FACIAL EXPRESSION RECOGNITION BASED ON EOG TOWARD EMOTION DETECTION FOR HUMAN-ROBOT INTERACTION

Aniana Cruz[1], Diogo Garcia[1], Gabriel Pires[1,3], Urbano Nunes[1,2]
[1]*Institute of Systems and Robotics, University of Coimbra, Coimbra, Portugal*
[2]*Department of Electrical and Computer Engineering, University of Coimbra, Coimbra, Portugal*
[3]*Department of Engineering, Polytechnic Institute of Tomar, Tomar, Portugal*
{anianabrito, diogojg, gpires, urbano}@isr.uc.pt

Keywords: EOG signal, facial expression, avatar, classification

Abstract: The ability of an intelligent system to recognize the user's emotional and mental states is of considerable interest for human-robot interaction and human-machine interfaces. This paper describes an automatic recognizer of the facial expression around the eyes and forehead based on electrooculographic (EOG) signals. Six movements of the eyes, namely, up, down, right, left, blink and frown, are detected and reproduced in an avatar, aiming to analyze how they can contribute for the characterization of facial expression. The recognition algorithm extracts time and frequency domain features from EOG, which are then classified in real-time by a multiclass LDA classifier. The offline and online classification results showed a sensitivity around 92% and 85%, respectively.

1 INTRODUCTION

Emotion is a complex process that characterizes the human feeling and it is associated with a specific pattern of physiological activity (Schacter, 2009). It is fundamental in human behaviour, since it has influence in the personality, disposition, motivation and interaction between people. Emotion can be expressed through: facial expressions such as surprise, fear, disgust, anger, happiness and sadness; the sound of the voice; the body posture and the arousal of the nervous system, for example, rapid heartbeat and breathing and muscle tension (Ekman and Friesen, 1975). Machine emotional intelligence, i.e., the ability of an intelligent system to recognize the user's emotional state and interact accordingly, is an interesting aspect that can improve the human-machine interaction. This topic has received increasing attention by the research community. Ekman and Friesen proposed an universal facial expression which is independent to human cultures and origins. The first computer-based recognition system of facial expression appeared later in 1990s (Mase, 1991; Terzopoulus and Waters, 1993). Most of these studies classify facial expression or vocal emotion based on a single data modality, such as static image or speech and video sequences (Black and Yacoob, 1997; Bartlett et al., 1999; Nwe et al., 2001; Cohen et al., 2003; Buenaposada et al., 2008; Verma and Singh, 2011). Bimodal approaches, combining the two modalities, image and speech, were also proposed in (Huang et al., 1998; De Silva and Ng, 2000; Emerich et al., 2009). Recognition of hand gestures, body pose and body motion can improve the robustness of emotion recognition (Busso et al., 2004; Castellano et al., 2008; Metri et al, 2011).

Image-based recognition of facial expressions is very sensitive to illumination, image quality, human's position and movements. Approaches based on biosignals such as electromyography (EMG) and electroencephalography (EEG) have been proposed recently. In (Hamedi et al, 2011), a method based on surface EMG (sEMG) is used to recognize five different facial gestures (rest, smile, frown, rage, and gesturing 'notch' by pulling up the eyebrows). In (Koelstra and Patras, 2013) a multi-modal approach combining facial expressions, recorded by a frontal camera, with EEG signals was proposed for the generation of affective tags. Electrooculography (EOG) can also be used for detection of eye movements, providing useful information to characterize facial expressions. Although the EOG is

used in a variety of applications including clinical and human machine interfaces (Barea et al., 2002; Shayegh and Erfanian, 2006; Duchowski, 2007; Banerjee et al., 2013), its use in emotion recognition has been up to now not very significant. Electrocardiography (ECG), galvanic skin response (GSR) are some other sensors useful to characterize emotion (Monajati et al., 2012; Kurniawan et al., 2013). Biosignals provide proprioceptive information that is impossible to detect with video/ speech/gesture, and therefore are a good complement to these sensing systems. Moreover, biosignal acquisition systems are affordable and can measure simultaneously several types of biosignals. Despite all these advantages, current biosignal electrodes are still somehow intrusive, uncomfortable, unaesthetic and difficult to setup, which justifies their low widespread use. Yet, new wearable devices with dry electrodes are emerging (Barea et al., 2011).

