DEVELOPMENT OF AN AUTOMATIC EXPRESSION RECOGNITION SYSTEM BASED ON
FACIAL ACTION CODING SYSTEM

HAMILTON RIVERA¹, CHRISTIANE GOULART², ALVARO FAVARATO³, CARLOS VALADÃO¹, ELIETE CALDEIRA¹,
TEODIANO BASTOS¹²

1. Postgraduate Program in Electrical Engineering, Assistive Technology Group
   E-mail: hamriver@gmail.com, carlostvaladao@gmail.com,
eliete.caldeira@ufes.br, teodiano.bastos@ufes.br

2. Postgraduate Program in Biotechnology, Assistive Technology Group
   E-mail: Christiane.ufes@gmail.com

3. Mechanical Engineering Department, Assistive Technology Group
   E-mail: alvarofavarato@gmail.com

Federal University of Espirito Santo
514 Av. Fernando Ferrari, Guaíbeiras 29075-910, Vitória, Brazil.

Abstract— Automated reading and analysis of human expressions have the potential to be a powerful tool to develop a wide variety of applications, such as human-computer interaction systems. Facial expression communication is especially effective because there are some emotions (called basic emotions), whose expressions are the same over the entire population, in contrast to communication by body gestures, speech or hand gestures, whose elements are different among the cultures throughout the world. Based on the Facial Action Coding System (FACS) in this work a system to map Action Units (AU) to six basic emotions is presented. To facilitate the design of this system, five modules are implemented, which allow: data acquisition, face detection, AU features extraction, expression classifier training and expression recognition. The results show that three emotion classes (negative, neutral and positive) were recognized by the computational system, with accuracy rates of 80.1% and 82.9% for LDA and KNN classifiers, respectively. For six classes, the accuracy rates for LDA was 62.3% and for KNN, 70.1%. The use of Kinect and the methods implemented in this work can benefit automatic emotion recognition applications, which require observation and evaluation of human expressions objectively and non-intrusively.

Keywords— Affective robot, facial expressions recognition, facial action coding system, action units.

1. Introduction

Facial expressions result from the contraction of facial muscles, making a temporary deformation of the neutral expression. These deformations are typically brief and last mostly between 250ms and 5s according to Fasel and Luettin (2003). Darwin (1965) is one of the early researchers to explore the evolutionary foundations of facial-expressions display. He argues that facial expressions are universal across humans; he contends that they are habitual movements associated with certain states of the mind. These habits have been favored through natural selection and inherited across generations. Ekman and Friesen (1978) worked on the idea of facial-expression universality to conceive the facial action coding system (FACS) that describes all possible perceivable facial muscle movements in terms of predefined action units (AUs). All AUs are numerically coded, and facial expressions correspond to one or more AUs. Although FACS is primarily employed to detect emotions, it can be used to describe facial muscle activation, regardless of the underlying cause.

Emotion recognition from facial cues based on FACS rules can be classified as: a) single-phase, where emotions are recognized directly; and b) two-phase, where the facial AUs, which are considered as building blocks of facial expressions, are detected first and then the output emotion is inferred from the detected AUs. This latter approach is found to be more practical than the former, as most of the facial expressions can be described using a sub-set of 44 AUs defined by Ekman (1978). Detecting AUs prior to emotion makes a recognition system more suited to a culture-independent interpretation. Ekman (2003) discovered that emotion expression depends only on part from human derivation who identified six basic emotions: anger, sadness, happiness, surprise, fear, disgust, whose derivations are described below.

Sadness: inner corners of eyebrows are raised, eyelids are loose and lip corners are pulled down.

Happiness: muscles around the eyes are tightened, crow’s feet wrinkles appear around the eyes, cheeks are raised and lip corners are raised diagonally.

Fear: eyebrows are pulled up and together, upper eyelids are pulled up and mouth is stretched.

Surprise: entire eyebrows are pulled up, eyelids are also pulled up and mouth is widely open.
Anger: eyebrows are pulled down, upper lids are pulled up, lower lids are pulled up and lips may be tightened.

