

Real Time Facial Expression Recognition Using Webcam and SDK Affectiva

M. Magdin*, F. Prikler

Constantine the Philosopher University in Nitra, Faculty of Natural Sciences, Department of Computer Science, Tr. A. Hlinku 1, 949 74 Nitra (Slovakia)

Received 23 June 2017 | Accepted 3 November 2017 | Published 9 November 2017



ABSTRACT

Facial expression is an essential part of communication. For this reason, the issue of human emotions evaluation using a computer is a very interesting topic, which has gained more and more attention in recent years. It is mainly related to the possibility of applying facial expression recognition in many fields such as HCI, video games, virtual reality, and analysing customer satisfaction etc. Emotions determination (recognition process) is often performed in 3 basic phases: face detection, facial features extraction, and last stage - expression classification. Most often you can meet the so-called Ekman's classification of 6 emotional expressions (or 7 - neutral expression) as well as other types of classification - the Russell circular model, which contains up to 24 or the Plutchik's Wheel of Emotions. The methods used in the three phases of the recognition process have not only improved over the last 60 years, but new methods and algorithms have also emerged that can determine the ViolaJones detector with greater accuracy and lower computational demands. Therefore, there are currently various solutions in the form of the Software Development Kit (SDK). In this publication, we point to the proposition and creation of our system for real-time emotion classification. Our intention was to create a system that would use all three phases of the recognition process, work fast and stable in real time. That's why we've decided to take advantage of existing Affectiva SDKs. By using the classic webcam we can detect facial landmarks on the image automatically using the Software Development Kit (SDK) from Affectiva. Geometric feature based approach is used for feature extraction. The distance between landmarks is used as a feature, and for selecting an optimal set of features, the brute force method is used. The proposed system uses neural network algorithm for classification. The proposed system recognizes 6 (respectively 7) facial expressions, namely anger, disgust, fear, happiness, sadness, surprise and neutral. We do not want to point only to the percentage of success of our solution. We want to point out the way we have determined this measurements and the results we have achieved and how these results have significantly influenced our future research direction.

KEYWORDS

Facial Expression,
Emotion Recognition,
Real Time Recognition,
Face Detection,
Classification.

DOI: 10.9781/ijimai.2017.11.002

I. INTRODUCTION

At present, an issue of great interest is the correspondence between information reflected in and conveyed by the human face and the person's concurrent emotional experience.

There are several recent studies reporting findings that facial signs and expressions can provide insights into the analysis and classification of emotional states [27], [28].

Many research works have been developed which are related to recognizing the actions of human. Human actions can be recognized in the form of skeleton [31], silhouettes, bio-metric gait [26] and in the form of images [17]. Visual surveillance technique is used for identifying packages of human actions [13]. Researchers used different techniques to recognize human actions such as hierarchical probabilistic approach [1], multi-modality representation of joints [18], [15].

Already long ago Darwin argued that facial expressions are universal, i.e. most emotions are expressed in the same way on the

human face regardless of race or culture [5], [11].

According to Darwin, facial expression can influence the course of communication with another person when we express our emotions, opinions, as well as our intentions in an efficient way. In addition, in his studies of human behavior he explicitly states that such expressions also provide information about the cognitive state of a person. This includes conditions such as boredom, stress, confusion and others.

Darwin's work is remarkable because at the time when he lived he had already assumed that emotional expressions are multimodal patterns of behavior, and thus established his own detailed descriptions of the manifestations of more than 40 emotional states [14].

Although face expressions doesn't necessarily express emotions (e.g. the physiological character of the face after stroke), when performing face detection (if we are talking about the intelligent systems research area), most of the authors of professional publications refer to the classification of facial features that was introduced by Ekman. This is because unlike other emotions (in his period Darwin identified more than 40 expressions), they are easily and unambiguously identifiable. Classification model of 6 emotions by Ekman [6]:

1. happiness,

* Corresponding author.

E-mail address: mmagdin@ukf.sk

2. sadness,
3. surprise,
4. fear,
5. anger,
6. disgust.

Later Ekman [7] added a neutral expression to the previously mentioned six emotions. This method of classifying emotions has gained a relatively high popularity. The main advantage is that facial expressions related to basic emotions are easily recognizable and they can be described even by ordinary people – non psychologists. This model of emotions is currently the most widespread one, used for their detection, and in the last four decades a number of facial recognition and facial parts recognition studies have been developed. Subsequent Russell's critical point of view on the issue (compiling the circular model) in 1994 raised questions about the degree of universality in recognizing these emotional (prototypical) facial states and what these expressions actually signify [24].

