Academic Science

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 7
July 2017

Facial Expression Recognition Augmented With Spatiotemporal Features

Shami Tripathi,
PG Student,
Department of EXTC
Thakur College of Engineering and Technology, Mumbai, India
Hemant Kasturiwale,
Associate Professor,
Department of Electronics
Thakur College of Engineering and Technology, Mumbai, India

ABSTRACT

Communication in any form i.e. nuncupative or nonverbally is essential to complete various daily routine tasks and plays a significant role in life. The most expressive way humans display emotions is through facial expressions. Humans detect and interpret faces and facial expressions in a scene with little or no effort. Facial appearance is the most effective form of non-verbal communication as it provides a hint about demonstrative state, mentality and intention. Generally facial expression recognition framework consists of three steps: feature extraction, processing and recognition. In order to develop a robust facial expression recognition framework that yields effective results, it is necessary to extract features that have strong discriminative abilities. Recently different methods for facial expression recognition have been proposed, but invariably they all spend reckoning time on whole face image or divides the facial image based on some mathematical or geometrical heuristic for features extraction and gives unreliable results.There are several related problems: detection of an image segment as a face, extraction of the facial expression information, and classification of the expression (e.g., in emotion categories). A system that performs these operations accurately and in real time would be a major step forward in achieving a human-like interaction between the man and machine. In this work a novel approach is proposed to recognize some facial expression i.e. happy, sad, surprise, disgust, neutral from time sequential depth images. FER in combination with LDP, optical flows, GDA using facial expression images is obtained using depth camera's image database to make features more robust and finally, these spatiotemporal features are fed Recursive Neural Networks to train and recognize different facial expressions successfully. The spatiotemporal features based facial expression recognition approach, also enhanced contrast level and with noise added in image is compared with conventional approaches such as LBP, PCA, ICA where the proposed method yields to out performs the other in terms of recognition rate.

KEYWORDS

Facial Expression, LBP, LDP, GDA, Optical Flows, RNN

INTRODUCTION

Improving life quality has always been one of the priorities amongst scientists, physicians and researchers. In recent years, incorporation of electronic, computer, mechanics with biology, medicine study fields has led to numerous invaluable achievements. One of the most promising results to assist the locked-in patients with crucial disabilities, amputees and the elderly is Human Computer Interaction (HCI) technology which is an approach to transmit the information between humans and computers. Human emotion analysis is a challenging machine learning task with a wide range of applications in human-computer interaction, e-learning, health care, advertising and gaming.In the present day era, lot of innovation is going in the field of face recognition. The aim of face recognition is to match the input image and the image stored in database. Face expression recognition (FER) has played a vital role in solving the problem of criminal identification, human computer interaction. Emotion analysis is particularly challenging as multiple input modalities, both

visual and auditory, play an important role in understanding it.Local Binary Patterns (LBP) has been used lately for FER. They produce long histograms, which slows down the recognition speed and also in some circumstances they don't consider the effect of centre pixel. LBP was improved by focusing on face pixel's gradient information and named as Local Directional Pattern (LDP). LDP represent much robust features than LBP due to considering the gradient information for each pixel. A robust discriminant analysis called Genera Discriminant Analysis (GDA) has recently been used in different applications where GDA significantly shows superiority over traditional feature extraction such as LBP.
In this paper, a novel approach is proposed for FER in combination of LDP, optical flows, GDA, and RNN using facial expression images obtained through a camera. The local LDP features are extracted from the facial expression image and all then are augmented with the optical flow features. The augmented LBP and motion features are then classified via GDA. Each of these local features are then compared to codebook vector generated from the training facial expression image and converted to discrete symbols. Then the symbols are utilized to train the RNNs of the facial expression to be later applied for recognition based likelihood.

II. LITERATURE SURVEY

The whole literature review is focused on the following literature work being done by an array of scholars and researchers in image processing. As face is a complex multidimensional visual model and for developing a model for face recognition is difficult task. Recognition plays a very important character when it comes to develop various culture visual communication systems for emotion transformation. Various researches have been done in this field.
Hamedi and Salleh [3] examined an analysis of neuromuscular signal activities to recognize eleven facial expressions for Muscle Computer Interfacing applications. A robust denoising protocol comprised of Wavelet transform and Kalman filtering is proposed to enhance the electromyogram (EMG) signal-to-noise ratio and improve classification performance. Fourteen pattern recognition-based algorithms are employed to classify the extracted features. The authors observed as a limitations that because of the low number of subjects along with the absence of patients, authors cannot safely generalize findings and conclude an effective MuCI approach. Small inter individual variability can be expected when analyzing other subjects facial EMGs which is partly due to differences in the morphology of the facial musculature. In addition, an extensive statistical analysis is needed to study the differences among the repetitions of the facial expressions. This requires recordings with more repetitions facial expressions.

