Keyboard Encoding of Facial Expressions

Nicoletta Adamo-Villani  
Department of Computer Graphics  
Technology, Purdue University  
nvillani@tech.purdue.edu

Gerardo Beni  
Department of Electrical Engineering,  
University of California, Riverside  
beni@ee.ucr.edu

Abstract

The paper focuses on the development of a human-computer communication method which utilizes the user’s typing skills to control the facial expression of a computer generated three-dimensional face. The method is based on the realization that the human face is a movable and deformable system with 26 degrees of freedom (the same number as the letters of the English alphabet). Therefore it is possible to create a parameterized graphical facial model confined to a set of 26 parameters each one controlled by a letter key. The method is an extension of the KUI technique [1][2][3] recently developed to encode hand gestures.

1. Introduction

The object of the paper is the demonstration that it is possible to create a simple human-computer communication of messages encoding facial expressions. The idea is based on the realization that any significant facial articulation can be represented with a set of 26 parameters, the same number as the letters of the English alphabet. Since touch typing is an easily acquired and widespread skill, it is possible to conveniently input messages encoding facial configurations if each letter key of the keyboard corresponds to one degree of freedom of the face.

Modeling and animation of the human face has been object of extensive research in the past thirty years. The first work in developing facial models was carried out in the early 70’s by Parke [4][5][6][7] who developed the first interpolated and the first parametric 3D facial models. Gillenson [8] developed the first interactive two dimensional face models and Platt and Badler [9][10][11] proposed the first muscle action based facial models. All these models made use of the Facial Action Coding System (FACS) [12][13] as the basis for facial configuration control.

In the last few years there has been considerable activity in the development of new facial models. Waters and Terzopoulos [14][15][16][17][18][19] proposed a series of physically based pseudo-muscle driven facial models. Magnenat-Thalmann, Primeau and Thalmann [20] presented the Abstract Muscle Action model and Nahas, Huitric and Sanintourens [21] developed a face model using B-spline surfaces.

Currently facial surfaces are controlled and manipulated using one of three basic techniques: (1) 3D surface interpolation; (2) ad hoc surface parameterization; (3) physically based techniques with pseudo-muscles.

In terms of human computer interaction (HCI), the emphasis of facial expression research has been on computer vision techniques for facial configuration input and processing, and categorization of facial expressions relevant to enhancing communication between man and machine [22][23][24].

In this paper we demonstrate that the human face, which consists of 44 bilaterally symmetrical muscles (muscles of facial expression and muscles of mastication [25] ), can be modeled with muscle (or group of muscles) actions totaling 22 degrees of freedom + 4 degrees of freedom required to control the direction of the gaze. Thus, it is possible to create a facial model confined to a parameter space not excessively large in terms, not only of computer representation, but also of human encoding. It is this characteristic that has suggested the approach to facial expression encoding described below.

A convenient facial configuration encoding is applicable to many practical tasks: (1) teaching the facial components (non-manual markers) of American Sign Language. The 26 facial parameter set could be easily optimized for keyboard encoding of facial expressions specific to the grammar of ASL; (2) Human Computer Interface, i.e., the possibility of building computer interfaces which understand and respond to the complexity of the information conveyed by the human face. Currently information has been conveyed from the computer to the user mainly textually or visually via ad hoc images; (3) Testing and quantitative calibration of vision algorithms for the analysis and recognition of video data involving faces; (4) Communication with patients suffering from textually impaired syndromes, e.g., severe dyslexia; (5) Development of socially adept interfaces for the communication of social displays in the acknowledgement of actions by other people, e.g., by smiling in response to intention to purchase a certain
item; (6) web deliverable 3D character animation. A simple set of (26) component vectors would represent a facial configuration and could be transmitted with very low bandwidth to animate complex face models held at the receiver site.

