

Towards an online Emotional Recognition System for Intelligent Tutoring Environment

Nouha Khediri

Northern Border University
Faculty of Computing and IT

Rafha, KSA

Nuha.khediri@nbu.edu.sa

Nouha.khediri@ieee.org

Mohamed Ben Ammar

Northern Border University
Faculty of Computing and IT

Rafha, KSA

Mohammed.Ammar@nbu.edu.sa

Mohamed benammar@ieee.org

Monji Kherallah

Faculty of Sciences
University of Sfax

Sfax, Tunisia

Monji.kherallah@fss.usf.tn

Abstract—Emotion recognition can be used in wide range of applications, we are interested in E-learning system because of several benefits of learning anywhere anytime and at anytime. The vision of affective computing is to impart to system, the possibility of understanding the user through recognizing his emotions and expressions. Obviously, the entire emotional state of a user is expressed and can be observed in different modalities. In this paper, a survey was conducted on the latest research in emotion recognition, first with unimodal system then with multimodal system, specifically applied in E-learning environment. Then, we propose the improvement of EMASPEL (Emotional Multi-Agents System for Peer-to-peer E-Learning) by the fusion of several real-time modalities of emotional communication.

Index Terms—Emotion Recognition, E-learning, Affective Computing, Fusion, Intelligent Tutoring System

I. INTRODUCTION

The proliferation of computers and electronic machines affect the Interaction of Human Machine which is an important field composed of 3 actors who are Human (with his perceptive modes and his memory), Machine (with its peripherals and technical capabilities) and Interface such as screen, menu, graphic, etc. Another branch of research on emotions in machines is named "Affective Computing" which is in turn a very interesting subject proposed by Picard [1] in 1994 that touches on a multitude of domains namely psychology, philosophy, computer science, cognitive science, etc. To better human-machine interaction, the emotion comes to allow machines to make decisions in a more "intelligent" way. In literature, we found that many authors have separated the emotion in (1) *Primary Emotions* and (2) *Secondary Emotions*. Primary emotions are a natural and automatic reactions, triggered by a rapid and imprecise treatment of a stimulus [2]. Sometimes called basic emotions or fundamental emotions. According to Ekman [3], primary emotions are only six; anger, fear, disgust, surprise, happiness and sadness. Secondary emotions are triggered by a cognitive treatment of information. It is a blend of primary emotions and provoke the same body reactions. Moreover, secondary emotions result from a mental evaluation process linked to experience and memory. Besides, another name for secondary emotions proposed by Tomkins in [4] is "derived emotions".

Emotions are applied in various scientific applications like psy-

chology and e-learning. In e-learning, the relationship between emotion and cognition is quite important and is influenced by the electronic learning environment. Development of affective computing leads to a new generation of Intelligent Tutoring Systems (ITSs) which is Affective Tutoring Systems (ATSs). ATS is a tutoring that accurately recognize learners' emotion and react to it.

The scope of this paper is to give a clear idea about the different methods used in which emotions can be recognized by a system. We primarily focus on application of affective computing methods in E-learning. Accordingly, we propose a novel approach of multimodal emotion recognition algorithms in E-learning. Our System is an extension of EMASPEL and it aims to improve E-learning. This paper is organized as follows. We first start by describing some related work in section 2. In section 3, we present the existent system EMASPEL, then we explain our proposed system. In section 4, we report the obtained results. Finally, in section 5, we conclude and outline future work.

II. STATE OF THE ART

Previous emotion recognition works in E-learning environment has mainly focused on methods using only one modality and has tried to detect learner emotion states using different signifiers such as facial expression, physiological signals, speech and so on. Within the scope of affective computing, various works have been investigated in regard to emotion recognition. In this paper, we aim to give readers an overview of emotion recognition that are applied to E-learning using unimodal and multimodal systems.