In this paper Facial representation based on statistical local features [4], Local Binary Patterns (LBP) is practically assessed. From this literature survey author gives an inclusive practical study of FER based on LBP features is presented. It primarily comprises of three parts, i.e. face representation, feature extraction, and classification. Face representation shows how to mock-up the face and resolves the succeeding algorithms of recognition and detection. The distinctive features of face images are extracted in feature extraction step. In the categorization or classification, the face image is evaluated among the images present in database. Face recognition gives approximately 90% of exactness; on the other hand face expression recognition also lies in the similar range of 90% accuracy.

T. Kiran [5] provided two class emotion detection and multi class facial expression classification using Support Vector Machine (SVM) is presented. Facial feature vectors in dual form are obtained using Local Binary Pattern (LBP) Histogram by tracing the bins in clockwise and anticlockwise direction. The Histogram feature descriptors are calculated from LBP images in dual form which are then concatenated to obtain features of complete face image. The proposed algorithm is tested using standard Japanese Female Facial Expression Database and Taiwanese facial Expression Database and results are verified using locally developed Indian face database of students. As per the author's proposed algorithm significantly outperforms the classical LBP based algorithm. The main contribution of this work include the development of simple strategy to encode textual information of face pattern. The experimental results [5] show increased accuracy in recognition rates in comparison with the other classical LBP based algorithms.The drawback of this work will include determination of more discriminative feature space for classification of highly correlated dataset.

In this paper S. Kumar [6] proposed method projection analysis evaluates the distribution of informative regions of a face. This is done by projecting the expressive face images onto their corresponding neutral images. Hence, the proposed face model can efficiently extract distinctive texture features from a face. Additionally, the proposed face model can extract geometrical features as well. The performance of the proposed face model is evaluated on MUG datasets which shows that the proposed face model outperforms several existing face models. Also, the proposed face model can give a recognition accuracy of 97.3% which is significantly better than the performance of state-of-the-art face models.

Facial expression recognition aims to classify facial expression as one of seven basic emotions including "neutral "[7]. This is a difficult problem due to the complexity and subtlety of human facial expressions, but the technique is needed in important applications such as social interaction research. Deep learning methods have achieved state-of-the-art performance in many tasks including face recognition and person re-identification. Here we present a deep learning method termed Deep Neural Networks with Relativity Learning (DNNRL), which directly learns a mapping from original images to a Euclidean space, where relative distances correspond to a measure of facial expression similarity.

Furthermore, the model is updated to a greater or lesser degree according to the sample difficulty, which leads to a more adjustable and robust model. Compared to traditional deep neural networks such as AlexNet, DNNRL is competitive and achieves excellent performance for FER.

III. PROPOSED METHODOLOGY

Generally, it is found that all the reviewed methods for automatic facial expression recognition are computationally expensive although they produce good results on different datasets but fail to work adequately on low resolution image.

