. While these methods produced satisfactory results, they were neither rapid nor automated.

The feature extraction method proposed in this study set feature points automatically and extracted few but crucial feature points. This process differs from that of past methods, where the entire face was used in processing. Compared to these methods, our approach is more efficient and automatic. Additionally, after extracting the images, the system also automatically set feature regions, positioned feature points, and calculated the generation of feature vectors and did not require manual labeling or selection. The results showed that the method proposed in this study can classify six human expressions effectively, namely neutral, happiness, surprise, anger, disgust, and fear.

ACKNOWLEDGMENT

Our thanks go to National Science Council, Taiwan, for supporting this project under grant no. NSC 99-2218-E-655-001 from Nov. 2010 to Oct. 2011.
REFERENCES