Recognizing emotion from facial expression is the earliest and most common technique. From this modality, we found in literature a system on e-learning facial expression recognition proposed by Loh et al. [5] that uses Gabor Wavelet for facial feature extraction and Neural Network for classification. However, their database contains only static images with four facial expressions. Another system in recognizing emotion through facial expression named EMASPEL [6] that is the combination of E-learning and peer-to-peer topology. It has achieved the 80% as a recognition rate and uses a standard webcam with lighting and brightness environment varied for facial features

TABLE I
SUMMARY OF UNIMODAL & MULTIMODAL EMOTION RECOGNITION IN E-LEARNING

Modalities	Approach				
	Authors & Year	Features	Classifier	Databases	Emotion
Unimodal					
Facial Expression	Loh et al. [5] 2006	34 facial feature points	Neural Network	eFEC	neutral, smile confuse, sleepy
Facial Expression	Ben Ammar et al. [6] 2010	eye, eyebrows mouth	distances computed on faces skeletons	MIT-CBCL facet Faces96 database	six basic emotions
Facial Expression	Yang et al. [7] 2011	eye, brows mouth, headx, heady	AdaBoost HMM, GMM	not mentioned	Blink, Wrinkle, Shake Nod, Yawn, Talk
Text	Rodriguez et al. [8] 2012	student profile	a developmental account	12 essays written by a fresher student	Joy, Anger Sadness, Fear
Text	Tian et al. [9] 2014	syntax features affective computing rules affective word	emotion words and syntax database emotion regulation casebase	Event Emotion regulation-inStance (EES)	Anxiety Anger Frustration
Text	Razek et al. [10] 2017	not mentioned	Dominant meaning technique	ISEAR dataset	Joy, Fear, Anger Sadness, Disgust Shame and Guilt
Speech	Chen et al. [11] 2007	speech Features	Neural Network	Not mentioned	Neutral, Joy Anger, Disgust Sadness, Teasing Fear, Surprise
Speech	Zhang et al. [12] 2007	speech Features	One-class-in-one Network	Not mentioned	Neutral, Joy Anger, Disgust Sadness, Teasing Fear, Surprise
Physiological Signals	Scotti et al. [13] 2006	skin conductance, blood volume pulse blood volume pulse electrocardiogram and electroencephalogram and electroencephalogram	Not mentioned mentioned	Not mentioned mentioned	Relax, Stress Engagement
Multimodal					
Facial Expression Verbalizations	Bahreini et al. [14] 2014	facial features voice intonation	FaceTracker software	10 persons	six basic emotions neutral emotion
Eyes Head Movement	Krithika et al. [15] 2016	visual features	Viola Jones LBP	5 students	Boredom Lack of interest

tracking. The classification method used in [6] is simple and based on the variation of some distances from the neutral face. In the work done by Yang et al. [7], a computer vision system that automatically analyze the learning status of their students during the learning has been implemented to enhance an online Virtual English Classroom called VEC3D. Facial features are extracted using AdaBoost classifier in a feature vector form, then, each feature vector sequence is recognized and trained using Hidden Markov Model (HMM). Finally, an expression vector, that contains six expressions, is analyzed and a Gaussian Mixture Model (GMM) is applied to evaluate three learning scores such as Understanding, Interaction, and Consciousness [6].

A few studies have been investigated in regard to textual emotion recognition. Rodriguez et al. confirm that the work proposed in [8] is very useful in the context of E-learning system. To extract emotions from text automatically, a different possibilities are presented and a developmental account is used for classifying emotions. Tian et al. [9] applied a text-oriented emotion classification method with active learning to identify the e-learners emotion. The syntax features point out the kind of emotion that a speaker has. The best classification performance is obtained on the dataset with the feature age and

Random Forest based classification algorithm. Another work by Razek and Frasson in [10] describe a text-based emotion recognition approach using Dominant Meaning Technique (DMT). This technique is based on the meaning of the word and allows to build the dominant meaning tree using ISEAR (International Survey on Emotion Antecedents and Reactions) dataset. Emotion Detection Algorithm was created to classify seven emotions which are joy, fear, anger, sadness, disgust, shame and guilt.

