Facial Expressions Using A Quadratic Deformation Model: Analysis and Synthesis

C. Tay*, M. Obaid†§, R. Mukundan†, A. Bainbridge-Smith*
†Computer Science and Software Engineering Department, University of Canterbury, Christchurch, New Zealand
§HITLab New Zealand, University of Canterbury, Christchurch, New Zealand
*Electrical and Computer Engineering Department, University of Canterbury, Christchurch, New Zealand
cht27@student.canterbury.ac.nz, {mohammad.obaid, mukund, andrew.bainbridge-smith}@canterbury.ac.nz

Abstract - In this paper we propose a novel method to generate facial expressions on two-dimensional sketch-models and images. The six facial expressions (smile, sad, angry, disgust, fear, and surprise) are represented using a set of quadratic deformation parameters defined on muscle-based regions. Quadratic deformation parameters are obtained in a preprocessing stage when motion capture data for several points on a human face is used. A general algorithm for mapping the transformation parameters onto two-dimensional images to generate the corresponding expressions is developed. We give the implementation aspects related to expression generation, and describe various stages involved in the process. Two user case studies, image-based and video-based, have been taken to study the performance of the developed method. Among 30 participants, 52.96% can correctly identify the image-based static expressions and 58.67% for video-based animated expressions. The correct response for facial expressions improves when the process of transformation is showing to the audience. Several limitations of the project are indentified and discussed; however, the overall result is still acceptable. Further development directions have been suggested for future research.

Keywords – Facial Expressions Representations, Quadratic Deformation Model, Image Recovery, Facial Grouping Method.

I. INTRODUCTION

The rapid growth of computer graphics hardware has accelerated the interest and activity in the research of facial expression representations for synthesising and animating faces, as the hardware is able to realise the required computer graphics algorithms. Facial expressions play an important role in many application areas including, virtual character animation, video conferencing, human-computer interaction and games. One of the main challenges in representing facial expressions is to be able to give a general model for generating different expressions. Careful measurement and construction for representing facial expressions are necessary for the high sensitivity in facial identification by human being. This paper consolidates the research developed by Obaid et. al. [1] by extending the proposed concept and implementing various methods for representing facial expression.

The structure of this paper is as follows. Section II describes the previous research and background related to the area of facial expression representations. Section III describes the method to generate facial expressions. Section IV discusses the primary results and outlines the advantages and the limitations of the method. Finally, section V concludes the paper and suggests future research directions.

II. BACKGROUND

The work presented by Obaid et al. [1] proposes a novel approach for representing facial expressions based on a quadratic deformation model applied to muscle regions. They subdivided the face into 16 facial regions to capture the non-linear nature of muscle deformations using the most general rubber-sheet transformation of second degree. They derived a set deformation parameters using a least-square minimization technique and used it to construct a Facial Deformation Tables (FDTs) to mathematically represent each of the main expression (smile, sad, anger, disgust, surprise, fear). The generalized nature of the transformations allows the mapping from one facial expression to another. The facial regions themselves where defined using the anatomy of the facial muscles [10] and the Facial Action Coding (FACS) description of facial expressions by Ekman et. al. [2] In the FACS system, facial expressions are modelled using a combination of 46 facial Action Units (AU) for the muscle movements.

Another commonly used standard is the MPEG-4 facial animation standard [3] which supports the definition, encoding, transmission of facial animation. The standard defines 68 Facial Animation Parameters (FAPs) to represent the main facial features. FAPs are represented in terms of Facial Animation Parameter Units (FAPU), which are derived based on distance ratios between key feature points.

In our work, we combine the Facial Deformation Tables (FDTs), developed in [1], with a new method that uses facial ratios based on the MPEG-4 FAPU to automatically generate facial expressions on 2-definition images or models.

III. METHOD

The method to simulate and synthesis facial expressions consist of five main steps: (1) automatically define the facial muscle regions, (2) point inclusion test, (3) facial transformation, (4) image synthesis, and (5) smoothing the
output image. The following sections describe the steps in more detail.