In psychological literature, emotions are defined as targeted responses to relevant individuals' suggestions, which include behavior, physiological and experiential components. Various techniques have been developed to enable computers to understand human affects or emotions, in order to predict human intention more precisely and provide better service to users to enhance user experience [16]. If we compare individual approaches, we find that ubiquitous hardware, available in most computer systems (classic webcam), allows to obtain important information about human facial expression. Webcam has become a de facto device thanks to the popularity of interactive social networking applications. A human can oftentimes deduce the emotion of a person sitting in front of a webcam to a certain degree of accuracy. Recent research (in last decade) in video processing and machine learning has demonstrated that human affects can be recognized via webcam video, noticeably via human facial features [32] and eye gaze behaviors [12], [16].

In the publication we will point out our designed and implemented solution, which with the help of a webcam can identify the subject's emotions in real time. When designing and building our system, we encountered many issues that will be described in detail.

II. RELATED WORK

Systems that could not only detect the face in the picture, but also evaluate the emotional states of the subject from individual facial features, have been the subject of research of psychologists as well as pedagogical and information technology specialists for a very long time.

We can say that their interest in this area dates back to the 1960s. One of the first documented studies is Woodrow W. Bledsoe's work for Panoramic Research, Inc. However, this software had been developed for face recognition and facial parts recognition for the US Department of Defense. Therefore, the work is not well accessible to the public. However, with the available materials we know that during 1964-1965, the Bledsoe, Helen Chan and Charles Bisson triad were able to create the first algorithm that ensured facial recognition. Bledsoe was the first to propose a semi-automatic system that worked on the principle of selecting and choosing the coordinates of some parts of the face (those were entered by a human operating a computer), and then these were used as recognition co-ordinates. As early as 1964, Bledsoe has described most of the problems faced even by contemporary facial recognition systems and their parts: change in illumination, turning the head in the recognition process, inappropriate expressions such as grimaces, ageing, and the like. That's why he was looking for other procedures and found that some parts of the face change in adulthood only for a minimum - ear size or distance between eye pupils. His

approaches and techniques were later used in Bell Laboratories, where Goldstein, Harmon and Lesk described a vector that can be used to identify 21 subjective functions such as ear lobe, eyebrow size, nose length, which was later the basis for designing other techniques (Extractions) used to recognize the face and its parts [23].

The first functional full-automatic system was implemented by Kanade in 1973. His algorithm was able to automatically evaluate 16 different facial parameters by comparing extractions obtained from the computer with extractions delivered by humans. His program, however, did correct identification in the range of only 45-75%.

In 1978, Ekman and Friesen developed Facial Action Coding System (FACS) to encode facial expressions, where the facial movements are described by a set of action units (AU) [8]. Each AU is based on the affinity of the muscles. This mimic coding system is manually executed using a set of rules. Entries are static images of the mimetic, often representing the peak of the expression. This process is time-consuming.

Ekman's work inspired many researchers (L. Davis, T. Otsuk, J. Ohyu, Y. Yacoob and others) to analyze facial expressions using an image and video processing. By tracking facial features and measuring the amount of movement on the face, he tried to sort out different facial expressions into individual categories [23].

Further work from this period drew the very ideas from the system developed by Kanade and tried to improve the results he had previously achieved, using methods that provide measurements using subjective functions. An example may be Nixon's work (Nixon in 1985), which proposed a method of measuring the geometric distribution of the eyes. This suggested approach has been improved several times by other authors (for example - Kobayashi & Hara in 1997) [23].

Pantic and Rothkrantz in their work provide a detailed, in-depth recapitulation of many researches on automatic face recognition from 1980 to 2000 [22].

As time went on, the methods for the individual phases of the recognition process (detection, extraction and classification) developed complete solutions (FILTWAM), the so-called SDK (Software Development Kit), various libraries (OpenCV, FaceTracker) and support api appeared [35].

III. AUTOMATIC SYSTEM FOR CLASSIFICATION EMOTION – THEORETICAL BACKGROUND

From the historical point we can say that there are various methods for face recognitions. Such a system for automatic facial expression recognition has to deal with such problems: facial detection and location in a chaotic environment, extraction of facial features and correct classification of the expression on face [34].