A great deal of works have been done on speech emotion recognition. As proposed by [11] a speech emotion recognition system in web learning achieved a recognition rate of approximately 50% with eight emotional states that are joy, teasing, fear, sadness, disgust, anger, surprise and neutral. Another work on the same domain approximately here [12], we have the same work as the previous one, but this time using a one-class-in-one network for emotion recognition module.

Recent trends in research emphasized the use of physiological signals in emotion recognition. These signals are very useful in this field because they are spontaneous and not controllable. The only work applicable in E-learning is [13]. It shows that a specific quantitative indexes extracted from four physiological signals which are Galvanic Skin Response (GSR),

Blood Volume Pulse (BVP), Electrocardiogram (EKG) and Electroencephalogram (EEG), evoke specific emotional states: stress, relaxation and engagement.

From literature, we observed that various results have been reported using only one modality, does not improve the performance of recognizing learner's emotions. Therefore, more attention has been paid so far to use more than one modality to recognize emotions. Many approaches have been introduced to classify emotion based on multimodal. But a few ones have focused on E-learning application. FILTWAM (Framework for Improving Learning Through Webcams And Microphones) [14] is a software that interprets user's emotional state from facial expression and verbalizations and provides appropriate feedbacks. Even when some parts of a face are covered, this system is able to recognize the emotion. Student Emotion Recognition System (SERS) for E-learning proposed by [15] detects the negative emotions of students through a method based on concentration level (high, low and medium) and it provides a real-time feedback mechanism.

Table I summarizes the recent approaches of emotion recognition using unimodal and multimodal inputs especially in E-learning environment. It is observed that current emotion recognition systems are not yet advanced enough to be used in E-learning.

III. PROPOSED FRAMEWORK

A. Existent System

EMASPEL is an affective framework for an Intelligent Affective System. It improves learning within a personalized virtual learning environment through recognizing the affective state of learners, with the aim of a pedagogical reacting. We choose EMASPEL for many reasons such as; EMASPEL is a facial expression recognition system and over the years, facial expression recognition gained a lot of popularity in the research and technique, since the psychologist Mehrabian [16] has found that 55% of the whole human expressions are conveyed through the facial expressions. On the one hand, EMASPEL enriches the communication in collaborative virtual environments through using emotionally expressive avatars. On the other hand, EMASPEL is capable of detecting certain failures in the pedagogical approach adopted. Another reason to choose EMASPEL is the use of multi-agent methodology. Due to this methodology, the communication between the learner and the affective system becomes more dynamic and flexible.

The agents used in this system are five:

- The emotional agent is responsible of emotion identification by extracting and categorizing the facial expression of the learner during a learning session.
- The Emotional Embodied Conversational Agent is based on PECS model (Physical Conditions, Emotional state, Cognitive capabilities and Social status) proposed by Schmidt [17], is responsible of analyzing the emotional state of the learner which allow the tutor to take the necessary teaching actions and communicates an affective response.

- The tutoring agent help and support learners with his pedagogical expertise to better exceed the difficulties and seeks to reinforce the learners intrinsic motivation.
- The curriculum agent is responsible of saving all the information of progressing of the learner during the learning activities.
- The interface agent is the intermediate between the human and the computer. He/she is considered as a service agent because he/she transmits the facial information and the learners affective state to the other agents of the multi-agent system.

B. Extended System

Our work can be considered as an extension of EMASPEL. We will interest on the behavioral and emotional interaction between a human and a machine. The goal is to automatically recognize users emotions by analyzing and processing multimodal inputs. Emotion recognition process has several steps presented in Fig.1. These steps are described in details below:

1) *Multimodal Inputs Channels*: In order to better recognize the emotions, we categorize them into four types: (1) *Vocal Channel*, (2) *Body Channel*, (3) *Sentimental Texture* and (4) *Physiological Signals*.

