FACIAL EMOTION EXPRESSIONS RECOGNITION WITH BRAIN ACTIVITIES USING KINECT SENSOR V2

Ph.D student Hesham A. ALABBASI, Doctoral School of Automatic Control and Computers, University POLITEHNICA of Bucharest, Bucharest, Romania.

Florica Moldoveanu, Faculty of Automatic Control and Computers, University POLITEHNICA of Bucharest, Bucharest, Romania.

Alin Moldoveanu, Faculty of Automatic Control and Computers, University POLITEHNICA of Bucharest, Bucharest, Romania.

Ph.D student Zaid Shhedi, Doctoral School of Automatic Control and Computers, University POLITEHNICA of Bucharest, Bucharest, Romania.

Abstract - One of the important non-verbal human communication along with other form like postures and gestures, is emotional facial expression. Facial expression, carries a huge amount of information which represent the emotional feelings of humanity, and by observing people's facial expressions with its cunning and moment movements can help us to understand what the face tell us because its reflect an immediate indicator of affective dispositions in other people. All the action and the reaction in the human body is controlled by the brain, and because of the major emotional significance of facial expression, many studies have used emotional faces to identify neural substrates of emotional processing. These studies have found that brain areas generally involved in the processing of emotional information are also activated during the processing of facial emotion. Microsoft company provides a new device technology for image and video acquisition with great high resolution for capturing the facial expressions, this device helps us to analyze and recognize the true facial emotions in real time, also it can be useful in many real time applications and fields, like airports, security, hospitals (patient monitoring), law enforcement, trading (the feeling for customer's about a product), sport training, and others. In this paper, we present an approach to analyze and recognize the emotional facial expressions with brain activities for a patient through Microsoft Kinect for Windows sensor V2 device with its face tracking SDK to get the features in the face by using visual Studio 2013 (C++) and Matlab 2015 to recognize eight expressions.

Key Words: Facial expressions, Facial features, Kinect sensor, Face tracking SDK, Neural network, Brain activities.

1. INTRODUCTION

Human can communicate with other through two main ways: verbal and non-verbal. In a conversation or in reviews, facial expressions play an important way to the explain and reflect the ideas beside the words. For this reason, in our lives, it happens in every relationship we care about, like friendships, in the family, in the work, and in our most intimate relationships. In many cases, emotions cause problems while at the same times it can save our lives. Facial Expression involves physical structures movement of the facial muscles, facial expression consists of certain positions of the muscles under the skin surface of the face. Facial muscles positions conveys the emotional state of an individual to observers, and to identify the emotional state of any individual it is necessary to understand his or her facial expressions. Many scientists confirmed about the importance of facial expressions in facing communication. One of the first of them was Charles Darwin. He explained and defines the facial expressions of emotions (happy, anger, fear, surprise, disgust, sadness and other more emotions) and the body language “the language of the emotions” [1]. Facial expressions can also provide information about the cognitive state of a person, such as interest, confusion, boredom, stress, and speech signal [2]. Researchers concerned to analyze and recognize the expressions of the face by using different computer technologies like camera, video camera and developed
many methods to discover facial expressions [3] [4] [5] [6]. Facial emotion recognition applications are numerous, such as virtual agents, robotics, games, etc. [7] [8] [9]. The aim of this research is to use facial emotion recognition in connection with the brain activity to detect the facial expressions and which part of the brain responsible for each emotion expression for a patient, who suffers from a stroke, the patient cannot communicate verbally or very little. This paper 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.

2. RELATED WORK BASED ON KINECT

In the field of Computer Vision, many researchers have reported about automatic facial expression recognition systems.

G R. Vineetha, C. Sreeji, and J. Lentin [10], used Microsoft Kinect sensor for the Xbox 360 video game console with its technique, described by MS Kinect for human face detection. They present a method to recognize six facial expressions in 3D from the input image acquired by the kinect device. First the human face was detected, performed edge detection, thinning, and token detection. The user has to give the input threshold value for the detection of tokens. It is a hard task to decide which is the best threshold value to generate the tokens for the expression detection. In their results, sadness and disgust expressions were more difficult than the others to recognize.

