IMEKO, IEEE, SICE
2nd International Symposium on Measurement, Analysis and Modeling of Human Functions
1st Mediterranean Conference on Measurement
June 14-16, 2004, Genova, Italy

AUTOMATIC FACIAL EXPRESSION RECOGNITION BY MEANS OF A NEURAL APPROACH

Giovanni Pioggia[a], Marcello Ferro[a], Janusz Kozlowski[c], Andrea Marchetti[a], Fabio Di Francesco[b], Arti Ahluwalia[a], Danilo De Rossi[a]

[a]Interdepartmental Research Centre "E. Piaggio", [b]CNR , [c]AILUN*

1. INTRODUCTION

René Descartes first suggested that facial expressions can be considered as a non-verbal communication of emotions and moods. As many illuminist followers, Descartes included facial expression as a part of the animal behaviour of human beings. In the nineteenth century, Charles Darwin first considered that facial expression and expression of emotions were adaptative, evolutionist. Furthermore, Darwin regarded non-verbal expression of emotions as being linked to physiological aspects. He took pictures and drew the facial expressions of his sons and studied the expression of emotions during growth. In the last century anatomical studies were added, to philosophical, behavioural, biological and physiological studies of emotion. In particular, in the 1970s, Ekman's research culminated in a system for the detailed description of facial expressions (FACS, Facial Action Coding System), and were an important starting point in the exploration of human visual communication [1]. Nowadays the non-verbal expression of emotional states has become an area of interest in computer science and robotics embracing important streams of thought in other fields, such as neurosciences, cognitive science, psychology, ethology. The importance of facial emotional expressions in human-human and human-machine communication has been recognised in applications such as robotics and Human-Computer Interaction (HCI). Several systems for automatic analysis, interpretation and categorisation of some basic human facial expressions have been developed. We at the University of Pisa are involved in an ambitious project to develop a believable facial display founded on the simulation of biological behaviour using materials, structures and control algorithms which can acquire and replicate some of the functions and responses of living systems [2]. The system is called FACE: Facial Automaton for Conveying Emotions. At present, the immediate objective is focused upon exploring its use in social skills and emotional therapy in individuals with autism. Among other things, FACE possesses a system for automatic facial expression recognition which is the topic of this paper. The underlying principle is based on measurement of the average curvature of the interlocutor using a 3D contouring system. Data are analysed by a hierarchical neural-network architecture based on self-organizing maps and a back-propagation perceptron.

2. METHOD

Recently, automatic face recognition has become a popular application field. A commercial example is the system "FaceIt", designed at the Computational Neuroscience Laboratory at Rockefeller University. Nevertheless. FACS represent an effective aid for the problem of automatic recognition of facial expressions; but resolving a given expression still remains a challenging topic. FACS classifies the movements of forty muscles in the face as action-units (AUs). The movements of a group of muscles in order to obtain a particular expression represents a single AU. Thus, a facial expression can be captured by an artificial vision system, or simply a camera, and by means of a suitable algorithm divided into its component AUs and subsequently identified by FACS coding. In the field of facial expression recognition, Waters and Terzopoulos developed a model of the skin and facial muscles based on FACS [3]. An artificial vision system acquires images and converts them into two dimensional potential functions whose local minima correspond to salient facial features (snakes). The snakes are tracked through each image at each time step. With the aim of representing muscle actions that lead to a skin deformation which matches with the image analysed, a 3D model of the face is deformed to approximate the AUs involved. By using this procedure, a facial expression can be identified. On the other hand, we adopted a neural approach to allow FACE to recognize the expression of a subject. It is based on the measurement of the average curvature that has been calculated using a 3D contouring system based on out-of-phase sinusoidal fringe pattern projection [4,5], built at AILUN*. This solution enables very fast image acquisition (about 40ms) in conditions of normal day-light ambient illumination, providing complete information about shape and texture of a (not necessarily immobile) human face. Once a curvature map of the head of a subject is obtained, a dedicated process detects a number of points (markers), which are used to divide the human face into four main areas (left eye, right eye, nose and mouth). The data of each area are processed by a Hierarchical Neural-Network (HNN) architecture based on Kohonen Self Organizing Maps (KSOMs) and Multi-layer Perceptron (MLP). The first module of the architecture is composed of four KSOMs that act as independent classification units. Curvature data of a zone is the input to only one map; in this way, each KSOM is trained (unsupervised learning) with the purpose of clustering data coming from the respective zones into crisp classes (pre-classification). The outputs of this pre-classification step are sent to the second module of the network, defined by a MLP, whose task is to perform the final classification. The outputs of this module are used to form the input pattern for the second module, which defines the group to which the facial expression belongs. For each facial expression image data is a curvature matrix at a resolution of 480x640 dots. The incoming curvature matrix is then split in four submatrices corresponding to the four selected areas. These submatrices are analysed by the HNN composed of as many identical KSOMs as the number of main areas and by a one-hidden-layered MLP. Several configurations have been tested to achieve the purpose of finding the optimal neural network's parameters. Here we show the results for 18 configurations obtained by combining 2 different resolutions for curvature submatrices of main areas, 3 different dimensions for KSOMs and 3 different dimensions for the MLP's hidden layer.

