IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 17, Issue 3, Ver. II (May – Jun. 2015), PP 61-68
www.iosrjournals.org

Real Time Facial Emotion Recognition using Kinect V2 Sensor

Hesham A. Alabbasi[1], Prof.Florica Moldoveanu[2,] Prof. Alin Moldoveanu[3,]
[1](*Doctoral School of Automatic Control and Computers, University POLITEHNICA of Bucharest, Romania, Ministry of Higher Education and Scientific Research / The University of Mustsnsiriyah/Baghdad IRAQ)*
[2](*Department of Computers/Faculty of Automatic Control and ComputersPOLITEHNICA of Bucharest ROMANIA*)
[3](*Department of Computers/Faculty of Automatic Control and ComputersPOLITEHNICA of Bucharest ROMANIA*)

Abstract: *The emotional facial expression is an important element in human communication, along with other forms of non-verbal communication, such as gestures and postures. The human facial expression, with its cunning and moment movements, carries an amazing amount of information that can reflect emotional feelings. Observing people's facial expressions can help a person understand their emotions. The new technology provided today for capturing the facial expressions, with rapid, high resolution image acquisition, helps us to analyze and recognize in real time the true facial emotions. This can be useful in many real time applications, like airport security, trading (the customer's feeling about a product), patient monitoring, and others.In this paper, we focus on the emotion recognition from facial expressions by using Microsoft Kinect for Windows sensor V2 and the face tracking SDK to recognize eight expressions. The implementation of our emotion recognition application was made with Visual Studio 2013 (C++) and Matlab 2014.*

Keywords: *Face expressions, facial features, kinect sensor, face tracking SDK, neural network.*

I. Introduction

Increased In our lives, emotions play a very important role. This happens in every relationship we care about in our friendships, in work, in the family, and in our most intimate relationships. Emotions sometimes can save our lives, but in other times can also cause many problems.

A facial expression is composed of specific positions of the muscles beneath the skin of the face. The movement of facial muscles in such positions conveys the emotional state of an individual to observers. Facial expressions are a form of nonverbal communication. They are a primary means of conveying social information between humans. One of the first scientists who stressed the importance of facial expression in face to face communication of human beings was Charles Darwin. He defines the facial expressions of emotions (anger, fear, surprise, disgust, joy, sadness and other more complex emotions) and the body language "the language of the emotions" [17].Facial expressions can also provide information about the cognitive state of a person, such as confusion, stress, boredom, interest, and conversational signal [18].

The applications for facial emotion recognition are numerous, such as robotics, virtual agents, games, etc. [1][2][3], and the need for automatic emotion recognition arises. The purpose of our research is to use facial emotion recognition in connection with the brain activity.The immediate application of the results of this research is in the monitoring of patients who suffered from a stroke, in the initial phase of their recuperation. But, these results can be extended for the monitoring of patients with other diseases of the brain, such as Alzheimer's or dementia.

This paper presents the results of using Microsoft Kinect for Windows sensor V2 for facial emotion recognition.

It is organized as follows. Section II describes the related work based on Kinect. Section III describes the Facial Action Coding System (FACS). Section IV shortly presents Kinect for Windows V2 sensor and SDK 2.0. Section V explains our work. Section VI shows the experimental results and section VII contains our conclusions and future work.

II. Related Work Based On Kinect

Many papers in the field of Computer Vision have reported about systems of automatic facial expression recognition.

G R. Vineetha, C. Sreeji, and J. Lentin [4], present a method for facial expression recognition by using MS Kinect in 3D from the input image. They used Microsoft Kinect sensor for the Xbox 360 video game console with its technique, described by MS Kinect for human face detection. After human face is detected, edge detection, thinning, and token detection is performed. The user has to give the input threshold value for the detection of tokens. It is a difficult task to decide the best threshold value to generate the tokens.Their results showed that the expression of sadness and disgust were more difficult than the others to recognize.