Drawing upon various published works [4], [10], [31], [2], [30], main system elements from various similar system have been identified. The main elements of each system to evaluate facial expression are:

1. facial detection,
2. extraction of face features,
3. classification, expression evaluation.

Our main goal was to design and implement a system for real-time recognition and evaluation of human emotions based on a facial expression using a webcam so that the resulting application provides the ability to analyze the collected data. In the future, such a system would be included (e.g. in the form of a plugin) in the LMS Moodle learning environment, so we could better understand the problems that students face while studying. We have already had experience with the design and implementation of such a system, since starting from 2015, the Department of Computer Science, FNS CPU in Nitra, has

been undertaking intensive research in this field. It resulted in a fully functional system which was described in detail from the perspective of the classification success degree in the International Journal of Interactive Multimedia and Artificial Intelligence [19]. However, it was impossible for this system to recognize and evaluate emotional states of more than one person simultaneously with 1 camera. The system worked Real Time, but stored the photos on the disk and there, on the basis of the ViolaJones detector (Haar Cascade Principle), detected the face in the image, extracted the facial features and eventually classified the emotions. In this way, we achieved relatively high accuracy (around 78%), but the system was slow, and the photos took up disk space (even if automatic photo deletion after the correct classification of the expression was set). That's why we decided to design and implement a completely new system that would be faster and work more accurately. At this point, we have got help from establishing the partnership with IBM Slovakia, which resulted in the realization of a project with a name *Pilot project with UKF Nitra University: Modelling the behavior of users based on data mining with a support of IBM Bluemix*.

To determine the level of face detection success and to determine the emotional state of the user with a help of our solution and SDK from Affectiva, we have established the following research question:

RQ1: Can our software unambiguously detect a face, regardless of the distance of a person, surrounding light conditions, rotation of the face or its part, or other non-standard conditions?

Based on the established research question, which is also a research problem for us, it is also possible to think about the hypothesis:

H1: The ability of our software to detect faces in the image and evaluate the user's emotional state is completely independent of the distance of the person being shot, the surrounding light conditions and the rotation of the person's face.

The principles of emotional states analysis through facial expression have a common basis with the principles used in face recognition. The whole analysis process can be divided into three consecutive phases [9]. Those are the detection and localisation of the face in the image with a complex background, including its normalisation, the extraction of appropriate symptoms describing the given expression in the face and finally the determination of the corresponding expression based on the extracted symptoms using the selected classifier. The choice of the observed symptoms (classifier) is always dependent on the approach used and the desired outputs.

As mentioned at the beginning of the publication, we do not need to use only the well-known procedures and methods for the detection, extraction, and classification phases when designing and implementing the system, but we can use available libraries, apis and SDKs that are verified. We decided to use the Affectiva SDK. The advantage of using such a solution is the fact that we do not have to deal with the training dataset in the final phase of the recognition process (classification).

Affectiva is a company that has been operating since 2009 at MIT University, Media Lab, and its main focus is "to teach computers to understand human emotions". Their solutions are mainly used in the advertising sector. During their research, for example, they exposed people to video ads and, with the help of the webcam, captured the people's reactions and analysed the emotions they were experiencing. Based on such observations, they could then evaluate responses to the same ad in different demographic backgrounds. Affectiva also introduced a wearable bio-sensor that can reveal emotions based on skin input.

For the first time, the company released the Software Development Kit for Developers in 2015. The current version of SDK 3.4.1 for the Windows operating system was released in April 2017. The SDK aims to detect facial expressions and classify emotions associated with them. The SDK includes components that enable integration with the.

NET framework and C++ written applications.

The recommended hardware requirements are as follows:

- Processor: 2GHz,
- RAM: 1 GB,
- Free disk space: 950 Mb.

SDK supports Windows 7 and above. For proper operation, Visual C ++ Redistributable runtime for VisualStudio 2013 and Microsoft .NET Framework 4.0 is required. The Affectiva SDK supports multiple platforms, including, in addition to Windows, Android (Java), iOS (Objective-C), Linux (C ++), MacOS (Objective-C), Unity and Web.

SDK solution from Affectiva works on the principle of the extended facial expression model at the following points:

1. Facial detection and detection of significant facial features,
2. Extraction of facial features,
3. Classification of facial action points,
4. Emotional prototype modeling using EMFACS.