- *Vocal Channel* contains all verbal communication like speech and video. Automatic emotion recognition from this channel is a difficult classification problem as sometimes even a human cannot easily classify natural emotions based on speech.
- *Body channel* contains the non-verbal communication such as gestures, facial expressions and head movement. This channel is an important mode of expressing and interpreting emotions especially the facial expression which is the most common communication.
- *Sentimental Texture* is focalized on the emotions expressed by texts and documents and when we say text, we mean sentences and the sentences are constituted of words. Emotion recognition using this channel is an implicit task since a sentence can have element of sadness without using the word sad or any of its synonyms.
- *Physiological Signals* often called bio-signals, applies to measures of peripheral nervous system functions such as Heart Beat, Blood Volume Pulse *BVP*, Skin Temperature *SKT*, Electrocardiogram *ECG*, Electromyography *EMG*, Respiration Rate, etc.

2) *Preprocessing*: Before extraction features, some of the input signals need a preprocessing like facial expression that includes detection, registration, normalization and verification of the face. Also, in the video case, a preprocessing is necessary to separate the audio from the images in order to treat them separately. We will consider the same preprocessing steps used in EMASPEL for facial expression recognition since we use this communication modality as first input in our process.

3) *Features Extraction*: An appropriate feature selection procedure is required before training a classifier. In literature,

