RECOGNITION OF FACIAL EXPRESSION BY IMAGE PROCESSING AND RADIAL BASIS FUNCTION NETWORK

Olímpio Murilo Capeli[1], Euvaldo Ferreira Cabral Junior[2], Sadao Isotani[3] e Antonio Roberto Pereira Leite de Albuquerque[4]

[1]Centro Universitário UNIFIEO, Rua Franz Voegeli, 300, 06020-190, Osasco, São Paulo

[2]Harmonic Science Institute., 211 Sugarwood Lane, Avon, IN 46123, Estados Unidos da América

[3]Instituto de Física, Universidade de São Paulo, Rua do Matão, Travessa R, 187, CEP 05508-090, São Paulo

[4]Universidade Paulista, Rua Dr. Bacelar, 1212, CEP 04026-002, São Paulo
capelliom@yahoo.com.br; euvaldojr@yahoo.com; sisotani@if.usp.br; antonio@albuquerque.pro.br

Abstract – In this work we report the application of a neural network-type Radial Basis Function in the recognition of facial expressions. The facial images were previously segmented to reduce the number of data. The results were satisfactory for the description of seven basic facial expressions of the human face.

Key-words – Facial expression, neural networks, image segmentation

Resumo – Neste trabalho nós reportamos a aplicação de uma rede neural do tipo Radial Basis Function no reconhecimento de expressões faciais. As imagens faciais foram previamente segmentadas para reduzir o número de dados. O resultado foi satisfatório para a descrição de sete expressões faciais básicas do rosto.

Palavras-chave – Expressão facial, redes neurais, segmentação de imagem

1. Introduction

The man-machine integration demands sophisticated abilities for facial recognition to be applied in the automation processes [Asada and Slotine 1986, Ito 1997, Kobayashi and Hara 1996, Terzopoulos 1993]. Several advanced research centers developed algorithms of facial recognition using parameters from cognitive psychology allied with mathematical models of signal processing. Also employs a technological branch of artificial intelligence known as man-machine communication.

The facial expression recognition is a complex field of investigation and was subject of considerable research [Hager and Eckman 1983, Faigin 1990, Freitas-Magalhães 2007]. There are many types of facial expressions, but there are six basic facial expressions for facial analysis [Ekman and Friesen 1978, Gil 1993]: happiness, fear, sadness, surprise, disgust, and anger. These six expressions together with the neutral emotion expression serve as means of communication in most of the cultures.

Artificial neural networks (ANN) have been one of the widely used computational tools for recognition of pattern and digital signals [Cabral Jr 2003, Bishop 2006, Theodoridis and Koutroumbas 2006]. Since facial expressions reflect emotions and other mental activities, social interaction and physiological signals [Fasel and Luettin 2003], artificial neural networks has been applied for their recognition [Kobayashi and Hara 1992, Rosenblum 1996, Lawrence et al 1997, Lisetti and Rumelhart, 1998, Cohen et al 2003, Ma and Khorasani 2004, Ioannou et al 2005].

The recognition of facial expressions using neural networks may require the use of large computational resources because of the increasing accuracy of images obtained from cameras and cell phones. In this case it is interesting to reduce the amount of data by removing the parts of images that are not of interest for the purpose of recognition of expressions. A good candidate for this, is for example a software for image processing that separates nuclei from cytoplasm and background

[Isotani et al 2007], because they are specialized in separation of delineated areas.

The interest in the RBF type neural networks has grown in several areas such as pattern recognition, digital processing signals, modeling and control of systems, because of its simple structure, well established theoretical basis and speed of learning [Timosczuk 2003].

We report in the present work, results of the application of image segmentation software to reduce the size of facial images and the use of a RBF network for the recognition process.

2. The image processing

To create a system capable of recognizing facial emotions it is natural that there is a concern with the image of the human face and the respective type of its computational treatment. Some restrictions should be observed before capturing the image. For example, movements of the head should be avoided, as well as glasses; the eyes should be kept wide open; hair, mustache or beard relatively short, and finally a stand still position should be held in front of the camera for the necessary period of time.