A. Youssef, S. F. Aly, A. Ibrahim, and A. Lynn [5], proposed a system that attempts to recognize facial expressions using a fast three-dimensional (3D) Kinect sensor. They constructed a training set containing 4Ddata (time is the 4th dimension) for 14 different persons performing the 6 basic facial expressions and used it with both SVM and k-NN classifiers. For individuals who did not participate in training the classifiers, the best accuracy levels were 38.8% (SVM) and 34.0% (k-NN). When considering only individuals who did participate in training, however, the best accuracy levels that they obtained raised to 78.6% (SVM) and 81.8% (k-NN). The authors also describe the potential to use such a system for treatment of children with autism spectrum disorders (ASD).

Mihaela Puica, focuses in her Ph.D thesis [6] on emotion recognition from facial expressions and the main tool for doing this was a Microsoft Kinect sensor with the Face Tracking SDK. She used 58 points that define the brows, eyes and mouth, returned by the Face Tracking SDK, to measure 18 distances between face elements. These distances were used as inputs to a feedforward backpropagation neural network, with 3 output neurons, each one grouping two emotions from the 6 basic emotions. The second experiment was by issuing directly the 58 coordinates to a neural network with 7 output neurons: one for each emotion, plus one for neutral state. The accuracy of emotion recognition with data outside the training set was off 80%.

P. Lemaire, L. Chen, M.Ardabilianand and M. Daoudi [7] proposed in their work an approach to 3D facial expression recognition based on differential mean curvature maps and histograms of oriented gradients. The aim of their work, like many other works, was to classify the face emotions from the 6 primary emotions that were mentioned by [8]: happy, surprise, sadness, anger, fear and disgust. Facial expression analysis systems which are using 3D data can be characterized as static or dynamic. In dynamic systems, time is the fourth dimension, for this reason they are sometimes called four-dimensional (4D) [9]. The techniques for detection vary greatly and include the use of Gabor wavelets [10], SIFT descriptors [11] and quad tree decomposition [12].

Many researchers used Kinect sensor for face recognition. Billy Y.L., Li1 Ajmal S., Mian Wanquan Liu1and Aneesh Krishnal [11] presented an algorithm that uses a low resolution 3D Kinect sensor for face recognition under challenging conditions. Their experiments were performed using a publicly available database containing over 5000 facial images (RGB-D) with varying poses,expressions, illumination and disguise, acquired using the Kinect sensor. The recognition rates were 96.7% for the RGB-Ddata and 88.7% for noisy depth data alone. Gaurav, G., Samarth, B., Mayank, V., and Richa, S. [14] describes a face recognition algorithm based on RGB-D images captured from a Kinect sensor. Their results demonstrate that using RGB-D information can improve face recognition performance compared to existing 2D and 3D approaches.

III. Facial Action Coding System (FACS)

The psychologists continue to ask what is an emotion, as it is a concept that is difficult to define. Psychologists have different opinions on the definition of emotion, some dispute that the definition of emotion should be limited to observable behaviors such as attack and escape, while others define it as a category of experiences that have something in common [15].

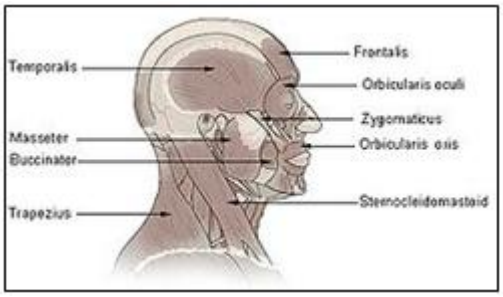

Fig. 1. Muscles of the head and neck (http://en.wikipedia.org)

Facial Action Coding System (FACS), is the earliest method of characterizing the physical expression of emotions. It was developed in 1978, by Paul Ekman, along with Wallace Friesen [8] and is still widely used today. Their system is used to measure all visually distinguishable facial movements and for encoding how movements of facial muscles result in changes in the appearance of the face. Ekman and Friesen studied anatomy and found the associations between the action of muscles, and the changes in facial appearance, Figure1.Some appearance changes are the outcome of movements of multiple muscles and some muscles can have more than one action. Because of this, they named the measurements of FACS action units (AUs).AUs are the actions performed by individual muscles or muscles in combination. FACS consists of 46 AUs of which 12 are for upper face, 18 are for lower face, and AUs 1 through 7 refer to brows, forehead or eyelids [16].The six basic emotions are: anger, disgust, fear, happiness, sadness, and surprise.

