Detection of the selected, basic emotions based on face expression using Kinect.

Adam Wyrembelski

Supervisor: Ph.D. Aneta Gądek – Moszczak

Abstract
Automation of reading emotion from facial expression is a milestone of communication between human and computer. This paper presents an application and way which allows to recognize four (anger, happiness, sadness and surprise) from six basic emotions defined by Paul Ekman. Application "Emotion Detector" use very simple method to recognize it. Thanks to data provided by Kinect, alghorithm (k – nearest neighbors) can classify user emotion in real time.

Keywords
Emotion detector, emotion, psychology, image analysis, kinect, action units.

1. Introduction

Human face is one of the best source of informations about him and his current emotional state. It is guaranteed by various face muscles movements. Furthermore, this is one of non – verbal communication methods.

This way of communication is especially effective, because there are some emotions (called basic emotions), which expression way on our face is the same over the entire population. In contrast to communication by body or hand gestures, which elements are different in different cultures. Good example of that element is thumbs up gesture. For example, in Poland it means something very positive, on the other hand in Near East and southern Europe this gesture is obscene. In the case of communication with basic emotion is very difficult for such a misunderstanding.

The issues related to emotions and showing it by face expression have fascinated scientist from many year. First research of this case have been performed in ancient times (inter alia in Greece, Babylon, China). But the most important foundations for emotion recognition through facial expression was made by Charles Darwi in XIX century. In his paper [1] he described position of facial muscles in animals and humans and grouped it by meaning.

In nowadays the greatest influence for these two topics has Paul Ekman. Most of projects which main aim is to make emotion recognition process automatic, bases its theoretical part on Ekman's research.

Thanks to many interdisciplinary studies over the centuries, it is possible to automate process of emotion recognition based on face expression.

2. Emotions in psychology

Emotions [2] are defined as strong, rush and relatively unstable mental processes which are followed by some events. Often emotions are directed to subject that caused it.

The simplest way to divide emotions is categorize it as negative, positive or neutral. In a set of negative emotions are situaded, for example sadness, anger or fear. The second set (positive

emotions) contains emotions such as happiness and positive surprise. The last one (neutral category) contains, for example bliss.

However, there is another, very important emotion category, namely **basic emotions** which were defined by Paul Ekman [3]. In his research, he discovered that emotion expression depends only in part from human derivation and he has identified six basic emotions. As mentioned earlier these emotions can be recognized over entire population. This group contains:

- anger,
- sadness,
- happiness,
- surprise,
- fear,
- disgust.

In his papers [3], Paul Ekman has also identified a set of facial features, which characterized an expression of each basic emotions (fig. 1).

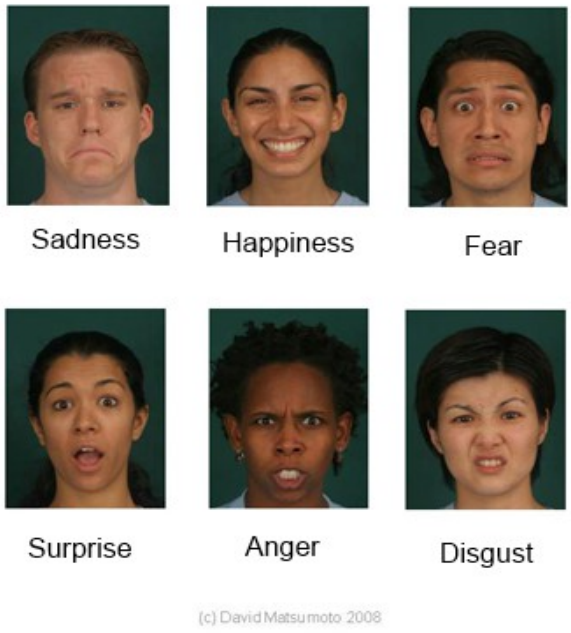

Fig.1. *Six basic emotions*

Source (20.01.2013): http://www.apa.org/ Images/PSA-2011-05-matsumoto-fig1˙tcm7-115934.jpg.

- **Sadness** – inner corner of eyebrowns are raised, eyelids are loose and lip corners are pulled down.
- **Happiness** – muscle around the eyes are tightened, crows feet wringkles appears aroud eyes, cheeks are raised and lip corners are raised diagonally.
- **Fear** – eyebrowns are pulled up and together, upper eyelids are pulled up and mouth are stretched.

