RECOGNIZING FACIAL EXPRESSIONS USING ACTIVE TEXTURES WITH WRINKLES

Lijun Yin†, Sergey Royt†, Matt T. Yourst‡ and Anup Basu‡‡

† Department of Computer Science, SUNY at Binghamton, Binghamton, NY 13902 USA
‡ Department of EE&CS, MIT, Cambridge, MA 02139 USA
‡‡ Department of Computing Science, University of Alberta, Edmonton, Alberta, Canada, T6G 2H1

ABSTRACT
This paper explores the use of facial wrinkle textures for recognizing the facial expressions. Based on the observation of the wrinkles appearance and change along with performed expressions, we propose to extract the partial texture information in both the facial organ areas (e.g., eyes and mouth) and the facial wrinkle areas, and use the texture dissimilarity between the neutral expression and the active expression to extract the active texture for the expression representation. We present a novel method using multiple levels of detail to measure the active texture dissimilarity. The rate of change between levels is used as the rule for discriminating 6 types of universal expressions. The experiments on video sequences demonstrate the simplicity and efficiency of the proposed method for recognizing expressions with a 82.8% correct recognition rate.

1. INTRODUCTION

The wide range of applications in human computer interaction, telecommunication and psychological research make the facial expression analysis and recognition an active research topic [9, 2]. The conventional methods on facial expression recognition concern themselves with extracting the expression data to describe the change of the facial features, such as Action Units (AUs) defined in Facial Action Coding System (FACS) [5, 10, 4] and its extension [7]. A number of techniques were successfully exploited for facial expression recognition, including feature movement estimation by optical flow [1, 12], eigen-mesh method [8], and neural networks [10]. Although the FACS is the most successful and commonly used technique for facial expression representation and recognition, the difficulty and complexity of the AUs extraction limit its application, especially in the case of real-time demands. To our knowledge, little investigation has been conducted on the wrinkle texture analysis for the facial expression recognition. Based on our previous work on facial expression analysis and synthesis with wrinkles [13], in this paper, we extend our study of the wrinkle information to the issue of facial expression recognition. The motivation of this work is to explore the active wrinkle texture information to design a simple and efficient facial expression recognition system. It has been studied from a psychological point of view that besides the shape of mouth and eyes, there is a strong correlation between facial expressions of certain basic emotions and the appearance of wrinkles in four main areas: forehead, in between eyebrows (glabella), outer corners of eyes (crows-feet), and mouth corners (nasolabial) [6], as shown in Figure 2(a). Facial wrinkle appearance plays an important role in representing a variety of facial expressions or emotions. Therefore, it is possible to classify a facial expression by noting the presence of wrinkles in the four areas, and by coding their combination in a higher level. Facial texture consists of static area and active area. The active area (so-called active texture) is changed along with the expression change. As shown in Figure 2(b-c), 9 active textures of interest (TOI) are defined on a face region. In this paper, we proposed a facial expression recognition system based on the information of the active texture variation. The system is composed by three major components: texture of interest detection, expression data extraction and expression classification, which is outlined in Figure 1. In the first stage, a model-based tracking

![Flow chart of the system composition](image)

Figure 1: Flow chart of the system composition

and detection algorithm is applied to extract the partial textures of interest. The image showing a neutral expression (e.g., the first frame as usual) can be a reference expression for the texture comparison. In the second stage, the texture similarity between the test TOI and the reference TOI is measured to discriminate the facial expressions. The
method outputs a score suggestive of the level of similarity. Based on the correlation measurement, we proposed a rate-of-change measurement for signifying the texture dissimilarity between the test TOI and the reference TOI. In the case of facial texture, the level of similarity may change due to the change of expression, lighting condition, and poses, etc. However, none of these except for the appearance of wrinkles can occur during the process of video sequence capture if the head rotation is limited. The lighting condition may be different in capturing different subjects’ expressions, however each video sequence can be ensured to have a fixed lighting condition. Therefore, within our problem domain, low texture similarity will indicate the presence of the active wrinkles. Essentially, this method attempts to qualitatively describe how saturated a facial texture is at any time, given the baseline texture. In the third stage, the detected TOIs are classified into active textures and non-active textures by thresholding the previous obtained dissimilarity scores. We encode the expressions according to the appearance of the active textures, a so-called expression code (EC) for representing the active texture distribution is defined as a classification rule for discriminating six universal expressions. Each part of the system will be described in the following sections, followed by experimental results.