IV. Kinect For Windows V2.0 Sensor And Sdk 2.0

In our system for facial emotion recognition we used the new Microsoft Kinect for Windows sensor V2, which adds simplicity to the facial feature extraction.

4.1 Microsoft Kinect for windows V2 sensor

Kinect for Windows v2 sensor is a new product from Microsoft. It is a device with depth sensing technology, a built-in color camera, an infrared (IR) emitter, and a microphone array, enabling it to sense the location and movements of people. And with up to 3x higher depth, fidelity, the v2 sensor provides significant improvements in visualizing small objects and all objects more clearly.

Microsoft provides, with this new Kinect sensor, a development kit (SDK 2.0) with new facilities, drivers, tools, APIs, device interface, and many sample code in C#, C++, and Java to help the application developers. With this new (SDK), body tracking is more stable and can track up to six persons, with 25 joints per person.

As compared to its predecessor, it features enhanced color depth, fidelity, video definition, depth perception, and skeletal tracking. The enhancements are rated to deliver better response time and an improved voice and gesture experience for users. The sensor additionally includes full-HD video for quality augmented reality scenarios and wider field of view, improved skeletal tracking and new active infra-reduction for better tracking in low-light. The sensor is also known to connect to a Windows PC via a power supply and computer interface hub that features a USB3.0 port. The bundled power supply with the Kinect for Windows V2 sensor will support voltages between 100-240 volts.

4.2 Features and distances from kinect

The Microsoft Face Tracking Software Development Kit for Kinect for Windows (Face Tracking SDK), together with the Kinect for Windows Software Development Kit (Kinect for Windows SDK), enables us to create applications that can track human faces in real time.

Face Tracking SDK contains a face tracking engine, which can analyze the input from the Kinect camera, it can detect the head pose and face features depending on the points that can be tracked, and generate an information to the application in real time. As an example, this information can be used in tracking person's head position. The Face Tracking SDK, tracks the 87 2D points, and 13 additional points that belong to the corners of the mouth, the center of each eye, the center of the nose, and for the bounding box around the head, Figure 2 shows the tracked points.

The 87points are:

16 points for the eyes(0-15,8 for the left eye and 8 for the right eye).
20 points for the brows (16-35,10 for the left brow and 10 for the right brow).
12 points for nose (36-47).
20 points for the lips (48-67,12 for the exterior lips, 8 for the interior lips)
19 points for the cheek (68-86).

These points are returned in an array, and are defined in the coordinate space of the RGB image (640 x 480 resolution) returned by the Kinect sensor

Fig. 2. Shows tracking points for kinect sensor V2

V. Prposed Method

As we mentioned in the introduction of this paper, the purpose of our research is to recognize facial emotions in connection with the brain activity of the patients after a stroke, in the initial phase of their recuperation. Until now, we have made experiments with our system using different kinds of persons, in order to establish its accuracy in recognizing emotions of different persons, in real time.

5.1 The steps of emotion recognition

Figure 3 illustrates the steps of our emotion recognition method.