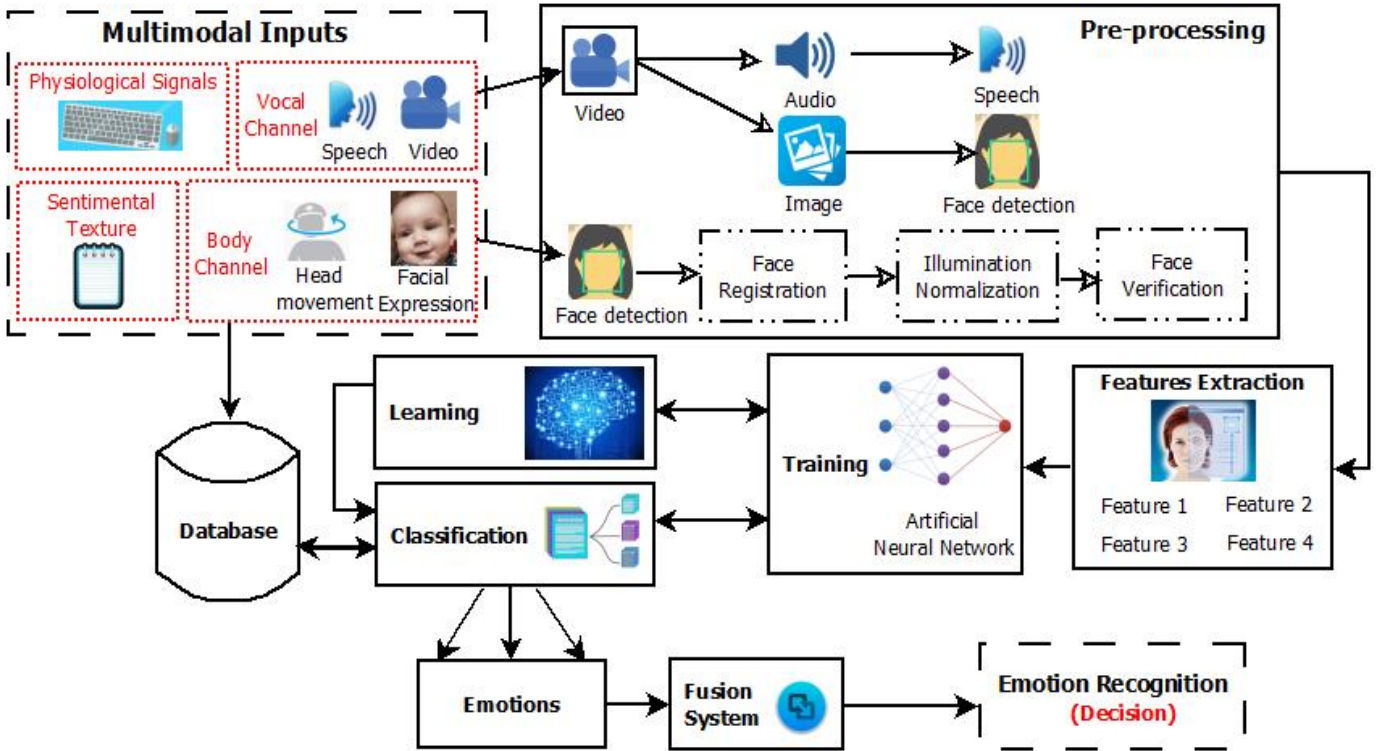


Fig. 1. Emotional Recognition Process

researchers use different set of features. From facial expression, there are geometric and appearance ones. Geometric features describe the characteristics points of face which are eyes, brows, lips and nose. Appearance features are usually based on the face colors and textures which includes wrinkles, bulges, and furrows. From speech, we found source features named also excitation source, spectral also known as segmental or system features, cepstral and time-domain features. From physiological signals, we found, for example, statistical features such as mean and standard deviation.

4) *Training*: Training the system comes before learning with the aim of recognizing or identifying the emotions from features extracted. There are many different ways of training the classifiers. Among them one can mark out examples of some algorithms used such as a machine learning algorithm, dominant meaning methods using in the case of text, a specific rule-based didactical approach for learners training. In our system, we will use the Artificial Neural Network which is popular to machine learning even with minimal input data.

5) *Classification*: Various classification algorithms have been used in recent studies like Naive Bayesian models, Artificial Neural Network (ANN), Decision Tree, K-nearest neighbours, Support Vector Machines (SVM) and Hidden Markov Models (HMM).

6) *Fusion*: Fusion means merge all the inputs of different modalities into a single representation of user's affect expression. The major issues of modality fusion are how and when to integrate the modalities. Typically, there are three methods of fusion. (1) *Single-level fusion*: is rarely used and applied only

when the modalities are similar in nature and have the same temporal resolution [18]. (2) *Feature-level fusion*: is frequently used and it proposes to combine the features obtained from the different modalities to obtain a single feature vector. (3) *Decision-level fusion*: the data from different modalities are treated independently and the obtained results at the output of the classifier are combined at decision level. This method is applied when the temporal characteristics of modalities are different.

7) *Databases*: The first requirement of developing an automatic emotion recognition system is to have a database that contains a variety and range of affective expressions. It is important to highlight that all available datasets for multimodal emotion recognition are acted, i.e. they contain strong emotional expressions. Examples of multimodal databases: HUMAINE database recorded audio-visual information and peripheral physiological signals [19] and The MAHNOB-HCI recorded physiological signals that included face videos, eye gaze data, ECG, respiration pattern, ST, EEG and GSR. [20]

IV. RESULTS & DISCUSSION

First of all, methods illustrated in Fig.2, were implemented in MATLAB and was tested on The MIT-CBCL face recognition database¹ and on the Faces96 database (C)Libor Spacek, 1996². Three methods are followed for face detection (cf.

¹B. Weyrauch, J. Huang, B. Heisele, and V. Blanz, Component-based face recognition with 3D morphable models. In First IEEE workshop on face processing in video, Washington, DC, 2004.

²<http://cswwww.essex.ac.uk/mv/allfaces/faces96.html>.

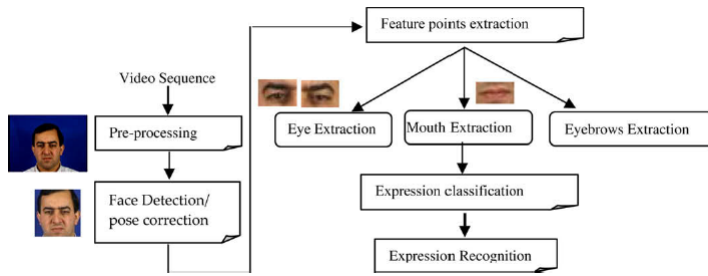


Fig. 2. Methodology of Facial Recognition in EMASPEL [6]

Fig.3) which are: (1) *Skin Color* [21] that needs firstly a preprocessing, (2) *SMA face detection* [22] that contains eye location, mouth location and face location, and (3) *Real time face detection* [23] based on AdaBoost which detects automatically the face of the learner under several constraints like distance, luminosity and varied background. It can also detect more faces in one image.

