Online Emotional Facial Expression Dictionary

Léon Rothkrantz

Abstract: Facial expressions play an important role in human communication. But there is a lot of disagreement about the semantic interpretation. In this paper we present an online facial expression dictionary. Users can upload a facial expression which will be interpreted automatically. All facial expressions are modelled by Action Units (AU's) as defined by Ekman. A request with an emotional label will result in a corresponding facial expression generated by a synthetic face modeller. The components of e-FED will be discussed and results of experiments will be reported.

Key words: Nonverbal dictionary, web-service, face recognition, facial animation, facial semantics.

INTRODUCTION

Facial expressions play an important role in human communication. Facial expressions are generated by changing contours of mouth, eyebrow and mouth by contraction or dilatation of underlying muscles. There is an ongoing discussion whether facial expressions are learned or coded in our genes. Related to the nature-nurture discussion the question is, if at least some facial expressions are universal. Friesen and Ekman [1] claimed that the facial expressions related to the emotions anger, fear, surprise, happiness, sadness and disgust are universal. They found that people all over the world show similar expressions listening to stories with one of the six basic emotions as dominant theme or telling stories after seeing pictures of emotional facial expressions expressing one of the basic emotions. Nowadays most researchers assume that facial expressions are cultural dependent and the semantic interpretation is also context dependent. Emotional facial expressions vary in intensity. And most facial expressions are usually not one of the standard universal emotions but blended emotions. As a result semantic interpretation of facial expression by humans or automated systems is a rather complex process with ambiguous results.

In 2004 [2], we started a project on designing an online facial expression dictionary (FED). The idea was that users upload a video with an emotional facial expression and the system replies with a semantic interpretations and a label. And the other way around if a user sends a label the FED generates an appropriate facial expression. But at that time our system was semiautomatic. The user has to mark fiducially points on the face such as corners of the eyes, mouth etc. to facilitate an automatic interpretation. In this paper we describe a fully automated version of the online facial expression dictionary.

An important module of e-FED is the automatic recognition of facial expressions. Ekman [1] developed a system called FACS which enables human annotators to describe facial expressions in terms of Action Units (AU’s), which are basic micro muscle movements. We developed an automated facial expression recognition system based on AU’s. The system is able to detect which AU’s are activated analysing the facial appearance. The e-FED should contain a database of a huge number of emotional facial expressions. It is almost impossible for one person to generate the required facial expression for many different combinations of AU’s of different intensities. For that reason we developed a system to generate a synthetic face. Given a list of activated AU’s the system generates the corresponding facial expression. Not every combination of AU’s generates a meaningful facial expression. Finally we had to label the generated facial expressions and find facial expressions corresponding to a given label.
We used the same methodology as Paul Ekman. Given a label we wrote a story with label as main topic and asked respondents to show the corresponding facial expressions. And the other way around given a facial expression we asked respondents to choose an appropriate label.

In this paper we provide more details about the e-FED system and composing modules. In the next section we report about related work. Then we present the general architecture of the e-FED system and a description of composing modules. In the last section we give a final conclusion.

**RELATED WORK**

In order to detect and analyse facial expressions we need first to define how to describe these emotional states expressed on our face. The prime example of the codified version of facial expressions is the Facial Action Coding System (FACS), widely used by psychologists. Another way of describing facial activity is to give some quantitative description in terms of geometrical changes of the face. Such geometry changes can be described by using an MPEG-4 standard with its Facial Animation Parameters (FAPs) [3]. Ekman and Friesen developed the original FACS in the 1970s by determining how the contraction of each facial muscle (singly and in combination with other muscles) changes the appearance of the face. This facial activity is described in terms of visually observable facial muscle actions (i.e., action units, AUs). Emotions can also be described by their coordinates on continuous scales. Two common scales are valence and arousal. Valence describes the pleasantness of the stimuli. The other dimension is arousal or activation [8].