This paper is focused on the detection of EOG signals to recognize facial expressions from the eyes and forehead region. The work herein described is part of a system to detect discrete emotional/mental states (Figure 1), which integrates EEG/EOG/ EMG and GSR signals, for human-machine interface/interaction purposes. The system will be used to adapt robot behaviour according to human emotional/mental state. In particular, the EOG detector recognizes the movements up, down, right, left, blink and frown, which are then reproduced in an Avatar. Most of the researches related with the eye's movement do not analyze frown movements. We introduce it here since it brings information for detecting anger or surprise.

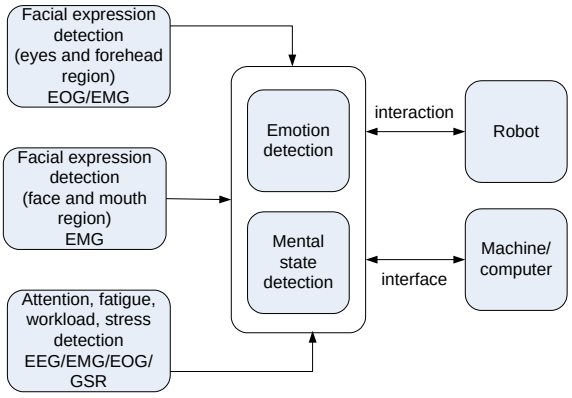


Figure 1: Overview of the system to detect discrete human emotion/mental states for human-robot interface/interaction.

2 METHODOLOGY

Figure 2 shows a block diagram of the proposed online classification system: 1) the raw EOG signals from vertical and horizontal channels are filtered in the band of interest; then 2) a sliding window is used to automatically detect the onset of a movement; 3) features are extracted; and finally 4) features are classified in up, down, right, left, blink or frown. A multiclass linear discriminant analysis (LDA) is used to classify the 6 movements.

2.1 Data Acquisition

EOG measures the potential difference between the cornea and the retina which varies from 0.4 to 1 mV changing with eye's orientation (Malmivuo and Plonsey, 1995). EOG signals can be used to measure vertical and horizontal eye movements by placing the electrodes in specific positions (see Figure 3). Four electrodes were mounted in a bipolar configuration: left and right electrodes in the outer canthus to detect horizontal movements (EOG_h) and below and above the eye to measure vertical movements (EOG_v). EOG signals were recorded with a g.MOBIlab bioamplifier, at a sampling rate of 256 Hz.

Five healthy subjects with ages between 23 and 28 years old performed a training session which consisted on the repetition of the six movements. Participants seated in front of a computer, and followed the movements of a moving ball that moved in four directions: right, left, up and down.

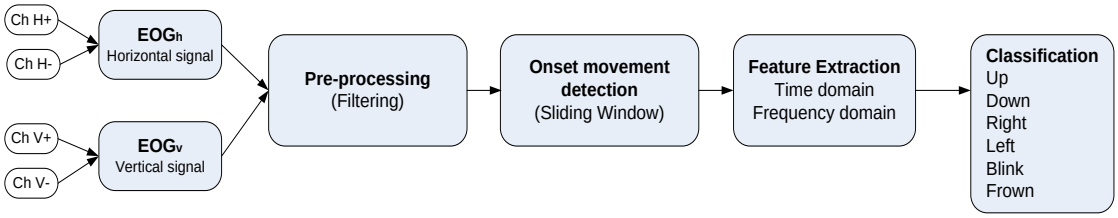


Figure 2: Algorithm structure of the proposed EOG online detector.