For feature points extraction (cf. Fig.2), there are a three



Fig. 3. Facial Detection [6]

main facial features that affect the nature of facial expression such as mouth, eyes and eyebrows. So, a detection of the contours of these different facial features is important (See Fig.4). Considering an image of learner' neutral expression, the six distances (D1 to D6) considered on face's skeletons and detailed in [6], leads to obtain the classification after analysis. Note that a facial expression is captured by all the distances, expression recognition will be obtained after calculating the similarity between a given expression and the neutral expression. Taking an example of the emotion "Disgust", it is obtained when the two distances D3 and D4 increases and the other four distances remain constant, with D3: opening of the mouth in width and D4: opening of the mouth in height.

This system is not free of some limitations including, the limitative case of facial expressions whose proposed method was developed without any animation of lips. On the one hand, the AGR meta-model has to be clarified in order to explain further the protocols of interaction and the deontic concepts that regulate the communication between the different agent of architecture. On the other hand, the calculation function of emotions intensity has a lack of robustness because the evalua-

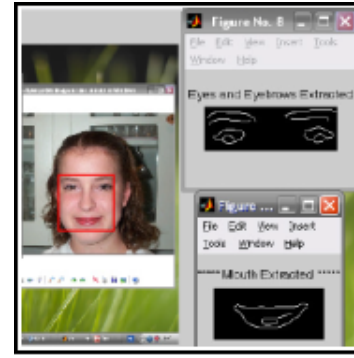


Fig. 4. Eyes, Eyebrows and Mouth Detection [6]

tion of the proposed model did not validate this function. Also, One modality by itself does not carry all the information to conveying the human affective state. Therefore, the drawbacks of one modality can be compensated by another modality. Bearing the aforementioned recognition problem, the combination of many modalities is the potential way suggested in our future work. We will try to merge some modalities that are complementary to each other and irreplaceable. The totality of current work dealing with emotions recognition are limited to two modalities which are facial expression and verbalizations or eyes and head movement. Depending on this principle, our research is focused on a multimodal approach based on the ability to analyze several communication channels simultaneously.

V. CONCLUSION

In this work, special focus is given on emotion recognition through one modality and with different modalities in the context of affective intelligent tutoring system with the aim to facilitate and improve the quality of E-learning. In the future, we plan to continue this work in the following directions. We would like to use an intelligent tutoring system which analyzes the situations that provoked users' emotions and give the adequate feedback. We will merge more than two modalities simultaneously, since fusing information derived from different communication modalities can correct various recognition errors of unimodal system. From multimodal input channels, we already use body channel, more precisely facial expression, as a first input in our model. The second input will be physiological signals since it has not yet been studied in triple with other modalities in E-learning system.

REFERENCES

- [1] R. Picard, "Affective Computing," Cambridge, MA: MIT Press, 1997.
- [2] A. Damasio, "Descartes error: Emotion, reason and the human mind", New York: Grossett/Putnam, 1994.
- [3] P. Ekman, "An argument for basic emotions". *Cognition and Emotion*, 6, pp.169-200, 1992.
- [4] S. S. Tomkins, "Affect theory", *Approaches to emotion*, 163:195, 1984.
- [5] M. Loh, Y. Wong, and C. Wong, "Facial Expression Recognition for E-learning System using Gabor Wavelet & Neural Network", In *International Conference on Advanced Learning Technologies* (2006)
- [6] M. Ben Ammar, M. Neji, A.M. Alimi, and G. Gouarderes, "The Affective Tutoring System", *Expert Systems with Applications* 37, 3013-3023, 2010.

