A NEURAL NETWORK-BASED FACIAL EXPRESSION ANALYSIS SYSTEM

Ioanna-Ourania Stathopoulou and George A. Tsihrintzis
Department of Informatics
University of Piraeus
Piraeus 185 34
GREECE

ABSTRACT

We develop a neural network-based system for face detection and expression analysis for use in multimedia interactive services. Our system consists of two modules, a face detection module based on features extracted from Sinha's template and an expression classification module based on shape parameters of parts of the human face. Preliminary evaluation of our system shows that, after training with only about 290 face/non-face images and about 250 facial expression images, our system managed to detect faces with very small error rate even in very bad quality images and to discriminate among the "scream" and "neutral" expressions with high accuracy.

Keywords: Facial expression analysis, face detection, feature extraction, artificial neural networks and multimedia interactive services.

1. INTRODUCTION

Automatic detection of human faces and facial expression recognition is one of the most difficult and important problems in the scientific areas of pattern recognition and computer vision. Images that contain faces are instrumental in the development of more effective and friendly methods in human computer interaction and multimedia interactive services. Vision-based human-computer interaction methods assume that information about a user's identity, state and intent can be extracted from images, and that computers can then react accordingly [1,2]. Similar information can also be used in a security control system or in criminology to uncover possible criminals Facial expressions play a communicative role in interpersonal relations because they can reveal the affective state, cognitive activity, personality, intention and psychological state of a person.

The first step of a fully automated system that analyzes the information contained in face images is to detect and locate the faces in the images. Given a single image, the goal of face detection is to determine whether or not there are any faces in the image and, if so, return the location and extent of each face.

After successful detection of the face, facial expression analysis must determine the psychological state of a person, so that the computer can react accordingly. This problem is quite challenging because faces are non-rigid and have a high degree of variability in size, shape, color and texture. Furthermore, variations in pose, image orientation and conditions add to the level of difficulty of the problem. Moreover, the variability in the ways people express themselves, depending on their culture, psychological state and habits, make it even more difficult to determine one's psychological condition through his/her face image.

To address this problem, a number of works have appeared in the literature [e.g., 3-11]. Most of them preprocess the image and use a classifier to determine the user's facial expression.

In this paper, a new and efficient human facial expression analysis system is proposed that combines artificial neural networks and image invariants approaches. Our proposed system may be particularly useful in human-computer interaction and multimedia interactive service applications, where simplicity, speed and efficiency are prerequisite.

More specifically, the paper is organized as follows: in Section 2, we describe our face detection module, a preliminary version of which is presented in [12], and illustrate its performance. In Section 3, we present the facial expression analysis module of our proposed system, whereas in Section 4 we make an evaluation of performance and present some results. Finally, in Section 5 we draw conclusions and point to future work.

2. FACE DETECTION SYSTEM

For our proposed face detection algorithm, we defined certain image invariants and used them to detect faces by feeding them to an Artificial Neural Network. These image invariants were defined on the basis of the 14 by

16 pixel ratio template proposed by P. Sinha, [13, 14] which we summarize below.

2.1 P. Sinha's Template

The method proposed by P. Sinha [13, 14] combines template matching and image invariant approaches. Specifically, P. Sinha aimed at finding a model that would satisfactorily represent some basic relationships between the regions of a human face and found out that while variations in illumination change the individual brightness of different parts of faces (such as eyes, cheeks, nose and forehead), the relative brightness of these parts remains unchanged. This relative brightness between facial parts is captured by an appropriate set of pairwise brighter - darker relationships between sub-regions of the face.

The proposed template is seen in Figure 1 below, where we observe 23 pairwise relationships represented by arrows. The darker and brighter parts of the face are represented by darker and brighter shades of grey, respectively.

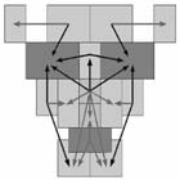

Figure 1: The P. Sinha's Template

2.2 The face detection algorithm and some results

Our proposed face detection algorithm is built on this model. We preprocess a candidate image to enhance the relationships mentioned above and then use the enhanced image as input to a fairly complex Artificial Neural Network to determine whether or not there is a face in the image. Our main goal is to form three clusters of the image using Principal Component Analysis with the Nystrom Algorithm [16, 17, 18, and 19] and the k-means algorithm so as to help the neural network detect faces. Typical results of our face detection system are depicted in the images in Figure 2.

3/5

2/2

Figure 2: Classification of two images by our system.
The first image is an old one of low quality. The system managed to the detect 3 out of the 5 faces. The second image is of better quality and our system detected 2 out of 2 faces.

3. FACIAL EXPRESSION ANALYSIS

After successful face detection, the system uses the determined image area to perform facial expression analysis. The basic algorithm for the facial expression analysis task can been seen in Figure 3.