2. Methods

2.1. Determination of the 26 facial parameters

The human face is a complex structure of muscles whose movements pull the skin, temporarily distorting the shape of the eyes, brows, and lips, and the appearance of folds, furrows and bulges in different areas of the skin [24]. Such muscle movements result in the production of rapid facial signals (facial expressions) which convey four types of messages: (1) emotions; (2) emblems - symbolic communicators, culture-specific (e.g., the wink); (3) manipulators - manipulative associated movements (e.g., lip-biting); (4) illustrators - movements that accompany and emphasize speech (e.g., a raised brow)[12].

Given the complexity of the human face, the first challenge faced by this research has been the determination of a relatively small set of facial parameters (26) able to encode any significant facial expression of a 3-dimensional computer generated face. There are several approaches to developing facial parameters including observation of the surface properties of the face and study of the underlying structure, or facial anatomy. However, which parameters are best included in a simple model of facial expression remains unresolved [26]. Below we propose a new set of parameters.

The eyes and mouth are of primary importance in facial expressions thus many of our facial parameters relate to these areas. We have modeled a 3-dimensional face as a continuous polygonal mesh and we have identified 22 regions on the mesh. The definition of the regions is based on the anatomy of the face and in particular on the location of the muscles of Facial Expression [23]. Figure 1 shows the 22 regions identified on the 3D face model. Each region of mesh is controlled by one joint. The translation of the joint produces a proportional (quasi linear) deformation of the corresponding skin region. The type of deformation has been determined based on the observation of the change of facial shape produced by the action of the muscle or set of muscles located in that region. The Facial Action Coding System (FACS) [12][13] has provided us with a complete list of possible muscle contractions or relaxations performable on a human face with relative induced deformations. FACS lists all the basic actions (called Action Units or AUs) that can occur on the human face (e.g., Inner Brow Raiser, Lip Tightener, Chin Raiser) and describes a facial expression as a combination of specific AUs. In the next section we demonstrate that it is possible to encode, via keyboard input, 94% of the FACS Action Units (we have not considered those relative to head orientation) with our 26 (22+4) parameter set.

2.2. Description of the coding

Using MEL (Maya Encrypted Language) we have created a program that encodes the facial expression of the above described three-dimensional face by mapping each letter key of the keyboard to a degree of freedom of the face (lower case letters induce positive translations of the joints and positive rotations of the eyes, upper case letters induce negative translations of the joints and negative rotations of the eyes). Figure 2 shows the locations of the 22 facial joints, on the left, and a rendered image of the face with the joints’ transformation (22) and eye rotation parameters (4) mapped to the 26 letters of the alphabet, on the right. Table 1 shows the deformation output produced by each letter key. We note that letter ‘z’ controls the deformation induced by the mentalis muscle as well as the rotation of the jaw.
Via keyboard input the face can be configured to attain any expression: by touching a letter key the user translates the corresponding joint a pre-specified number of units along an axis. The letters “G H I J” control the rotation of the eyes and therefore the direction of the gaze. The eyes have been modeled as two separate spheres with procedural mapped pupils. The rotation of each sphere around the Y axis causes the eye to look left or right; the rotation of each sphere around the X axis causes the eye to look up or down. The transformation “step” induced by each key touch can be changed to increase or decrease precision. Figure 3 shows an example of keyboard encoding of the six basic facial expressions commonly used in animation (anger, joy, surprise, fear, disgust, sadness), table 2 shows the letter codes corresponding to each expression.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Letter Code</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sadness</td>
<td>AABBccddklZPPQWWXYtuvv</td>
</tr>
<tr>
<td>Joy</td>
<td>pppqqrrrrsstttuuuvvZZWWXXYYYmoo klcddabbb</td>
</tr>
<tr>
<td>Fear</td>
<td>KLefccccddAAABBPPQQQZZZZWWW XXXYYYYYtu</td>
</tr>
<tr>
<td>Surprise</td>
<td>KLefccccddAAABBPPQZZWWXXYYYYRStuvv</td>
</tr>
<tr>
<td>Disgust</td>
<td>CCCDDDDAAABBklllttttuvvuvvRRSSZY XWno</td>
</tr>
<tr>
<td>Anger</td>
<td>CCCDDDKiZZZWWWXXYYYYYtt uuvvab</td>
</tr>
</tbody>
</table>

Table 2. Keyboard encoding of the six basic facial expressions.