1. Using interface to track a person's face, we saved the 23 face animation feature values in a matrix (using C++). Figure 4 shows the 23 face animation values. These are: headpivx, headpivy, headpivz, jawopenn, jawsliderightt, leftcheekpufff,lefteyebrowloweredd, lefteyeclosedd, righteyebrowloweredd, righteyeclosedd, lipcornerepressedleftt, lipcornerdepressedrightt, lipcornerpulledleftt, lipcornerpulledrightt,lipspuckeredd, lip stretchleftt, lipcornerstretchrightt,lowerlipdepressedleftt,lowerlipdepressedrightt, rightcheekpuffedd, facepitchh, faceroll and faceyaww.
2. Using the Matlab engine to provide face features distances to the neural network, by changing the previous matrix into a mat file.
3. Using the Neural Network to classify and recognize the emotion expression based on the input of 17 out of these 23 Aus.
4. Write the name of the emotional expression in a text file, after recognizing the emotion.
5. Using C++ to read the text file and display the facial emotional expression on the screen.

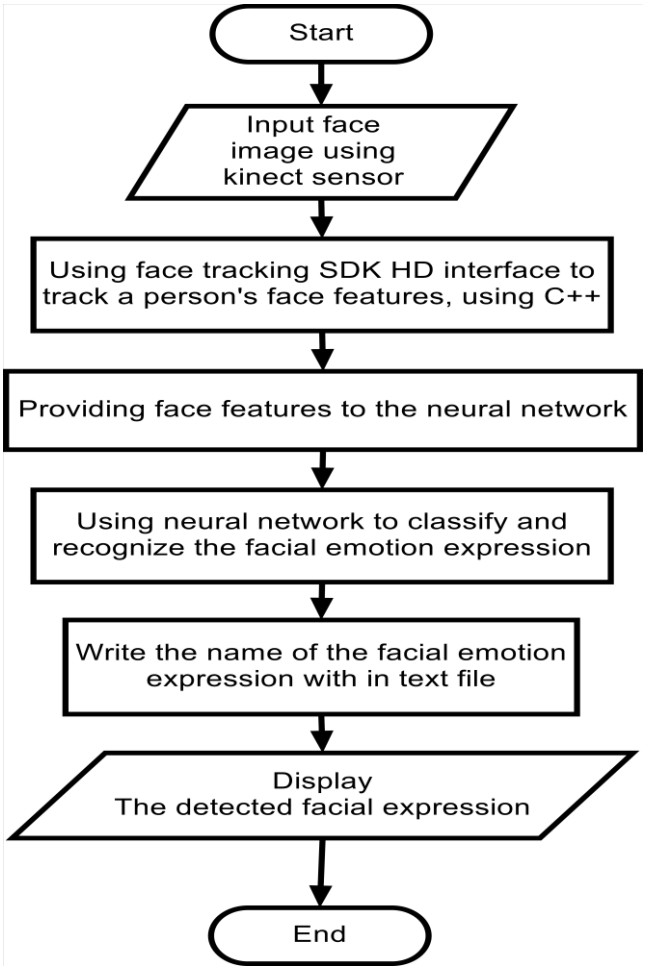

Fig. 3 Flowchart of our proposed method

5.2 Building the database

The Kinect sensor camera gives us 30 frames in one second and a set of 23 face animation unit values for each frame, which are mentioned in the previous paragraph and shown in Figure 4.

We used kinect sensor to build the database. All the values for the face animation units were taken from the subject and updated on each frame. We recorded 8 expressions, and took 25 samples of each expression, for each 12 persons.

The person looks at the kinect camera and makes any expression, we took such snapshots for all the 8 expressions (25 for each). The total number of frames for each expression is 200 and the total number of frames for all persons is 2400 frames. All this data above were then saved in a .Mat file from C++, by using the Matlab engine. This file represents the database.

5.3 Neural network used

After a few tests, we decided that the best neural network, which gives us a high performance, is a multilayer feedforward neural network that can be implemented using the simple NN tool present in Matlab.

We took 70% of the database as training data, 15% for validation and 15% for testing. In Matlab 2014, there is an option to save a neural network in the form of a standalone function.

In our application, NN was used to classify and recognize the expression in each frame. With 30 frames per second, it would have been extremely time consuming if we had opted to train the NN in each frame.