- **Surprise** – entire eyebrowns are pulled up, eyelids are also pulled up and mouth are widely open.
- **Anger** – eyebrowns are pulled down, upper lids are pulled up, lower lids are pulled up and lips may be tightened.
- **Disgust** – eyebrowns are pulled down, nose is wrinkled and the upper lip is pulled up.

The ability to recognize and express these emotions has been developing in the process of evolution from thousends of years, therefore are completely natural for human being. Their reading ability allow to take appropriate decisions and reactions when will be noticed one of theme. Moreover ability to express emotions allows to notificate surrounding about our mental state.

3. Kinect as vision system.

Emotion detection based on facial expression analysis is associated with obtain data from environment by a vision system. In this case, as vision system has been used a Kinect. Thanks to which the informations are extracted from environment, and then filtered and processed in methods which are come from *Kinect for Windows SDK*.

3.1 Kinect description

Kinect [4] is an input device created by *Microsoft*. There are two different version: *Kinect for Xbox 360* and *Kinect for Windows*. Thanks to it, users can interact with computer or console by means of their voice or many gestures.

Device with so many features also have application to industry not related to entertainment, such as image analysis or mobile robotic. All usage way is possible thanks to library package (developed by Microsoft) better known as *Kinect for windows SDK*.

3.2 Kinect construction

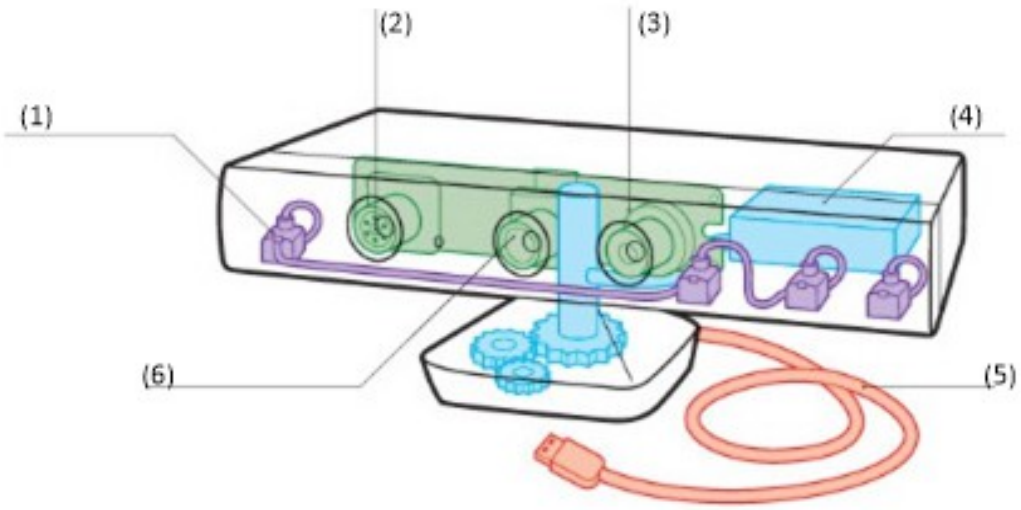

Fig.2. *Kinect construction*

Source (19.01.2013): http:// www.wired.com/magazine/ wp-content/ images/ 19-07/ mf˙kinect2˙f.jpg

(1) Microphone array.

(2) IR Emitter.

(3) IR Sensor.

(4) Tilt motor.

(5) USB Cable.

(6) RGB Sensor.

3.3 Data provided by Kinect

Video data which is collected from environment by Kinect is processed by libraries from SDK, and in the next step, it is send to application as independent streams [5]. Descriptions of streams with information about video data was presented in table 1.

Table 1. - *Descriptions of video streams provided by Kinect*

Name	Features
ColorStream	• Information from this stream is obtain by RGB sensor. • Available formats: RGB, YUV. • Available resolutions: up to 1280x1024
DepthStream	• Data from this stream is obtain thanks to IR sensor and emitter. • All informations contained in this stream include information about distance of each pixel from Kinect. • Available resolutions: 80x60, 320x240, 640x480 px.
SkeletonStream	• Contains information about position of 20 body parts of users which are track in active mode. • Contains information about position of users which are track in passive mode. • Store ID of each tracking user.