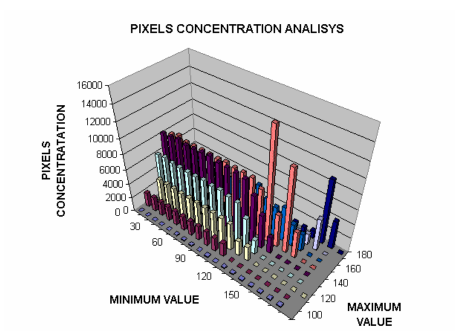

Figure 1. Empirical concentration of pixels.

Yacoob and Davis (1994) explain asymmetric facial muscle movements using a construction of lips during image capture, analysis and expression recognition by ANN. This showed that recognition depends on the good facial image computational process strategy. Mathematically, an image can be spread to a multispectral image and represented by a group of components $f_1(x,y)$, $f_2(x,y)$,....., $f_n(x,y)$ [Pratt 1978, Marion 1991, Russ 1995]. Then images are digitalized for a monochrome 2-D system, and for that reason, it is necessary to define a representation of function using three classic colors RGB (red, green, and blue) for the three functions $f_G(x,y)$, $f_R(x,y)$ and $f_B(x,y)$ which express the intensities for each point (x,y).

The main algorithms of this work were developed for processing microscopic images in morphologic clinical analyses [Isotani et al 2007]. Since this software collect images from a camera CCD, we adapted to capture images by other devices. In the present a work we used a video camera that sends the image to a computer frame-grabber interface in real time.

The dimension of facial image can reach approximately 200x300 pixels. In this case the neural networks demand long time processing. Since strategic regions for recognition are the two eyes, mouth, and nose regions, the facial images were processed to isolate these areas.

The values of segmentation are controlled using two pallets in the range [0,255]. As we can carry "max-min" values of pallets we can increase or decrease the dimension of images. A test applied to a typical facial image gives for max=150 and min=10 a segmented image with about 17.000 pixels, while for max=140 and min=60 the dimension decrease to 9000 pixels.

Figure 1 shows an example of variation of the concentration of pixels.

Figure 2 shows variation of the pixels number for each basic facial expression. An inspection in this figure show that the best values are around max=140 and min=80.

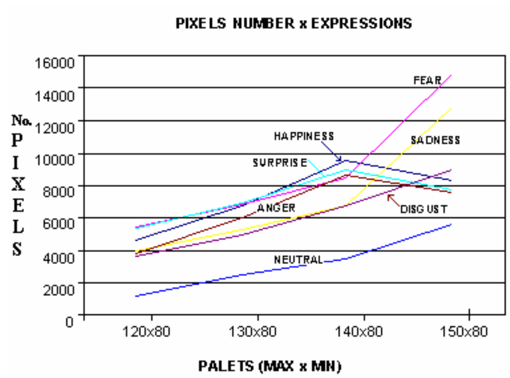

Figure 2. Pixels concentration and expressions

Figure 3 show the histogram of the input image. For choosing the threshold, we used the medium values (μ) and (σ) as the histogram deviation. The medium value μ is close to the peak of the light tones (more intense). In fact, the bi-modal characteristics of the histogram allow the adoption of (μ-σ) as threshold, and certainly this value will be in the intermediary area of the histogram. As a matter of fact, the main difficulty is the definition of the threshold. An outlet to the problem is through the **min/max** calculation based on statistical functions. The decision of adopting (μ-σ) is due to the simplicity and efficiency allied to the characteristics of the problem of image processing. This method allows us to make erosion/dilation on pixels regions, then the image dimension could be reduced for segmentation routine. The **max** is named ζ and **min** is τ, shows small histogram region for segmentation analysis on pixel range value.

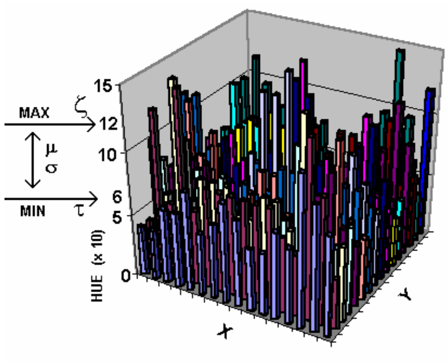

Figure 3. Hue spectrum histogram.

