Computación y Sistemas Número Especial pp. 019 - 025
© 2002, CIC - IPN. *ISSN 1405-5546* Impreso en México

Facial Communication for Human-Computer Interaction

Comunicación Facial para Interacción Humano-Computadora

Homero V. Ríos[1-3], Ana Luisa Solís[2], Emilio Aguirre[3], Lourdes Guerrero[2] and Joaquín Peña[3]
[1]Maestría en Inteligencia Artificial, Universidad Veracruzana
Sebastián Camacho No. 5, C.P. 9000, Xalapa, Veracruz, México
[2]Departamento de Matemáticas, Facultad de Ciencias-UNAM
[3]Laboratorio Nacional de Informática Avanzada A.C.
Rébsamen 80, C.P. 91090, Xalapa, Veracruz, México
E-mails: hvrf@yahoo.com, alsgc@yahoo.com

Article received on July 28, 2001; accepted on February 19, 2002

Abstract

This paper describes a new interface for human-computer interaction based on recognition and synthesis of facial expressions. This interface seeks to sense the emotional state of the user, or his/her degree of attention, and communicate more naturally through face animation.

Keywords: Human-Computer Interaction, Computer Vision, Facial Communication, Facial Feature Extraction.

Resumen

Este artículo describe una nueva interfaz para interacción humano-computadora basada en reconocimiento y síntesis de expresiones faciales. Esta interfaz busca detectar el estado emocional del usuario, o su grado de atención y comunicarse más naturalmente a través de animación facial.

Palabras clave: Interacción Humano-Computadora, Visión por Computadora, Comunicación Facial, Extracción de Rasgos Faciales.

1 Introduction

The work described here is the result of the joint research project "Gesture recognition interfaces and intelligent agents for virtual environments", funded by the Mexican National Council of Science and Technology (CONACYT) as project ref. C098-A and C100-A. The problem addressed by the project is human-computer interaction through face and gesture recognition, and animation, to expand current interfaces, to make them more natural.

Our approach consists in analyzing images of the user, to extract relevant features such as position, orientation, shape and emotions of the face, or other parts of his body. Then, to make gesture recognition or behaviour understanding, to give the user feedback, and communicate through a synthetic animated face or avatar (virtual human) with synthetic voice.

The problem is so complex, that we have only suceeded in facial image analysis and in facial animation of a 3D face. However, the approach that we have selected can be expanded in the future.

In this paper, we describe the results that we have obtained for face analysis and synthesis, and how they could be integrated as part of a user interface system, such as an intelligent tutoring system.

This paper describes a new interaction technique for human-computer interaction based on the integration of results from computer graphics and computer vision. In the last few years this integration has shown important results

and applications (Eisert, 98; Pentland, 96). For example, Richard Szeliski described the use of image mosaics for virtual environments in 1996 and, in the following year for combining multiple images into a single panoramic image. H. Ohzu et al. described hyper-realistic communications for computer supported cooperative work.

Facial expression understanding is a good example of the rich middle ground between graphics and vision. Computer vision provides an excellent input device, particularly for the shapes and motions of complex changing shapes of faces when expressing emotions (Pentland, 96; Parke et al, 96).

We have been studying how to analyze efficiently video sequences for capturing gestures and emotions. Relevant expressions and their interpretations may indeed vary depending upon the chosen type of application (Rios et al., 98).

Facial expression recognition is useful for example, for adapting interactive feedback in a tutoring system based on the student's level of interest. The type of expressions associated with these applications are: degree of interest, degree of doubt of the information presented, boredom among other expression, or to assess the time of interest or lack of interest presented by an application.

The work mentioned here also strives to capture the high resolution motion and appearance of an individual face. The goal is to use this information to animate and render synthetic faces and to have interaction with the user.

2 Analysis and Interpretation of Facial Expressions

The communicative power of faces makes it a focus of attention during social interaction. Facial expressions and the related changes in facial patterns inform us on the emotional state of people and help to regulate both social interactions and spoken conversation. To fully understand the subtlety and expressive power of the face, considering the complexity of the movements involved, one must study face perception and related information processing.

For this reason, face perception and face processing have become major topics of research by cognitive scientists, sociologists and more recently by researchers in computer vision and computer graphics.

The automation of human face processing by computer will be a significant step towards developing an effective human-machine interface. We must consider the ways in which systems with this ability understand facial gestures (analysis), and the means of automating this interpretation and/or production (synthesis) to enhance human-computer interaction.

