TOWARDS AUTOMATED INFERENCING OF EMOTIONAL STATE FROM FACE IMAGES

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

Keywords: Facial Expression Classification, Human Emotion, Knowledge Representation, Human-Computer Interaction.

Abstract: Automated facial expression classification is very important in the design of new human-computer interaction modes and multimedia interactive services and arises as a difficult, yet crucial, pattern recognition problem. Recently, we have been building such a system, called NEU-FACES, which processes multiple camera images of computer user faces with the ultimate goal of determining their affective state. In here, we present results from an empirical study we conducted on how humans classify facial expressions, corresponding error rates, and to which degree a face image can provide emotion recognition from the perspective of a human observer. This study lays related system design requirements, quantifies statistical expression recognition performance of humans, and identifies quantitative facial features of high expression discrimination and classification power.

1 INTRODUCTION

Facial expressions are particularly significant in communicating information in human-to-human interaction and interpersonal relations, as they reveal information about the affective state, cognitive activity, personality, intention and psychological state of a person and this information may, in fact, be difficult to mask.

When mimicking communication between humans, 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 friendlier methods in multimedia interactive services and human computer interaction systems. Vision-based human-computer interactive systems 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. Studies have concluded to six facial expressions which arise very commonly during a typical human-computer interaction session and, thus, vision-based human-computer interaction systems that recognize them could guide the computer to "react" accordingly and attempt to better satisfy its user needs. Specifically, these expressions are: "neutral", "happy", "sad", "surprised", "angry", "disgusted" and "bored-sleepy".

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 and 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. These two facts make the analysis of the facial expressions of another person difficult and often ambiguous. This problem is even more severe in *automated facial expression classification*, as face images are non-rigid, have a high degree of variability in size, shape, color and texture and variations in pose, facial expression, image orientation and conditions add to the level of difficulty of the problem.

Towards achieving the automated facial image processing goal, we have been developing an automated facial expression classification system (Stathopoulou, I.-O. and Tsihrintzis, G.A.), called NEU-FACES, in which features extracted as deviations from the neutral to other common expressions are fed into neural network-based classifiers. Specifically, NEU-FACES is a two-module system, which automates both the face detection and the facial expression process.

Stathopoulou I. and A. Tsihrintzis G. (2007).
TOWARDS AUTOMATED INFERENCING OF EMOTIONAL STATE FROM FACE IMAGES.
In *Proceedings of the Second International Conference on Software and Data Technologies - PL/DPS/KE/WsMUSE*, pages 206-211
DOI: 10.5220/0001329802060211

To start specifying requirements and building NEU_FACES, we needed to conduct an empirical study first on how humans classify facial expressions, corresponding error rates, and to which degree a face image can provide emotion recognition from the perspective of a human observer. This study lays related system design requirements, quantifies statistical expression recognition performance of humans, and identifies quantitative facial features of high expression discrimination and classification power. The present work is the outcome of the participants' responses to our questionnaires.

An extensive search of the literature revealed a relative shortage of empirical studies of human ability to recognize someone else's emotion from his/her face image. The most significant of these studies are summarized next. Ekman and Friesen first defined a set of universal rules to "manage the appearance of particular emotions in particular situations" (Ekman, P., 1999; Ekman, P. & Friesen, W, 1975; Ekman, P., 1982; Ekman, P. et al., 2003; Ekman, P. & Rosenberg, E.L.). Unrestrained expressions of anger or grief are strongly discouraged in most cultures and may be replaced by an attempted smile rather than a neutral expression; detecting those emotions depends on recognizing signs other than the universally recognized archetypal expressions. Reeves and Nass (Reeves, B. and Nass, C.) have already shown that people's interactions with computers, TV and similar machines/media are fundamentally social and natural, just like interactions in real life. Picard in her work in the area of affective computing states that "emotions play an essential role in rational decision-making, perception, learning, and a variety of other cognitive functions" (Picard, R. et al., 1997, Picard, R.W., 2003). De Silva et al. (De Silva, L. C., Miyasato, T., and Nakatsu, P.) also performed an empirical study and reported results on human subjects' ability to recognize emotions. Video clips of facial expressions and corresponding synchronised emotional speech clips were shown to human subjects not familiar with the languages used in the video clips (Spanish and Sinhala). Then, human recognition results were compared in three tests: video only, audio only, and combined audio and video. Finally, M. Pantic et al. performed a survey of the past work in solving emotion recognition problems by a computer and provided a set of recommendations for developing the first part of an intelligent multimodal HCI (Pantic, M. et al., 2003).