Facial detection and extraction of significant facial features - performed using the Viola-Jones Detection algorithm [32]. Landmark detection is then applied to each rectangle bounding the face by 34 identified landmarks. If the benchmark detection confidence is below the threshold, the boundary is ignored. Extracted facial points, head rotation, and a center point between the eyes for each face are displayed in the SDK.

Facial actions - Histogram of Oriented Gradient (HOG) is extracted from the image area of interest defined by the face orientation points. Support Vector Machine Algorithm (SVM), trained on 10,000 manually encoded face images collected from around the world, is used to generate a score from 0 to 100 for each face action. For details on the training and testing system, see [25].

Emotional expressions - in this case, anger, fear, joy, sadness, surprise, and contempt are based on combinations of face action. This encoding was based on the emotional EMFACS face coding system. Emotional expressions get a score of 0 (the expression is missing) to 100 (the expression is present).

Computer classification of facial expressions requires a large amount of data to reflect the diversity of conditions occurring in the real world. Public databases (listed in the Success Rating section) help speed up research progress by providing a reference source for researchers. They represent a complexly identified set of valid spontaneous face reactions recorded in the natural environment over the Internet. In the case of Affectiva's SDK for data collection, viewers watched one of three deliberately entertaining ads on the "Super Bowl" over the Internet while uploading them using a webcam. They answered three evaluation questions about their experience. In addition, some audiences have given their consent to make their data publicly shared with other researchers. This subgroup consists of 242 video footage (168,359 images) recorded in real conditions. The database is completely marked with the following data:

1. Single-frame labels that represent 10 symmetric FACS action units, 4 asymmetric (single-sided) FACS action units, 2 head movements, smile, general expressivity, face and gender facial extraction faults.
2. Placement of 22 automatically recognized landmarks.
3. Participants' own answers to questions: Did you see the video before? Do you want to watch stimulation videos again?
4. Index of basic performance of detection algorithms on this database.

This data is available for distributing to researchers through the Internet [20], [21].

Our user interface (Fig. 1) is very simple, but it is sufficient for the determination of emotional states of the subject.

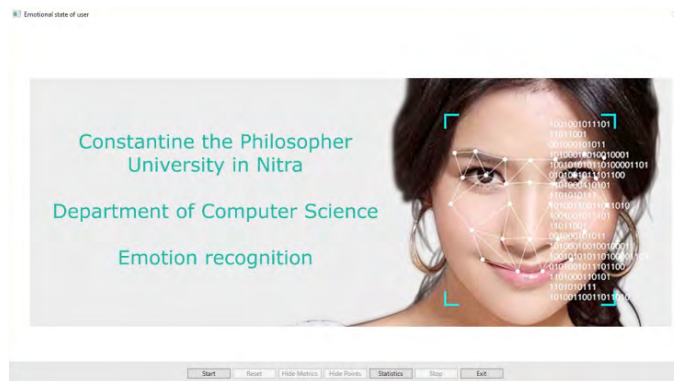


Fig. 1. The preview of the application's user interface (own creation).

The system is capable, under normal lighting conditions, of detecting a face correctly at a distance of 7.5 meters from the recording device, while setting the parameter `Affdex.FaceDetectorMode` to `SMALL_FACES`. The minimum distance for successful face detection is 20 cm.

All controls are located in the main window (Fig. 2). The largest part of the content is the webcam output. In addition to the "Start" and "Stop" buttons, the following elements were added to the beta testing experience:

- "Reset" button - to resume image capture in the event of an error.
- "Show / Hide Metric" button - allows you to view and turn off real-time emotional status evaluation.
- "Show / Hide Points" button - to switch between displaying and turning off the display of points that show facial features.
- "Statistics" button - this element allows you to look at the statistics that the program recorded during the shooting.
- "Exit" button - to turn off the program.

The program is set to start scanning immediately after startup and initial initialization. The graphical interface was defined using XAML (eXtensible Application Markup Language). In the picture above there is the application's user interface in the paused state.

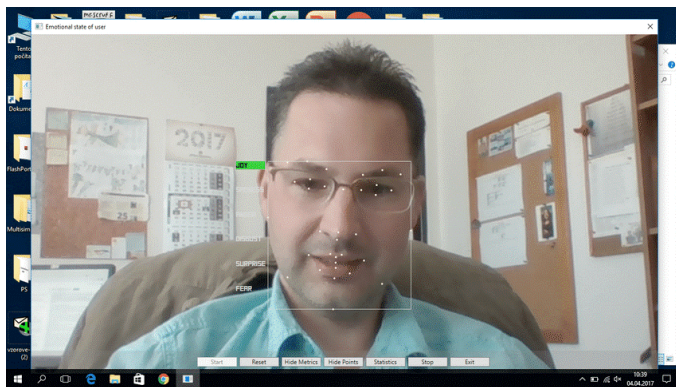


Fig. 2. Demonstration of the application's user interface (own creation).