A. Define Muscle Regions

The facial deformation parameters are unique for each muscle region; in other word, every region has its specific deformation characteristics. Figure 1 shows the automatic muscle groupings, which imitates the muscle region described in [1]. The algorithm for automatic muscle grouping method uses ratio definitions, which is similar to the FAPU. The muscle boundary regions are approximately based on the distance between the eye pupils and lips as shown in Figure 2. The coordinates of two eye pupils give the origin of the face; symmetrical points of Lip A and Lip B depict the outline of the lips. Various studies have indicated human face is not symmetric; however, the difference is insignificant in this paper. Therefore the grouping method can still approximate the facial region.

The automatic grouping method uses the divine proportion of the human face to generate the outline of the face, allowing the facial regions to be effectively specified by six facial feature points as shown in Figure 2. This is very effective for a neutral face.

As an example, Figure 3 shows the details of one side of a face in which we describe the muscle boundary points for the automatic muscle grouping method. Table I. summaries the details of these boundary points and full details can be found in [4]

![Figure 1: defined by auto-ratio definition.](image)

![Figure 2: Principal feature points.](image)

![Figure 3: The facial boundary points for the right side of the face](image)

<table>
<thead>
<tr>
<th>Point</th>
<th>x</th>
<th>y</th>
</tr>
</thead>
<tbody>
<tr>
<td>P1</td>
<td>origin x - k</td>
<td>origin y + 2k</td>
</tr>
<tr>
<td>P2</td>
<td>origin x</td>
<td>origin y + 2k</td>
</tr>
<tr>
<td>P3</td>
<td>P17 x</td>
<td>origin y + k</td>
</tr>
<tr>
<td>P4</td>
<td>origin x - 0.75k</td>
<td>origin y + 0.75k</td>
</tr>
<tr>
<td>P5</td>
<td>P17 x</td>
<td>P7 y</td>
</tr>
<tr>
<td>P6</td>
<td>2(P1 x) - P7 x</td>
<td>P7 y</td>
</tr>
<tr>
<td>P7</td>
<td>origin x - 0.75k</td>
<td>origin y + 0.75r</td>
</tr>
<tr>
<td>P8</td>
<td>origin x</td>
<td>P7 y</td>
</tr>
<tr>
<td>P9</td>
<td>2(P1 x) - P10 x</td>
<td>origin y</td>
</tr>
<tr>
<td>P10</td>
<td>(P1 x + origin x) / 2</td>
<td>origin y</td>
</tr>
<tr>
<td>P11</td>
<td>origin x</td>
<td>origin y</td>
</tr>
<tr>
<td>P12</td>
<td>P10 x</td>
<td>origin y - 0.75r</td>
</tr>
<tr>
<td>P13</td>
<td>P6 x</td>
<td>origin y - 1.5r</td>
</tr>
<tr>
<td>P14</td>
<td>P7 x</td>
<td>origin y - 1.5r</td>
</tr>
<tr>
<td>P15</td>
<td>P17 x</td>
<td>origin y - 0.75k</td>
</tr>
<tr>
<td>P16</td>
<td>origin x - (origin x - lipB x) / (\Phi)</td>
<td>origin y - 0.75k</td>
</tr>
<tr>
<td>P17</td>
<td>origin x - 2k</td>
<td>lipB y + 0.5k</td>
</tr>
<tr>
<td>P18</td>
<td>origin x + ((\Phi - 1))k</td>
<td>origin y - ((\Phi * k))</td>
</tr>
<tr>
<td>P19</td>
<td>0.9(P17 x) + 0.1(P23 x)</td>
<td>lipA y</td>
</tr>
<tr>
<td>P20</td>
<td>lipA x</td>
<td>lipA y</td>
</tr>
<tr>
<td>P21</td>
<td>lipB x</td>
<td>lipB y</td>
</tr>
<tr>
<td>P22</td>
<td>origin x - 0.375k</td>
<td>2(lipA y) - P18 y</td>
</tr>
<tr>
<td>P23</td>
<td>(lipA x + lipB x) / 2</td>
<td>2(P22 y) - P18 y</td>
</tr>
</tbody>
</table>
The following list described the variables used in Table 1:

1. The *origin* is the centre of the face which is aligned with the eye pupils.
2. Scalar $k$ is the distance from the eye pupils to the origin in horizontal direction.
3. Scalar $r$ is the radius of the eye pupil.
4. The Facial feature points *lipA* and *lipB* are described in Figure 2.
5. The ratio $\Phi = 1.618$, which we will call the golden ratio.