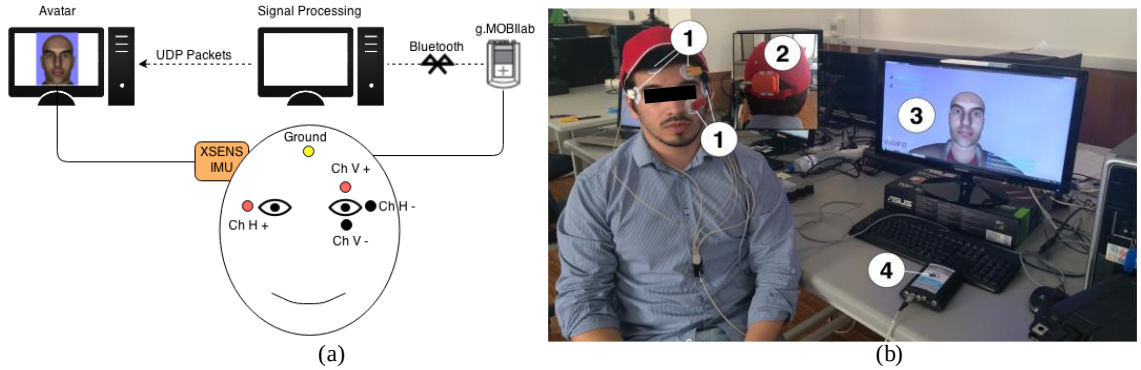


Figure 3: a) Setup for data acquisition using the g.MOBIlab + system with vertical channel (Ch V+/-) and horizontal channel (Ch H +/-). b) Picture of experimental setup with: (1) electrodes (2) IMU (3) avatar, (4) g.MOBIlab +.

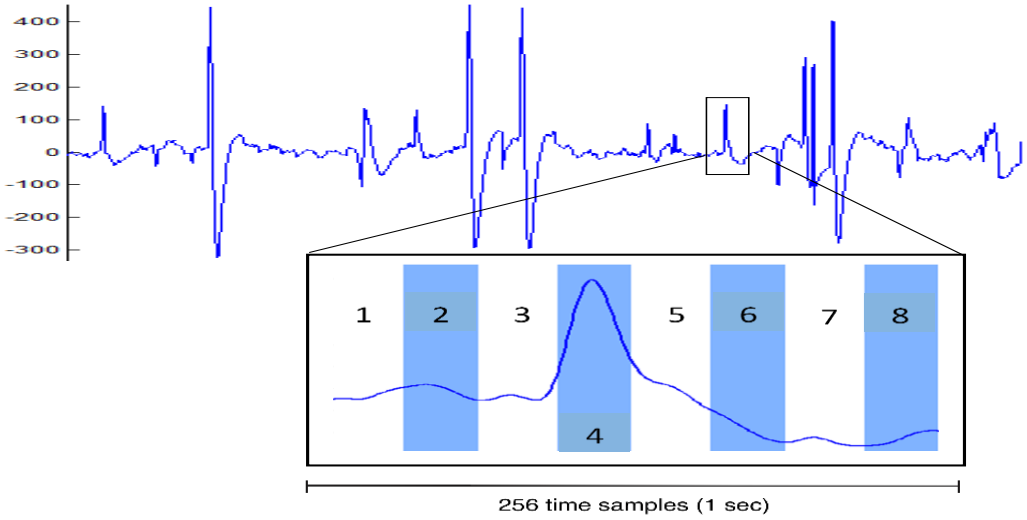
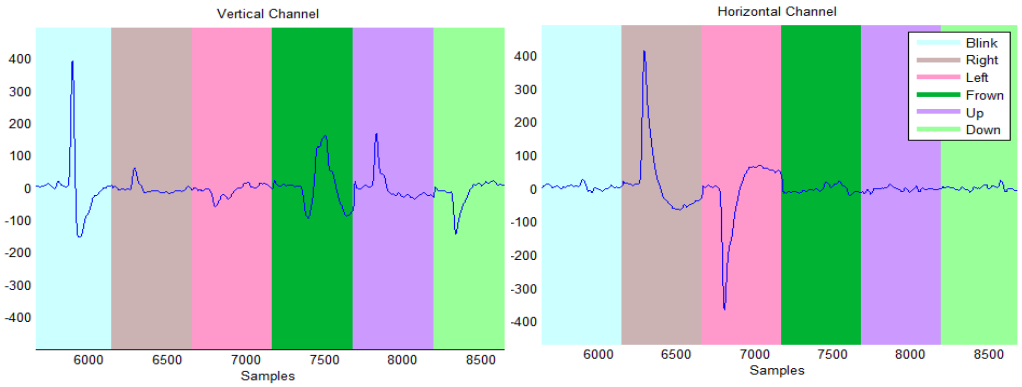


Figure 4: Sliding window to detect up movement (vertical EOG signals).