2.1 Facial Displays as a New Modality in Human-Computer Interaction

Facial expressions are viewed in either of two ways. One regards facial expressions as expressions of emotional states. The other view, facial expressions related to communication. The term "facial displays" is equivalent to "facial expressions", but does not have the connotation of emotion.

The present paper assumes the second view. A face is an independent communication channel. A facial display can be also seen as a modality because the human brain has specialized neurons dedicated to the necessary processing.

Part of this research is an attempt to computationally capture the communicative power of the human faces and to apply it to user interfaces.

2.1.1 Theory of Communicative Facial Displays

First of all, facial displays are primarily communicative. They are used to convey information to other people. The information that is conveyed may be emotional information, or other kinds of information, indications that the speaker is being understood, listener responses, etc.

Facial displays can function in an interaction as means of communication on their own. That is, they can send a message independently of other communicative behavior. Facial emblems such winks, facial shrugs, and listener's comments (agreement or disagreement, disbelief or surprise) are typical examples. Facial displays can also work in conjunction with other communicative behavior (both verbal and nonverbal) to provide information.

2.2 Facial Expressions as Emotional Signals

Facial expressions convey emotions and there are ongoing debates about their discreteness and universality. One of the most documented research efforts led by Ekman has permitted to identify six basic universal emotions: fear, anger, surprise, disgust, happiness, and sadness (Ekman and Friesen, 75), (Fig. 1). Others like Russel prefer to think that facial expressions and labels are probably associated, but the association may vary with culture (Russel, 94).

2.3 Face Analysis

The analysis of faces by computer is difficult since there are several factors that influence the shape and appearance of faces on images. Some of these factors are illumination, viewpoint, color, facial hair and the variability of faces. In addition, we still do not know the exact mechanisms used by humans for face processing. For instance is face

processing a holistic or feature analysis process?. The brain itself has specialized structures like IT cells on the visual areas to handle face detection (Bruce & Green, 89).

The problem of face analysis can be divided in face detection and face recognition. In the first case the goal is just to locate the general position of faces on images. On the latter, the purpose is to recognize faces using extracted features. This recognition can take place on the following conditions (Chellapa et al., 95): a) static images, b) range images, and c) video sequences.

Figure 1. The universal expressions: fear, anger, happiness, surprise, disgust, sadness

Since faces are non-rigid objects the best way to model them is through the use of deformable models which should have enough parameters to accommodate most of the variations in shape. Active contours have been used to detect and track facial features like head contour, lips, eyebrows and eyes (Lam & Yang, 96; Terzopoulos & Szeliski, 92). The problem is that since "snakes" can adapt any shape, sometimes they take nonvalid shapes. One solution is the use of trainable deformable models from examples, like the point distribution model proposed by Cootes (Cootes et al. 92). This model has been improved to learn grey and color variations, and the search of the model on images can be optimized as shown by active appearance models (Cootes et al. 98).

In this work, we have used a trainable model similar to Cootes et al., 92, with a difference. First, we have used in our model, more landmarks per face, so as to capture more lines of expression: around the mouth, below the eyes, and on the forehead. Second, our training set consisted of less pictures per individual, but more individuals. Third, we included individuals showing all different universal expressions. In constrat, the mentioned work, included mainly neutral and happy faces. Fourth, the backgrounds that we used were more diverse, and this introduced more complexity when automatically adjusting faces.

To train the deformable model for face detection we used 230 faces and 163 landmarks per face (Figure 2). After aligning the training faces to reduce the effect of translation, rotation and scaling, principal component analysis is used to obtain the main modes of variation (Cootes et al. 92). Twenty three eigenvectors account for more thant 90% of shape variation. Any shape in the training set can be approximated using the mean shape and a weighted sum of the first t eigenvectors (ordered from the most significant eigenvalue to the least significant) as follows:

$$\mathbf{x} = \bar{\mathbf{x}} + \mathbf{Pb}$$

where $\mathbf{x}$ is a shape in the training set, $\bar{\mathbf{x}}$ is the mean, $\mathbf{P}=(p_1,\ldots, p_t)$ is the matrix of the first t eigenvectors, and $\mathbf{b}=(b_1, b_2, \ldots, b_t)^T$ is a vector of weights for each eigenvector (Figure 3).