Emotion Detector has been used all these streams. Thanks to information contained in theme, it is possible to detect characteristic points on face and track theme in real time. This operations (tracking and detecting) can be done by methods from library *Microsoft.Toolkit.FaceTracking* [6].

Information about characteristic points can be converted by libraries from *Kinect for Windows SDK* to Action Units (AUs) [7], which are the most important information in emotion classification process. The AUs were described thanks to research over FACS (Facial Action Coding System) which have been conducted by two, well known psychologists: Paul Ekman and Wallace V. Friesen.

FACS [8][9] is a system of coding facial muscle movements by their appearance on the face. In paper about this system were described 44 Action Units. However, libraries from *Kinect for Windows SDK* shares only six of theme. Their values (which are beetwen 1 and -1) described position of face muscle groups. All available AUs (with code from FACS) were presented on the figure 3.

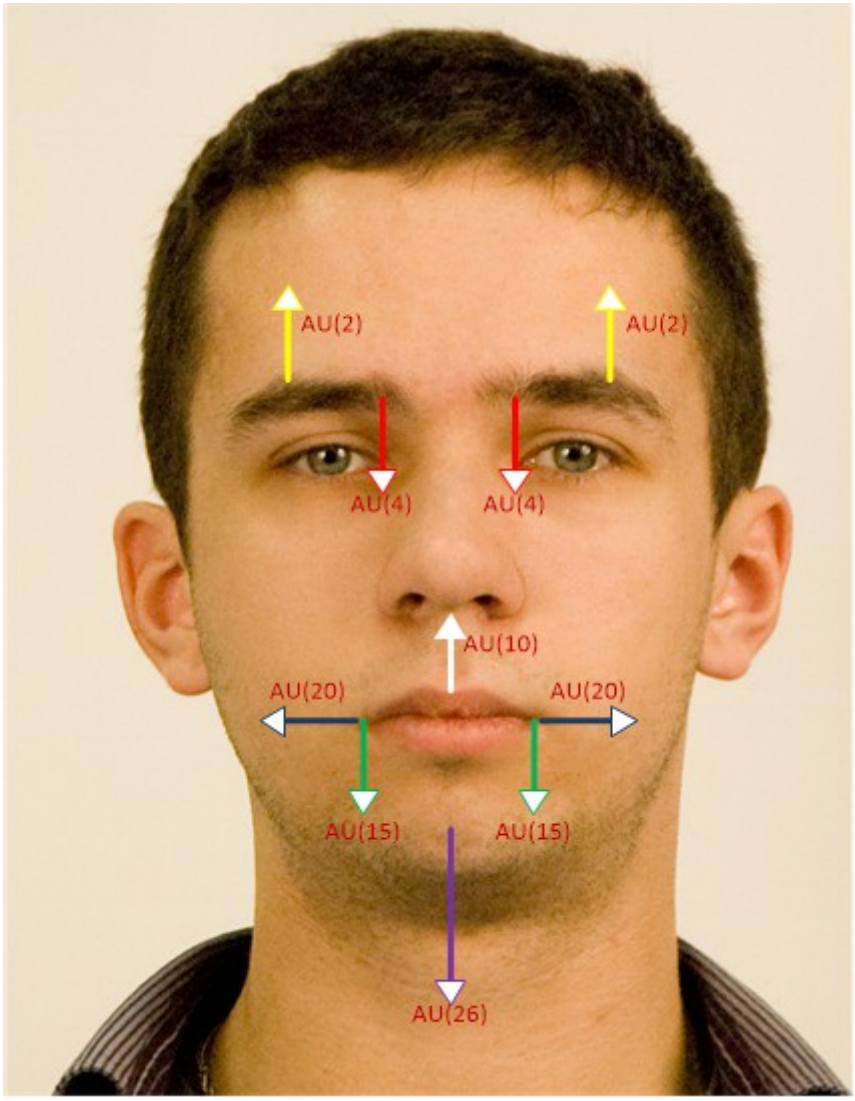

Fig.3*.Position of Action Units on human face*

AU(2) OuterBrowRaiser

AU(4) BrowLowerer

AU(10) UpperLipRaiser

AU(20) LipStretcher

AU(15) LipCornerDepressor

AU(26) JawLowerer

This set of Action Units is completely enough to classify emotion based on user facial expression.