The remaining movements were instructed by messages displayed at the screen, namely "frown forehead" and "blink". Each movement was performed during 2 seconds with a rest interval of 2 seconds. A full sequence of movements is completed in 24 seconds. A training dataset containing 40 repetitions of the same movement (40x6 data segments), was used to train the classifier for online operation. The computational time for obtaining the classification models is less than 10 seconds

2.2 Pre-processing and Onset Movement Detection

EOG signals are often affected by noise coming from the electrode-skin contact, muscular artifacts and powerline. To reduce these interferences, EOG signals were filtered in the band of interest using a notch filter at 50 Hz and a 4th-order Butterworth band-pass filter with lower cutoff frequency of 0.2 Hz and a higher cutoff frequency of 30 Hz. While in the training session, the user is instructed to perform a specific movement, during the free online operation, the onset of each movement must be automatically detected before being classified. This is achieved by applying a sliding window approach, dividing the EOG signal in non-overlapped segments of 1-second. Each segment is sampled yielding a data vector $X = X_1 \cup X_2 \ldots \cup X_8$ composed of eight X_j subintervals, each with 32 samples (Figure 4). To adjust its position to capture the entire movement, the absolute maximums are computed for each subinterval. If the maximum value of a subinterval (vertical and horizontal



(a) (b)

Figure 5: Vertical (a) and horizontal (b) EOG signals for a sequence of six eye movements: blink, right movement, left movement, frown, up movement and down movement, recorded during a training session.

channels) exceeds a given threshold (adjusted after the training session of each individual) and it is more than the other subinterval, the center of the window is shifted to this subinterval. The rules to detect a movement (Mov) are formally presented below.

$$Mov = \begin{cases} yes, & if \quad AM_v > T_v \cup AM_h > T_h \\ no, & otherwise \end{cases} \quad (1)$$

where T_v is the vertical threshold, T_h is the horizontal threshold, AM_v and AM_h are the absolute maximum of vertical and horizontal channel respectively calculated as:

$$AM_{vj} = \max(|Xv_j|) \quad j\epsilon \{1, \dots, NS\} \quad (2)$$

$$AM_{hj} = \max(|Xh_j|) \quad j\epsilon \{1, \dots, NS\} \quad (3)$$

$$T_v = 0.7 \times \left[\frac{1}{N_{vm}} \sum_{i=1}^{N_{vm}} \max(|Xv_i|) \right] \quad (4)$$

$$T_h = 0.8 \times \left[\frac{1}{N_{hm}} \sum_{i=1}^{N_{hm}} \max(|Xh_i|) \right] \quad (5)$$

where $NS = 8$ is the number of subintervals, N_{vm} is the number of vertical movements and N_{hm} is the number of horizontal movements performed during training sessions. The vertical threshold (equation 4) is 70% of the mean of the absolute maximum value of the up and down eye movement recorded from vertical channel. The horizontal threshold (equation 5) is 80% of the mean of the absolute maximum value of the right and left eye movement recorded from horizontal channel. The center of the window is the subinterval with the maximum absolute value.

2.3 Feature Extraction

After the detection of the onset of a movement, a feature extractor is applied to the segment of 256 samples. As we can see in Figure 5, when the subject blinks his/her eyes, there is a higher positive peak and a weaker negative peak in the vertical channel. The same occurs when the subject frowns his/her forehead, but with a smaller amplitude. When the subject moves the eyes to the right, a large positive peak and a small positive peak occur respectively in the horizontal and the vertical channel. The opposite effect appears when the subject moves the eyes to the left. There is a positive peak in the vertical channel when the subject moves the eyes up and a negative peak when the subject moves the eyes down. These time domain features are extracted using the maximum (Max), minimum (Min), total and partial average values. The total average (MedT) is the mean of the epoch and the two partial averages (MedP1 and MedP2) are respectively the means of the segment taking into account only the samples with amplitudes that are higher and lower than a given threshold. The thresholds were empirically set to +20 and -16, by experimentation. For each segment X the time domain features are computed as:

$$Max = \max(X) \quad (6)$$

$$Min = \min(X) \quad (7)$$

$$MedT = \frac{1}{L}\sum_{i=1}^{L} X_i \quad (8)$$

$$MedP1 = \frac{1}{L_1}\sum_{i=1}^{L_1} X_i \quad \forall\, X_i > 20 \quad (9)$$

$$MedP2 = \frac{1}{L_2}\sum_{i=1}^{L_2} X_i \quad \forall\, X_i < -16 \quad (10)$$