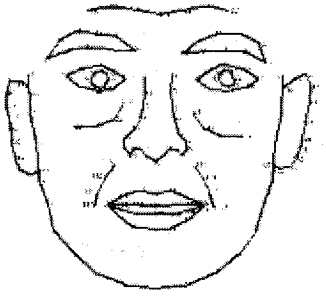

Figure 2. Landmarks used for face representation

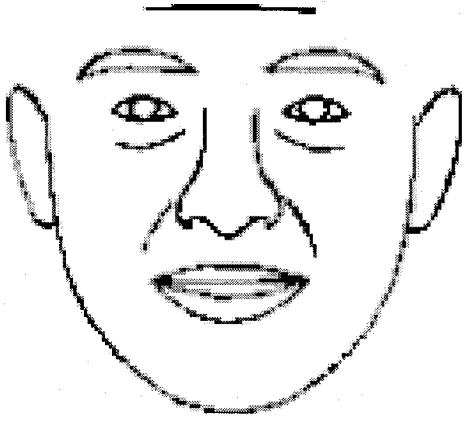

Figure 3. Mean shape (face)

We have implemented search techniques to locate the deformable model on images. Some results are shown on figure 4. We have found that it is best to use a combination of gradient based and gray level based snakes to locate facial features. For instance, on the face contours, it is best to used gradient information, since the color, shape, and background around the faces varies a lot. In contrast, in the inner features of the face, we found better results, by using the grey level information learned from the training set.

One of the main results of our work is to expand previous deformable model of faces reported in the literature, with more points and features associated with expression recognition. Also, we have found that many more eigenvectors are needed to explain facial shape, when expressions are introduced. In contrast, to was reported before, when training sets included only a few expressions. Finally, we have found which eigenvectors influence more in the generation of each of the so called "universal expressions".

Figure 4. Initial and final position for model fitting for image analysis

3 Modelling and Animation of Facial Expressions

3.1 Facial Action Coding System

Ekman and Friesen have produced a system describing all visually distinguishable facial movements. The system, called the Facial Action Coding System or FACS, is based on the enumeration of all "Action units" of a face that cause facial movements. As some muscles give rise to more than one action unit, the correspondence between action units and muscle units is approximate

The Facial Action Coding System (FACS) describes the set of all possible basic action unit (AU's) performable by the human face. Each action unit is a minimal action that cannot be divided into smaller actions. According to Ekman, "FACS allows the description of all facial behavior we have observed, and every facial action we have attempted".

The primary goal of the FACS was to develop a comprehensive system which could reliably describe all possible visually distinguishable facial movements.

The use of FACS in facial animation goes beyond what was originally intended. In a number of facial animation systems, FACS is used as a way to control facial movement by specifying the muscle actions needed to achieve desired expression changes.

FACS was derived by analysis of the anatomical basis for facial movements. Since every facial movements is the result of some muscular action, the FACS system was developed by determining how each muscle of the face acts to change appearance.

Ekman´s FACS model has 46 major Action Units and divides the face into three areas: browns, eyes and lower face.

3.2 Synthesis and Animation of Expressions

Facial animation typically involves execution of a sequence of a set of basic facial actions. We use action units (AU) of the Facial Action Coding System (FACS) as atomic action units and as basic facial motion parameters.

Each AU has a corresponding set of visible movements of different parts of the face resulting from muscle contraction. Muscular activity is simulated using a parameterized facial muscle process. We can define atomic action units similar to AU and definite expressions and phonemes. These can be used for defining emotion and sentences for speech.

For the driven facial animation the input parameters to the facial animation are the AUs. The facial expression recognition module provides a two way dialog and requires that the person on the other side to have a camera (Figure 5, 6).

This module would perform much more task than just merely copying other's facial expression. We are

developing a prototype for a virtual dialog system where a virtual actor communicates with a real person by the analysis of facial expression.

Figure 5. Facial Communication

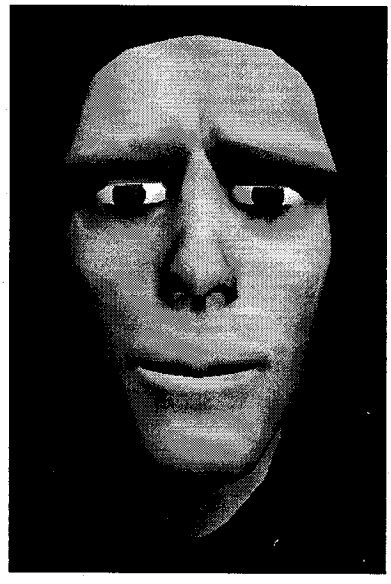

Figure 6. A facial expression of doubt

4 Integration of Facial Analysis and Synthesis in a Intelligent Tutoring Systems (ITS)

4.1 Face expression input for the ITS

One possible application of the facial analysis and synthesis modules is for developing an autonomous intelligent tutor that can communicate and exchange a dialog with a real person through expression recognition.