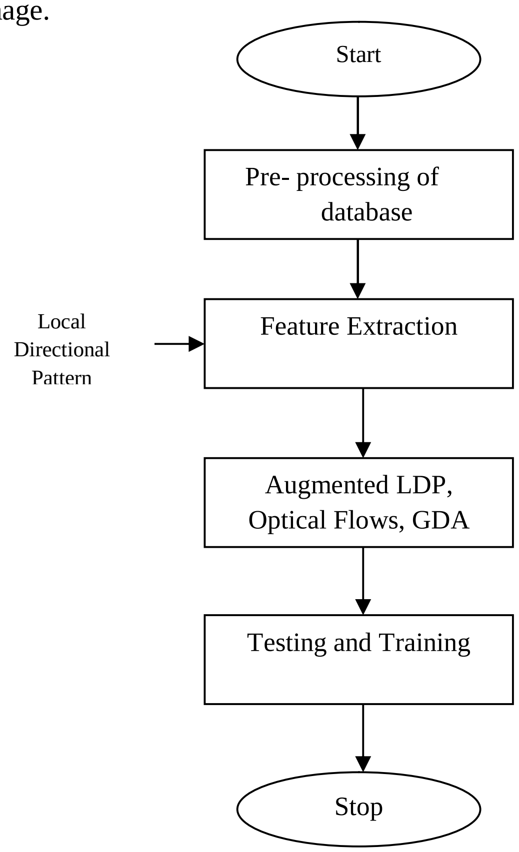

Figure 3: Flowchart of the implementation

Academic Science

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 7
July 2017

More research effort is required to be put forth for recognizing more complex facial expressions than the six classical, such as fatigue, pain, and mental states such as agreeing, disagreeing, lie, frustration, thinking as they have numerous application areas. Other problems include expression intensity estimation, spontaneous expression recognition, micro expression recognition, misalignment problem, illumination, and face pose variation.

m_3	m_2	m_1

First, a strong feature space is generated via LDP, Optical motion flows, and GDA to project all the feature vectors including training and testing. Then discrete symbol sequences are generated from the features using vector quantization and then further applied to corresponding expression RNN. For testing an expression image sequence, the testing sequence is applied on all trained RNNs and one is chosen with highest likelihood. Figure 3 shows the block diagram for the FER approach.

A. Pre-processing through Database

A very significant step in facial expression recognition is preprocessing procedure. The initial step in processing is to acquire pure facial expression images, which has normalized intensity, uniform shape and size. The effect of illumination and lighting should also be removed by it. First the original images are transformed into gray scale images. Positioning the centers of eyes on every face, the entire images are appropriately rotated, scaled, translated and cropped to 100×100 pixels. Images are subsequently subjected to a few image pre-processing operations. The image pre-processing step consists of contrast and illumination equalization, fuzzy filtering and histogram equalization which are described below

Face Detection:

Detecting a face in an image has to decide which pixels in the image is part of the face and which are not. Methods that focus on face (such as eyes, nose etc) that detect face-like colors in facial region or that use standard feature templates, are used to detect faces. In face detection, it has to choose which pixels in images are the part of face and which are not the part of it

Feature Extraction

The extraction of the feature matrix is the primary step in any face recognition system. A characteristics feature extraction algorithm, construct a computational model with a few linear or non-linear data. Hence, to make the extracted feature as representative as possible or when the input data to an algorithm is very big to be processed and is supposed to be infamously outmoded then the input data will be changed to a concentrated representation set of features. Transformation of input data in set of features is known as feature extraction. When the extracted features are cautiously selected, then it is estimated that features set will distort the appropriate detail from input data in order to carry out preferred work using the reduced representation in place of full-size input. However, if there is non-availability of specialist knowledge, general dimensionality reduction methods may be used.

B. LDP

The LDP descriptor is an eight bit binary code assigned to each pixel of an input image that can be calculated by comparing the relative edge response value of a pixel in different directions. So that eight directional edge response values $\{m_i\}$, i=0,1,2…..7 of a particular pixel are computed using Kirsch masks [2] in eight different orientations Mi centered on its own position. Figure below shows eight directional edge response positions and LDP binary bit positions. Because different importance of the response values, the k most prominent directions are considered to generate the LDP. So the top k values $|m_j|$are set to 1, and other positions are set to 0. Finally, the LDP code is derived by formula (1), where m_k is the k-*th* is the most significant directional response value. Figure shows an example LDP code with k=3

$$L_k = \sum_{i=0}^{7} b_i(m_i - m_k).2^i, b_i(a) = \begin{cases} 1, & a \geq 0 \\ 0, & a \leq 0 \end{cases} \dots\dots\dots (1)$$

Academic Science

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 7
July 2017

m_4	x	m_0
m_5	m_6	m_7

b_3	b_2	b_1
b_4	X	b_0
b_5	b_6	b_7

85	32	26
53	50	10
60	38	45

313	971	503
537	x	393
161	97	161

0	0	1
1	x	1
0	0	0

LDP Binary Code 00010011

The input image of size $M * N$ can be represented by LDP histogram using (2) after computing all the LDP code for each pixel $(r,)$, where i is the LDP code value.

$$H(i) = \sum_{r=1}^{M} \sum_{c=1}^{N} f(L_k(r,c), i); f(a,i) = \begin{cases} 1, & a = i \\ 0, & a \neq i \end{cases} \quad ...(2)$$

For a particular value of k, there has C_8^k different number of bins for the histogram H. In essence, a resulting histogram vector size of $1 \square C_8^k$ is produced for the image. Whereas computing LDP over the whole face image only considers the occurrences of micro-pattern without any information of their location and spatial relationship which usually represents the image content better [3].