The automatic grouping method defines the boundary points based on the origin, Lip A, and Lip B feature points. Variables $k$, $r$, and $\Phi$ play an important role in this method; they are defined to enclose the facial feature points and suit the outline of the face. The asterisks in Table 1 are adjustable feature points. This design ensures the algorithm can be used for any natural facial image.

**B. Point inclusion test**

The muscle regions described in Figure 1 are polynomial shapes. Hence, the point inclusion test is suggested for finding if a pixel point is located within the enclosed boundary [5]. This method gives accurate result, albeit for a long computational time. To optimise the performance of assigning the pixel points to their corresponding muscle regions, the axis aligned bounding boxes is [6] introduced before the point inclusion test. A pixel point may have more than one corresponding muscle region, which means that this point falls at the edge or the corner of more than one region.

**C. Facial Transformation**

The initial step to transforming the facial regions is to normalise the facial image. The normalisation process places the centre of the right pupil is normalised at (-1, 0), the centre of the left pupil to (1, 0) and the origin position to (0, 0). The quadratic deformation models are in normalised coordinates and therefore every pixel within the boundary regions need to be normalised before the transformation process is carried out using equations (1) and (2).

\[
\begin{align*}
    x'_i &= a_1 x_i^2 + a_2 x_i y_i + a_3 y_i^2 + a_4 x_i + a_5 y_i + a_6 \\
    y'_i &= b_1 x_i^2 + b_2 x_i y_i + b_3 y_i^2 + b_4 x_i + b_5 y_i + b_6
\end{align*}
\] (1)

The reverse normalisation is then carried out to retrieve the final facial expression. Since every pixel is treated as a feature point, the transformation process simply shifts all of the pixels to a new position representing the facial expression. This introduces gaps in the output as it is shown in Figure 4. A smoothing algorithm is applied to overcome this problem as described in section III (E).

**D. Image Synthesis**

1) **Defining the oral cavity**

The shape for the oral cavity of an opened mouth can be calculated from the known lips positions. Transforming the centre point of the lips with quadratic deformation models returns the middle position of the open mouth. The height of the vertical lips distance can be calculated by comparing the upper and lower boundary of the lips (Figure 5).

2) **Adding Extra Features**

The implementation for constructing the missing facial features uses the divine proportion, described in [4] to approximate the width of the upper teeth. In [4] the central incisors, lateral incisors, canines, and the first and second premolars define the acceptable oral cavity details using equations (3) to (5).

\[
\begin{align*}
    width &= k_1 + \Phi k_4 = k_1 + k_2 \\
    k_5 &= k_3 + \Phi k_3 = k_3 + k_4 + k_5 \\
    k_5 &= \Phi k_4
\end{align*}
\] (3-5)

where $width$ is the distance between the centre of the eye pupil and the centre of the mouth, $k_4$ is the width of incisor, $k_5$ is the width of lateral, and $k_3$ is the width of canine. The following describe the function $\text{feature}(x, y)$, which defines the content of the oral cavity (Figure 6).
where, \( x \) and \( y \) are the coordinates of the gap pixel, \( \mu(x) \) is the upper boundary of the lips, \( \alpha(x) \) describes the parabola shape of the gum, \( \beta \) is the edge of the incisor, \( m \) is the horizontal centre line of the mouth, \( \gamma(x) \) describes height of the lower teeth which has a constant height above the lower lips boundary, \( l(x) \).

