EVALUATION OF THE DISCRIMINATION POWER OF FEATURES EXTRACTED FROM 2-D AND 3-D FACIAL IMAGES FOR FACIAL EXPRESSION ANALYSIS

Ioanna-Ourania Stathopoulou and George A. Tsihrintzis

Department of Informatics, University of Piraeus
Karaoli & Dimitriou 80, Piraeus 185 34
GREECE
phone: + 30 210 4142322, fax: + 30 210 4142264, email: {iostath, geoatsi}@unipi.gr

ABSTRACT

In previous works of ours [1-3], we proposed a neural network-based face detection and facial expression analysis system, which was able to classify three expressions in frontal view face images. In the present work, we examine the possibility of classifying these expressions in side view face images. Specifically, we evaluate the extracted facial feature discrimination power of three image acquisition techniques, namely acquisition in (1) frontal view, (2) side view and (3) stereoscopic pairs. Our findings are important in the design of human-computer interaction systems and multimedia interactive services.

Keywords: *Facial expression analysis, face detection, feature extraction, artificial neural networks.*

1. INTRODUCTION

Facial expressions play a significant communicative role in human-to-human interaction and interpersonal relations, because they can reveal information about the affective state, cognitive activity, personality, intention and psychological state of a person. It is common experience that the variety in facial expressions of humans is large and, furthermore, the mapping from psychological state to facial expression varies significantly from human to human. These two facts can make the analysis of the facial expressions of another person difficult and often ambiguous.

When mimicking human-to-human communication, human-computer interaction systems must determine the psychological state of a person, so that the computer can react accordingly. Indeed, images that contain faces are instrumental in the development of more effective and friendly methods in multimedia interactive services and human computer interaction systems. 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. Similar information can also be used in security control systems or in criminology to uncover possible criminals.

The task of processing facial images with the intent to analyze facial expressions 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, facial expression, image orientation and conditions add to the level of difficulty of the problem. The task is complicated further by the problem of *pretence*, i.e. the case of someone's facial expression not corresponding to his/her true psychological state.

To address this problem, a number of relevant works have appeared in the literature [e.g., 4-15]. Most of them preprocess a facial image, extract facial features and use a classifier to determine the corresponding person's facial expression. The automatic feature extraction and classification tasks are considered extremely difficult and important problems in the scientific areas of pattern recognition and computer vision. However, no reported research has addressed the question of determining which image acquisition technique achieves better results in classifying facial expressions.

In this paper, we introduce and evaluate three techniques for facial image acquisition, namely acquisition of images of faces in (1) frontal view, (2) side view, and, (3) in stereoscopic pairs. We compare these acquisition techniques and rank them according to their facial expression classification performance. More specifically, the paper is organized as follows: in Section 2, we present the facial image acquisition setups and the extracted facial features. In Section 3, we present the facial feature extraction algorithm and we extend this algorithm for the other two techniques in Section 4. Finally, in Section 5 we draw conclusions and point to future work.

2. FACIAL IMAGE ACQUISITION SETUPS

2.1 Geometry

In order to acquire the image data, we have built a three camera system, the architecture of which is presented in Figure 1. Specifically, three identical web cameras of 352*288 pixel resolution, are placed with their optical axes on the same horizontal plane. Successive axes are separated by 30 degree angles and subjects are asked to seat in front of the cameras and form facial expressions. All three cameras photograph the face of each subject simultaneously, so as to create a database of facial expressions as viewed from three different angles. To ensure spontaneity, the subject is presented with pictures on a screen behind the central camera. These pictures are expected to generate such emotional states that map on the subject's face as the desired facial expression. For example, to have a subject assume a "smile" expression, we show him/her a picture with funny content. We photograph the resulting facial expression and only then ask him/her to classify this expression. If the image shown to him/her has resulted in the desired facial expression, the corresponding photographs are saved to the disk and labelled; otherwise, the procedure is repeated with other pictures.

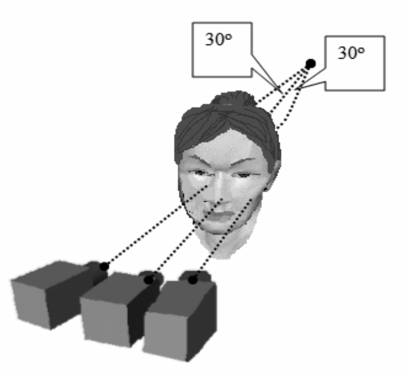


Figure 1: The geometry of the data acquisition setup

2.2 Facial Features

The main goal of feature extraction is to 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 terms of facial expression classification.