IV. EXPERIMENT – EVALUATION OF SUCCESS RATE (WITH CONTINUOUS DISCUSSION)

One of the most important aspects of creating a new system to detect or recognize emotions is a database choice that would be used for new system testing. If only one database was used for each research, then new system testing, compared to other top-class systems and efficiency

testing, would be rather trivial tasks [3].

Therefore, we have used various databases that are generally respected by professionals [36]:

1. Bao Face Database,
2. CMU/VASC Image Database,
3. Caltech Faces 1999 Database,
4. NIST Mug-shot Images Database,
5. Yale Face Database,
6. Vision Group of Essex University Face Database.

The testing environment (Fig. 3) was constructed in the following way: on one side, a notebook with software designed to classify emotions was placed at a distance of 1 meter (which is the standard distance of the subject by the webcam). On the other side, a monitor was placed on which test photographs were displayed.



Fig. 3. Testing environment.

Our proposed solution was subjected to evaluation using individual databases. Since each of these databases specializes in different conditions, such as light, turning of the subject's face, the amount of simultaneously captured faces in the image, the result of such an experiment should be a comprehensive evaluation not only of the success rate of face detection but also of the subsequent classification of the emotional state. The evaluation was carried out gradually through the mentioned databases.

Bao Face Database - This database includes lots of face images, mostly of people from Asia (Fig. 4). It is divided into two subdirectories: single face pictures and picture with many faces. The total number of images with one face is 149. Photographs are created in low resolution in bad light conditions. The resolution is most often from 149x207 to 192x273 pixels with DPI 96. Out of this number of images our software detected all 149 faces in the image, but could not classify emotional status with 21 photos. The subjects tested on these 21 photographs had a face rotated by more than 15 degrees. We can say that the software in this case worked with 86% detection success.



Fig. 4. Selection of photos on which a face was detected but the emotional state was not classified.

The second part of the database contained 221 photos. There was a lot of people in each photo. Our goal was to determine if the software can detect faces in a group. Photos are created in different resolutions (from 348x432 to 800x600, DPI 72) in poor lighting conditions. Out of a total of 221 photos, the software was able to detect all the faces on each of the photos and at the same time classify the emotional state of the subject. So he worked with 100% success (Fig. 5).



Fig. 5. Example of face detection and emotional status classification in a picture with multiple people (own creation).

CMU/VASC Image Database - The image dataset is used by the CMU Face Detection Project and is provided for evaluating algorithms for detecting frontal views of human faces. This particular test set was originally assembled as part of work in Neural Network Based Face Detection. It combines images collected at CMU and MIT. The database contains 4 subdirectory: *newtest*, *rotated*, *test*, *test-low*.

The *newtest* subdirectory contains - 65 pictures of movies and fairytales, as well as covers from CDs, cartoons or cards. Images are created in different resolutions (from, for example, 60x75 to the high resolution 1280x1024, DPI 600) in different lighting conditions. The software was able to detect the face and evaluate the emotional state even in non-standard images (Fig. 6).



Fig. 6. Face detection of nonstandard subjects.

We only noticed the problem with two pictures where the system did not detect the face and could not determine the emotional state of the subject (Fig. 7).



Fig. 7. Problem of face detection.

So the software was able to detect the subject's face with a 96% success rate.

The directory *rotated* contains 50 different images (resolution from 109x145 to 2615x1986, DPI 600). These pictures were not rotated. Rotated was the face of a scanned subject. The system, out of a total 50 shots, was able to detect face and classify emotional status in only two shots (Fig. 8). The software only worked with 4% success.