Disgust: eyebrows are pulled down, nose is wrinkled and the upper lip is pulled up.

Figure 1 shows the six basic expressions described by Paul Ekman, who has also identified a set of facial features, that can characterize an expression of each basic emotion.

Ying-Li (2001) discovered that from 44 AUs, 30 AUs are anatomically related to the contractions of specific facial muscles: 12 are from upper face, and 18 are from lower face. AUs can occur either singly or in combination. When AUs occurs in combination, they may be additive, situation in which the combination does not change the appearance of the constituent AU, or non-additive, which is the opposite situation, when the appearance of the constituents does change. Although the number of AU is relatively small, more than 7,000 different AUs combinations have been observed by Scherer and Ekman (1982).

Commonly occurring AUs and some of the additive and non-additive AUs combinations are shown in Table 1. As an example of a non-additive effect, AU 4 appears differently, depending on whether it occurs alone or in combination with AU 1 (as in AU 1+4). When AU 4 occurs alone, the brows are drawn together and lowered. In AU 1+4, the brows are drawn together but are raised due to the action of AU 1. AU 1+2 is another example of non-additive combinations.

Automatic recognition of FACS-AU is a difficult problem and relatively few works have been reported. AUs have no quantitative definitions and, as noted, they can appear in complex combinations.

Mase (1991) and Essa (1997) described patterns of optical flow that corresponded to several AUs, but did not attempt to recognize them. Bartlett et al. (1999) reported some of the most extensive experimental results of upper and lower face AUs recognition.
2. Proposals

The goal of this work contemplates the development of an automated system for emotion recognition based on facial features (FACS-AUs) to facilitate human-robot social interaction, for a robot with playful friendly form, used as a pedagogical tool for social development of children with Autistic Spectrum Disorder (ASD). This robot named New-MARIA (New-Mobile Autonomous Robot for Interaction with Autistics) is being built in the Federal University of Espirito Santo (UFES), Brazil.

In New-MARIA, five sub-systems are used, which allow autonomous navigation, robot control, multimedia interaction, social interaction, therapeutic-robot-child approach and automatic emotion recognition. Figure 2 shows a block diagram of the automatic emotion recognition subsystem which allows human-robot emotional interaction process.

![Figure 2. Affective robot advocates the idea of emotionally intelligent machines that can recognize and simulate emotions.](image)

Figure 2. Affective robot advocates the idea of emotionally intelligent machines that can recognize and simulate emotions.

Figure 3 shows the block diagram of the method for emotion recognition based on facial expressions developed in this study, which is based on the FACS-AU system. To facilitate the design of this interface, five modules were implemented, which consist of: data acquisition module for acquiring and managing data from Kinect; face detection module for detecting the face and FACS points; AU features extraction module, to allow facial AU recognition; a trained classifier and an expression recognition module, for expression detection and classification.

![Figure 3 Block diagram of the proposed expression recognition system.](image)

Figure 3 Block diagram of the proposed expression recognition system.

3. Materials and Methods

The experimental platform implemented in this work is composed of one Microsoft Kinect 2.0 device, which provides high quality color and depth images to obtain facial points FACS-AU; and a minicomputer for image and data processing. For the operation platform, different software are used. The programming languages used in this work are: Matlab 2013b, C# 6.0, and Kinect for Windows SDK 2.0.1, which allows building desktop applications for Windows. In addition, Brekel Pro Face v2, a Windows application that enables 3D animators is used to record and stream 3D face tracking using a Kinect.

The procedure in each test consist in asked each volunteer to sit comfortably in a chair positioned in front of a 19-inches computer screen and the Kinect sensor, with his/her eyes at 90 cm away from the screen and the sensor. The screen displays the six pictures relative to human facial expressions for ten seconds. The volunteer is asked to imitate each emotional expression three times, as shown in Figures 4 and 5.