A. Youssef, S. F. Aly, A. Ibrahim, and A. Lynn [11], proposed a system that attempts to recognize facial expressions using a fast three-dimensional (3D) Kinect sensor. They constructed a training set containing 4D data (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 [12] 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 [13] 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 [14]: 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) [15]. The techniques for detection vary greatly and include the use of Gabor wavelets [16], SIFT descriptors [17] and quadtrees decomposition [18].

Many researchers used Kinect sensor for face recognition. Billy Y. L., Li1 Ajmal S., Mian Wanquan LiuLand Aneesh Krishnan [19] 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-D data and 88.7% of noisy depth data alone. Gaurav, G., Samarth, B., Mayank, V., and Richa, S. [20] 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.

3. 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 [21].

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 [20] 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 shows muscles of the head and neck. 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 [22]. The six basic emotions are: anger, disgust, fear, happiness, sadness, and surprise.

Fig -1: Muscles of the head and neck  
http://mewarnaius/634655-gallery-of-muscles-of-head-and-neck

4. KINECT FOR WINDOWS V2.0 SENSOR AND SDK 2.0
We used the Microsoft Kinect for Windows sensor V2 in our system to detect facial emotion expressions, this device adds simplicity to the facial feature extraction.

4.1 Microsoft Kinect for windows V2 sensor
Microsoft Company, Produced the latest Kinect sensor for windows V2, this device with has a modified technology than the older versions, we list some of its specification and features:

1- A built in color camera captures full, great 1080p video which can be re displayed in the viewing screen with the same resolution.
2- A new infrared (IR) emitter, for allowing the sensor to be light-independent viewed.
3- New Depth sensing, it has a higher depth, fidelity and a significantly improved noise floor, and by providing 3D visualization, we have the ability to see smaller objects and all objects more clearly.
4- Microphone array, enabling it to sense the location and movements of people.

For the new Kinect sensor V2, Microsoft provides windows 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.
The 87 points are:
• 16 points for the eyes (0-15, 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.

5. Human brain and emotions
Since the 1970’s, almost the researches in the emotional field based on Ekman and his colleague Wallace Friesen theory that when an emotion appear on the face, a series of electrical impulses stemming from the emotion center in the brain, leads to a specific facial expressions and other physiological changes such as increased or decreased heart rate or increase blood pressure. The question arises, why do we have emotions? If we were happy, we relax; if we feel afraid, we try to escape from the danger, if we were disgusted, we may feel sick. Our emotion affects our behavior. Our ancestors adopted on their emotions for survival, but in these days in our life, we use our emotion to conduct the lifestyle decisions simply staying alive. Emotions make us
react to situations, as an example, when we feeling happy this will make us smile, when we feel fear or anger this will set our heart racing.

We need to understand which area of the brain controls emotions?. One of the key areas of the brain that deals with showing, recognizing and controlling the body's reactions to emotions is known as the limbic system. The limbic system responds to sensory stimuli detected to the body surface, Figure 2 shows the limbic system. It's also associated with pain and pleasure as well as many emotional aspects of behavior [23].

![Limbic System](http://alinenewton.com/neuroscience-of-touch-touch-and-the-brain/)

**Fig -2**: Limbic system

6. PROPOSED 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.

6.1 The steps of emotion recognition

Figure 3 illustrates the steps of our emotion recognition method.

1. Using an 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, left cheekpufff, left eye brow loweredd, left eye closedd, right eye brow loweredd, right eye closedd, right eye brow loweredd, right eye closedd, right eye brow loweredd, right eye closedd, right eye brow loweredd, right eye closedd, right eye brow loweredd, right eye closedd, right eye closepufff, left cheekpuffed, right cheekpuffed, and facepitchh.

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 23 Aus with its active brain area.

4. Write the name of the emotional expression and its active brain area with visitor detection in text files.

5. Using C++ to read the text files and display the following on the screen
   1-Detected facial expression.
   2-Active brain area.
   3- Recognize the visitor.

![Flowchart of our proposed method](http://alinenewton.com/neuroscience-of-touch-touch-and-the-brain/)

**Fig -3**: Flowchart of our proposed method

6.2 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 30 samples of each expression, for each 14 persons.

The person looks at the Kinect camera and makes any expression, we took such snapshots for all the 8 expressions (30 for each). The total number of frames for
6.3 Neural network used

After a few tests, we decided that the best neural network, which gave us a high performance, was 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 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.