\[
\text{feature}(x,y) = \begin{cases} 
\text{gum}, & \alpha(x) < y \leq \mu(x) \\
\text{upper teeth}, & \beta < y \leq \alpha(x) \\
\text{cavity}, & m < y \leq \beta \\
\text{tongue}, & \gamma(x) < y \leq m \\
\text{lower teeth}, & l(x) < y \leq \gamma(x)
\end{cases}
\]  

(6)

**IV. RESULTS AND DISCUSSION**

Figure 7. Generated facial expressions from a sketched image (a) angry (b) fear (c) smile (d) disgust (e) sad (f) surprise

The Quadratic Deformation Models have shown that such approach is proven for representing facial expressions. Figure 7 shows the simulated facial expressions from a natural sketched image. Figure 8 shows the result of generated facial expressions with a colour photo.

To test the developed method for generating facial expression, two separate user studies were carried out: using still image frames, and a short movie sequence.

Figure 8. Generated facial expressions from a colour photo (a) angry (b) fear (c) smile (d) disgust (e) sad (f) surprise

\[
r_u = \frac{y - \text{down}}{d_y}
\]  

(7)

\[
r_i = \frac{\text{right} - x}{d_x}
\]  

(8)

\[
c_h = r_i c_i + (1 - r_i) c_v
\]  

(9)

\[
c_v = r_u c_u + (1 - r_u) c_v
\]  

(10)

\[
c = \frac{c_h + c_v}{2}
\]  

(11)

where \( x \) and \( y \) are the coordinates of the pixel, \( \text{down} \) is the bottom index position with colour, \( \text{right} \) is the closest right index with colour, \( d_y \) is the vertical distance between indexes with colour, \( d_x \) is the horizontal distance from the blank pixel to the neighbouring coloured pixel, \( r_u \) and \( r_i \) are the ratio distances to the upper and the right neighbours.

The symbol \( c \) is the average colour of the horizontal and vertical colours, in this case \( c_h \) and \( c_v \) are the average horizontal and vertical colour, where \( c_l \) is the colour of the left neighbour, \( c_r \) is the colour of the right neighbour, \( c_t \) is the colour of top colour, and \( c_b \) is the colour of the bottom neighbour.
The user studies were taken at different intervals and with different group of people. 30 users participated in the imaged-based study and 25 users participated in the video-based study. In both studies the user was shown a series of facial expressions, and they were asked to select the best corresponding expression from the provided six expressions.

<table>
<thead>
<tr>
<th>Correct answer</th>
<th>smile</th>
<th>sad</th>
<th>angry</th>
<th>disgust</th>
<th>fear</th>
<th>surprise</th>
</tr>
</thead>
<tbody>
<tr>
<td>smile</td>
<td>91.7%</td>
<td>1.7%</td>
<td>0.8%</td>
<td>1.7%</td>
<td>1.7%</td>
<td>2.5%</td>
</tr>
<tr>
<td>sad</td>
<td>0.8%</td>
<td>52.5%</td>
<td>24.2%</td>
<td>13.3%</td>
<td>7.5%</td>
<td>1.7%</td>
</tr>
<tr>
<td>angry</td>
<td>0.0%</td>
<td>23.3%</td>
<td>41.7%</td>
<td>30.8%</td>
<td>4.2%</td>
<td>0.0%</td>
</tr>
<tr>
<td>disgust</td>
<td>0.8%</td>
<td>6.7%</td>
<td>45.0%</td>
<td>26.7%</td>
<td>6.7%</td>
<td>14.2%</td>
</tr>
<tr>
<td>fear</td>
<td>3.3%</td>
<td>10.0%</td>
<td>4.2%</td>
<td>3.3%</td>
<td>30.8%</td>
<td>48.3%</td>
</tr>
<tr>
<td>surprise</td>
<td>2.5%</td>
<td>3.3%</td>
<td>1.7%</td>
<td>0.8%</td>
<td>17.5%</td>
<td>74.2%</td>
</tr>
</tbody>
</table>

