Facial Emoticons – Facial Expressions as a Means of Interaction

Sandra Gama, Daniel Gonçalves
Instituto Superior Técnico, Universidade Técnica de Lisboa, Portugal
sandra.gama@ist.utl.pt, daniel.goncalves@inesc-id.pt

Abstract

Facial expressions are highly effective non-verbal means to share emotions among humans. Nowadays, in the cybernetic panorama, emoticons are used as surrogate facial expressions in electronic-mediated communication. As a new modality, it provides a natural means of interaction, opening a feedback channel that makes it possible to interact without using a keyboard or a mouse.

We present FacialEmoticons, a facial expression recognition library that can be used as the basis for meaningful interaction. After face tracking and feature detection, Bayesian Classifiers are used to infer the expression on the user's face. Our methodology requires low computational power yielding recognition rates comparable to those of humans. Our results show the viability for facial-expression-based real-time interaction.

1. Introduction

For humans, recognizing facial expressions is fairly straightforward: they consist of a series of specific features that often make them easy to understand, even among people of different social and ethnical backgrounds. Ideally, human-machine interaction should be similar to communication among humans. In the context of facial expressions, it would be desirable that computer systems classified facial expressions and used this information to enhance interaction.

In order to use facial expressions as a means of interaction, we must adopt image processing and pattern recognition methodologies. A wide range of research has been done over the last two decades: Pantic and Rothkranz [6], Bartlett et al. [1] and Lu et al. [4], among others, process facial images and associate these with corresponding states of emotion.

FacialEmoticons takes advantage of the vast information of facial expressions, providing a rich non-conventional modality for interaction. Also, being extremely versatile, it can be used synergistically with any application. We will succinctly describe how expression recognition takes place, as well as illustrate two prototypes we have created.

2. Recognizing Facial Expressions

We have followed a top-down approach to facial expression recognition, dividing the main problem of recognizing facial expressions into five sub-problems:

Facial Detection: In the context of our research, the user's environment is arbitrary. As such, the system must process context-independent images. We adopted Viola and Jones' method [7] for face detection, since it uses Haar classifier cascades, providing a means for robust real-time facial detection.

Normalization: So that facial feature extraction is done with precision, and since facial areas may present a variety of dimensions (due to device resolution, etc.), we normalized the face area for further processing. We adopted a facial feature model, which consists of facial features of an *average face*, to infer unknown facial values, granting our system higher flexibility and robustness.

Face Feature Extraction: Based on FACS [2], a technical norm that classifies facial behaviors depending on the muscles that are responsible for each facial action, we considered three main facial features: eyes, eyebrows and mouth. Accordingly to FACS, facial parameters and variations are predominantly related to these elements. Our feature extraction methodology consists of:

(1) *Detection of a region of interest*, where we take advantage of OpenCV library functions to restrict the areas for further processing;

(2) *Grayscale conversion*, which consists of a pre-processing to the Canny operator, which manipulates single-channeled images. It also simplifies Gaussian Blur processing, since applying its convolution matrix to only one channel is more efficient;

(3) *Gaussian Blur* which, although the Canny operator by default applies a Gaussian Blur filter before detection, is useful for reducing noise and eliminating additional artifacts such as dark circles and lines;

(4) *Canny* algorithm, which is computationally fast and noise-robust, providing a means for facial feature extraction with increased precision; and

(5) *Coordinate Extraction*, for obtaining values corresponding to detected features by analysis of the previous step results. Extreme points define both maximum and minimum values in x and y axes and the average between them makes it possible to generate a geometric form.

Feature Transformation: In order to maximize the expressiveness of the features that have been collected at former stages, we transformed these into data structures corresponding to features that would provide the classifier with more relevant information. These new features, based on the concept of Action Units [2], are: Vertical distance between eyes and eyebrows, eye aperture, mouth aperture, mouth width, average vertical distance from mouth corners to eyes, and average vertical distance from mouth corners to mouth center.

Classification: We adopted Bayesian classifiers because, despite their low computational weight, they make it possible to obtain good results. We started with a Bayesian classifier based on a discreet set of intervals. These intervals are associated with the values of samples' features, being the whole classification process based on a discreet set of intervals. However, since its performance wasn't sufficiently satisfactory, we decided to follow another approach: A Gaussian Bayesian classifier. This classifier uses an estimation of values for infinite samples, assuming that these follow a normal distribution. The likelihood of a certain value belonging to a given class is calculated through the cumulative distribution function of the Gaussian distribution that is generated from training samples. The training of the classifier is done both through positive and negative samples, so that it is possible to determine the likelihood of a sample belonging or not to a given class. The classifier associates the sample with the class that presents the highest likelihood for that sample.

3. Prototype Applications

We created two prototype applications in order to demonstrate the feasibility of FacialEmoticons as a means of interaction. The first (Figure 1), whose main objective is to illustrate the usage of the FacialEmoticons library as e means for easy insertion of emoticons in any active window, runs in background and allows users to insert emoticons into the current application only by pressing the F12 key. Their expression at that time is automatically classified and inserted in the active window. The second prototype, E-motional Jukebox (Figure 2), consists of a multimodal interface with hand gesture recognition to control basic audio functions (using the HandVU library [3]), and facial expressions for track classification. The player infers the user's taste based on previous classifications and automatically selects new tracks to play.