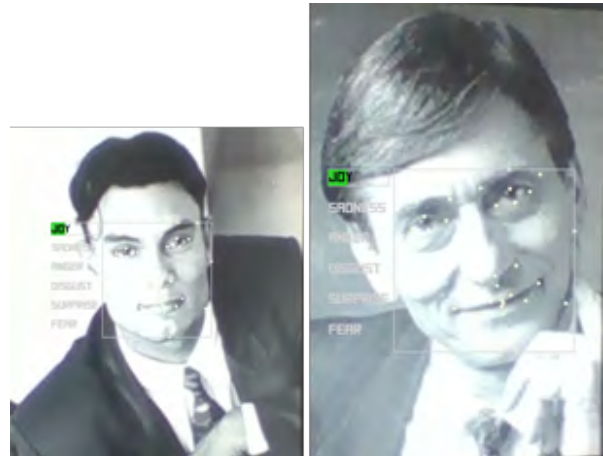


Fig. 8. Determination of the slightly inclined face of the subject and determination of the emotional state.

The directory *test* contains 42 various pictures (resolution from e.g. 142x192 to 716x684, DPI 600). System out of a total of 48 frames failed to detect face and classify emotional status in 3 frames (Fig. 9). The software in this case worked with a 93.75% success rate.



Fig. 9. Demonstration of problem detection of the subject's face.

The directory *testlow* contains 23 various pictures (resolution from e.g. 160x100 to 628x454, DPI 600). The system was able to correctly detect the face and classify the emotional state for each frame (and non-standard, Fig. 10). The software was 100% successful.

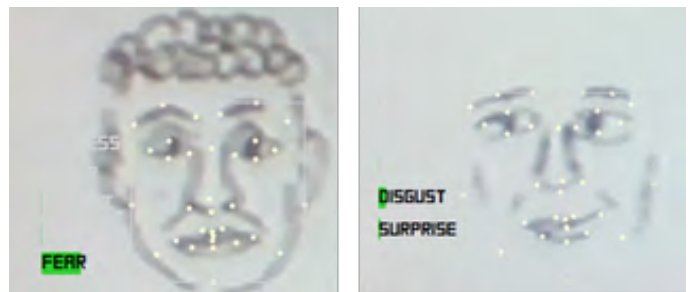


Fig. 10. Demonstration of subject's non-standard face detection.

Caltech Faces 1999 Database – Frontal face dataset. Collected by Markus Weber at California Institute of Technology. Contains 450 face images in great resolution e.g. 896 x 592 pixels, bad lights conditions, Jpeg format. This folder contains 27 jpg files of unique people under with different lighting, expressions and backgrounds (Fig. 11). On all 450 photos, the system was able to detect the face and classify the emotional state of the subject, so we achieved 100% success.



Fig. 11. Demonstration of subject's face detection under poor lighting conditions.

NIST Mug-shot Images Database – contains 2 subdirectory: *single* and *multiple*. For a better idea of the internal structure of these subdirectories, we present the chart in Fig. 12.

The subdirectory *f1_p1* contains 24 subdirectories under the label *sing01* to *sing24*. Each of these subdirectories contains various count of pictures:

sing01 – 68 frontal face pictures, 58 right rotated face pictures, 10 left rotated face pictures

sing02 – 50 frontal face pictures, 42 right rotated face pictures, 8 left rotated face pictures

sing03 – 50 frontal face pictures, 46 right rotated face pictures, 4 left rotated face pictures

sing04 – 50 frontal face pictures, 47 right rotated face pictures, 3 left rotated face pictures

sing05 – 50 frontal face pictures, 44 right rotated face pictures, 6 left rotated face pictures

sing06 – 50 frontal face pictures, 48 right rotated face pictures, 2 left rotated face pictures

sing07 – 50 frontal face pictures, 46 right rotated face pictures, 6 left rotated face pictures

sing08 – 50 frontal face pictures, 48 right rotated face pictures, 2 left rotated face pictures

sing09 – 50 frontal face pictures, 40 right rotated face pictures, 10 left rotated face pictures

sing10 – 50 frontal face pictures, 45 right rotated face pictures, 5 left rotated face pictures

sing11 – 50 frontal face pictures, 43 right rotated face pictures, 7 left rotated face pictures

sing12 – 50 frontal face pictures, 40 right rotated face pictures, 10 left rotated face pictures

sing13 – 50 frontal face pictures, 45 right rotated face pictures, 5 left rotated face pictures

sing14 – 50 frontal face pictures, 48 right rotated face pictures, 2 left rotated face pictures

sing15 – 50 frontal face pictures, 43 right rotated face pictures, 7 left rotated face pictures

sing16 – 50 frontal face pictures, 42 right rotated face pictures, 8 left rotated face pictures

sing17 – 50 frontal face pictures, 45 right rotated face pictures, 5 left rotated face pictures

sing18 – 50 frontal face pictures, 44 right rotated face pictures, 6 left rotated face pictures

sing19 – 50 frontal face pictures, 42 right rotated face pictures, 8 left rotated face pictures