Table 3 shows the keyboard encoding of the Action Units of the Facial Action Coding System (the AUs relative to head orientation are not included). In the example below the eyes rotation is quantized in steps of 5 degrees and the joints translation is quantized in steps of 0.15 units.

3. Discussion of Results

The keyboard encoding method presents several advantages including: (1) Simplicity of user input requiring no additional input hardware (e.g. video cameras or motion capture devices); (2) Familiarity of the input method which requires no additional skills or learning time; (3) Accuracy: although the method uses a discretized representation of joints translation and eye rotation, the resolution of the quantization can be adjusted to configure the face with high precision; (4) Low bandwidth for storage and transmission: facial configuration/animation data can be stored in text files of minimum size, exported cross platform or transmitted via internet; (5) Easy extension to voice input.

There are some limitations to the method presented here. The first limitation is the restriction to a particular facial skeletal structure. While the method is applicable to any polygonal facial model rigged with a 22-joint skeleton, we have left to future developments the extension of the method to different facial skeletal setups.

Another limitation is the fact that the 22 regions discussed above, with relative deformations, need to be manually specified when the face is constructed. Future work involves the implementation of a method of automatically applying the 22 regions with relative deformations to any polygonal facial model. Such method would involve the development of a categorization of face models based on geometrical characteristics and skeletal structures.
Table 3. Keyboard encoding of the FACS Action Units

Another limitation so far is the restriction to a static head and face. Although the model of the head can be dynamic while retaining the encoded facial expression, other expressions obtainable by re-orientation of the head are not included in the method. The motion/inclination of the head also conveys emotions, feelings and meaning.

The extension to include this motion in the interface is straightforward and will be considered in a future publication where keyboard encoding of facial expressions, hand gestures and body motions are considered.