In this paper, we present our empirical study on identifying those face parts that may lead to correct facial expression classification and on determining the facial features that are more significant in recognizing each expression. Specifically, in Section 2, we present emotion perception principles from the psychologist's perspective. In Section 3, we describe the questionnaire we used in our study. In Section 4, we show statistical results of our study. Finally, we summarize and draw conclusions in Section 5 and point to future work in Section 6.

2 EMOTION PERCEPTION

The question of how to best characterize perception of facial expressions has clearly become an important concern for many researchers in affective computing. Ironically, this growing interest is coming at a time when the established knowledge on human facial affect is being strongly challenged in the basic psychology research literature. In particular, recent studies have thrown suspicion on a large body of long-accepted data, even on studies previously conducted by the same people.

In the past, two main studies regarding facial expression perception have appeared in the literature. The first study is the classic research by psychologist Paul Ekman and colleagues (Ekman, P., 1999; Ekman, P. & Friesen, W, 1975; Ekman, P., 1982; Ekman, P. et al., 2003; Ekman, P. & Rosenberg, E.L.) in the early 1960s, which resulted in the identification of a small number of so-called "basic" emotions, namely *anger, disgust, fear, happiness, sadness and surprise (contempt was added only recently).* In Ekman's theory, the basic emotions were considered to be the building blocks of more complex feeling states (Ekman, P., 1999), although in newer studies he is sceptical about the possibility of two basic emotions occurring simultaneously (Ekman, P. & Rosenberg, E.L). Following these studies, Ekman and Friesen (Ekman, P. & Friesen, W, 1975) developed the, so-called, ''facial action coding system (FACS),'' which quantifies facial movement in terms of component muscle actions. Recently automated, the FACS remains the one of the most comprehensive and commonly accepted methods for measuring emotion from the visual observation of faces.

In the past few years, a second study by psychologist James Russell and colleagues summarizes previous works on human emotion perception (Russell, J. A., 1994) and challenges strongly the classic data (Russell, J. A., 2003), largely on methodological grounds. Russell argues that emotion in general (and facial expression of emotion in particular) can be best characterized in terms of a multidimensional affect space, rather than discrete emotion categories. More specifically, Russell claims that two dimensions, namely “pleasure” and “arousal,” are sufficient to characterize facial affect space.

Despite the fact that divergent studies have appeared in the literature, most scientists agree that:

- Human experience emotions in subjective ways.
- The “basic emotions” deal with fundamental life tasks.
- The “basic emotions” mostly occur during interpersonal relationships, but this does not exclude the possibility of their occurring in the absence of other humans.
- Facial expressions are important in revealing emotions and informing other people about a person’s emotional state. Indeed, studies have shown that people with congenital (Mobius Syndrome) or other (e.g. from a stroke) facial paralysis report great difficulty in maintaining and developing interpersonal relationships.
- Each time an emotion occurs, a signal will not necessarily be present. Emotions may occur without any evident signal, because humans are, to a very large extent, capable of suppressing such signals. Also, a threshold may need to be exceeded to bring about an expressive signal and this threshold may vary across individuals.
- Usually, emotions are influenced by two factors, namely *social learning* and *evolution*. Thus, similarities across different cultures arise in the way emotions are expressed because of past evolution of the human species, but differences also arise which are due to culture and social learning.
- Facial expressions are emotional signals that result into movements of facial skin and connective tissue caused by the contraction of one or more of the forty four bilaterally symmetrical facial muscles. These striated muscles fall into two groups:
  - four of these muscles, innervated by the trigeminal (5th cranial) nerve, are attached to and move skeletal structures (e.g., the jaw) in mastication
  - forty of these muscles, innervated by the facial (7th cranial) nerve, are attached to bone, facial skin, or fascia and do not operate directly by moving skeletal structures but rather arrange facial features in meaningful configurations.

Based on these studies and by observing human reactions, we identified differences between the “neutral” expression of a model and its deformation into other expressions. We quantified these differences into measurements of the face (such as size ratio, distance ratio, texture, or orientation), so as to convert pixel data into a higher-level representation of shape, motion, color, texture and spatial configuration of the face and its components. Specifically, we locate and extract the corner points of specific regions of the face, such as the eyes, the mouth and the brows, and compute their variations in size, orientation or texture between the neutral and some other expression. This constitutes the feature extraction process and reduces the dimensionality of the input space significantly, while retaining essential information of high discrimination power and stability.