The results for imaged-based survey are shown in Table II. Users were asked to complete an online questioner which has 24 different sketched facial expressions, where the expressions were a mixture of the six expressions with different exaggeration levels. An example of exaggerated facial expressions can be seen in Figure 10, where the level of expression is increased by interpolating the parameters from 0.1 to 2. The results, from Table II show that the overall rate of recognising the facial expressions is 52.92%. In addition, from Figure 9, we found that that recognising expressions improves from 48.33% to 62.22% when expressions are exaggerated. The positive results obtained from the user study show that facial expressions can be generated using Quadratic Deformation Models.

The results of the video-based user study are shown in Table III, the majority participants found that the smile expression was easy to be identified. 74% of the users can tell the surprise expression, while 20% thought it was a fear expression. On the other hand, more than half of the users believed that the given fear expression was a surprised expression. In addition, 42% answered disgust when an angry expression was showing to them, while only 40% recognised it. 64% users selected the correct answer for the sad expression when 16% believed it is angry expression.

Comparing the result of video-based and image-based user studies, we can clearly see that there are some improvements in recognising the smile, sad, and disgust expressions. This suggests that facial expression generated by Quadratic Deformation model can be identified easier when the intermediate transformations is showing to the participants. The overall recognition results show that 52.92% of participants recognised the expression in the image based study; while, this result has improved to 58.67% in the video-based user study.

The disgust expression has improved significantly in the video-based study. The result for the surprise expression remains the same for both video-based and image-based studies. Comparing the imaged-based and video-based results, the correct respond for fear expression decreases from 31% to 22%. This can be explained because the number of users who answered for surprise has increased.
from 48% to 66%. In other words, the difference between fear and surprise expression is difficult to distinguish by the users.

According to previous studies in [8] and [9] the results show that interpreting facial expressions may differ for participants who come from a different cultural background. From the study, East Asian participants focus mainly on the eyes, but Westerners scanned the whole face when reading an expression. The study also points out that Asian were more likely to read the fear expression as surprise and disgust expression as anger. The majority of the participants in both studies were Asian. Thus, it can be argued that the chosen answers by the participants for fear and surprise, and for disgust and anger have higher rates than any other expression options. Further investigation is needed to argue such statement is true for the developed transformation model.

It has been noticed that some of the expressions are hard to distinguish by the majority of users during the user studies. The quadratic models were developed and given in [1] by asking several participants to mimic the expressions. The parameters are generated by comparing the neutral face and the expression face with least square minimization method. Although the collected parameter values have been averaged to provide a general model, some of the participants may have introduced some false results during their acting of the expressions. Hence the collected parameters can be slightly ambiguous for some expression models especially for the fear expression. Introducing professional performers for the training phase may improve the quality of the Quadratic Deformation Parameters.

As it is noticed in Figure 10, distortion is formed between the nose and the upper lip. This is a consequence of each muscle region having unique transformation characteristics, and therefore significantly different deformations either side of the region boundaries. This can lead to overlapping and holes. Further work is required in this area.

V. CONCLUSION AND FUTURE WORK

The project has shown that the Quadratic Deformation Model representations of facial expressions, defined in [1] is a valid approach to synthesis and simulate facial expressions. We described a method to synthesis facial images and gave its implementations aspects. To evaluate the performance of the developed method, we conducted two user studies, image-based and video-based. The general respond from the public shows that the developed method is acceptable for generating facial expressions. Overall, the correctness for identifying facial expressions for image-based is 52.96% and 58.67% for video-based studies. Although the number of collected data survey is small, the quality of the data is reasonable for this project. Equivocal collected mimicked data source maybe the main reason which causes some facial expressions to be undistinguishable. Further developments for improving the performance of the program are considered.

The following lists several considered future directions:

- Taking a more proper user study with more accurate measurement
- Advancing the auto muscle grouping method
- Minimising the number of quadratic parameters
- Preventing overlaps between transformed muscle regions
- Improving the additional extra facial feature

REFERENCES