A possible architecture for the system is shown on Figure 7. The input to the analyzer is the facial expression recognition module. The result of the analysis can provide the expressions recognition of the user. The virtual actor responds to a real person, a data base is used with content words with facial expression states. These are defined in terms of constituent phonemes and facial expressions.

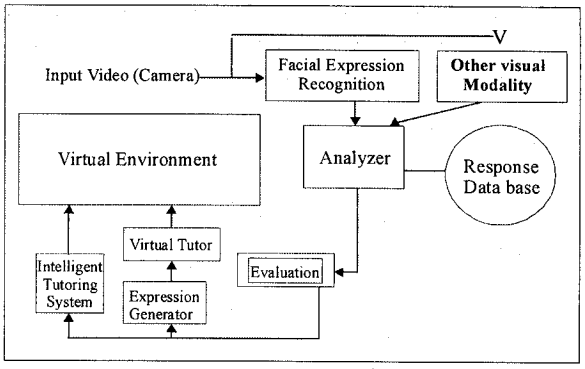

Figure 7. Global structure of an intelligent tutoring system which uses facial analysis and synthesis

The series of timed phonemes and expressions are then compiled and decomposed into the basic action units to produce the visual changes to be done on the virtual tutor's face. The words to be spoken are transmitted to the vocal synthesizer module.

4.2 Interacting with the ITS Through Face Expressions

Input from facial expression may be used to interact with the intelligent tutoring system. Relevant expressions are assigned with a virtual tutorial system when it is important to know the user interest on the information that is displayed or when the user is interacting in a virtual environment.

4.3 Interacting with Hand Movements

Other possibility of interaction, is to complement facial expressions, with hand movements allowing the user a more natural interaction. It is possible to define a special gesture language to define the meaning of hand movements or use sign language (Starner et al. 98).

Conclusion and future work

In this paper we have described a human-computer interface based on face analysis and synthesis, that enhances the communicative power of intelligent tutoring systems. The analysis allows facial expression recognition, while synthesis renders a realistic virtual tutor which could be complemented with synthetic speech.

We are currently working with the deformable model of the face to test face and emotion recognition, which is relevant for people identification as well as human-computer interaction.

Acknowledgments

This work has been funded by the Mexican National Council of Science and Technology (CONACYT) as project ref. C098-A and C100-A, "Gesture recognition interfaces and intelligent agents for virtual environments".

References

Bruce, V. and Green, P. (1989), *Visual Perception: Physiology, Psychology and Ecology*. Lawrence Erlbaum Associates, London.

Campbell, L.W., Becker, D.A., Azarbayejani, A., Bobick, A., and Pentland, A. (1996). Invariant features for **3-D gesture recognition.** *MIT Media Laboratory Perceptual Computing Section,* Techni cal Report no. 379.

Cassell, J., Pelachaud, C. Badler, N. et. al (1994) Animated Conversation: Rule-based Generation of Facial Expression, Gesture and Spoken Intonation for Multiple Conversational Agents. *Computer Graphics Proceedings*, pp.413-420.

Chellapa, R., Wilson, C.L. and Sirohey, S. (1995). Human and machine recognition of faces: a survey. *Proceedings of the IEEE,* Vol. 83, No.5, pp. 705-740

Cohen, M. and Massaro, D. (1993) Modeling Coarticulation in Synthetic Visual Speech, *Models and Techniques in Computer Animation,* Springer Verlag, pp. 139-156.

Cootes, T.F., et al. (1992). Training models of shape from sets of examples. *Proceedings of the British Machine Vision Conference.*

Cootes, T.F., Edwards, G.J. and Taylor, C.J. (1998). *Proceedings of the European Conference on Computer Vision,* Burkhardt, H. and Neumann, B. (Eds.)., Vol. 2, pp. 484-498, Springer-Verlag.

Eisert, P. and Girod, B. (1998). Analyzing facial expressions for virtual conferencing. *IEEE Computer Graphics and Applications*, Vol. 18, No. 5, pp. 70-78.

Ekman, P. and Friesen, W.V. (1975). *Unmasking the Face: A Guide to Recognizing Emotions from Facial Expressions*, Englewood Cliffs, New Jersey; Prentice Hall, Inc.

Ekman, P. and Friesen, W.V. (1978), *Manual for the Facial Action Coding System.* Consulting Psychologists Press, Inc. Palo Alto, CA.