![Figure 4. Experimental procedure. Volunteer imitating the model of emotion facial expression displayed on the screen.](image)

Figure 4. Experimental procedure. Volunteer imitating the model of emotion facial expression displayed on the screen.

The Kinect device records images of each emotional facial expression performed by the volunteer, and an algorithm identifies the set of features related to expressions of each emotion, based on the AUs. The test involves the participation of eight healthy adults, aged between 24 and 33 years (M: 26, SD: ±3.81).

![Figure 5. The volunteer imitating each emotional expression.](image)

Figure 5. The volunteer imitating each emotional expression.
3.1 Data acquisition

The data acquisition module used the Microsoft Kinect 2.0 device, which provides high quality color, infrared and depth images (Figure 6) that allows obtaining a 3D-positions facial model. The algorithms for data acquisition, filtering and preprocessing were developed by methods from Microsoft ToolKit. FaceTracking library (Kinect for Windows SDK, 2013).

![Figure 6. Kinect data acquisition: A) Depth image; B) Infra-red image; C) Color image.](image)

3.2 Facial feature extraction: Action Units

Functions provided by Brekel proFace 2 Software applications were used for face detection and AUs features extractions.

For face detection automatic functions provided by Brekel proFace 2 Software are used to detect a 3D facial model, which is based on Colombo (2006) and Nair (2009) methods, using curvature features to detect high curvature areas, such as the nose tip and eye cavities (Alyuz, 2012). This technique allows the matched 3D model to deform. In Mao (2004), a correspondence is established between landmarks of the model and the captured data from face, using a model to deform the shape to match 3D points to the FACS model. Figure 7 shows an example of 3D points detection, extraction and matching to 3D model FACS using the Brekel.

The AU features module allows obtaining 20 AU features, which are the value of 20 action units. Those values are represented as a set of values that ranges between 1 and -1, which can be treated as a vector from a 20-dimensional space. Figure 7 shows the module for AU feature extraction.

![Figure 7. Module for AU feature extraction, Face detection and 3D facial model creation.](image)

Table 3 describes the AUs extracted by the features detection module. This module allows obtaining 9 bilateral (left/right face side) and 2 unilateral AUs, corresponding thus to 20 action units.

<table>
<thead>
<tr>
<th>N°</th>
<th>AU description</th>
<th>Quantity</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-2</td>
<td>AU(1) InnerBrowRaiser</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>3-4</td>
<td>AU(2) OuterBrowRaiser</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>5-6</td>
<td>AU(4) BrowLowerer</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>7</td>
<td>AU(10) UpperLipRaiser</td>
<td>1 (Unilateral)</td>
</tr>
<tr>
<td>8-9</td>
<td>AU(12) LipCornerPuller</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>10-11</td>
<td>AU(13) CheekPuffer</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>12-13</td>
<td>AU(20) LipStretcher</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>14-15</td>
<td>AU(15) LipCornerDepressor</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>16</td>
<td>AU(26) JawLowerer</td>
<td>1 (Unilateral)</td>
</tr>
<tr>
<td>17-18</td>
<td>AU(43) EyesClosed</td>
<td>2 (Left/Right)</td>
</tr>
<tr>
<td>19-20</td>
<td>AU(47) JawLeftRight</td>
<td>2 (Left/Right)</td>
</tr>
</tbody>
</table>

3.3 Expression recognition

In the emotion detector module, in order to classify the user emotions based on AU, K− Nearest Neighbors (KNN) and Linear Discriminant Analysis (LDA) algorithms were used, which are based on the work of Jiawei et al. (2012).

KNN algorithm allows predicting a value of variables (20-dimentional AU vector), and classifying them into different classes (six basic facial expressions).