The number of pixels was reduced once more replacing RGB coordinates by intensity [Gomes and Velho 1994, Russ 1995], which is given by:

$$NEW\_HUE = \left(\frac{R+G+B}{3}\right) \div 1000 \qquad (1)$$

Figure 4 represents the three basic phases during the facial image processing. Figure 4A shows the original image, 4B represents the four strategic regions, the two eyes, nose and mouth, for the segmentation, and 4C the segmentation itself using pallets set with ζ=140 and τ=80 values. The numbers of figure 4C are sub regions detected using erosion/dilatation algorithm.

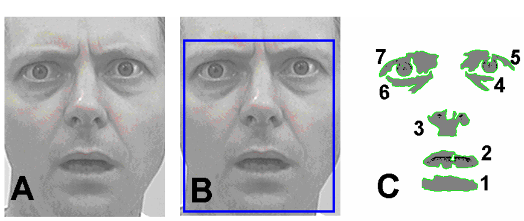

Figure 4. Facial image processing phases.

3. Artificial neural network RBF adapted to the facial model

Since RBF nets simulated cerebral human cortex, for active hide neurons $\mathbf{z_j}$ of ANN, this algorithm use gaussian function [Hassoun 1995] on Euclidean distance:

$$z_j(x) = exp\left(-\frac{\left\|x-\mu_j\right\|^2}{2\sigma_j^2}\right) \qquad (2)$$

Where the parameters **x** represent inputs, **μ** is medium and **σ** deviation of receptive fields on architecture of RBF [Timoszczuk 2003], and we calculated the centers of receptive field using our inputs, that is, on image data.

Figure 5 show the basic topology of a Radial Basis Function (RBF). It is able to receive input $\mathbf{x_i}$ signals, calculate $\mathbf{W_{zy}}$ weights in each hide neuron $\mathbf{Z_i}$, to use sum and transfer functions and its respective neurons $\mathbf{y_L}$ (or $\mathbf{d_L}$) of output layer. Finally, it provides desired training $\mathbf{d_L}$ and compare with test $\mathbf{y_L}$ information in the output, [Timoszczuk 2003] refers and gives in C^{++} programs main RBF algorithms in use for solutions pattern recognition.

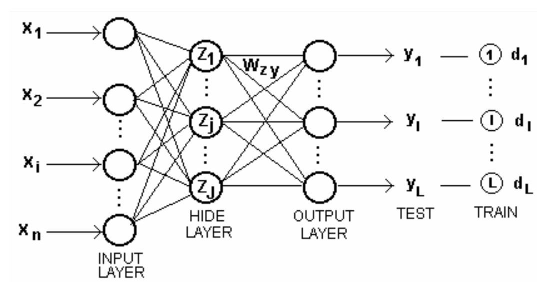

Figure 5. RBF topology.

Figure 6 exhibits the input signals of the image using RGB system going through the topology of a network.

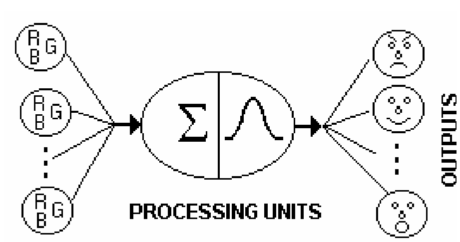

Figure 6. Simplified scheme of the ANN processing RGB signals.

4. Results

After the image processing, the total pixel number were reduced to around 20-30%.

A preliminary study of the application described above was conducted by using only two output channels (y_1, y_2). Figure 7 show the results of training in about 50 images. We observed that the outputs are grouped in clusters associated with facial expressions. This result, although preliminary, has given support to consider this a promising method.

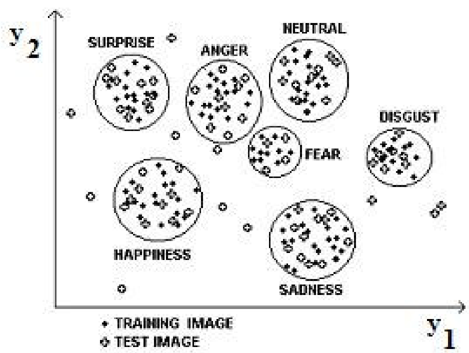

Figure 7. Radial Base function distribution.

Increasing the number of images up to 350 and through trial and error we have obtained the best results with 17000 inputs and 8 outputs. The values min-max threshold function were chosen in agreement with the local illumination, reflection of the light, face type and also considering small variations of the distance between camera and the analyzed person.