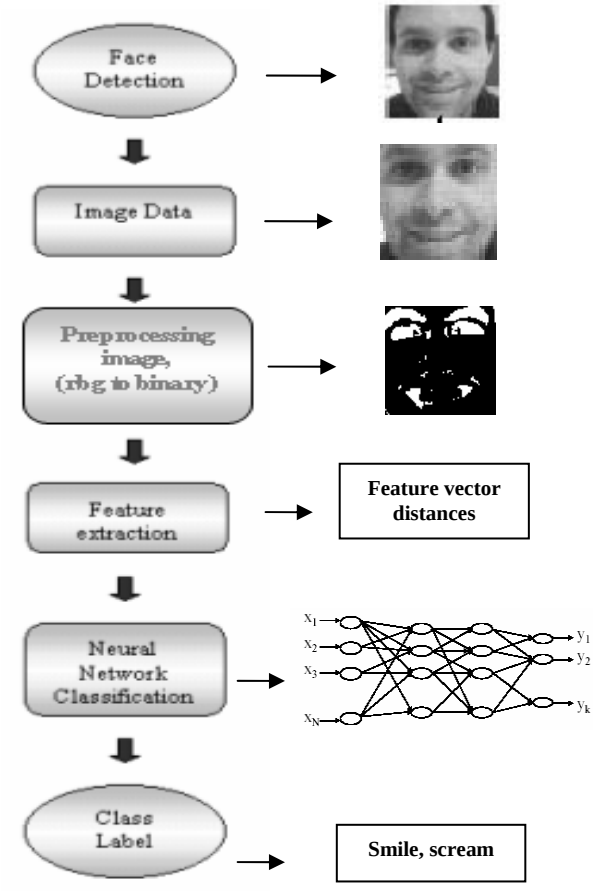

Figure 2: The Facial Expression Analysis Module

The face detection step determines the location and extent of the face in the image, the "window pattern", which is used by the facial expression analysis module for expression classification. In detail, our algorithm works as follows:

1. We *detect* the front view of the face. The image region defined by the face detection step constitutes the "window pattern" for our facial expression analysis module, which will be examined to determine the psychological state of the person.
2. We *preprocess* the "window pattern":
   2.1 We apply Histogram Equalization techniques to enhance the contrast within the "window pattern".
   2.2 We convert the image to binary and fill any holes in the binary image.
3. *Feature Extraction [21]*: We convert pixel data into a higher-level representation of shape, motion, color, texture and spatial configuration of the face or its components. This representation is used for subsequent expression categorization. Feature extraction generally reduces the dimensionality of the input space. The reduction procedure retains essential information with high discrimination power and stability. In Figure 4, we show facial points on which we base the extraction of expression classification features. We observe that, for the "scream" and "neutral" facial expressions, the location and shape of face parts vary significantly, which allows for the definition of expression classification features of high discrimination power.

Scream:

Neutral:

Figure 4: The extracted features
In the first column there are the input images and in the second column, the extracted features, depicted with gray points.

To define such features, we compute the Euclidean distances between the points in face parts and certain specific ratios of these distances.

4. We *apply* the resulting feature vector to an artificial perceptron-based neural network, which classifies the window pattern into one of two class labels which depict the psychological state of the person as either "scream" or "neutral".

5. EVALUATION OF PERFORMANCE

The feature extraction process aims at reducing the information fed into the neural network and categorize each facial expression according to some particular feature vector. In this way, we help the neural network in its expression classification task. Direct feeding of the entire face image into the network would require a significantly larger training set to achieve an acceptable performance level.

To train the neural network, we used a set of 250 images, 125 images for each expression. These images where gathered from sources of the World Wide Web [15, 20] and preprocessed before entered into the neural network. After training with this image set, the neural network achieved an error rate of 0.016. In fact, this error rate was gradually reduced as the number of the images in the training set increased. Given the relatively small size of the training set, this error rate is quite satisfying.

According to the requirements set, when the window pattern represented a neutral facial expression, the neural network should produce an output value of "one" (or something close to this value). On the other hand, if it represented a screaming face, the output value should be "zero". The output value can be regarded as the degree of membership of the face image in each of the "neutral" and "scream" classes. Typical results of the two neural networks can be seen in Figure 5.

0riginal Image:	***Extracted Features:***	***Network's Response:***
Scream:		
		**0**

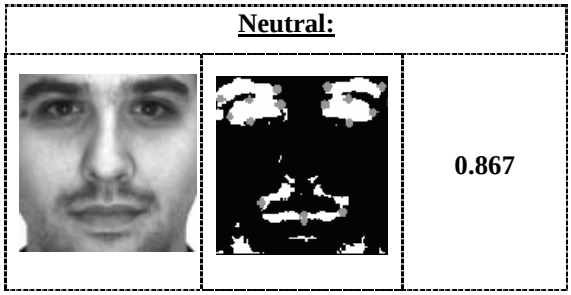

Figure 5: Classification of two images by our system.