3 THE QUESTIONNAIRE

In order to validate these *facial features* and decide whether these features are used by humans when attempting to recognize someone else’s emotion from his/her facial expression, we developed a questionnaire where the participants were asked to determine which *facial features* helped them in the classification task. In the questionnaire, we used images of subjects of a facial expression database which we had developed at the University of Piraeus (Stathopoulou, I.O. & Tsihrintzis, G. A., October 2006). Our aim was to identify the facial features that help humans in classifying a facial expression. Moreover, we wanted to know if it is possible to map a facial expression into an emotion. Finally, another goal was to determine if a human observer can recognize a facial expression from isolated parts of a face, as we expect computer-classifiers to do.

3.1 The Questionnaire Structure

In order to understand how a human classifies someone else’s facial expression and set a target error rate for automated systems, we developed a questionnaire in which each we asked 300 participants to state their thoughts on a number of facial expression-related questions and images.

Specifically, the questionnaire consisted of three different parts:

1. In the first part, the observer was asked to identify an emotion from the facial expressions that appeared in 14 images. Each participant could choose from the 7 of the most common emotions that we pointed out earlier, such as: “anger”, “happiness”, “neutral”, “surprise”, “sadness”, “disgust”,

"boredom–sleepiness", or specify any other emotion that he/she thought appropriate. Next, the participant had to state the degree of certainty (from 0-100%) of his/her answer. Finally, he/she had to state which features (such as the eyes, the nose, the mouth, the cheeks etc.), had helped him/her make that decision. A typical question of the first part of the questionnaire is depicted in Figure 1.

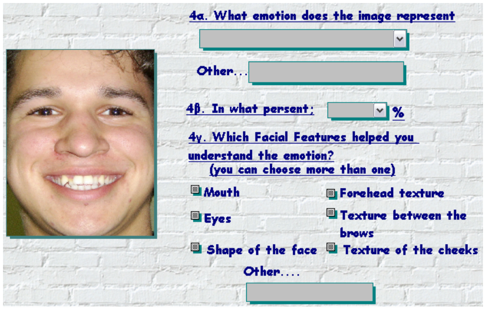

Figure 1: The first part of the questionnaire.

2. When filling the second part of the questionnaire, each participant had to identify an emotion from parts of a face. Specifically, we showed them the "neutral" facial image of a subject and the corresponding image of some other expression. In this latter image pieces were cut out, leaving only certain parts of the face, namely the "eyes", the "mouth", the "forehead", the "cheeks", the "chin" and the "brows." This is typically shown in Figure 2. Again, each participant could choose from the 7 of the most common emotions "anger", "happiness", "neutral", "surprise", "sadness", "disgust", "boredom –sleepiness", or specify any other emotion that he/she thought appropriate. Next, the participant had to state the degree of certainty (from 0-100%) of his/her answer. Finally, the participant had to specify which features had helped him/her make that decision.

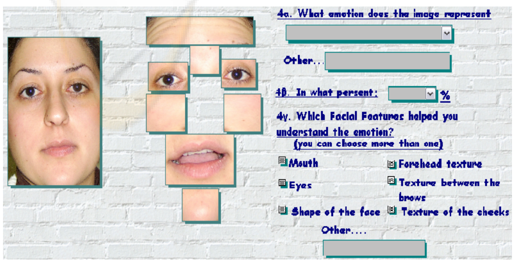

Figure 2: The second part of the questionnaire.

3. In the final (third) part of our study, we asked the participants to supply information about their background (e.g. age, interests, etc.). Additionally, each participant was asked to provide information about:
   - The level of difficulty of the questionnaire with regards to the task of emotion recognition from face images
   - Which emotion he/she though was the most difficult to classify
   - Which emotion he/she though was the easiest to classify
   - The percentage to which a facial expression maps into an emotion (0-100%).

3.2 The Participant and Subject Backgrounds

There were 300 participants in our study. All the participants were Greek, thus familiar with the greek culture and the greek ways of expressing emotions. They were mostly undergraduate or graduate students and faculty in our university and there age varied between 19 and 45 years.

4 STATISTICAL RESULTS

4.1 Test Data Acquisition

Most users agreed that a facial expression represents the equivalent emotion with a percentage of 70% or higher. The results are shown in Table 1.