C. Optical flow

The Optical flows of the facial expression images from the consecutive facial expression frames are obtained using Lucas-Kanade method. Then, the flow region is divided into sub blocks to compare the average flow vector. Fig 2 shows two consecutive images and corresponding optical flows. The vectors are then converted to a single row vector as

$$J = [\overline{O_1}, \overline{O_1}, \ldots, \overline{O_1}]$$

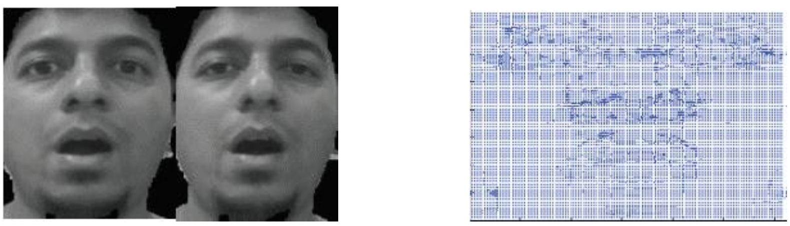

Figure 3.2 a) Two Surprised images b) Optical Flows

D. GDA

Generalized Discriminant Analysis is a discriminant analysis approach that produces an optimal discriminant function that maps the input samples into the classification feature space on which the class identification of the samples is determined. Thus, the goal of GDA is to maximize following equation as

$$G_G = \left|\frac{V^T \theta_B V}{V^T \theta_T V}\right|$$

Where θ_B and θ_T are the between-class and total scatter matrices of the depth face features after mapping them into a Gaussian Kernel Function. Thus, the depth face features of different expressions can be extended by GDA as

$$U = G_G^T \quad M.$$

E. RNN

A recurrent neural network (RNN) is a class ofartificial neural network. RNNs are called recurrent network because they perform the same task for every element of a input sequence, with the output being depended on the previous computations. It uses internal memory to process sequence of inputs. This makes them applicable to the task like speech or face recognition.

IV. EXPERIMENTAL RESULTS

For the recognition of the face different datasets are used. There are two types of datasets used by FER community. First ones are the datasets of images and the others are the datasets of videos of facial expression. For real-time facial expression analysis, the image datasets are of no use as they do not simulate the real-time environment. The Cohn–Kanade database, Cohn–Kanade extended database , Japanese female facial expression (JAFFE) [8] ,GEMEP-FERA2011 dataset, static facial expressions in the wild (SFEW) etc. are the datasets of sequence of the facial images for an expression. Here in this paper we have used JAFFE dataset. JAFFE contains the image sequence of facial expression images. There are images with five universal emotions (disgust, surprise, happy, neutral, sad).

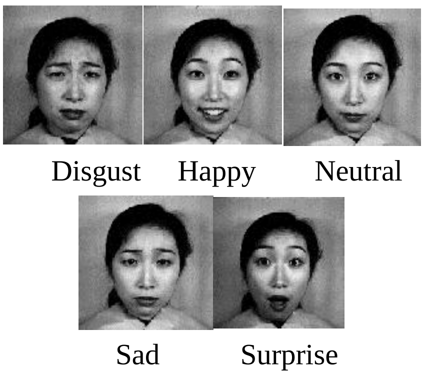

Figure 4.1: Five basic expressions

For training and testing facial expression model, few image sequence were applied. To compare the proposed features with the other feature extraction methods, all methods were implemented using RNN to recognize aforementioned five different facial expressions. First of all, RGB datasets were tried and then they were followed by the videos with the same experiment setups. Regarding RGB experiments, all the images were converted to grayscale first. For the PCA feature case, the eigenvectors were computed from all the dataset and selected 100 eigenvectors to train the RNNs. The average recognition rate using PCA is 58%, the lowest recognition rate in the experiments. Then, ICA features were employed with improved recognition rate (i.e. 76.25%) using ICA. Then LBP is applied for FER with 80.42% which is higher than PCA and ICA. Furthermore, LDP-GDA augmented with optical flow and RNN (proposed) showed its superiority over the other features extraction methods by achieving the highest recognition rate (95%) as shown in table below.