6.4 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, is 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. 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:

<table>
<thead>
<tr>
<th></th>
<th>Expression</th>
<th>Confidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.0000</td>
<td>Happy</td>
<td>0% sure</td>
</tr>
<tr>
<td>0.0000</td>
<td>Sad</td>
<td>0% sure</td>
</tr>
<tr>
<td>0.0000</td>
<td>Disgust</td>
<td>0% sure</td>
</tr>
<tr>
<td>0.1000</td>
<td>Fear</td>
<td>10% sure</td>
</tr>
<tr>
<td>0.9000</td>
<td>Surprise</td>
<td>90% sure</td>
</tr>
</tbody>
</table>

Then, the maximum surety of the expression is “Fear”, because the highest confidence value lies at position 6, which corresponds to fear.

6.5 Brain activities

Many researches associated the emotional expression to a specific area in the brain by using functional MRI (fMRI) or Electroencefalografia (EEG) [24] [25] [26] [27] [28]. To find out the brain activity that goes on behind the scenes, during these facial expressions, we use the information provided in the previous reference. The facial expression has been translated into particular emotions. For any particular emotion, a particular region of the brain gets active. Table 1: Represent emotion and its associated active brain area.

Table 1. Emotions and its associated active brain area

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Active brain area</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neutral</td>
<td>Dorsal Anterior Cingulated</td>
</tr>
<tr>
<td>Happiness</td>
<td>Left frontal lobe / Ventral tegmental</td>
</tr>
<tr>
<td></td>
<td>(TVA)</td>
</tr>
<tr>
<td>Sadness</td>
<td>Anterior Limbic System</td>
</tr>
<tr>
<td>Angry</td>
<td>Limbic center</td>
</tr>
<tr>
<td>Disgust</td>
<td>Insular Cortex</td>
</tr>
<tr>
<td>Contempt</td>
<td>Insular Cortex / Amygdala</td>
</tr>
<tr>
<td>Fear</td>
<td>Insular Cortex / Amygdala</td>
</tr>
<tr>
<td>Surprise</td>
<td>Dorsolateral Prefrontal Cortex</td>
</tr>
</tbody>
</table>

To get brain activity information using facial expressions, we also incorporate this information in our code, so that our program tells us
1. Which emotion our patient is experiencing.
2. Which center in his/her brain is active corresponding to this emotion.

The hippocampus deals with the formation of long-term memories and spatial navigation. In diseases such as Alzheimer’s disease, the hippocampus is one of the first regions of the brain to become damaged and this leads to the memory loss and disorientation associated with the condition [29].

The hippocampus is a small region of the brain, which activates when a person recognizes someone. However, if a person does not recognize anyone, any of the 3 cases will arise:

Fig 4: The 23 HD face animation values
1. His face will remain expressionless. For example, if he sees a person approaching him/her, then his face will not be showing any emotion, so his Dorsal anterior cingulated will be active.

2. Or he will look expectantly at the approaching visitor, with no facial expression other than the raising of the inner eyebrow. In this case frontalis (pars medialis) will be active.

3. Or if the patient wants to be polite and attempts a smile towards the, the smile will not be a heartfelt one (it will be a voluntary smile). It will be due to a contraction of the zygomatic major alone.

7. EXPERIMENTAL RESULTS

1- We tested the system on many persons, the facial features for them were stored in the database, we’ve got a very good result and the identification rate was more than 96%. Figure 5 shows the snapshot of our automatic real time facial emotion recognition. When we used other models of Neural Network with different layers, didn’t give us the same error histogram for 15 layers which was so close to 0, figure 6. By using this simple NN, 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 7. The true positive rate is very high and false positive rate is very low.

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: $IR=1- (N/T)$, where $N$ represents the number of false cases and $T$ represents the number of trails in each case. The identification rate ($IR$) for all expressions was 93%.

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.
8. CONCLUSIONS AND FUTURE WORK
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.

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.

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 to make the recognition more general.

ACKNOWLEDGEMENT
This work was co-founded by the Ministry of Higher Education & Scientific Research with a program implemented at College of Engineering, The University of Mustansiriyah.

REFERENCES


19. 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.


27. Michael S. Fanselow and Hong-wei Dong,” Are the Dorsal and Ventral Hippocampus Functionally Distinct Structures?”, Neuron. Author manuscript; available in PMC 2011 Jan 14.