The extracted feature vector consists of the corner points of the eyes, mouth and brows, respectively. The extracted features as well as the distances used to classify the expressions can be seen in Figure 2.

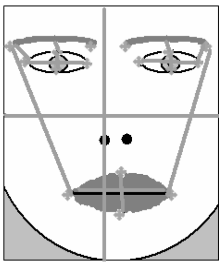


Figure 2: The extracted features depicted by gray points and the calculated distances (gray lines)

3. FEATURE EXTRACTION ALGORITHM

3.1 The Algorithm

Our feature extraction algorithm aims at finding the feature points, mentioned above and computing some basic distances between them. These distances vary significantly among the expressions. We use the discrimination power of these distances, in order to compute the performance of each image acquisition technique. Specifically, 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 Since the web camera's resolution isn't too good, we preprocess the image by applying various techniques [16], 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. We search the binary image of the face and extract each part of the face (eyes, mouth and brows) into a new image of the same size and coordinates as the original image.
4. In each image of a part of the face, we locate corner points using relationships between neighboring pixel values. This results in the determination of 16 points which form the feature vector. Typical results of the feature extraction algorithm in good quality images are seen in Figure 3 for the "neutral", "smile" and "surprise" facial expressions. In the first column, we can observe these expressions of a given person and, in the second column, the preprocessed image and the corresponding extracted features for each expression.
5. We compute the Euclidean distances between these points, depicted with gray lines in Figure 2, and certain ratios of these distances.
6. We feed the computed distances to an artificial neural network trained with images from face databases available over the web [17,18]

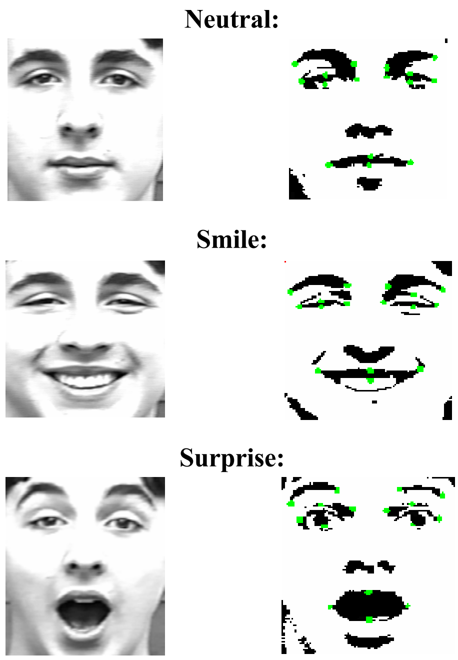


Figure 3: Typical extracted features.

3.2. Variations between facial expressions

The location and shape of face parts vary significantly for the three facial expressions: "neutral", "smile" and "surprise", which allows for the definition of expression classification features of high discrimination power. The distribution of these classification features (i.e., their values over a broad range of images of different persons in same expressions) is shown in Figures 4 and 5, respectively. Specifically in Figure 4, we plot the distribution of the mouth dimensions ratio for each of the examined facial expressions. Similarly in Figure 5, we plot the distribution of the face dimensions ratio relatively to the "neutral" expression.