International Journal of Electronics, Electrical and Computational System
IJEECS
ISSN 2348-117X
Volume 6, Issue 7
July 2017

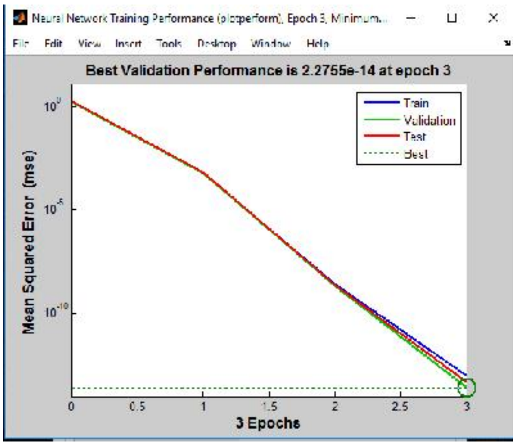

Figure 4.2 Performance analysis of Training

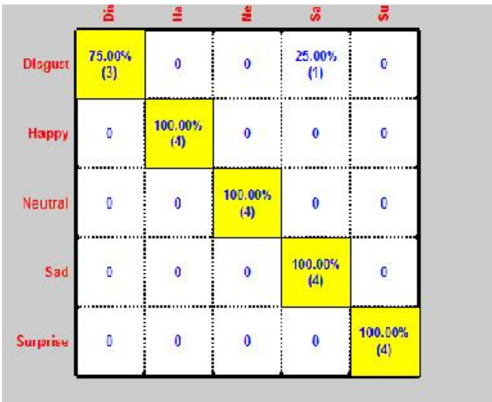

Figure 4.4 Confusion Matrix with proposed work

Also in this proposed work, Facial expression recognition in noisy environment is obtained. Applying Salt and Pepper noise to JAFFE database as shown below, the recognition rate obtained is 65% which is better than other techniques. Then for further enhancement, we applied adaptive histogram equalization to image and observed that accuracy of the system is 85% which proves boon to proposed work.

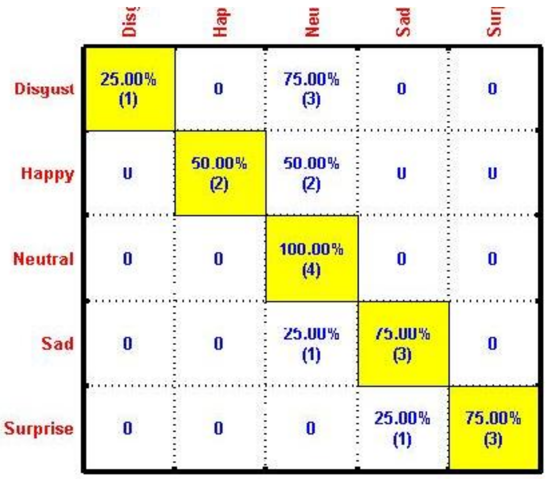

Figure 4.5 Confusion matrix for Noise added image

	Disg	Hap	Neu	Sad	Sur
Disgust	50.00% (2)	25.00% (1)	0	25.00% (1)	0
Happy	0	100.00% (4)	0	0	0
Neutral	0	0	100.00% (4)	0	0
Sad	0	0	0	100.00% (4)	0
Surprise	0	0	50.00% (2)	0	50.00% (2)

Figure 4.6 Confusion matrix for Contrast Enhanced image

V.FUTURE SCOPE

There are several areas of potential future work in this area that could be explored. This study attempted to test as many types of FER as possible but left out several that are in use or will continue to grow in the future. In future work, an attempt will be made to work in real time applications.

VI. CONCLUSIONS

The proposed method has been compared with other traditional approaches including RGB video based and recognition result shows superiority over others. A recognition accuracy of 95% is achieved on JAFFE database, exceeding the results of earlier studies. Also when observed in noisy environment and adaptive histogram contrasted image the accuracy obtained is 65% and 80% respectively. Evaluation over a range of image resolutions and frame rates shows that our method outperforms the state-of-the-art, which makes our approach promising for real-time recognition in poor image acquisition conditions.