4. Classification mechanism

In *Emotion Detector* in order to classify user emotion based on action units, has been used k – nearest neighbors (KNN) algorithm [10]. It allows to predict a value of variables or classify an object. Main assumption of this algorithm is that, the state of the observed object can be classified based on previous observations of similar objects which were classified with the same features. Object is classified by a majority vote of its neighbors, with the object being assigned to the class most common amongst its **K** nearest neighbors, where **K** is positive number, greater or equal 1.

K – nearest neighbor alghoritm is a type of lazy learing classification method. It means that all computations of similarity objects from training examples to object which is currently classified are making during classification process. The training set was allways created by learing process. In this case, this process based on remember all previous observations.

Very interesting issue in this alghorithm is how to find a similarity of object to another object based on their features. In *Emotion Detector* similarity can be compute with using a Euclidean distance [10], because the emotion features (which are the value of six action units) are represent as set of value between 1 and -1. It can be treated as a vector from six-dimensional space. Basic Euclidean distance formula looks as follows:

$$D\left(p_{1,}p_2\right)=\sqrt{\left(\left(x_1-x_2\right)^2+\left(y_1-y_2\right)^2\right)}$$

But in alghorithm which should find distance (similarity) between six-dimension vectors, this formula looks as follows:

$$D\left(p_1\left(x_{1,}\ldots x_n\right),\ldots,p_n\left(y_{1,}\ldots y_n\right)\right)=\sqrt{\sum_{i=1}^{N}\left(x_i-y_i\right)^2}$$

D – distance (similarity)
p_n – object
$x_1,...,x_n$ and $y_1,...,y_n$ – features sets

By using it, algorithm can find a similarity of each object from training set to currently classified object and choose **K** the most similar objects.

In classification mechanism also has been used a normalization formula for all results afeter similarity computing. Normalization is a process to adjust values which are mesured on differenc scale to strictly specified range. Often it is made in order to allow easier data comparison. In this application has been used a Min – Max normalization method [10]. This method uses following formula:

$$N=\frac{W-min_v}{max_v-min_v}\cdot\left(B-A\right)+A$$

N – value after normalization
W – value to normalize
min_v – minimum value from normalized set
max_v – maximum value from normalized set
[A,B] – range for value after normalization

Thanks to normalization, results of calculating similarity are easier to compare and present to user.

5. Application specification

Emotion Detector has been written in *Windows Presentation Foundation* technology with using C# language based on .NET 4.0 framework. To write and compile application code was used Visual Studio Professional 2012. Another component used in process of developing *Emotion Detector* was *Kinect for Windows SDK* 1.6. The entire process of developing take place on computer with *Windows 8* system.

5.1 Application requiremenst

Application minimal requirements:

- Windows 8 (x64).
- Dual core processor, 2.66 GHZ or faster.
- 4GB RAM.
- 60 MB space on hard drive.
- *Kinect for Windows* or *Kinect for Xbox 360.*

5.2 Application design

Current graphical design of *Emotion Detector* based on modern interface better known as *ModerUX.* Apperance of this application is clear and rich in informations also do not weigh user down by its enormity.
Main window desing of application was presented on figure 4.