There are many different methods for recognising facial expressions. The approaches can be roughly divided into two main groups, the feature based methods and the template based methods. The feature based methods use a set of texture and geometrical information as features where the template based methods use 2D or 3D head and facial models as templates for facial expression recognition. Many different research groups follow different approaches and claim to get the best results, but most of the methods have only been tested on a single facial expression database. One of the feature based methods is Dynamic Bayesian Networks (DBNs). Probably the most used method for facial expression recognition is the Support Vector Machines (SVMs). For an overview of all possible methods see [4]. Recent attempts in facial expression recognition try to recognise a small set of prototypic expressions. The developed systems are trained on datasets of acted emotions. These ‘acted’ expressions may differ in appearance and timing from spontaneous occurring emotion. Automatic recognition of facial expressions is a complex topic. The main problems are lighting conditions, posture and occlusion. In most research only frontal images are considered.

The FACS model has recently inspired the derivation of facial animation and definition parameters in the framework of the ISO MPEG-4 standard. In particular, the facial definition parameter set (FDP) and the facial animation parameter set (FAP) were designed in the MPEG-4 framework to allow the definition of a facial shape and texture, as well as the animation of faces reproducing expressions, emotions, and speech pronunciation [3]. There are two major approaches to facial modelling and animation. The first one is based on image manipulation, and the second one is based on geometric manipulation.

In this paper we concentrate exclusively on geometric modelling based on deformation of 3D human face. Synthesis of 3D facial expressions involves two aspects: accurate representation of a human face, and modelling of facial movements. The methods of representing and displaying a detailed 3D geometry of human face are the same as for any other graphical object. We developed a Facial Expression Modeller (FEM) [3] for the generation of facial expressions on a synthetic face.
The e-FED main website enables users to issue queries into e-FED. It consists of static HTML pages and Java applets. The HTML pages provide the user with information and are used to structure the layout of the website. The applets implement the GUI for issuing a query into e-FED. For each query, there exists an applet that implements the GUI for that query. When a user selects to issue a certain query, the appropriate applet is loaded from the server. The Communication Layer of the e-FED system resides on the server and handles all data traffic between the client and the server. The communication layer consists of a collection of Java servlets. The Query Processing Module or QPM of the e-FED system also resides on the server and consists of several modules, each of which has the ability to process a specific type of query. Each module is implemented through one or more static Java classes. The FED admin website provides the GUI for the management part of the FED system. Like the main website, it consists of static HTML pages and Java applets. The differences with the main website are that the admin website is only accessible through user authentication, and that the applets don’t provide the GUI for a certain query, but the GUI for a certain FED management function. The Admin Processing Module or APM implements the functionality needed to manage the FED system. Like the QPM, the APM consists of several modules that in turn consist of one or more static Java classes. The FED Database contains all the entries in the dictionary, admin user information, and log info. The PostgreSQL database management system is used to implement the database.

RECOGNITION OF FACIAL EXPRESSIONS
The input of our system is a picture showing an emotional facial expression. In this section we describe the procedure how to extract AU’s automatically. We used an adaptive version of the system described in [4]. To test and train our system we used the Cohn-Kanade Facial Expression database [5]. We completed this set with our own recordings. In total we had 600 grey scale recordings of 150 subjects displaying facial expressions, labelled according to the FACS standard.

The pre-processing step consists of detecting the facial region. We adopted a commonly used face detection method, introduced by [4], based on Haar-wavelet features and a cascade of weak classifiers trained by Adaboost. Once the facial region has been selected, the next step consists of localization of fiducial points such as the corners of the eyes, mouth and eyebrows. To detect the special points we used Active Appearance Model (AAM) extraction as we described in [6]. These fiducial points are used to localise 12 blocks around special facial regions. These blocks are called Regions of Interest.
(ROI’s). The size and location of each ROI was determined using the entropy maximum algorithm which captures the regions with the most complex motion. In order to perform this search, we used the motion flow estimators calculated for the whole facial region between every two consecutive frames, using the Lucas-Kanade algorithm.