Lam, K.M. and Yang, H. (1996). ``Locating and extracting the eye in human face images". *Pattern Recognition*, Vol. 29, No. 5, pp. 771-779.

Ohzu, H. and Habara, K. (1996). Behind the scenes of virtual reality: vision and motion. *Proceedings of the IEEE*, Vol. 84, No. 5, pp. 782-798.

Oliver, N., Pentland, A. and Berard, F. (1997). Lafter: lips and face real time tracker. *MIT, Media Laboratory Perceptual Computing Section,* Technical report no. 396.

Maggioni, C. and Kammerer, B. (1998). GestureComputer - History, Design and Applications. *In Computer Vision for Human-Machine Interaction*, Cipolla, R. and Pentland, A. (Eds.)., pp. 23-52, Cambridge University Press.

Parke, F.I. and Waters, K. (1996). *Computer Facial Animation.* A K Peters.

Pentland, A.P. (1996). Smart rooms. *Scientific American.* April 1996, pp. 54-62.

Pentland, A.P. (1998). Smart rooms: Machine understanding of human behavior. *In Computer Vision for Human-Machine Interaction*, Cipolla, R. and Pentland, A. (Eds.)., pp. 3-22, Cambridge University Press.

Rios, H.V. and Peña, J. (1998). Computer Vision interaction for Virtual Reality, In *Progress in Artificial Intelligence*, Helder Coelho (Ed.), *Lecture Notes in Artificial Intelligence, Subseries of Lecture Notes in Computer Science*, No. 1484, pp. 262-273. Springer-Verlag.

Russel, J. (1994). Is There Universal Recognition of Emotion From Facial Expression?. *A Review of Cross-Cultural Studies Psychological Bulletin* 115(1) 102-141.

Schwartz, E.I. (1995). A face of one's own. *Discover* the world of Science. Vol. 16, No. 2, pp. 78-87.

Starner, T., Weaver, J., Pentland, A. (1998). Real-time American sign language recognition using desk and wearable computer based video. *MIT, Media Laboratory Perceptual Computing Section,* Technical report no. 466.

Sucar, L.E. and Gillies, D.F. (1994). Probabilistic reasoning in high-level vision. *Image and Vision Computing,* Vol. 12, No. 1, pp.42-60.

Terzopoulos, D. And Szeliski, R. (1992). Tracking with Kalman Snakes. In *Active Vision,* Blake, A. and Yuille, A. (Eds.), MIT Press.

Waters, K. (1987). A Muscle Model for Animating Three Dimensional Facial Expression. *Computer Graphics Proceedings* Vol. 21, No 4 , pp.17-23.

Waters, K. and Levergood, T.(1994) An Automatic Lip-Synchronization Algorithm for Synthetic Faces. *Proceedings of Multimedia 94, ACM*, pp149-156.

Homero V. Rios-Figueroa, *received the B.Sc. degree in Mathematics and the M.Sc. degree in Computer Science from the National Autonomous University of Mexico (UNAM) in 1987 and 1989, respectively. The Ph.D. degree in Computer Science and Artificial Intelligence from the University of Sussex, England, in 1994. From 1988 to 1990 he was a Lecturer with the UNAM. From 1994 to 2000, he was a full time researcher at the National Laboratory for Advanced Informatics (LANIA) in Xalapa, Mexico. From 2000 to 2001, he was the academic head of the Master Program in Artificial Intelligence of the University of Veracruz (UV). He is currently the Central Administrator of Information Technology Planning of the Internal Revenue Service (SAT) of the Government of Mexico. His research interests include computer vision, virtual reality and information technology.*

Ana L. Solís-González-Cosio, *received the B.Sc. degree in Mathematics from the National Autonomous University of Mexico (UNAM), and also holds a speciality on computer graphics obtained at Japan. She is currently a lecturer at UNAM.*

Emilio Aguirre, *received a B.Sc. in Computer Science from Anahuac University, and a Masters in Artificial Intelligence, from the University of Veracruz in 2000. Currently, he is working as research assistant at Montreal University.*

Lourdes Guerrero, *received a B.Sc. in Mathematics, and M.Sc. degree in Computer Science from the National Autonomous University of Mexico (UNAM). She is currently a lecturer at Faculty of Sciences at UNAM.*

Joaquín Peña-Acevedo, *received a B.Sc. in Mathematics from the University of Veracruz. Currently he is studying for a M.Sc. degree in Applied Mathematics, at the Research Center in Mathematics (CIMAT).*