where L is the length of the segment X, L_1 and L_2 are respectively the number of time samples satisfying $X_i > 20$ and $X_i < -16$. The frown movement is also characterized by frequencies resulting from muscular contraction in the forehead. Therefore, features were also extracted in the frequency domain through a relative power measure for the frequency bands {10-15; 15-20; 20-25 and 25-30 Hz}, according to:

$$RP_j = 100 \times \left[\frac{P_j}{\sum_{i=10}^{30} P_i}\right] \quad j\epsilon\ \{1,\dots,\mathrm{NB}\} \quad (11)$$

where RP_j is the relative power for each frequency band, $\mathrm{NB} = 4$ is the number of frequency bands, P_j is the power of band j, and P_i is the power from 10 to 30 Hz, i.e., the total power. The feature vector (FV) used for classification has a dimension of 18, corresponding to 9 features for each EOG channel:

$$FV = [Max_v\ Min_v\ MedT_v\ MedP1_v\ MedP2_v\ RP_{jv}\ Max_h\ Min_h\ MedT_h\ MedP1_h\ MedP2_h\ RP_{jh}] \quad (12)$$

where the subindex v and h represent vertical and horizontal channel respectively.

2.4 Classifier and Performance Measures

EOG patterns representing each one of the 6 classes are classified by a multiclass LDA (Duda et al., 2000). LDA is a generative classifier that finds a linear combination of features that separates the 6 classes. To evaluate the performance of the classification, the following parameters were computed: sensitivity (Sens), specificity (Spec) and accuracy (Acc) (Zhu et al., 2010):

$$Sens = \frac{TP}{TP + FN} \times 100 \quad (13)$$

$$Spec = \frac{TN}{TN + FP} \times 100 \quad (14)$$

$$Acc = \frac{TN + TP}{TN + TP + FN + FP} \times 100 \quad (15)$$

where TP is the true positive, TN is the true negative, FN is the false negative and FP is the false positive.

3 SYSTEM FRAMEWORK

In the current stage of the work, we used an avatar to mimic (replicate) the movements of the user. The virtual avatar was developed in Vizard™ software. The 3D model of the head and expressions of the avatar are designed on Maya Autodesk™. After importing the 3D models built on Maya to Vizard, the expressions are represented through a mix of different faces, reproducing the subject movements. The avatar receives a trigger code via UDP/IP each time a movement is recognized in real-time (the number of movements was limited to a maximum of one per second). A wired XSENS sensor IMU (inertial measurement unit) is used to detect and replicate the movements of the head/neck of the user on the Avatar.

4 RESULTS AND DISCUSSION

The first step of the recognition system is the detection of the onset of a movement through the sliding window. False positive and false negative rates of 4.9% and 15.4 % were obtained.

Table 1 shows the confusion matrix obtained for the offline classification of the six movements using the features in the time domain: maximum, minimum, total and partial average values. The results reveal that blink movement has the highest number of true positives, followed by the down and right movements. The greatest number of false positives and false negatives appears in the down and frown movements, respectively. Frown movement is mainly confused by down movements. Table 2 shows the confusion matrix using the combination of the features in time and frequency domains. Adding the relative power feature increases the true positive values of frown movement. Table 3 and 4 presents the accuracy, specificity and sensitivity values using the features

Table 1: Confusion matrix of the offline classification system using only the features in time domain: maximum, minimum, and total and partial averages.

		Movements					
Automatic		Blink	Frown	Right	Left	Up	Down
	Blink	191	11	2	3	1	2
	Frown	7	153	2	3	14	2
	Right	0	3	190	1	1	0
	Left	0	1	0	189	0	1
	Up	1	13	1	4	175	5
	Down	1	19	5	0	9	190

Table 2: Confusion matrix of the offline classification system using relative power features in addition to maximum, minimum, and total and partial averages.

		Movements					
Automatic		Blink	Frown	Right	Left	Up	Down
	Blink	191	11	2	2	1	2
	Frown	7	176	4	7	10	7
	Right	0	4	192	1	1	0
	Left	0	1	0	189	0	1
	Up	1	5	1	1	176	7
	Down	1	3	1	0	12	183

Table 3: Overall offline classification results for maximum, minimum, total and partial average values as features.

	Blink	Frown	Right	Left	Up	Down	Average
Sens	95.5	76.5	95.0	94.5	87.5	95.0	90.7
Spec	98.1	97.2	99.5	98.9	97.6	96.6	98.1
Acc	97.7	93.8	98.8	98.9	95.9	96.3	96.9

Table 4: Overall offline classification results for maximum, minimum, total and partial average values and relative power as features.