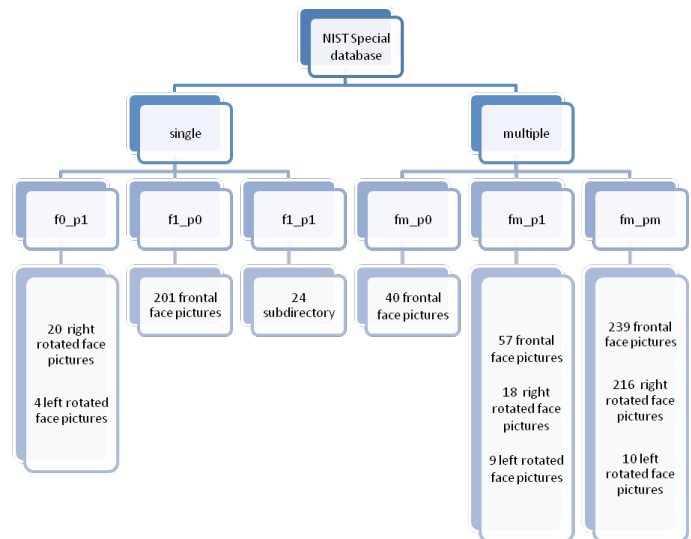


Fig. 12. Internal structure of NIST Special database.

sing20 – 50 frontal face pictures, 45 right rotated face pictures, 5 left rotated face pictures

sing21 – 50 frontal face pictures, 45 right rotated face pictures, 5 left rotated face pictures

sing22 – 50 frontal face pictures, 41 right rotated face pictures, 9 left rotated face pictures

sing23 – 50 frontal face pictures, 56 right rotated face pictures, 4 left rotated face pictures

sing24 – 49 frontal face pictures, 45 right rotated face pictures, 4 left rotated face pictures

With a detailed evaluation of the images from this database we found that the software can detect the face and also determine the emotional state of the subject only in frontal face (100% success rate). In the case of rotated the head to the right or to the left, the software was unable to detect the face and thus not even determine the emotional state (0% success rate).

Yale Face Database – this directory database contains 3 subdirectories: *centered*, *faces* and *rotated*. This database concentrates to the illumination variations. All pictures in subdirectory with a name *centered* are in gray scale, resolution is 195x231 pixels and 600 DPI. Count of pictures is 165. These pictures are divided to the 3 sections: *centerlight*, *glasses/noglasses* and type of *emotional state*. All pictures in subdirectory with a name *faces* is in gray scale, resolution is 320x234 pixels and 600 DPI. Count of pictures is 165. The face of the subject itself is not rotated, rotated is only picture.

With a detailed evaluation of the images from this database we found that the software can detect the face and also determine the right emotional state of the subject in all cases from 3 subdirectories: *centered*, *faces* and *rotated*. The software worked with 100% success rate (Fig. 13).

Vision Group of Essex University Face Database - this robust database is created by Dr. Libor Spacek. Database contains 4 subdirectories: *face94*, *face95*, *face96* and *grimace*. The subdirectory *face94* is divided to 3 subdirectories with name *female*, *male* and *malestaff*. Subdirectory with name *female* contains 20 subdirectories. In each of these subdirectories there are 20 face pictures. We tested therefore 400 face pictures. Subdirectory with name *male* contains 113 subdirectories. In each of these subdirectories there are 20 face pictures. We tested therefore 2260 face pictures. Subdirectory with name *malestaff* contains 20 subdirectories. In each of these subdirectories there are 20 face pictures. We tested therefore 400 face pictures. With

a detailed evaluation of the images from this database we found that the software can detect the face and also determine the emotional state of the subjects in all cases. The software worked with 100% success rate (Fig. 14).



Fig. 13. Face Detection of a subject with a classified emotional state.

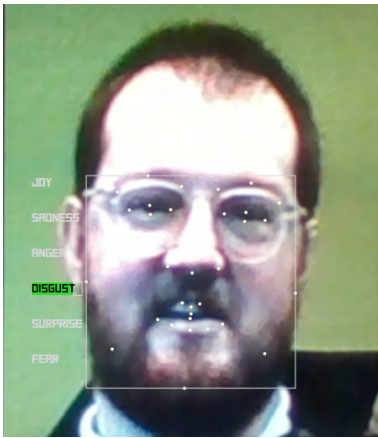


Fig. 14. Demonstration of subject's face detection (directory face94).