The training set was always created by a learning process, which is based on all previous observations. A very interesting issue in this algorithm is how to find a similarity between two objects based on their features. In Jiawei et al. (2012), KNN emotion detector is used to compute the emotion using Euclidean distance, the emotion features (which are the value of the 20 AUs) are represented as set of values between 1 and -1. The basic Euclidean distance for the two-dimensional space is represented in Equation (1):

\[ D(p_1, p_2) = \sqrt{\left( x_1 - x_2 \right)^2 + \left( y_1 - y_2 \right)^2} \]  

However, in this work, the algorithm used to find the distance between the 20-dimension vectors uses the formula represented by Equation (2):

\[ D(p_1, x_1, \ldots x_n, \ldots p_n, y_1, \ldots y_n) = \sqrt{\sum_{i=1}^{N} (x_i - y_i)^2} \],  

where: D is the distance (similarity); p is the object; x1,...,xn and y1,...,yn are the features sets.

By using it, the algorithm can find a similarity of each object from the training set to the currently classified object and choose the K with most similar objects. In the classification mechanism a normalization for all results after computing the similarity among the objects was used. In this
research, a Min–Max normalization method was used, which is based on the work of Jiawei et al. (2012).

Linear Discriminant Analysis (LDA) algorithm or Fisherfaces method (Belhumeur et al., 1997) overcomes the limitations of the eigenfaces method by applying the Fisher’s linear discriminant criterion. This criterion tries to maximize the ratio of the determinant of the between-class scatter matrix of the projected samples to the determinant of the within-class scatter matrix of the projected samples.

The LDA method tries to find the subspace that best discriminates different facial expressions classes. The within-class scatter matrix, also called intrapersonal, represents variations in appearance of the same individual due to different lighting and face expression, while the between-class scatter matrix, also called the extra-personal, represents variations in classes.

In this research, the objective is to maximize the between-class scatter (SB), while minimizing the within-class scatter matrix (SW) in the projective subspace. The within-class scatter matrix (SW) and the between-class scatter matrix (SB) are defined as in Equation (3).

$$S_B = \sum_{j=1}^{C} N_j \left( \mu_j - \mu \right)^T \left( \mu_j - \mu \right),$$

where $\mu_j$ is the $i_{th}$ sample of class $j$, $\mu$ is the mean of class $j$, $C$ is the number of classes, $N_j$ is the number of samples in class $j$. In Equation (4), is defined how the scatter matrix (SB) is calculated:

$$S_W = \sum_{j=1}^{C} N_j \left( \mu_j - \mu \right)^T \left( \mu_j - \mu \right)^T,$$

where $\mu$ represents the mean of all classes. The subspace for LDA is spanned by a set of discriminant vectors $W=[W_1, W_2, \ldots, W_d]$, satisfying Equation (5):

$$W = \arg \max W^T S_B W.$$

The within-class scatter matrix expresses how closely facial AUs are distributed within the classes, while the between-class scatter matrix quantifies how separated the classes are from each other. When face AUs are projected onto the discriminant vectors $W$, facial AUs should be distributed closely within the classes and separately between the classes, as much as possible.

4. Results

The number of samples per class was 100, and the data was split in 60% for training sets and 40% for test sets. For KNN different $k$ were evaluated (3, 5, 7, 9), with the best accuracy was for $k=7$.

4.1 AU features

Figure 8 shows an example of AUs from the eight volunteers imitating the six basic expressions.

Regarding the identification of facial emotions by the Kinect device, using the LDA classifier for the three emotional classes positive (happiness, surprise), negative (sadness, anger, disgust, fear) and neutral (neutral), they were recognized with an overall accuracy of 80.1%, being 83.6% for neutral, 81.1% for negative class and 75.8% for positive. In turn, the KNN classifier recognized an overall accuracy of 82.9%, being 87.5% for neutral, 84.2% for negative class and 77.7% for positive, as shown in Table 4. Thus, the recognition system of emotional classes was able to identify a large quantity of specific facial features related to the neutral emotional class.

4.2 Expressions recognition

Regarding the identification of facial emotions by Kinect, but in this case considering the six emotional classes (anger, fear, sadness, happiness, surprise, disgust), the LDA classifier identified them with an overall accuracy of 62%, while the KNN classifier recognized the classes with an overall accuracy of 70%, as shown in Tables 5 and 6. Thus, the recognition system of emotional classes was also able to identify a large quantity of specific facial features.