2. EXTRACTION OF TOI: BRIEF OVERVIEW

The ability to detect a major change in facial TOI areas (e.g., wrinkle areas) is a first step towards the success of our texture-based facial expression recognition system. We developed an algorithm using deformable template and energy-oriented dynamic mesh to detect the facial fiducial points as shown in Figure 2(a) (see detail in [13]) . The texture of interest can be determined by the geometric relationship of these fiducial points. For example, a nasolabial area is a polygon which is formed by the side of the nose, outside corner of the mouth, and outside corner of the eye. To reduce the computation time, the size and shape of the extracted textures are re-normalized to a standard size and shape, which is smaller than the actual size by using the geometric transformation. Figure 2(d) shows the normalized TOI in wrinkle, eyes and mouth areas. (see [13] for the detailed algorithm). Figure 2(b-c) shows examples of the extracted TOI.

3. EXTRACTION OF FACIAL EXPRESSION DATA

Although the textures of interest are extracted in each frame, only the active textures (AT), which are the typical textures representing significant facial expression, are used for classifying the facial expressions. The active texture detection conforms to the following criteria: if we view a specific TOI area along consecutive frames — for example, the nasolabial wrinkle texture — each nasolabial wrinkle texture of the successive frame is correlated with the reference frame to a certain degree. The active wrinkle generated from the different expressions can be estimated by a lower correlation between the two frames’ textures. Similarly, the active textures in the mouth and eye areas can also be estimated using the same principle.

3.1. Rate-of-change on multi-level of detail similarity

The correlation is the reflection of the two TOIs’ similarity, however, the threshold to determining the active texture is person-dependent and lighting-dependent. It is infeasible to classify the active texture in a large video database. To overcome this drawback, we propose to compute the correlation in different levels of detail, instead of one level calculation. The following simple idea prevailed: compare the two textures several times (e.g., N times) as they gradually lose detail. The rate-of-change of correlation values will reflect the dissimilarity of the two TOIs. The rate-of-change is less dependent to the imaging condition and individual’s appearance. We choose to use the Gaussian blurring operation to decrease the texture detail [11]. The intuitive assumption is that as the level of detail decreases texture similarity will rise, approaching maximum value (e.g., 100%), which is the equal similarity. Two facts are of interest to us: what the initial correlation is and how the correlation increases. High initial correlation indicates high similarity. Additionally, if the correlation has not changed by much across the levels of detail, then the two textures are very similar to begin with. We will measure the rate of change by summing the differences between levels. However, the sum is also affected by how a given set of textures changes after blurring. The value of summation depends on the detail complexity of the initial reference texture. To account for this, we introduce a scaling factor that depends on the initial and final (Nth level) standard deviations of the reference texture, which is mathematically expressed as follows:

\[
\text{Score}(x, y) = \frac{\sigma_0}{\sigma_n - \sigma_0} \times \sum_{j=1}^{N} (r_j - r_{j-1}) + d_0 + d_1
\]

\[d_0 = 1 - r_0, \quad d_1 = (1 - r_1)/2, \quad \sigma_j = \sqrt{\sum_{i=1}^{N} (x_i^j - x_j^j)^2}
\]

where \(N\) is the number of levels of detail, level 0 signifies the initial input before blurring, \(\sigma_j\) is a standard deviation of the reference texture at detail level \(j\). The correlation between the reference texture and the current frame texture at level \(j\) is:

\[
r_j = \frac{\sum_{i=1}^{N} (y_i - \bar{x}_j^i)(x_i^j - \bar{x}_j^j)}{\sqrt{\sum_{i=1}^{N}(x_i^j - \bar{x}_j^j)^2} \sqrt{\sum_{i=1}^{N}(y_i - \bar{x}_j^i)^2}}
\]
where \( n \) is the number of pixels in a texture patch, \( x_j \) is the reference texture at level \( j \) (blurred \( j \) times); \( y_j \) is the input subsequent test texture at level \( j \). The more the blurring level, the more incremental correlation between the reference texture and the test texture. Based on our intensive experiment, it is found that the correlation value has no significant increasing when the number of level is beyond 4, therefore the number of level \( N=4 \) (plus the original reference level 0) is suggested in our algorithm. Separated-kernel Gaussian blurring by using finite-state machine method [11] is employed. We apply the \( 9 \times 9 \) Gaussian blurring operator incrementally to create levels. The smoothing operation removes the texture details in both the reference texture and the test texture, and repeatedly apply the smoothing operation on the previous blurred texture result. Figure 4(A) shows the examples of the dissimilarity scores obtained from three persons' forehead regions with different lighting conditions and different texture details. The dissimilarity scores between the active wrinkle textures and the non-active textures have significant difference, they are more distinguishable than the correlation values, a simple thresholding operation can achieve the task to detect the active wrinkle textures. Figure 4(B) shows examples of dissimilarity scores obtained from all other texture areas. In order to select the threshold values for the dissimilarity score, we took a set of video sequences as the training set, which consist of six universal expressions performed by 20 subjects. The nine areas of TOIs are extracted for measuring the dissimilarity in all frames. The dissimilarity scores for each TOI area are classified into two classes by the K-means classifier, the bisector value between the two classes is selected as the threshold for classifying the active texture. Table 1 shows the thresholds obtained for all TOI areas.