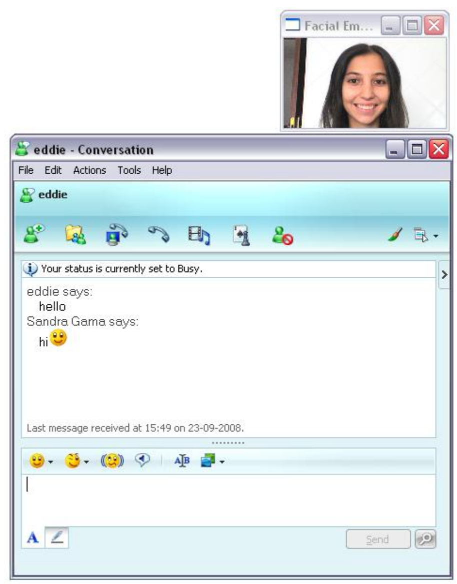

Figure 1. FacialEmoticons prototype interface.

Figure 2. E-motional Jukebox prototype interface.

4. Evaluating Expression Recognition

We conducted several tests, considering two scenarios: (1) distinguishing between happy and sad faces; and (2) distinguishing among five facial expressions: happy, sad, neutral, angry and surprised. The first considers the two most common opposed expressions, and the second considers all expressions recognized by our system. Both scenarios have been run for each test setting.

We ran tests on 936 database images from a University of Dallas' Database [5], obtaining hit-rates of 81% and 56% for scenarios (1) and (2), respectively, as summarized in Figure 3.

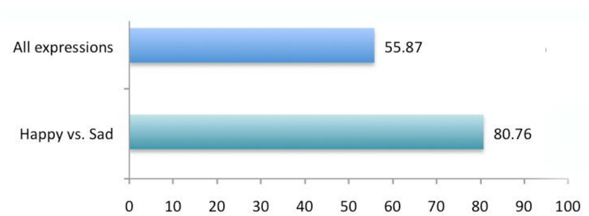

Figure 3. Hit rates for the first test setting.

To simulate a real-life setting, we ran additional tests in a random environment outside the lab, gathering 10 frames per expression, for each of 10 users. Results improved to 85% and 59%, respectively, as depicted in Figure 4.

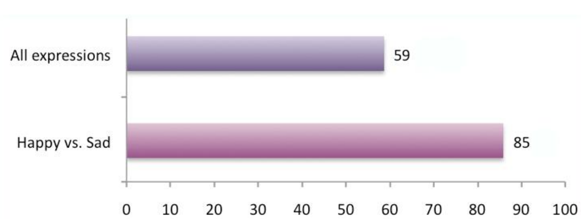

Figure 4. Hit rates for the second test setting.

Realizing that some expressions may be hard to identify or ambiguous, we asked 10 people to classify the same set of 335 facial images, obtaining hit rates of 96% and 64%, respectively. Taking these values as a baseline and comparing them with FacialEmoticons' recognition rates we can see that, considering happy and sad expressions, our library correctly recognizes nearly 90% of human-recognized expressions, while for all classes of expressions, the classifier recognizes 90% of them, as summarized in Figure 5. This shows that our system performs close to human annotation.

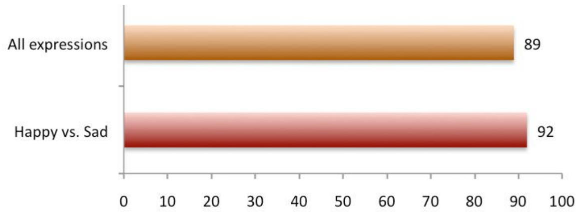

Figure 5. Results of automatic classification compared with human performance.

5. Conclusions

Facial expressions, being effective means for sharing emotions among humans, provide a way of interaction that FacialEmoticons takes advantage of. We provide a library that, along with being viable to use with other applications, presents relevant means for non-conventional interaction. We intend to enhance our results by improving feature detection. We also plan to further explore the potential of this modality by performing system evaluation to prove its adequacy and by considering new interaction scenarios.

6. References

[1] M. Bartlett, G. Littlewort, I. Fasel and J. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction", *CVPRW Proceedings*, pp. 53, 2003.

[2] P. Ekman and W. Friessen, *Facial Actions Coding System (FACS): Manual*, Palo Alto: Consulting Phsychologists Press, 1978.

[3] M. Kolsch and M. Turk, "Robust Hand Detection", in *IEEE Intl. Conference on Automatic Face and Gesture Recognition Proceedings*, 2004.

[4] H. Lu, Y. Huang, Y. Chen and D. Yang, "Real- time Facial Expression Recognition Based on Pixel-Pattern-Based Texture Feature", *Electronic Letters Proceedings*, 2007, pp. 916-918.

[5] M. Minear and D. Park, "A Lifespan Database of Adult Facial Stimuli", in *Behavior Research Methods, Instruments and Computers*, 2004.

[6] M. Pantic and L.J.M. Rothkrantz, "Expert System for Automatic Analysis of Facial Expressions", in *Image and Vision Computing Journal Proceedings*, 2000, pp. 881-905.

[7] P. Viola and M. Jones, "Robust Real-Time Object Detection", *Technical Report*, University of Cambridge, 2001.