### Table 4. Accuracy of the emotion recognition for three class.

<table>
<thead>
<tr>
<th>Emotion recognition</th>
<th>LDA Accuracy</th>
<th>KNN Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive Class</td>
<td>75.8%</td>
<td>77.7%</td>
</tr>
<tr>
<td>Negative Class</td>
<td>81.1%</td>
<td>84.2%</td>
</tr>
<tr>
<td>Neutral Class</td>
<td>83.6%</td>
<td>87.5%</td>
</tr>
<tr>
<td>Total</td>
<td>80.1%</td>
<td>82.9%</td>
</tr>
</tbody>
</table>

### Table 5. Confusion matrix for recognition using LDA.

<table>
<thead>
<tr>
<th>Expression</th>
<th>Anger</th>
<th>Fear</th>
<th>Sadness</th>
<th>Happiness</th>
<th>Surprise</th>
<th>Disgust</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anger</td>
<td>49</td>
<td>12</td>
<td>15</td>
<td>15</td>
<td>2</td>
<td>17</td>
</tr>
<tr>
<td>Fear</td>
<td>6</td>
<td>55</td>
<td>12</td>
<td>1</td>
<td>12</td>
<td>2</td>
</tr>
<tr>
<td>Sadness</td>
<td>3</td>
<td>4</td>
<td>48</td>
<td>3</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>Happiness</td>
<td>24</td>
<td>3</td>
<td>7</td>
<td>70</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Surprise</td>
<td>7</td>
<td>25</td>
<td>6</td>
<td>5</td>
<td>80</td>
<td>0</td>
</tr>
<tr>
<td>Disgust</td>
<td>11</td>
<td>1</td>
<td>12</td>
<td>6</td>
<td>0</td>
<td>72</td>
</tr>
</tbody>
</table>
5. Conclusions

This work shows the development of a system for expression recognition based on the FACS-AU system to classify six basic expressions. The developed system allows: image acquisition (IR, depth and color), face detection (FACS 3D model), features extraction (20-dimensional AU vector), classifiers training (LDA or KNN) for six expression recognition (anger, fear, sadness, happiness, surprise, disgust and neutral). This system and the implemented functions are the basis for the emotion recognition subsystem that are being implemented in New-MARIA robot.

The image acquisition module allows connecting the Brekel application with the Kinect device using the Microsoft Toolkit FaceTracking library from Kinect for Windows SDK 2. Then, a 3D facial model is obtained and 20 AU features are extracted using the face detection and feature extraction module of Brekel. The quality of the AUs depends on the illumination, head movements, distance to the sensor, facial characteristics and accessories (hair length, bangs, beard, mustache, glasses).

The expression recognition module is based on trained classifiers (LDA or KNN). Facial expressions for three classes (negative, neutral and positive) were recognized by the computational system, with accuracy rates of 80.1% and 82.9% for LDA and KNN classifiers, respectively. For six classes, the accuracy rate for LDA was 62.3% and for KNN, it was 70.1%. This success rate can be improved by implementing algorithms to detect more AUs, due the system based on Brekel only allows the detection of 20 AUs. However, such success rate can be improved if more AUs could be used, since the FACS system has more than 44 AUs. The results show the possibility to detect more AUs, and, consequently, improve the accuracy, other works shows accuracy rates between 74% to 95% for different AUs and classifiers in Sebe (2004).

The next steps for future works are to embed the expression recognition subsystem in the New-MARIA robot, integrate the emotions recognition system to a multimedia system (animated face, sound and video) able to create affective computing applications and test the total system in experimental therapy with children with autism.

Acknowledgments

The authors thank CAPES/CNPq and FAPES for the scholarships and UFES for technical and scientific support.

References


Mao, Z.; Siebert, J.; Cockshott, W.; Ayoub, A. (2004). Constructing dense correspondences to analyze 3D facial change. ICPR.