To find the best results for the number of hidden layers used in NN, that gives minimum mean square error (mse),a different number of layers were tested with the calculation of the mean square error for each case. We found that 15 hidden layers give the best result. We could use more layers, but a pretty good model was obtained at 15 layers.

We made this standalone NN function which gives us greater freedom even with 15 hidden layers, that's makes our model very advanced.Figure 5 shows the used neural network.

Fig. 4 The HD face animation values

5.4 Testing the neural network and the system

The face features were provided to the neural network in a 1x23 matrix, which contains all the 23 values stated before. Its 1st 3 values are the head pivot x y z, and last three values are the roll pitch and yaw. These 6 values are not used in expression recognition; we used the rest 17 features.

The result of the Matlab neural network engine is passed to the 'outputs' variable, in the form of an 8x1 numeric matrix. For example, if it is this:

0.0001 000.01% sure that the expression is happy
0.0000 0% sure that the expression is sad
0.8999 89.99% sure that the expression is disgust
0.0000 0% sure that the expression is angry
0.0000 0% sure that the expression is fear
0.0000 0% sure that the expression is a surprise
0.0000 0% sure that the expression is contempt
0.1000 10% sure that the expression is neutral

Then, the maximum surety of the expression is "Disgust",because the highest confidence value lies at position 3, which corresponds to disgust

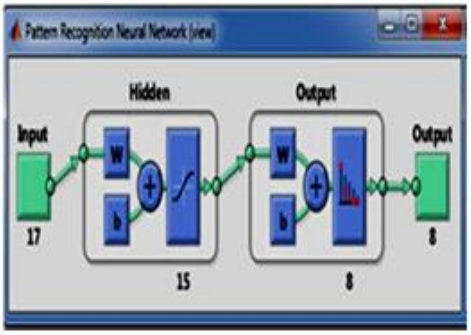

Fig. 5 Neural network used

VI. Experemental Results

1- When we tested the system on many persons stored in the database, we obtained a very good result and the identification rate was about 96%. Figure 6 shows the snapshot of our automatic real time facial emotion recognition.The error histogram was so close to 0, the other models of neural network with different layers, were not given that same error histogram (Figure7). We have used a simple NN which makes our project simple and efficient, this is a plus point for our proposal that we avoided any complex NN methods and gave us the mean square error equal to 0.17 (Figure 8).The true positive rate is very high and false positive rate is very low (Figure 9).

2- We also tested the system with other persons (not from the database) and calculated the identification rate individually for each expression, using this simple formula: IR=1- (C/N), where C represents the number of false cases and N represents the number of trails in each case. The identification rate (IR) for all expressions was 92%.

3- We used 17 features from the face to recognize eight emotion expressions to reduce the time and storage.

We suffered from some problems with the Kinect that can be summarized as:

a. This kind of Kinect needs a specific requirements: 64bit x64 processor, physical dual-core 3.1 GHz - 2 logical cores per physical or faster processor, USB 3.0 controller dedicated to the Kinect for Windows v2 sensor, 4 GB of RAM, graphics card that supports DirectX 11, Windows 8 or 8.1, or Windows Embedded 8.

b. It is not compatible with all kinds of USB3.0 even we tried to use USB3.0 switch hub with high power.

c. Sometimes the frame per second is down in (18-24 FPS) and the recognition became little bit slow. This also happened with the older version of Kinect, V1.8.

Fig. 6 Automatic real time facial emotion recognition. From top to bottom, left to right: neutral, happy, angry, fear, disgust, contempt, surprise and the labels are written in each screenshot.