Table 1. Result of training with $\mathbf{C_1}$.

File	Age. Sex M=Male F=Female	Images	Threshold		Rate (%)
1	40 M	74	145	80	100
2	40 F	29	120	140	100
3	5 M	20	120	140	100
4	14 M	20	120	140	100
5	5 F	30	80	100	100
6	6 M	34	90	100	100
7	6 F	34	80	100	100
8	35 F	13	80	145	100
9	6 F	28	80	100	100
10	5 F	22	120	140	100
11	8 M	10	120	140	100
12	25 F	6	80	145	100
13	4 F	5	90	100	100
14	20 F	4	90	100	100
15	2 F	6	80	110	100
16	2 F	4	80	110	100
17	5 M	7	90	100	100
18	5 M	4	120	80	100

Three sets were prepared: $\mathbf{C_1}$ with total **n**=350 expressions images using, $\mathbf{C_2}$ using **n**=175 expressions images and $\mathbf{C_3}$ using **n**=175 complementary images, where $\mathbf{C_1 = C_2 \cup C_3}$, but $\mathbf{C_2}$ and $\mathbf{C_3}$ are disjoint sets.

Table 2. Result Crossing-Test for C_3 set.

File	Test Image Number	Image Reconized	Rate Test
1	38	29	76%
2	13	8	61%
3	10	8	80%
4	10	5	50%
5	15	9	60%
6	17	10	59%
7	17	9	53%
8	6	4	67%
9	15	12	80%
10	11	7	64%
11	5	2	40%
12	3	2	67%
13	2	2	100%
14	2	2	100%
15	3	2	67%
16	2	2	100%
17	4	2	50%
18	2	2	100%

After training the RBF network, with all the images of set C1, we performed a recognition test procedure using the same image set C1. Table 1 shows that the results of this test was 100%, showing that the training was successful.

Table 2 show a mean recognition ratio of 70% for a crossing-test, where $\mathbf{C_3}$ was put to be recognized using the parameters resulting from the training of $\mathbf{C_2}$.

A possible source of such imprecision can be associated to the low quality of the images, since they were obtained using a video camera without any further care, like precise distance, light source and other factor.

Despite some rate below that is expected, the results obtained by this method are satisfactory. Using RBF and expanding data base, we expect to improve the recognition rate of the Crossing-Test.

Finally, there are possibilities for use this modular ANN program for automatization of input data and use on other applications, for example: face morphology [Acton and Mukherjee 2000], facial paralysis [Ahrens et al 1999], face disturb and schizophrenia [Iwase et al 1999], and others.

Acknowledgements

This work was supported by grants from CNPq, FAPESP and FINEP.

References

Acton, S. T., Mukherjee, D. P. (2000) Scale space classification using area morphology. *IEEE Transactions on Image Processing*, **9**: 623-635, Apr..

Ahrens, A., Skarada, D., Wallace, M., Cheung, J. Y. and Neely, J. G. (1999) Rapid simultaneous comparison system for subjective grading scales for facial paralysis. *American Journal of Otology*, **20**: 667-71.

Asada, H. and Slotine, J. J. E. (1986) Robotic analysis and control. New York, John Wiley.

Bishop, C. M. (2006) Patern recognition and machine learning, Springer.

Cabral, Jr., E. F. (2003) Redes neurais artificiais: modelos em C, Grupo Reneart, São Paulo.

Cohen, I, Sebe, N., Cozman, F. G., Cirelo, M. C. and Huang, T. S. (2003) Learning Bayesian Network Classifiers for Facial Expression Recognition using both Labeled and Unlabeled Data, *Proceedings of 2003 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**: 595.

Ekman, P. and Friesen, W. V. (1978) The facial action coding system, Consulting Psychologists Press, Palo Alto.

Faigin, G. (1990)The Artist's Complete Guide to Facial Expression, Watson-Guptill, Publications, New York.

Fasel, B. and Luettin, J. (2003) Automatic facial expression analysis: a survey, *Pattern Recognition*, **36**(1): 259-275.

Freitas-Magalhães, A. (2007). The Psychology of Emotions: The Allure of Human Face. Oporto: University Fernando Pessoa Press.