Based on the participants' answers in the second part of our questionnaire it was observed that smaller error rates could be achieved if parts rather than the entire face image were displayed. The differences in error rates are quite significant and show that the extracted facial parts are well chosen.

An exception to this observation occurred with the "angry" and "disgusted" emotions where we observed a 6,44% and 5,10% increase in the error rate in the second part of our questionnaire. This is expected to be observed in the performance of automated expression classification systems when shown a face forming an expression of anger of disgust. More specifically, these differences in the error rates are shown in Table 2. As shown in the last column (P-value) these results are statistically significant.

Table 1: Percentage to which a facial expression represents an emotion.

Percentage to which an expression represents an emotion (%)	Percentage of user answers (%)
0	0,00
10	0,00
20	0,76
30	2,27
40	1,52
50	9,85
60	14,39
70	31,06
80	21,97
90	15,91
100	2,27

Table 2: Error rates in the two parts of the questionnaire.

Emotion	Error rates		Difference	P-value
	1st Part	2nd Part		
Neutral	61,74	----------	**61,74**	----------
Happiness	31,06	3,79	**27,27**	0,000000003747
Sadness	65,91	17,42	**48,48**	0,000000000035
Disgust	81,26	86,36	**-5,10**	0,029324580032
Boredom	49,24	21,97	**27,27**	0,000012193203
Angry	23,86	30,30	**-6,44**	0,026319945845
Surprise	10,23	4,55	**5,68**	0,001390518291
Other	9,47	18,18	**-8,71**	

The facial features that helped the users to understand the emotions are mostly the eyes, the mouth, and the cheeks. In some expressions, e.g. the "angry", there were some other features very important, for example the texture between the brows in this case. The most important facial features are shown in Table 3.

Table 3: Important features for each facial expression.

	A	B	C	D	E	F	G
1	66,3	81,6	63,6	82,6	77,3	55,7	83,7
2	84,5	67,8	76,1	81,1	79,9	81,4	88,8
3	10,2	22,7	4,2	6,1	4,9	30,9	46,4
4	20,8	14,4	31,1	7,6	14,4	10,0	11,4
5	18,2	59,5	8,7	3,0	4,2	8,9	23,7
6	46,6	8,1	30,7	28,8	60,6	21,4	5,1
7	0,0	2,5	3,0	3,0	3,0	2,3	1,5

1	**Eyes**
2	**Mouth**
3	**Texture of the Forehead**
4	**Shape of the Face**
5	**Texture between the brows**
6	**Texture of the cheeks**
7	**Other**

A	**Neutral**
B	**Angry**
C	**Bored-Sleepy**
D	**Disgusted**
E	**Happy**
F	**Sad**
G	**Surprised**

5 SUMMARY AND CONCLUSIONS

Automated expression classification in face images is a prerequisite to the development of novel human-computer interaction and multimedia interactive service systems. However, the development of integrated, fully operational such automated systems is non-trivial. Towards building such systems, we have been developing a novel automated facial expression classification system (Stathopoulou, I.-O. and Tsihrintzis, G.A.), called NEU-FACES, in which features extracted as deviations from the neutral to other common expressions are fed into neural network-based expression classifiers. In order to establish the correct feature selection, in this paper, we conducted an empirical study of the facial expression classification problem in images, from the human's perspective. This study allows us to identify those face parts of the face that may lead to correct facial expression classification. Moreover, the study determines those facial features that are more significant in recognizing each expression. We found that the isolation of parts of the face resulted to better expression recognition than looking at the entire face image.

6 FUTURE WORK

In the future, we will extend this work in the following directions: (1) we will improve our NEU-FACES system by applying techniques based on multi-criteria decision theory for the facial expression classification task, (2) we will investigate the application of quality enhancement techniques to our image dataset and seek to extract additional classification features from them, and (3) we will extend our database so as to contain *sequences of images of facial expression formation* rather than simple static images of formed expressions and seek in them additional features of high classification power.

ACKNOWLEDGEMENTS

Support for this work was provided by the General Secretariat of Research and Technology, Greek Ministry of Development, under the auspices of the PENED-2003 basic research program.

REFERENCES

Csikszentmihalyi M., (1994) Flow: The Psychology of Optimal Experience, Harper and Row, New York.

De Silva L.C., Miyasato T., and Nakatsu R., (1997) Facial Emotion Recognition Using Multimodal Information, in Proc. IEEE Int. Conf. on Information, Communications and Signal Processing - ICICS Singapore, pp397-401, Sept. 1997.