![Display of Region of Interest (ROI).](image1)

Figure 2: Display of Region of Interest (ROI). Figure 3: Examples of optical flow on faces.

Our goal is to capture movements which can be associated with activation of AU’s. For every region of interest we calculated the average displacement on x and y axes. In our approach, we modelled the relations between AUs and the corresponding parameters (displacement on x and y axes) by employing a causal model in which the hidden variable AU causes the observations (Vx, Vy), as depicted in Figure 4. In order to approximate the continuous distribution of the Vxi and Vyi variables we used a mixture of Gaussian distributions, depicted in the model by the node M.

![Dynamic Bayesian network (DBN) model for Action Units (AU’s).](image2)

Figure 4: Dynamic Bayesian network (DBN) model for Action Units (AU’s).

In the implementation process we employed an open source Matlab toolbox: Bayes Net Toolbox. We used the Expectation-Maximization algorithm (EM), to maximize the posterior probability of the (Vx_i; Vy_i) i=1:n parameters, given the observed training data (AU, Vx, Vy), in the presence of the hidden parameters (Mi) i=2;4;2n. We found that the overall recognition rate of AUs was 93%. For more details we refer to [4].

**FACE GENERATION**

FEMSFacial Expression Modeler (FEM) [3] is a system for the generation of facial expressions. It is implemented in C++ language on a PC platform. It uses multiplatform OpenGL and Qt GUI toolkits, and so it is available on both Windows and Linux operating systems (with the possibility of porting it to other systems as well). The 3D face model is built from triangular mesh modelled and textured in 3D Studio Max. The shape of the wireframe was built on the basis of an existing person’s face. In order to create a texture two pictures of this specific person have been: a frontal and lateral view of the face. Both pictures were orthogonally projected on a cylindrical texture, and blended together. This software includes a parser to read “.ase” files exported from 3D Studio Max, and builds an
internal 3D model which is displayed in the main window. The user interface consists of a window with 3D facial model and two controls for accessing facial expressions from the library, and editing them by editing the different AU controllers. The facial model is based on FACS, but the user of this system does not have to be an expert in FACS in order to use the system itself. For the user a facial expressions script language was designed that wraps up the AU’s in more intuitive terms. While designing a new facial expression, a user can make use of pre-defined facial expressions, which can be loaded from the library. It is assumed that each expression in the library is defined by a unique name, and that any given expression has a fixed set of AU’s with their intensities. The user can also create facial expressions and save them into the library. There is also an editor to modify an existing facial expression. In order to edit or to create new facial expressions, a user can access lower-level animation controls (using GUI elements that control all of the parameters corresponding to each AU). He can interactively move sliders and observe changes on the face resulting from activation of a given AU. These controls are divided into 5 groups of AU’s: Upper face AU’s, Lower face linear AU’s, Lower face orbital AU’s, Head AU’s, Eyes AU’s. In Figure 5 we show the GUI of our facial animation system (FEM) and in Figure 6 some of the generated facial expressions.

Figure 5: A screenshot of the implemented facial animation system.

Figure 6: Facial expressions (from left to right: admiration, like, happy, dislike, and disagreement.

Table 1: Examples of labels of emotional facial expression and correspondent activated AU’s.
The database should ideally contain only genuine expressions of emotions. However, as the database should also consist of high-quality video samples (with constant illumination, background, head pose, etc…) to be useful for practical applications, the choice that was made was to get as close as possible to spontaneous emotions, while keeping at the same time a fully controlled recording environment. To obtain the participant’s facial expressions the experimenter should instruct the participant to perform the pure facial expressions. To provide the right context the experimenter tells a story and at the end the participant has to show a facial expression.

Example scenario Sadness It was assumed that the participant took part in a difficult exam on mathematics. This exam was very important for the participant and he spend a lot of time and effort. He gets a phone call of the student counsellor with the result of the exam. “Good afternoon, the student counsellor is speaking to you, unfortunately I have to tell you that you failed for the exam”. It was assumed that the participant shows a facial expression of sadness. In Figure 6 we show some of the generated facial expressions, labelled using scenario’s. Next we computed the activation of AU’s using the recognition module and finally generated facial expressions using FEM (see Figure 7).