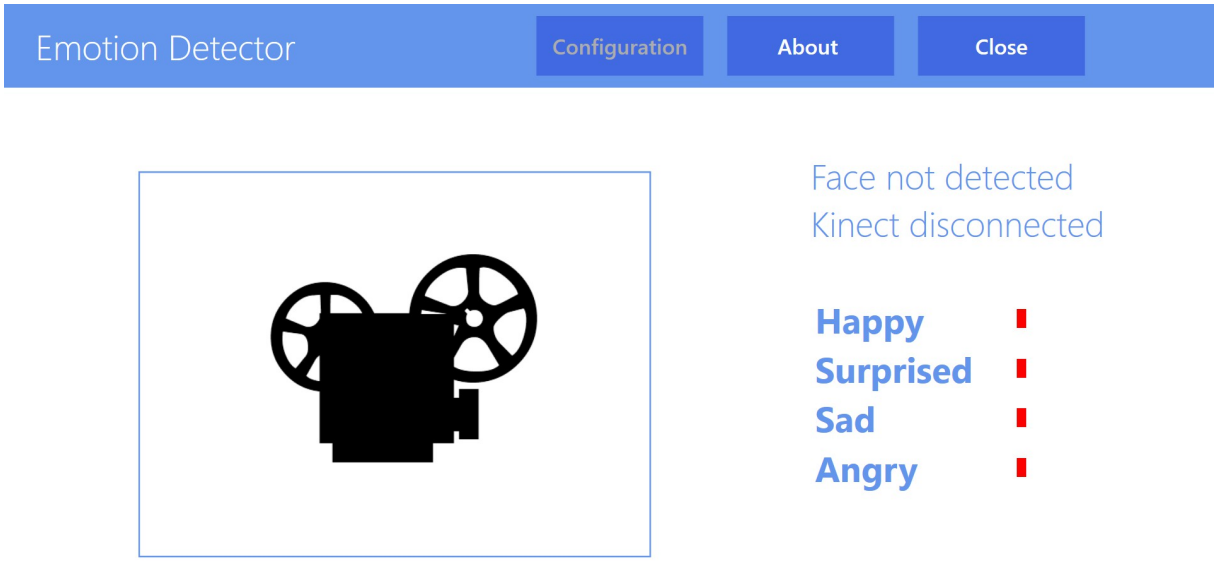

Fig.4*. Main application window*

6. Summary

Emotion Detector has been implemented with using *Microsoft* technology, which provides high performance of using it on system from Windows series, and allows to easily port this application to another platforms developed by *Microsoft,* such as Xbox. As mentioned earlier it is able to detect four basic emotions: anger, happiness, sadness and surprise.

The results obtained with *Emotion Detector* are satisfying. Emotion detection, in people whose data are located in the training set, proceed without any trouble. Small problems can appear when user face expression data is out of training set. But thanks to enlargement training examples by minimum dozens of new observation, this problem can be duly eliminate.

The usage of Kinect as vision system, which provides data from surroundings, characterized by many pros and cons. One of the most significant advantage was library package: *Kinect for Windows SDK,* which directly allowed to detect and track face in real time. However, this solution was not free from defects. The most significant disadvantage was susceptibility to noise from user surrounding which may cause serious problem with proper emotion classification.

7. Bibliography

[1] Charles Darwin, *„The Expression of the Emotions in Man and Animals”*, England, 1904.

[2] Kristina Hook, *„Knowing, Communicating, and Experiencing through Body and Emotion”*, IEEE Transactions on Learning Technologies, vol. 1, no. 4, 2008.

[3] P.Ekman, „*Emotions Revealed. Recognizing Faces and Feelings to Improve Communication and Emotional Life*”, Times Books, USA, First edition 2003.

[4] *„Kinect”*, *http:// en.wikipedia.org/ wiki/Kinect*, Wikipedia, 2013.

[5] *„Kinect for Windows SDK: Natural User Interface for Kinect for Windows”*, *http://msdn.microsoft.com/en-us/ library/ hh855352.aspx*, Microsoft MSDN, 2013.

[6] *„Kinect for Windows SDK, Programming Guide: Face Tracking”*, *http:// msdn. microsoft.com/en-us/ library/ jj130970.aspx*, Microsoft MSDN, 2013.

[7] Lien, Kanade, Cohn, Li, *„Detection, Tracking, and Classification of Action Units in Facial Expression”*, Journal of Robotics and Autonomous Systems, 1999.

[8] J. Przybyło, *„Automatyczne rozpoznawanie elementów mimiki w obrazie twarzy i analiza ich przydatnosci do sterowania”*, AGH University of Science and Technology, Faculty of Computer Science, Electronics and Telecommunications, 2008.

[9] P.Ekman, W.V. Friesen, „*Facial Action Coding System*”, *http://face-and-emotion.com/dataface/facs/description.jsp*, 1970.

[10] Jiawei Han ,Micheline Kamber, Jian Pei, *„Data Mining: Concepts and Techniques.”*, Elsevier, USA, Third edition 2012.