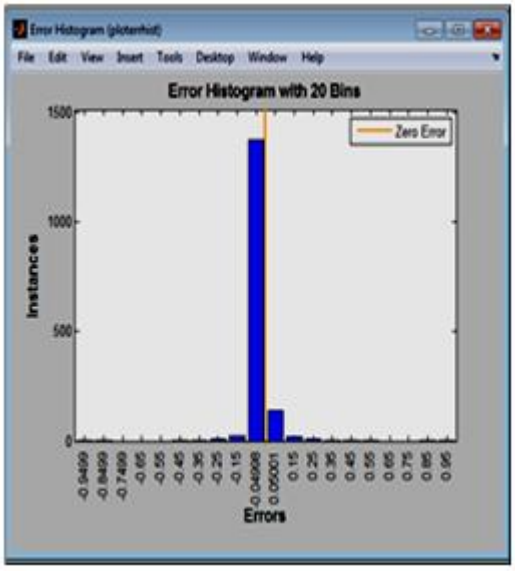

Fig. 7 Error Histogram ,the error close to 0

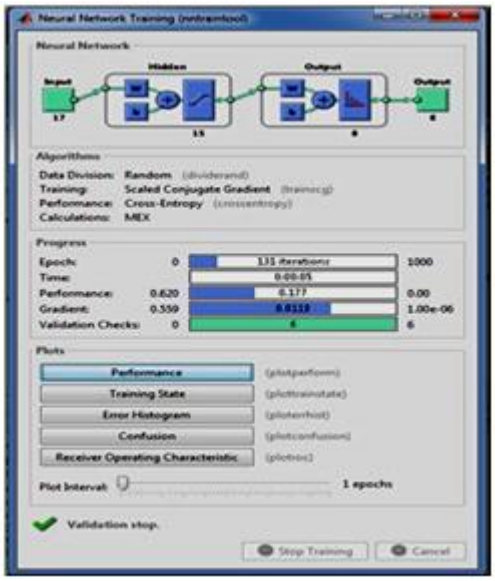

Fig. 8 Running of the used neural network

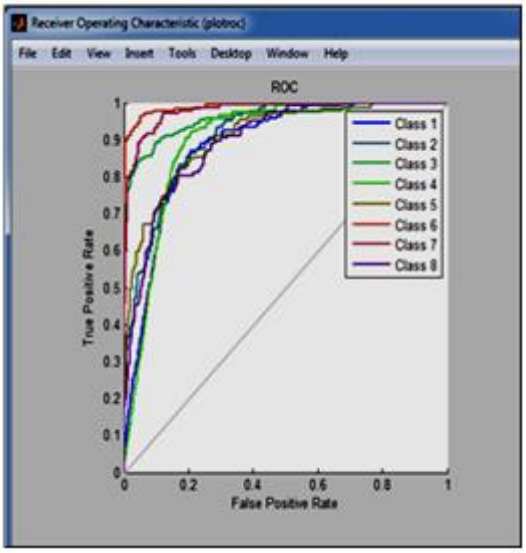

Fig. 9 True positive rate

VII. Conclusion

Our proposed method is simple. The main application is Visual Studio 2013 (C++). The face of the person was tracked and the recorded features coordinate values provided to the Matlab Engine neural network. In Matlab2014, the neural network classifies and recognizes the facial expression and passes back the result to the main application, which displays it on the screen.

We used minimum features (17) to recognize 8 emotion expressions to reduce time and storage. We didn't remove the other six features of the head, because they could be useful for next researches.

Some difficulties appeared when building the database; the persons should learn how doing each expression and this could take too much time to populate the database.

For the future work, we consider that is necessary to build a bigger database, using people with different ages and ethnicities and also, adding the head orientation to make the recognition more general.

References

[1]. F. Malawskil,B. Kwolekl,and S. Shinji, "Using Kinect for Facial Expression Recognition under Varying Poses and Illumination," AGH University of Science and Technology 30-059 Krakow, Poland, Nagoya Institute of Technology, Japan.