STRUCTURE DATABASE

In a verbal dictionary words are ordered in an alphabetic order. In the project WordNet [7] researchers try to reveal the semantic structure between words. In our nonverbal dictionary facial expressions can be ordered using AU’s. But the Euclidean distance between vectors of AU’s is a bad measure for semantic similarity. If for example the corners of a smiling mouth moving downwards, the meaning of the facial expression changes from happiness to sadness, two opposing emotions. To solve that problem we use the verbal labels of the facial expressions. Ekman claims that six emotions are universal, but most emotions are dependent on the context, words, character, etc. But even universal emotions show variation in facial appearance. Some people laugh with open mouth other with closed mouth and there is an enormous difference in intensity. The
semantic interpretation of the facial expression of Mona Lisa is still ambiguous. To reduce the ambiguity we generated and labeled our facial expressions context dependent. In a user test (see Table 2) it proves that without information about the context there is a significant confusion in the interpretation/labelling, of facial expressions. Our hypothesis is that the different labels of semantic similar facial expressions are more or less synonyms. To test this hypothesis we ranked every label on a valence and arousal scale using DAL. Whissell’s dictionary of affect in language (DAL) is an annotated dictionary of words, which all have been given a value for activation and evaluation. The term valence usually refers to the positive and negative character of an emotion. At present, valence is often used as affect valence it refers to how good or bad an emotion experience, or affect, feels “Arousal” stands for the level of activation of the emotion, and it is characterised as a range of affective responses extending from “passive” to “active”. Meaning the subject is in a condition of sensory alertness and readiness to respond.