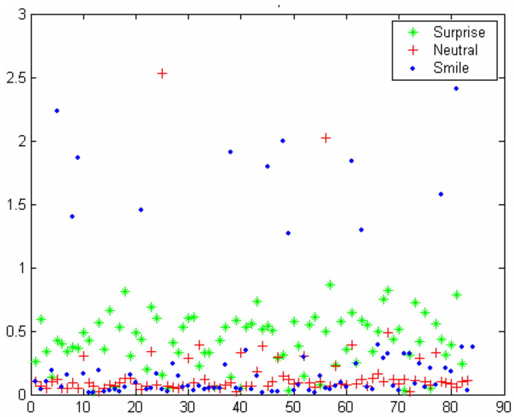


Figure 4: Distribution of mouth dimensions ratio

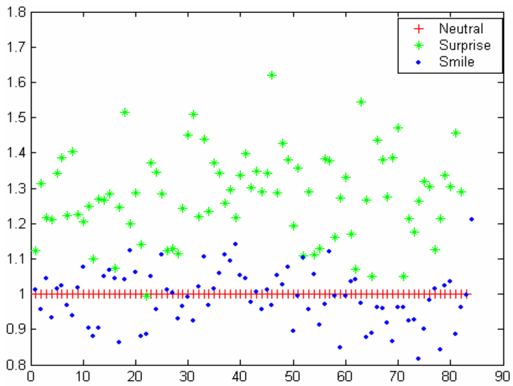


Figure 5: Distribution of face dimensions ratio

4. EXTENSION TO SIDE VIEW IMAGES

We apply the facial expression classification algorithm to faces in side and frontal view. Typical results of the algorithm are presented in Table 1. In the first column, we see the preproccessed 'enhanced' image. In the second and third columns, we show the corresponding binary image and extracted features, respectively. For side view images, features are extracted using only the visible side of the face and assume the same feature values for the other side. Finally, in the fourth column, we see the network's response.

Clearly, we achieve better classification in front view images for the "neutral" and "smile" expression, whereas the "surprise" expression is better classified in side view images. Better results in facial expression classification can be achieved by using alternative or additional views of a person's face. For example, the cheeks are more visible in side than in frontal view images.

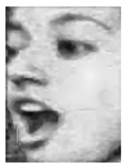
	Window Pattern	*Binary Image*	*Extracted Features*	*Network's response*
		Frontal View		
Neutral				**[0.765;0.232; 0.003]**
Smile				**[0.128;0.805; 0.067]**
Surprise				**[0;0.186 ;0.81 4]**
		Side View		
Neutral				**[0.622; 0.490; 0.052]**
Smile				**[0.395; 0.6; 0.005]**
Surprise				**[0; 0.060; 0.940]**

Table 1: Facial expression classification

5. FUTURE WORK

We plan to extend this work in the following directions: (1) we plan to extend our facial expression classification system [1-3] so as to cover more than the three facial expressions covered by the present system. 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 applications, such as those that arise in mobile telephony. (4) Finally, the use of fully 3-D (stereoscopic) vision techniques reveals depth-related features of the face, such as, for example, backwards motion of the head when certain facial expressions are formed.

Another extension of the present work of longer term interest will address several problems of ambiguity concerning the emotional meaning of facial expressions by processing contextual information that a multi-modal human-computer interface may provide. For example, complementary research projects are being developed [19-21] that address the problem of emotion perception of users through their actions (mouse, keyboard, commands, system feedback) and through voice words. This and other related work will be presented on future occasions.

6. REFERENCES

[1] Stathopoulou, I.-O. and Tsihrintzis, G.A., "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

[2] Stathopoulou, I.-O. and Tsihrintzis, G.A., "A neural network-based facial analysis system," *5th International Workshop on Image Analysis for Multimedia Interactive Services*, Lisboa, Portugal, April 21-23, 2004

[3] Stathopoulou, I.-O. and Tsihrintzis, G.A., "An Improved Neural Network-Based Face Detection and Facial Expression Classification System," *IEEE International Conference on Systems, Man, and Cybernetics 2004,* The Hague, The Netherlands, October 10-13, 2004.

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

[5] 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.

[6] 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.

[7] 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

[8] 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.

[9] 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*

[10] 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*

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

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

[13] 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

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

[15] 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

[16] Stathopoulou, I.-O. and Tsihrintzis, G.A., "Pre-processing and expression classification in low quality face images", *5th EURASIP Conference on Speech and Image Processing, Multimedia Communications and Services*, Smolenice, Slovak Republic, June 29 – July 2, 2005

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

[18] T. Kanade, J.F. Cohn, and Y. Tian, "Comprehensive Database for Facial Expression Analysis", Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition, Grenoble, France, 2000

[19] M. Virvou and E. Alepis, "Mobile educational features in authoring tools for personalised tutoring", In the journal *Computers & Education*, Elsevier, to appear (2004).

[20] M. Virvou and G. Katsionis, "Relating Error Diagnosis and Performance Characteristics for Affect Perception and Empathy in an Educational Software Application," *Proceedings of the 10th International Conference on Human Computer Interaction (HCII) 2003*, June 22-27 2003, Crete, Greece

[21] M. Virvou and E. Alepis, "Creating tutoring characters through a Web-based authoring tool for educational software," *Proceedings of the IEEE International Conference on Systems Man & Cybernetics,* 2003, Washington D.C., U.S.A.