	Blink	Frown	Right	Left	Up	Down	Average
Sens	95.5	88.0	96.0	94.5	88.0	91.5	92.3
Spec	98.2	96.5	99.4	99.8	98.5	98.3	98.5
Acc	97.8	95.1	98.8	98.9	96.8	97.2	97.4

in time domain and combining time and frequency features, respectively. Analyzing the results presented in Table 3 we observe that the blink movement has the highest sensitivity detection. On the other hand, the frown movement is the less accurately detected. The use of relative band power increases the sensitivity of the frown movement to 12% and the average sensitivity about 2%. All movements have specificity values above 88%. The average sensitivity, specificity and accuracy are 92.3, 98.5 and 97.4, respectively. From the five participants, the three with the highest scores completed also the online experiments. The performance of online classification is presented in Table 5. Subject 3 has the highest performance with sensitivity close to 90%. These results reflect also the false positive rate of the detection of movements' onset, thereby slightly decreasing the overall classification performance. The sliding window was adjusted using a subject-dependent thresholds. We aim to improve the system in the near future to include generic thresholds obtained from a database of several subjects, thus improving

Table 5: Online classification performance for each subject.

	Sens	Spec	Acc
Subject 1	86.9	97.5	98.0
Subject 2	77.7	97.5	97.9
Subject 3	88.1	98.8	98.9
Average	84.7	97.9	98.3

the robustness of the classification model.

Blinks give relevant information for user state and emotion characterization, since activities that need thought and attention causes a decrease on blink frequency. Usually, greater blink rate indicates lower attention and fatigue (Andreassi, 2000). Frown movement is an expression that characterizes emotions like anger or surprise. These movements are accurately detected, thus the EOG signal can provide important cues for detecting emotions like fatigue, anger or surprise. Moreover, vertical and horizontal movements provide useful information to detect stress.

5 CONCLUSION

In this paper, six eye movements (up, down, right, left, blink and frown) are classified from EOG patterns and reproduced in an Avatar. This is an integrated part of a system being developed toward the recognition of human's emotion for human-robot interaction. Offline and online sensitivity of the EOG classifier were around 92% and 85%, respectively, which are promising results.

The next research steps will be the integration of EMG for facial expressions like smile, open/close mouth, and then the implementation of an emotion recognizer obtained from the combination of all detected facial expressions.

ACKNOWLEDGEMENTS

This work has been supported by the FCT project "AMS-HMI2012 - RECI/EEI-AUT/0181/2012" and project "ProjB-Diagnosis and Assisted Mobility - Centro-07-ST24-FEDER-002028" with FEDER funding, programs QREN and COMPETE.

REFERENCES

Andreassi, J. L., 2000. *Psychophysiology: Human Behavior and Physiological Response*. Lawrence Erlbaum Associates. London, 4th edition.

Banerjee, A., Datta, S., Pai, M., Konar, A., Tibarewala, D. N., Janarthanan, R., 2013. Classifying Electrooculogram to Detect Directional Eye Movements. *International Conference on Computational Intelligence: Modeling Techniques and Applications (CIMTA)*, (10) 67–75.

Barea R., Boquete L., Mazo M., Lopez E., 2002. System for assisted mobility using eye movements based on electrooculography. *IEEE Transactions on Neural Systems and Rehabilitation Engineering*, 10(4):209-218.

Barea, R., Boquete, L., Rodriguez-Ascariz, J. M., Ortega, S., López, E., 2011. Sensory System for Implementing a Human-Computer Interface Based on Electrooculography. *Sensors*, 11 (1), 310-328.

Bartlett, M. S., Hager, J. C., Ekman, P., Sejnowski, T. J., 1999. Measuring facial expressions by computer image analysis. *Psychophysiology*, 36(2):253–263.

Black, M. J., Yacoob, Y., 1997. Recognizing Facial Expressions in Image Sequences Using Local Parameterized Models of Image Motion. *International Journal of Computer Vision*, 25(1), 23–48.

Buenaposada, J., M., Muñoz, E., Baumela, L., 2008. Recognising facial expressions in video sequences. *Pattern Analysis and Applications*, 11:101-116.