<table>
<thead>
<tr>
<th>AU</th>
<th>Description</th>
<th>Code</th>
<th>Image</th>
<th>AU</th>
<th>Description</th>
<th>Code</th>
<th>Image</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Inner Brow Raiser</td>
<td>cddAB</td>
<td><img src="1" alt="Image" /></td>
<td>20</td>
<td>Lip Stretcher</td>
<td>PQrsstuvvwwxxyyyyy</td>
<td><img src="2" alt="Image" /></td>
</tr>
<tr>
<td>2</td>
<td>Outer Brow Raiser</td>
<td>aabbbCDEF</td>
<td><img src="3" alt="Image" /></td>
<td>22</td>
<td>Lip Funneler</td>
<td>PQQRRSStuvvvwwYYY</td>
<td><img src="4" alt="Image" /></td>
</tr>
<tr>
<td>4</td>
<td>Brow Lowerer</td>
<td>CCDD</td>
<td><img src="5" alt="Image" /></td>
<td>23</td>
<td>Lip Tightener</td>
<td>PQQRRSSTUvwxwy</td>
<td><img src="6" alt="Image" /></td>
</tr>
<tr>
<td>5</td>
<td>Upper Lid Raiser</td>
<td>cdef</td>
<td><img src="7" alt="Image" /></td>
<td>24</td>
<td>Lip Pressor</td>
<td>PQQQQQRSSSTTUUvVwxx</td>
<td><img src="8" alt="Image" /></td>
</tr>
<tr>
<td>6</td>
<td>Cheek Raiser</td>
<td>CDEFlkllnn</td>
<td><img src="9" alt="Image" /></td>
<td>25</td>
<td>Lips part</td>
<td>PQtvvYZZZ</td>
<td><img src="10" alt="Image" /></td>
</tr>
<tr>
<td>7</td>
<td>Lid Tightener</td>
<td>CDEFlk</td>
<td><img src="11" alt="Image" /></td>
<td>26</td>
<td>Jaw Drop</td>
<td>PPQQQTuvvwWWXYYWYYZZ</td>
<td><img src="12" alt="Image" /></td>
</tr>
<tr>
<td>9</td>
<td>Nose wrinkler</td>
<td>Unable to encode</td>
<td><img src="13" alt="Image" /></td>
<td>27</td>
<td>Mouth Stretch</td>
<td>ZZZZZZZYYYYYYWVWXXXvvttvyy</td>
<td><img src="14" alt="Image" /></td>
</tr>
<tr>
<td>10</td>
<td>Upper Lip Raiser</td>
<td>notttuuvvwwxxyyyy</td>
<td><img src="15" alt="Image" /></td>
<td>28</td>
<td>Lip Suck</td>
<td>PQrsSTUVwxy</td>
<td><img src="16" alt="Image" /></td>
</tr>
<tr>
<td>11</td>
<td>Nasolabial Deepener</td>
<td>PPQQrswwxxy</td>
<td><img src="17" alt="Image" /></td>
<td>41</td>
<td>Lid Droop</td>
<td>EEEFF</td>
<td><img src="18" alt="Image" /></td>
</tr>
<tr>
<td>12</td>
<td>Puller</td>
<td>PQwxystvu</td>
<td><img src="19" alt="Image" /></td>
<td>42</td>
<td>Slit</td>
<td>EEEFFkl</td>
<td><img src="20" alt="Image" /></td>
</tr>
<tr>
<td>13</td>
<td>Cheek Puffer</td>
<td>NORS</td>
<td><img src="21" alt="Image" /></td>
<td>43</td>
<td>Eyes Closed</td>
<td>EEEFFFFkl</td>
<td><img src="22" alt="Image" /></td>
</tr>
<tr>
<td>14</td>
<td>Dimpler</td>
<td>Unable to encode</td>
<td><img src="23" alt="Image" /></td>
<td>44</td>
<td>Squint</td>
<td>CCDDEEFFkkllnn</td>
<td><img src="24" alt="Image" /></td>
</tr>
<tr>
<td>15</td>
<td>Lip Corner Depressor</td>
<td>PPPQQQQQRRSTUvyy</td>
<td><img src="25" alt="Image" /></td>
<td>61</td>
<td>Eyes turn left</td>
<td>GHIJJG</td>
<td><img src="26" alt="Image" /></td>
</tr>
<tr>
<td>16</td>
<td>Lower Lip Depressor</td>
<td>PPQQTUVYyzzz</td>
<td><img src="27" alt="Image" /></td>
<td>62</td>
<td>Eyes turn right</td>
<td>GHIijjG</td>
<td><img src="28" alt="Image" /></td>
</tr>
<tr>
<td>17</td>
<td>Chin Raiser</td>
<td>PPPQQQQTUyyzzz</td>
<td><img src="29" alt="Image" /></td>
<td>63</td>
<td>Eyes up</td>
<td>efgghh</td>
<td><img src="30" alt="Image" /></td>
</tr>
<tr>
<td>18</td>
<td>Lip Puckerer</td>
<td>PQRsvwwwwxwxxxyyyyyy</td>
<td><img src="31" alt="Image" /></td>
<td>64</td>
<td>Eyes down</td>
<td>EFGGGHHHKL</td>
<td><img src="32" alt="Image" /></td>
</tr>
</tbody>
</table>

Another limitation so far is the restriction to a static head and face. Although the model of the head can be dynamic while retaining the encoded facial expression, other expressions obtainable by re-orientation of the head are not included in the method. The motion/inclination of the head also conveys emotions, feelings and meaning.

The extension to include this motion in the interface is straightforward and will be considered in a future publication where keyboard encoding of facial expressions, hand gestures and body motions are considered.
combined to provide a complete human body language representation.

Apart from the development of these current limitations, future applications of the method can conceivably include client-server operation via the internet.

References


[16] Waters, K. The computer synthesis of expressive three-dimensional facial character animation. Middlesex Polytechnic, Faculty of Art and Design, Cat Hill Barnet Herts, June, 1988, EN4 8HT.