### 4. Expression Classification Rule

After the active textures are determined by the dissimilarity score, the TOI patches are labeled by 1 for the active texture patch, and 0 for the non-active texture patch. A 9-bits expression code is then created for representing the expression. Although \( 2^9 = 512 \) possible expressions can be represented in total, we only use six universal expressions in our current study. There may be a wide range of extents of these expressions, but all variations are associated with the three facial regions (eyes, eyebrows, and mouth) and their associated facial wrinkles. Table 2 shows the expression codes as the classification rules for the six universal expressions. For example, the light smile expression is represented as: 000000001 (smile-1), only the mouth corner has a small movement; The normal smile is coded as 000011001.
(smile-2), the significant texture change will occur in the nasolabial area and the mouth area. The big laughing is coded as 011111111 (smile-3), the significant active texture (wrinkles) will appear in all the areas except the forehead area. The surprise has forehead wrinkle appeared and eye raised, with mouth either open (100000110) or closed(100000111). Based on the classification rule, the expression recognition is simply realized by a look-up table in our implementation, given the active textures being detected.

<table>
<thead>
<tr>
<th>EC</th>
<th>b8</th>
<th>b7</th>
<th>b6</th>
<th>b5</th>
<th>b4</th>
<th>b3</th>
<th>b2</th>
<th>b1</th>
<th>b0</th>
</tr>
</thead>
<tbody>
<tr>
<td>smile-1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>smile-2</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>surprise</td>
<td>0</td>
<td>x</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>x</td>
</tr>
<tr>
<td>disgust</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>x</td>
<td>1</td>
<td>1</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>anger</td>
<td>0</td>
<td>x</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>sad</td>
<td>x</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>fear</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>neutral</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
</tbody>
</table>

Table 2: Expression codes (EC) of universal expressions plus neutral expression. From left to right (b8 - b0): forehead, glabella, crow-feet(L), crow-feet(R), nasolabial(L), nasolabial(R), eye(L), eye(R) and mouth. 1 - appeared; 0 - not appeared; x - either 0 or 1; \( \overline{x} \) - the complement of x.

5. EXPERIMENTS

Two sets of video sequences showing the subjects’ head-shoulder scene in total of 40 video sequences are captured. First 20 sequences are used for training the threshold values. Second 20 sequences are used for the test. In each sequence, the individual subject performed the expressions twice for each of six expressions, with the transition of neutral expression. The image size of each frame is 512*460 pixels (taken at 10 frames per second, 150 frames in total for each sequence). The Figure 3 shows one example clipped video sequence, in which the active textures are detected in the nine local facial areas. The dissimilarity values above the thresholds signify the appearance of the active textures. Tested on 200 image sequences displaying 6 types of universal expressions with total 3000 frames, the system achieved an average correct recognition rate of 82.8%. Table 3 shows the details of our experimental results. Note that due to the eye blinking, it will confuse the system to distinguish the performed expression and neutral expression. In order to reduce this affection, we will check every 3 consecutive frames in the sequence, the expression is said to be determined only in the condition of the 3 consistent expression codes being found. The recognition results are encouraging in view of such a condition that we use only nine active areas to encode the expression, our representation range is much smaller than the FACS system, where 46 AUs are required. The possible extension of this work is to use the dissimilarity score instead of the binary values for classifying more rich range of fine expressions.

Table 3: Correct expression recognition rate (RR). The controlled expressions are performed by the performers as follows: each expression is performed twice, the transition in between any two expressions is the neutral expression. 20 individuals create 40 expressions for each of six universal expressions, and 260 neutral expressions.

<table>
<thead>
<tr>
<th>Expression</th>
<th>RR (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Smile</td>
<td>92.5</td>
</tr>
<tr>
<td>Surprise</td>
<td>87.5</td>
</tr>
<tr>
<td>Sad</td>
<td>67.5</td>
</tr>
<tr>
<td>Fear</td>
<td>70.0</td>
</tr>
<tr>
<td>Happy</td>
<td>82.5</td>
</tr>
<tr>
<td>Neutral</td>
<td>80.0</td>
</tr>
</tbody>
</table>

Average RR: 82.8%

6. CONCLUSION

We presented a new scheme to recognizing facial expressions based on the appearance of the active textures including wrinkles and the organs’ textures. This work can be the first step for the coarse classification of the expressions. The fine classification will be investigated as the second step by taking the probability of the appearance of the active textures into account. In the future, the dissimilarity score can be the input to design the classifier to deal with the large number of fine expressions recognition, for example, by using the Bayesian recognition framework [3].

7. REFERENCES