3D Contouring System

The initial data is composed of two images of the measured surface illuminated twice with sinusoidal fringe patterns shifted in phase mutually by ??? out-of-phase). The surface was illuminated and observed from a distance of 1.4 m, and with the distance between illuminating system and CCD camera of 80 mm (the base of triangulation of 3.3°). The resolution of the applied camera (Dragon Fly type) was of 480x640 pixels, with Bayer mosaic, and of 10 bit dynamics.

First, the distribution of the normalised difference between two registered images, with respect to their sum, has been calculated. In the next step the Hilbert transform of this (real-value) distribution generates the complex-value fringe-like pattern, the phase of which is related to the shape of the analysed surface.

The sensitivity of the applied system enables us to detect changes/variations of the surface shape of order of 0.2 mm and is satisfactory for the present application. The average curvature of the analysed surface in point P may be defined as difference between its height at this point, and the average height in the r-radius neighbourhood of P.

In this work, instead of height, the phase of the intermediate complex distribution has been considered. This approach enables one to skip (otherwise necessary) the phase unwrapping process, and to speed-up calculus (full 3D form in 6 sec., curvature in 2 sec.). To simplify this procedure the average (complex) value over the r-radius neighbourhood (ring) was calculated and then, the phase of this average value was subtracted from the phase of the central point complex-value. The result was calculated as 16 bit integer (short) and then nonlinearly transformed and stretched into the range of 784 levels, enabling brightness-independent false colour representation.

Facial Expression Estimator (FEE)

In FACE the module representing the head of the facial expression analysis is the Facial Expression Estimator (FEE), that receives as input the features extracted from the pattern preprocessing which are output by the acquisition system. Once the input patterns have been received, in order to classify the facial expression, the hierarchical neural network needs a preprocessing step that consists in locating key points in the input patterns. These points allow the algorithm to recognize the triplet "eyes-nose-mouth" and subsequently send the patterns corresponding to the four areas that form the triplet (right eye, left eye, nose, mouth) to the neural network.

The sequence of images of the human subject's face is first acquired and then, for each frame, four key points are located: right side of the right eye, left side of the left eye, tip of the nose and bottom of the mouth. Starting from such information it is a simple task to calculate the line joining the eyes and to recognize the four main areas. A high precision in locating key points is not needed since the committed error during position calculation will be easily compensated up by the neural network.

The Hierarchical Neural Network (HNN)

The neural network consists of a hierarchical architecture with two different modules. The first module is a parallel classifier composed of Kohonen Self Organizing Maps (KSOMs), which receive the input data from the pre-processing unit. The analysis of these data permits a pre-classification by means of extraction of characteristic features. The outputs of this pre-classification step are sent to the second module of the network, defined by a Multilayer Perceptron (MLP), whose task is to combine the inputs properly and perform the final classification.

Each KSOM of the first module acts to separate the input data into crisp classes. This pre-classification is performed without any information on the facial expression to be recognised (i.e. unsupervised learning), but using only the appropriate features for each of the KSOMs. The outputs of this first module are used to form the input pattern for the second module, which refines the classification and gives the group to which the expression belongs as a final response. Implementation by means of two independent modules implies rapid and efficient training.