REFERENCES

[1] T. Jabid, M. H. Kabir and O. Chae, "Facial expression recognition using Local Directional Pattern (LDP)," *2010 IEEE International Conference on Image Processing*, Hong Kong, 2010, pp.1605-1608 doi: 10.1109/ICIP.2010.5652374

[2] Jabid, T., Kabir, M.H., Chae, O.S.: Local Directional Pattern (LDP) for Face Recognition.In: IEEE Int. Conf. Consum. Electron. 2010, pp. 329–330 (2010)

[3] M. Hamedi; S. H. Salleh; C. M. Ting; M. Astaraki; A. Noor, "Robust Facial Expression Recognition for MuCI: A Comprehensive Neuromuscular Signal Analysis," in *IEEE Transactions on Affective Computing*, vol.PP, no.99, pp.1-1doi: 10.1109/TAFFC.2016.2569098

[4] Nivedita Chitra and Geeta Nijhawan, "Facial Expression Recognition Using Local Binary Pattern," in *International Journal of Innovative Research in Advanced Engineering (IJIRAE)*ISSN: 2349-2763 Issue 06, Volume 3 (June 2016).

[5] T. Kiran and T. Kushal, "Facial expression classification using Support Vector Machine based on bidirectional Local Binary Pattern Histogram feature descriptor," *2016 17th IEEE/ACIS International Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD)*, Shanghai, 2016, pp. 115-120 doi: 10.1109/SNPD.2016.7515888.

[6] S. Kumar, M. K. Bhuyan and B. K. Chakraborty, "An efficient face model for facial expression recognition," *2016 Twenty Second National Conference on Communication (NCC)*, Guwahati, 2016, pp. 1-6. doi: 10.1109/NCC.2016.7561187

[7] Yanan Guo, Dapeng Tao, Jun Yu, Hao Xiong, Yaotang Li and Dacheng Tao, "Deep Neural Networks with Relativity Learning for facial expression recognition," *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, Seattle, WA, 2016, pp. 1-6. doi: 10.1109/ICMEW.2016.7574736

[8] http://www.kasrl.org/jaffe.html

[9] G. Gupta and N. Rathee, "Performance comparison of Support Vector Regression and Relevance Vector Regression for facial expression recognition," *2015 International Conference on Soft Computing Techniques and Implementations (ICSCTI)*, Faridabad, 2015,pp.1-6.doi: 10.1109/ICSCTI.2015.7489548

[10] C. Turan, K. M. Lam and X. He, "Facial expression recognition with emotion-based feature fusion,"*Asia-Pacific Signal and Information Processing Association Annual Summit and Conference(APSIPA)*,Hong Kong, 2015, pp.1-6.doi: 10.1109/APSIPA.2015.7415453.

[11] K. R. Kulkarni and S. B. Bagal, "Facial Expression Recognition," *2015 Annual IEEE India Conference (INDICON)*, New Delhi,2015,pp.1-5. doi: 10.1109/INDICON.2015.7443572

[12] Min Hu, Xiaoyin Huang, Xiaohua Wang, Kun Li, Fuji Ren and Liangfeng Xu, "Adaptive facial expression recognition method based on MBP and HMOG feature," *2014 IEEE 3rd International Conference on Cloud Computing and Intelligence Systems*, Shenzhen, 2014,pp.108-112. doi: 10.1109/CCIS.2014.7175712.

[13] Dong-Ju Kim, Sang-Heon Lee and Myoung-Kyu Sohn, "Face Recognition via Local Directional Pattern" *International Journal of Security and Its Applications Vol. 7, No. 2, March, 2013*

[14] Ekman P., Friesen, W. V., "Unmasking the face", *Malor Books, Cambridge, MA 02238, Division of the Institute for the Study of Human Knowledge, Copyright © 2003 by Ekman P*

[15] Ekman P., Friesen, W. V., & Joseph C. Hager, 2002, "Facial Action Coding System, The Manual", *Published by Research Nexus division of Network Information Research Corporation,Copyright © 2002 Paul Ekman, Wallace V. Friesen, & Joseph C. Hager*

[16] Galateia Latraki, "Emotional Facial Expression Recognition and Classification," Master of Science in Media and Knowledge Engineering thesis report of Delft University of Technology, 2009

[17] Bartlett, M.; Hager, J.; Ekman, P. and Sejnowski T. *Measuring Facial Expressions by Computer Image Analysis*, Psychophysiology, vol. 36, pages 253-264, 1999

[18] Bartlett, M.S; Littlewort, G.; Lainscsek, C.; Fasel, I. and Movellan, J. *Machine Learnin Methods for Fully Automatic Recognition of Facial Expressions and Facial Actions*, IEEE International Conference on Systems, Men and Cybernetics, pages 592-597, 2004