Busso C., Deng, Z., Yildirim S., Bulut, M., Lee, C. M., Kazemzadeh, A., Lee, S., Neumann, U., Narayanan S., 2004. Analysis of Emotion Recognition using Facial Expressions, Speech and Multimodal Information. *Proceedings of the 6th international conference on Multimodal interfaces*, 205-211.

Castellano, G., Kessous, L., Caridakis, G., 2008. Emotion Recognition through Multiple Modalities: Face, Body Gesture, Speech, *Affect and Emotion in Human-Computer Interaction*, 92-103.

Cohen, I., Sebe, N., Garg, A., Chen, L., Huang, T.S., 2003. Facial expression recognition from video sequences: Temporal and static modeling. *Computer Vision Image Understand*. 91: 160-187.

De Silva, L. C., Ng, P. C., 2000. Bimodal emotion recognition, *In: IEEE International Conference on Automatic Face and Gesture Recognition*, 332–335.

Duchowski, A., 2007. *Eye Tracking Methodology: Theory and Practice*, Springer. 2nd edition.

Duda R. O., Hart, P. E., Stork, D., G., 2000. *Pattern Classification*. John Wiley and Sons Ltd. 2nd edition.

Ekman, P., Friesen, W. V., 1975. *Unmasking the face. A guide to recognizing emotions from facial clues*. Englewood Cliffs, New Jersey: Prentice-Hall.

Emerich S., Lupu, E., Apatean, A., 2009. Emotions recognition by speech and facial expressions analysis. *17th European Signal Processing Conference*.

Hamedi, M., Rezazadeh, I. M., Firoozabadi M., 2011. Facial Gesture Recognition Using Two-Channel Bio-Sensors Configuration and Fuzzy Classifier: A Pilot Study. *International Conference on Electrical, Control and Computer Engineering*, 338-343.

Huang, T. S., Chen L. S., Tao, H., Miyasato, T., Nakatsu, R., 1998. Bimodal Emotion Recognition by Man and Machine. *ATR Workshop on Virtual Communication Environments*.

Koelstra, S., Patras, I., 2013. Fusion of facial expressions and EEG for implicit affective tagging. *Image and Vision Computing*, 31(2) 164 –174.

Kurniawan, H., Maslov A. V., Pechenizkiy, M., 2013. Stress detection from speech and Galvanic Skin

Response signals. *International Symposium on Computer-Based Medical Systems (CBMS)*, 209-214.

Malmivuo, J., Plonsey, R., 1995. *Principles and Applications of Bioelectric and Biomagnetic Fields*. New York, Oxford, Oxford University Press, Inc.

Mase, K., 1991. Recognition of facial expressions for optical flow. *IEICE Transactions, Special Issue on Computer Vision and its Applications*, E 74(10).

Metri, P., Ghorpade, J., Butalia, A., 2011. Facial Emotion Recognition using Context Based Multimodal Approach. *International journal on interactive multimedia and artificial intelligence*, 2(1), 171-182.

Monajati, M., Abbasi, S. H., Shabaninia, F., Shamekhi, S., 2012. Emotions States Recognition Based on Physiological Parameters by Employing of Fuzzy-Adaptive Resonance Theory. *International Journal of Intelligence Science*, 2, 166-175.

Nwe, T. L., Wei, F. S., De Silva, L. C., 2001. Speech based emotion classification. *Electrical and Electronic Technology*, (1) 297–301.

Schacter, D. S., Gilbert, D. T., Wegner, D. M., 2009. *Psychology*. New York: Worth.

Shayegh, F., Erfanian, A., 2006. Real–time ocular artifacts suppression from EEG signals using an unsupervised adaptive blind source separation. *Engineering in Medicine and Biology society, 28th Annual International Conference of the IEEE*, 5269-5272.

Terzopoulus, D., Waters, K., 1993. Analysis and synthesis of facial image sequences using physical and anatomical models. *IEEE Trans. Pattern Analysis and Machine Intelligence*, 15(6):569–579.

Verma, G. K., Singh, B. K., 2011. Emotion Recognition based on Texture Analysis of Facial Expression. *International Conference on Image Information Processing*, 1-6.

Zhu W., Zeng N., Wang N., 2010. Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC Analysis with Practical SAS® Implementations. *NESUG proceedings: Health Care and Life Sciences, Baltimore, Maryland*.