Gil, I. A. (1993) Relações entre ações faciais e relatos verbais de estados subjetivos de emoções e eventos correlatos. São Paulo, 183p. Tese (Doutorado) Instituto de Psicologia, Universidade de São Paulo.

Gomes, J. and Velho, L. (1994) Computação Gráfica: Imagem. Instituto de Matemática Pura e Aplicada. Sociedade Brasileira de Matemática. Rio de Janeiro.

Hager, J. C. and Ekman, P. (1983) The Inner and Outer Meanings of Facial Expressions. In Social Psycho-physiology: A Sourcebook, Eds. Cacioppo, J. T. and Petty, R.E., Chap. 10, The Guilford Press, New York.

Hassoun, M. H. (1995) Fundamental of Artificial Neural Networks. MIT Press, Cambridge, Massachusetts.

Ioannou, S. V., Raouzaiou, A. T., Tzouvaras, V. A., Mailis, T. PKarpouzis, K. C. and Kollias, S. D. (2005) Emotion recognition through facial expression analysis based on a neurofuzzy network, *Neural Networks*, **18**(4): 423-435.

Isotani, S., Brasil Fo, N., Capeli, O. M., Sakai, Y. I. and Albuquerque, A. R. P. L. (2007) A system for the determination of image density and morphometry developed for the analysis of malignant nuclei", *Proc. Twentieh IEEE Int. Symposium on Computer-Based Medical Systems*, Maribor, pp. 263-268.

Ito, M., Miyashita, Y. and Rolls, E. T. (1997) Cognition, Computation and Consciousness, Oxford University Press, New York, Tokyo.

Iwase, M., Yamashita, K., Takahashi, K., Kajimoto, O., Shimizu, A., Nishikawa, T., Shinosaki, K., Sugita, Y. and Takeda, M. (1999) Diminished facial expression despite the existence of pleasant emotional experience in schizophrenia. *Methods and Findings in Experimental and Clinical Pharmacology,* **21**: 189-194.

Kobayashi, H. and Hara, F. (1992) Recognition of Six basic facial expression and their strength by neural network. *Proceedings of the IEEE International Workshop on Robot and Human Communication*, Tokyo, pp. 381-386.

Kobayashi, H. , Hara, F. (1996) Analysis of the Neural Network Recognition Characteristics of 6 Basic Facial Expression, *JSME International Journal. Series C (Dynamics, Control, Robotics, Design and Manufacturing)*, **39**(2): 323-31.

Lawrence, S., Giles, C.L., Tsoi, A. C. and Back, A.D. (1997) Face recognition: a convolutional neural-network approach. *IEEE Transactions on Neural Networks*, **8**(1): 98-113.

Lisetti, C. L. and Rumelhart, D. E. (1998) Facial expression recognition using a neural network, *Proceedings of the 1998 International Florida Artificial Intelligence Research Symposium Conference (FLAIRS'98),* Menlo Park, pp 328-332.

Ma, L. and Khorasani, K. (2004) Facial Expression Recognition Using Constructive Feedforward Neural Networks. *IEEE Transactions on Systems, Man and Cybernetics-Part B: Cybernetics*, **34**(3): 1588-1595.

Marion, A. (1991) An introduction to image processing. Chapman and Hall, Great Britain.

Pratt, W. K. (1978) Digital image processing. John Wiley & Sons.

Rosenblum, M., Yacoob, Y. and Davis, L.S. (1996) Human expression recognition from motion using a radial basisfunction network architecture. *IEEE Transactions on Neural Networks*, **7**(5): 1121-1138.

Russ, J. C. (1995) The image processing handbook – 2^{nd} edition , IEEE Press.

Terzopoulos, D. (1993) Analysis and synthesis of facial image sequences using physical and automatical models. *IEEE Transactions on Pattern Analysis and machine Intelligence*, **15**(6): 569-579.

Theodoridis, S. and Koutroumbas, K. (2006) Pattern Recognition, Academic Press.

Timoszczuk, A. P. (2003) As redes artificiais do tipo Radial Basis Function (RBF). In Redes neurais artificiais: modelos em C. Editor Cabral Jr., E. Grupo Reneart, São Paulo.

Yaccob, Y. and Davis, L. S. (1994) Recognizing human facial expression. Computer Vision Laboratory Pub., Center for Automation Research University of Maryland, College Park.