6. SUMMARY, CONCLUSIONS, FUTURE WORK

We developed a neural network-based system for face detection and expression analysis for use in multimedia interactive services. Our system consists of two modules, a face detection module based on features extracted from Sinha's template and an expression classification module based on shape parameters of parts of the human face. Preliminary evaluation of our system shows that, after training with only about 290 face/non-face images and about 250 facial expression images, our system managed to detect faces with very small error rate even in very bad quality images and to discriminate among the "scream" and "neutral" expressions with high accuracy.

We plan to extend our work in the following three directions: (1) we plan to extend our system so as to cover more than the facial expressions "scream" and "neutral". These two expressions are more clearly distinct than other expressions and their classification may be easier. The classification of other facial expressions may require the extraction and tracing of additional facial points and corresponding features. (2) We will improve our system by using a wider training set so as to cover a wider range of poses and cases of low quality of images. (3) We plan to apply our system in human computer interaction techniques. We will develop a system, which will react according to the psychological state of the user and will be installed in computers and mobile phones. This and other related work is currently in progress and will be announced shortly.

6. REFERENCES

[1] Christine L. Lisetti, Diane J. Schiano, "Automatic Facial Expression Interpretation: Where Human-Computer Interaction, Artificial Intelligence and Cognitive Science Intersect", *Pragmatics and Cognition (Special Issue on Facial Information Processing: Multidisciplinary Perspective)*, Vol. 8(1): 185-235, 2000.

[2] Diane J. Schiano, Sheryl M. Ehrlich, Krisnawan Rahardja & Kyle Sheridan, "Face to Interface: Facial Affect in (Hu) Man and Machine", *Proceedings of ACM CHI 2000 Conference on Human Factors in Computing Systems* (pp. 193-200). NY: ACM.

[3] Essa. I. and A. Pentland, "Coding, Analysis, Interpretation and Recognition of Facial Expressions", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Volume 19 (7), IEEE Computer Society Press, July, 1997

[4] Dailey M.N., Cottrell G.W. and Adolphs R. (2000). A six-unit network is all you need to discover happiness. In *Proceedings of the Twenty-Second Annual Conference of the Cognitive Science Society*, Erlbaum, Mahwah NJ, pp. 101-106.

[5] Michael J. Black and Yaser Yacoob, "Recognizing Facial Expressions in Image Sequences using Local Parameterized Models of Image Motion", *Int. Journal of Computer Vision, 25(1), 1997, 23-48*

[6] Yaser Yacoob and Larry Davis, "Computing Spatio-temporal representation of human faces", *IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, June, 1994, 70-75, and IEEE Trans. on PAMI, 18(6), 636-642*

[7] Fabrice Bourel, Claude C. Chibelushi and Adrian A. Low, "Recognition of Facial Expressions in the Presence of Occlusion"

[8] M.S. Bartlett, G. Littlewort, B. Braathen, T.J. Sejnowski, and J.R. Movellan, "An Approach to Automatic Analysis of Spontaneous Facial Expressions"

[9] Irfan A. Essa and Alex P. Pentland, "Facial Expression Recognition using a Dynamic Model and Motion Energy", *International Conference on Computer Vision '95*, Cambridge, MA, June 20-23, 1995

[10] Shaogang Gong, Stephen J. McKenna, Alexandra Psarrou, "*Dynamic Vision :From Images to Face Recognition*", Imperial College Press,2000

[11] Hallinan P.W., Gordon G.G., Yuille A.L., Giblin P., Mumford D., "*Two- and Three- Dimensional Patterns of the Face*", A. K. Peters, Natick , Massachusetts

[12] Ioanna-Ourania Stathopoulou and George A. Tsihrintzis, "A new neural network-based method for face detection in images and applications in bioinformatics", *Proceedings of the 6th International Workshop on Mathematical Methods in Scattering Theory and Biomedical Engineering* , September 17-21, 2003

[13] P. Sinha, "Object Recognition via image-invariants"

[14] M.–H. Yang, N. Ahuja, *Face Detection and Gesture Recognition for` Human-Computer Interaction*, Kluwer Academic Publishers

[15] Gender Classification (Databases)
http://ise0.stanford.edu/class/ee368a_proj00/project15/intro.html
http://ise0.stanford.edu/class/ee368a_proj00/project15/append_a.html

[16] C. Fowlkes, S. Belongie and J. Malik," Efficient Spatiotemporal Grouping Using the Nystrom Method"

[17] S. Belongie, "Notes on Clustering Pointsets with Normalized Cuts" (2000)

[18] J. Shi and J. Malik, "Normalized Cuts and Image Segmentation", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 8, AUGUST 2000

[19] Image Segmentation using the Nystrom Method
http://rick.ucsd.edu/~bleong/

[20] A.M. Martinez and R. Benavente, "The AR face database", CVC Tech. Report #24, 1998.

[21] Paul Ekman and Erika Rosenberg, "*What The Face Reveals*", Oxford University Press 1997