- [7] M.-T. Yang, Y.-J. Cheng, Y.-C. Shih, "Facial Expression Recognition for Learning Status Analysis". In Jacko, J.A. (ed.), *Human-Computer Interaction, Part IV, HCII 2011*, LNCS 6764, pp. 131–138, SpringerVerlag, Berlin Heidelberg, 2011.
- [8] P. Rodriguez, A. Ortigosa, R. M. Carro, "Extracting Emotions from Texts in E-learning Environments", *Sixth International Conference on Complex, Intelligent, and Software Intensive Systems*, 2012.
- [9] F. Tian, P. Gao, L. Li, W. Zhang, H. Liang, Y. Qian and R. Zhao, "Recognizing and regulating e-learners emotions based on interactive Chinese texts in e-learning systems", *Knowledge-Based Systems* 55, 148164, 2014.
- [10] M.A.Razek, and C. Frasson, "Text-Based Intelligent Learning Emotion System". *Journal of Intelligent Learning Systems and Applications* ,9, 17-26, 2017.
- [11] K. Chen, GX Yue, F. Yu, Y. Shen and A.Q. Zhu, "Research on speech emotion recognition system in elearning", *Lecture Notes in Computer Science*. Berlin, vol.4489, pp. 555-558, 2007.
- [12] Q. Zhang, Y. Wang, L. Wang, and G. Wang, "Research on speech emotion recognition in elearning by using neural networks method". In *IEEE International Conference on Control and Automation* (pp. 2605-2608). Los Alamitos, 2007.
- [13] S. Scotti, M. Mauri, R. Barbieri, B. Jawad, S. Cerutti, L. Mainardi, E. N. Brown, M. A. Villamira, "Automatic Quantitative Evaluation of Emotions in E-learning Applications", *Proceedings of the 28th IEEE EMBS Annual International Conference New York City, USA, Aug 30-Sept 3, 2006*.
- [14] K. Bahreini, R. Nadolski and W. Westera, "Towards multimodal emotion recognition in e-learning environments", *Interactive Learning Environments*, 2014.
- [15] Krithika and GG. Lakshmi Priya, Student Emotion Recognition System (SERS) for e-learning improvement based on learner concentration metric, *International Conference on Computational Modeling and Security (CMS 2016)*, *Procedia Computer Science* 85, 767–776, 2016.
- [16] A. Mehrabian, "Communication without words", *Psychol. Today* 2 (9), 5255, 1968.
- [17] C. Urban, and B. Schmidt, "PECS agent-based modelling of human behaviour". In *Emotional and intelligent, The tangled knot of social cognition*, AAAI fall symposium series, North Falmouth, MA, 2001.
- [18] H. Hamdi, "Multimodal platform for the recognition of emotions through the analysis of physiological signals: application to the simulation of job interviews". *Modeling and Simulation*. University of Angers, 2012.
- [19] D. Cowie, E., Cowie, R., Sneddon, I., Cox, C., Lowry, O., McRorie, M. et al. "The HUMAINE database: addressing the collection and annotation of naturalistic and induced emotional data". In A. R. 2007. Paiva, R. Prada and R. Picard (Eds), *Affective computing and intelligent interaction* (pp. 488-500). Berlin, Heidelberg: Springer.
- [20] M. Soleymani, J. Lichtenauer, T. Pun, and M. Pantic, "A multimodal database for affect recognition and implicit tagging". *Affective Computing, IEEE Transactions On*, 3, 4255, 2012.
- [21] S. NEIFER, M. BEN AMMAR and A. M. ALIM, "Dtection de visages dans des images couleur", *GEI, Mahdia*, 2003.
- [22] M. NEJI, M. BEN AMMAR et A. M. ALIM, "Emotion expression function in human computer interaction", *Medianet, Tozeur, Tunisia*, 2004.
- [23] M. Ben Ammar, "MASER (Multi-Agents System for Facial Expression Recognition)", *International Arab Conference of e-Technology (IACe-T)*, October 15-16, Amman, Jordan, 2008.