Ekman P., (1999) “The Handbook of Cognition and Emotion”, T. Dalgleish and T. Power (Eds.) Pp. 45-60. Sussex, U.K.: John Wiley & Sons, Ltd.

Ekman P. and Friesen W., (1975) “Unmasking the Face”, Englewood Cliffs, NJ: Prentice-Hall.

Ekman P., (1982) “Emotion In the Human Face” Cambridge: Cambridge University Press (1982).

Ekman, P., Campos, J., Davidson R.J., De Waals, F., (2003) “Darwin, Deception, and Facial Expression”, Emotions Inside Out, Volume 1000. New York: Annals of the New York Academy of Sciences.

Ekman, P., & Rosenberg, E.L, “What the Face Reveals: Basic and applied studies of spontaneous expression using the Facial Action Coding System (FACS)”, New York: Oxford University Press.

Ortony A., Clore G. L., & Collins A., (1988) The Cognitive Structure of Emotions, Cambridge University Press.

Pantic, M. & Rothkrantz, L.J.M. (2000) Automatic Analysis of Facial Expressions: The State of the Art. IEEE Transactions on Pattern Analysis and Machine Intelligence, 22(12), 1424–1445.

Pantic M. & Rothkrantz .L.J.M, (2003) “Toward an affect-sensitive multimodal HCI”, Proceedings of the IEEE, vol. 91, no. 9, pp.1370-1390.

Pantic M., Valstar M.F., Rademaker R. and Maat L., (2005) “Web-based Database for Facial Expression Analysis”, Proc. IEEE Int'l Conf. Multmedia and Expo (ICME'05), Amsterdam, The Netherlands, July 2005

Picard R.W., (1997) Affective Computing, Cambridge, The MIT Press.

Picard R.W., (2003), "Affective Computing: Challenges," International Journal of Human-Computer Studies, Volume 59, Issues 1-2, July 2003, pp. 55-64.

Reeves, B. and Nass, C., Social and Natural Interfaces: Theory and Design. CHI Extended Abstracts 1997: 192-193.

Reeves, B., and Nass, C. The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places, Cambridge University Press and CSLI, New York.

Rosenberg, M., (1979) Conceiving the Self, Basic Books, New York.

Russell, J. A., (2003) Core affect and the psychological construction of emotion. Psychological Review, 110, 145-172.

Russell, J. A., (1994) “Is there universal recognition of emotion from facial expression?: A review of the cross-cultural studies”, Psychological Bulletin, 115, 102-14.

Stathopoulou I.-O. and Tsihrintzis G.A.(2004), “A neural network-based facial analysis system,” 5th International Workshop on Image Analysis for Multimedia Interactive Services, Lisboa, Portugal, April 21-23, 2004.

Stathopoulou I.-O. and Tsihrintzis G.A.(2004), “An Improved Neural Network-Based Face Detection and Facial Expression Classification System,” IEEE International Conference on Systems, Man, and Cybernetics 2004, The Hague, Netherlands, October 10-13, 2004.

Stathopoulou I.-O. and Tsihrintzis G.A.(2005), “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.

Stathopoulou I.-O. and Tsihrintzis G.A.(2005), Evaluation of the Discrimination Power of Features Extracted from 2-D and 3-D Facial Images for Facial Expression Analysis, 13th European Signal Processing Conference, Antalya, Turkey, September 4-8, 2005

Stathopoulou I.-O. and Tsihrintzis G.A.(2005), Detection and Expression Classification Systems for Face Images (FADECS), 2005 IEEE Workshop on Signal Processing Systems (SiPS’05), Athens, Greece, November 2 – 4, 2005.

Stathopoulou I.-O. and Tsihrintzis G.A.(2006), An Accurate Method for eye detection and feature extraction in face color images, 13th International Conference on Signals, Systems, and Image Processing, Budapest, Hungary, September 21-23, 2006

Stathopoulou I.-O. and Tsihrintzis G.A.(2006), Facial Expression Classification: Specifying Requirements for an Automated System, 10th International Conference on Knowledge-Based & Intelligent Information & Engineering Systems, Bournemouth, United Kingdom, October 9-11, 2006.

Stathopoulou I.-O. and Tsihrintzis G.A.(2007), NEU-FACES: A Neural Network-based Face Image Analysis System, 8th International Conference on Adaptive and Natural Computing Systems, Warsaw, Poland, April 11-14, 2007.