[2]. Moldoveanu A., Morar A., Asavei V. 3DUPB - The Mixed Reality Campus: A glimpse at how mixed reality systems can shape the future. Revista Română de Interacţiune Om-Calculator 6 (1) 2013, pp. 35-56. ISSN 1843-4460. 2013

[3]. Alin Moldoveanu, Florica Moldoveanu, Victor Asavei, Anca Morar, Alexandru Egner, From HTML to3DMMO - a Roadmap Full of Challenges, CSCS 18 - The 18th International Conference On Control Systems And Computer Science, 24-27 May 2011, Bucharest
[4]. G R. Vineetha, C. Sreeji ,and J.Lentin ,"Face Expression Detection Using Microsoft Kinect with the Help of Artificial Neural Network", Trends in Innovative Computing 2012 - Intelligent Systems Design.
[5]. A.Youssef, S. F. Aly, A. Ibrahim, and A. Lynn ," Auto-Optimized Multimodal Expression RecognitionFramework Using 3D Kinect Data for ASD TherapeuticAid", International Journal of Modeling and Optimization, Vol. 3, No. 2, April 2013.
[6]. M. Puica ,"Towards a Computational Model of Emotions for Enhanced Agent Performance", Ph.D thesis, University Politehnica of
[7]. P. Lemaire, L. Chen, M. Ardabilianand M. Daoudi, " Fully Automatic 3D Facial Expression Recognition using Differential Mean Curvature Maps and Histograms of Oriented Gradients**",** Workshop 3D Face Biometrics, IEEE Automatic Facial and Gesture Recognition, Apr 2013, Shanghai, China from :https://hal.archives-ouvertes.fr/hal-00823903/document.
[8]. P. Ekman and W.V. Friesen, "Manual for the Facial Action Coding System", Consulting Psychologists Press, 1977.
[9]. G. Sandbach, S. Zafeiriou, M. Pantic, and L. Yin. "Static and dynamic3D facial expression recognition: A comprehensive survey," Image and Vision Computing, Oct. 2012.
[10]. M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with Gabor wavelets," in Proc. Third IEEE Conf. Face and Gesture Recognition, pp. 200-205, Nara, Japan, Apr. 1998
[11]. S. Berretti, B. Ben Amor, M. Daoudi, and A. del Bimbo, "3D facial expression recognition using SIFT descriptors of automatically detected key points", The Visual Computer, vol. 27, no. 11, 2011.
[12]. G. Sandbach, S. Zafeiriou, M. Pantic, and D. Rueckert, "A dynamic approach to the recognition of 3D facial expressions and their temporal models," in Proc. 9th IEEE International Conference on Automatic Face and Gesture Recognition, March 2011.
[13]. Billy Y.L., Li1 Ajmal S., Mian2 Wanquan Liu1and Aneesh Krishna1, "Using Kinect for Face Recognition Under Varying Poses, Expressions, Illumination and Disguise",Curtin University, The University of Western Australia, Bentley, Western Australia Crawley, Western Australia.
[14]. Gaurav, G., Samarth, B., Mayank, V. and Richa, S.,"On RGB-D Face Recognition using Kinect",IIIT Delhi.
[15]. Kalat, J., & Shiota, M. (2007). Belmont, CA: Thomson Wadsworth from : http://en.wikiversity.org/wiki/Motivation_and_emotion/Book/2014/Facial_Action_Coding_System.
[16]. Ekman, P., Friesen, W. V., & Hager, J. C. (2002). Facial Action Coding System Investigator's Guide. Retrieved from: http://face-and- emotion.com/ dataface/ facs/guide / FACSIVTi.html.
[17]. Darwin Charles, Ekman Paul, Prodger Phillip (1998) The Expression of the Emotions in Man and Animals, 3rd edn, London: Harper Collins.
[18]. Marian Stewart Bartlett, Gwen Littlewort, Ian Fasel, Jvier R. Movellan, "Real Time Face Detection and Facial Expression Recognition: Development and Applications to Human Computer Interaction", Computer Vision and Pattern Recognition Workshop, 2003 (CVPRW '03), DOI: 10.1109/CVPRW.2003.10057.