The first module is composed of a set of k classifiers. Each unit is trained with the purpose of clustering each input pattern into specific classes of characteristics, without using any information related to the facial expression to which the input belongs. Each specific input pattern is the input to only one classifier of the first module; in this way, each KSOM can be singularly optimised without affecting the other modules of the global architecture, in order to reduce the computational complexity locally and, at the same time, to implement a flexible system.

3. RESULTS

In order to generate the seven basic facial expressions, the subject was asked to express in turn neutrality, happiness, surprise, anger, disgust, sadness and fear. Each of these was then altered by translating and rotating randomly the original curvature matrix. A total of 200 different images for each expression was thus generated, obtaining a total of 1400 input patterns. The first half were used to train the HNN and the second half to test it. The artificial vision system is still undergoing optimisation, for instance the acquisition of 1400 real expressions is yet a highly time consuming task.

The information needed by the neural network to understand the input patterns is hown in table 1.

Description	*Variable*	*Value*
Dimension of the input matrix to FEE	$W_0 x H_0$	480 x 640
Range of the acquisition system	$[c_{\min}, c_{\max}]$	[-1024, 1023]
Scale factor of FEE	Q_c	256

Table 1 – Input variables to the HNN

To identify optimal values for network's characteristic parameters we tested many configurations, as shown in tables 2 and 3.

Conf. number	*Right eye map*	*Left eye map*	*Nose map*	*Mouth map*
	$W_1 x H_1$	$W_2 x H_2$	$W_3 x H_3$	$W_4 x H_4$
1	15 x 15	15 x 15	15 x 20	30 x 10
2	15 x 15	15 x 15	15 x 20	30 x 10
3	15 x 15	15 x 15	15 x 20	30 x 10
4	15 x 15	15 x 15	15 x 20	30 x 10
5	15 x 15	15 x 15	15 x 20	30 x 10
6	15 x 15	15 x 15	15 x 20	30 x 10
7	15 x 15	15 x 15	15 x 20	30 x 10
8	15 x 15	15 x 15	15 x 20	30 x 10
9	15 x 15	15 x 15	15 x 20	30 x 10
10	25 x 25	25 x 25	25 x 30	40 x 20
11	25 x 25	25 x 25	25 x 30	40 x 20
12	25 x 25	25 x 25	25 x 30	40 x 20
13	25 x 25	25 x 25	25 x 30	40 x 20
14	25 x 25	25 x 25	25 x 30	40 x 20
15	25 x 25	25 x 25	25 x 30	40 x 20
16	25 x 25	25 x 25	25 x 30	40 x 20
17	25 x 25	25 x 25	25 x 30	40 x 20
18	25 x 25	25 x 25	25 x 30	40 x 20

Table 2 – Dimensions of the input matrices for each configuration

Conf. number	*1st classification level*	*2st classification level IL | HL | OL*
	$KSOM_{1..4}$	MLP
1	3 x 3	28 | 5 | 7
2	3 x 3	28 | 10 | 7
3	3 x 3	28 | 15 | 7
4	7 x 7	28 | 5 | 7
5	7 x 7	28 | 10 | 7
6	7 x 7	28 | 15 | 7
7	10 x 10	28 | 5 | 7
8	10 x 10	28 | 10 | 7
9	10 x 10	28 | 15 | 7
10	3 x 3	28 | 5 | 7
11	3 x 3	28 | 10 | 7
12	3 x 3	28 | 15 | 7
13	7 x 7	28 | 5 | 7
14	7 x 7	28 | 10 | 7
15	7 x 7	28 | 15 | 7
16	10 x 10	28 | 5 | 7
17	10 x 10	28 | 10 | 7
18	10 x 10	28 | 15 | 7

Table 3 – Parameters of the HNN

In table 4 the success percentage for each configuration is reported. The figure shows the recognition percentage chart.