**Table 2: Confusion matrix of 12 emotions attached to FEM expressions on a 5-point scale.**

<table>
<thead>
<tr>
<th></th>
<th>admire</th>
<th>afraid</th>
<th>angry</th>
<th>cheerful</th>
<th>desire</th>
<th>disgust</th>
<th>happy</th>
<th>irritated</th>
<th>moved</th>
<th>sad</th>
<th>smiling</th>
<th>surprised</th>
</tr>
</thead>
<tbody>
<tr>
<td>admire</td>
<td>0</td>
<td>4.2</td>
<td>4.2</td>
<td>3.2</td>
<td>3.4</td>
<td>4.2</td>
<td>3</td>
<td>5</td>
<td>5</td>
<td>4.6</td>
<td>3.2</td>
<td>3</td>
</tr>
<tr>
<td>afraid</td>
<td>4.2</td>
<td>0</td>
<td>3.4</td>
<td>3.8</td>
<td>2</td>
<td>4.4</td>
<td>4.2</td>
<td>4.4</td>
<td>2.6</td>
<td>3.2</td>
<td>4.2</td>
<td>3</td>
</tr>
<tr>
<td>angry</td>
<td>4.2</td>
<td>3.4</td>
<td>0</td>
<td>4</td>
<td>4</td>
<td>3.8</td>
<td>4</td>
<td>4</td>
<td>4.4</td>
<td>3.6</td>
<td>4</td>
<td>3.6</td>
</tr>
<tr>
<td>cheerful</td>
<td>3.2</td>
<td>3.8</td>
<td>4</td>
<td>0</td>
<td>3.6</td>
<td>4.2</td>
<td>2.6</td>
<td>4.6</td>
<td>4</td>
<td>5</td>
<td>1</td>
<td>4.2</td>
</tr>
<tr>
<td>desire</td>
<td>3.4</td>
<td>2</td>
<td>4</td>
<td>3.6</td>
<td>0</td>
<td>4.2</td>
<td>3.4</td>
<td>4.4</td>
<td>2.8</td>
<td>3.8</td>
<td>2.6</td>
<td>3</td>
</tr>
<tr>
<td>disgust</td>
<td>4.2</td>
<td>4.4</td>
<td>3.8</td>
<td>4.2</td>
<td>4.2</td>
<td>0</td>
<td>4.4</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>4.6</td>
<td>4.4</td>
</tr>
<tr>
<td>happy</td>
<td>3</td>
<td>4.2</td>
<td>4</td>
<td>2.6</td>
<td>3.4</td>
<td>4.4</td>
<td>0</td>
<td>4.6</td>
<td>4.8</td>
<td>4.4</td>
<td>1.8</td>
<td>3</td>
</tr>
<tr>
<td>irritated</td>
<td>5</td>
<td>4.4</td>
<td>4</td>
<td>4.6</td>
<td>4.4</td>
<td>3</td>
<td>4.6</td>
<td>0</td>
<td>3.4</td>
<td>4</td>
<td>4.8</td>
<td>4</td>
</tr>
<tr>
<td>moved</td>
<td>5</td>
<td>2.6</td>
<td>4.4</td>
<td>4</td>
<td>2.8</td>
<td>4</td>
<td>4.8</td>
<td>3.4</td>
<td>0</td>
<td>2.2</td>
<td>4.2</td>
<td>3.8</td>
</tr>
<tr>
<td>sad</td>
<td>4.6</td>
<td>3.2</td>
<td>3.6</td>
<td>5</td>
<td>3.8</td>
<td>3</td>
<td>4.4</td>
<td>4</td>
<td>2.2</td>
<td>0</td>
<td>4.6</td>
<td>4.8</td>
</tr>
<tr>
<td>smiling</td>
<td>3.2</td>
<td>4.2</td>
<td>4</td>
<td>1</td>
<td>2.6</td>
<td>4.6</td>
<td>1.8</td>
<td>4.8</td>
<td>4.2</td>
<td>4.6</td>
<td>0</td>
<td>3.4</td>
</tr>
<tr>
<td>surprised</td>
<td>3</td>
<td>3</td>
<td>3.6</td>
<td>4.2</td>
<td>3</td>
<td>4.4</td>
<td>3</td>
<td>4</td>
<td>3.8</td>
<td>4.8</td>
<td>3.4</td>
<td>0</td>
</tr>
</tbody>
</table>

**Figure 8: Plot of 38 emotional labels in 2D valence and arousal plane.**
We decided to divide the list of sixty emotional words (labels) into groups. For performing this division we initially plot all the emotional words in 2D space (see Figure 8) according to their pleasantness and activation value given by Whissell [8]. In a later step we implemented the fuzzy c-means (FCM) algorithm for initializing the clusters of the words in the 2D space. As a prototype facial expression of the specific cluster we selected the barycentre of the cluster after removal of possible outliers.

**SUMMARY, CONCLUSIONS AND FUTURE WORK**

Similar to a verbal dictionary we developed a nonverbal dictionary composed of facial expressions as “words”. The spelling of the facial words is by Action Units from the facial annotation tool FACS as developed by Ekman. Emotional facial expressions are labelled by emotional words using a context sensitive procedure. The valence and arousal coordinates from the DAL dictionary provides a semantic meaning to the facial expression and a distance measure. At this moment the e-FED is composed of 60 facial expressions. In the near future the number of facial expressions will grow. The e-FED system has an automated tool for the recognition and generation of facial expressions in terms of AU’s.

The system has been used by many students from Delft University of Technology [3,4,6]. The recognition of AU’s is about 90%, but the semantic interpretation can be ambiguous. Unfortunately there are only a few universal facial expressions; most expressions are cultural, context and individual dependent. The e-FED system is supposed to provide a standard in the set of facial expressions. Currently the system has been used in an adaptation course of foreign students at Delft University of Technology.

**REFERENCES**


**ABOUT THE AUTHOR**

Prof. Leon Rothkrantz, PhD, Multimodal Communication, Intelligent Interaction, Delft University of Technology, Mekelweg 4, 2628 CD Delft, The Netherlands, E-mail: L.J.M.Rothkrantz@TUDelft.nl.