	N.	*H.*	*Su.*	*A.*	*D.*	*Sa.*	*F.*	*Av.*
1	87	99	50	88	93	94	58	***81***
2	94	99	54	93	98	98	66	***86***
3	90	98	63	91	96	90	58	***84***
4	100	100	100	100	100	99	90	***98***
5	100	100	100	100	100	100	99	***99,8***
6	100	100	100	100	100	98	95	***99***
7	98	98	100	100	100	100	98	***99***
8	100	100	100	100	100	100	100	***100***
9	100	100	100	100	100	99	100	***99,8***
10	90	98	43	95	99	93	60	***83***
11	96	99	44	92	97	93	35	***80***
12	92	99	67	94	99	92	50	***85***
13	100	99	97	100	100	100	96	***99***
14	100	100	100	100	99	100	96	***99***
15	100	100	99	100	100	99	99	***99,5***
16	100	100	100	100	100	100	99	***99,8***
17	100	100	100	100	100	100	99	***99,8***
18	100	100	100	100	100	100	99	***99,8***

Table 4 – Facial expression recognition percentage

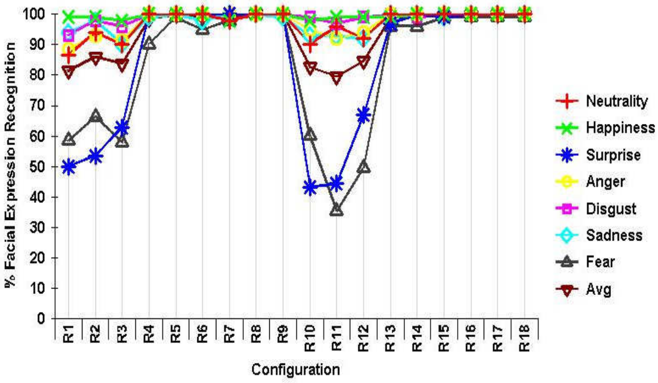

Figure – Facial expression recognition percentage chart

4. CONCLUSION

The development of FACE is a long-term project focused on the realization of biomimetic believable systems, which will evolve as new technological breakthroughs in materials engineering, control and other fields are made. In the immediate future, attention will be focused on the use of the FACE system in therapy for people on the autistic spectrum. Autism is a wide spectrum developmental disorder in which one of the characteristics is the inability to perceive mental states [6]. FACE could be used to train people with high functioning autism to recognise emotions and interpret and react to them appropriately. Although these results have to be considered as preliminary, our neural approach, leads to an automatic facial expression recognition for FACE, poses the basis to introduce new channels of interactivity in "intelligent" artificial systems.

REFERENCES

1. P. Ekman, E.L. Rosenberg, "What the face reveals: basic and applied studies of spontaneous expression using the facial action coding system (FACS)", *Oxford University Press*, New York, 1997.
2. G. Pioggia, A. Ahluwalia, F. Carpi, A. Marchetti, M. Ferro, W. Rocchia, D. De Rossi, "FACE: Facial Automaton for Conveying Emotions", *Applied Bio. and Biomech.*, in press.
3. D. Terzopoulos, K. Waters, "Analysis and synthesis of facial image sequences using physical and anatomical models", *IEEE Trans. on Pat. Anal. and Mac. Int., Special Issue on 3D Mod. in Image Analysis and Synthesis*, 15(6): 569-579, 1993.
4. J. Kozlowski, G. Serra, "A new modified PLL method for fringe pattern demodulation", *Opt. Eng.*, 36, 2025-30, 1997.
5. J. Kozlowski, G. Serra, "The Complex Phase Tracing method for fringe pattern analysis", *Appl. Opt.*, 38, 2256-62, 1999.
6. S. Baron-Cohen, "Mindblindness: An Essay on Autism and Theory of Mind", *MIT Press*, Cambridge (MA), USA, 1997.

Authors: Dr. G. Pioggia, M. Ferro, A. Marchetti, Prof. A. Ahluwalia, Prof. D. De Rossi, Interdepartmental Research Centre "E. Piaggio" University of Pisa, Via Diotisalvi 2, Pisa, Italy, Ph: +39050553639, Fax: +39050550650, pioggia@piaggio.ccii.unipi.it
Dr. F. Di Francesco, Istituto di Fisiologia Clinica, Area della Ricerca CNR, Via Moruzzi 1, Pisa, Italy
Dr. J. Kozlowski, *AILUN, Associazione per Istituzione Libera Università Nuorese, viale della Resistenza, 39, 08100 Nuoro, Italy