The subdirectory face95 contains 72 subdirectories. In each of these subdirectories there are 20 face pictures. We tested therefore 1440 face pictures. These pictures are not divided to the subdirectory female or male. Our software solution recognised all pictures also with slight tilt of head subject (to 15%). The software worked with 100% success rate (Fig. 15).

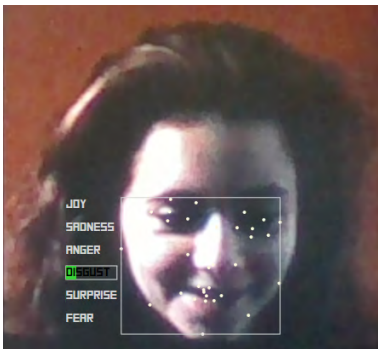


Fig. 15. Demonstration of subject's face detection (directory face95).

The subdirectory face96 contains 152 subdirectories. In each of these subdirectories there are 20 face pictures. We tested therefore 3040 face pictures. These pictures are not divided into the subdirectory female or male. Resolution of each picture is 196x196 pixels with 600

DPI, bad light conditions. With a detailed evaluation of the images from this database we found that the software can detect the face and also determine the emotional state of the subject in all cases. The software worked with 100% success rate (Fig. 16).

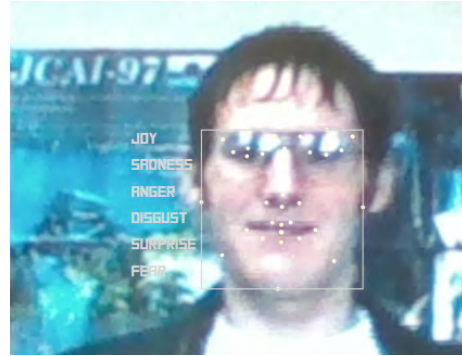


Fig. 16. Demonstration of subject's face detection (directory face96).

The subdirectory *grimace* contains 18 subdirectories. In each of these subdirectories there are 20 face pictures. We tested therefore 360 face pictures. Resolution of each picture is 180x200 pixels with 600 DPI, bad light conditions. With a detailed evaluation of the images from this database we found that the software can detect the face and also determine the emotional state of the subject in all cases. The software worked with 100% success rate (Fig. 17).

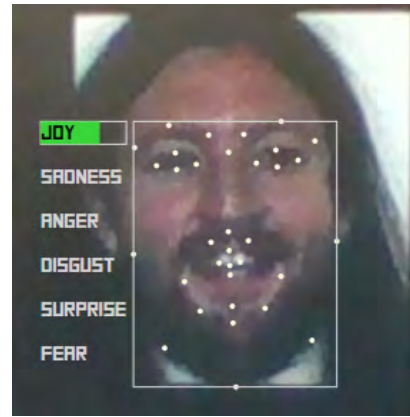


Fig. 17. Demonstration of subject's face detection (directory grimace).

V. GENERAL DISCUSSION

Testing our solution designed using the Affectiva SDK [37] was thoroughly tested with 6 robust databases [36].

1. Bao Face Database,
2. CMU/VASC Image Database,
3. Caltech Faces 1999 Database,
4. NIST Mug-shot Images Database,
5. Yale Face Database,
6. Vision Group of Essex University Face Database.

While experiment, we focused not only on determining the success rate of our solution, but we were wondering if this solution can correctly detect the face in the image under different lighting conditions, regardless of the distance or the face of the subject being shot. At the same time, we were interested, whether the software can not only detect but also evaluate the emotional state of the subject. From these listed databases, only one (Yale) contained images with rated emotional states of the subject. Therefore, other databases have

been used mainly to determine the ability to detect face in the image under different conditions: low resolution, different DPI, poor lighting conditions, head rotation and so on...

We determined the following research question to determine the level of face detection success and determine the user's eco-condition using our Affectiva solution and SDK solution:

RQ1: Can the software designed by us uniquely detect the face, regardless of the subject's distance, ambient light, face or facial part, or other nonstandard conditions?

Based on this research question, which was also a research problem for us, we considered the hypothesis:

H1: The ability of our software to detect faces in the image and evaluate the user's emotional state is completely independent of the distance of the subject being shot, ambient light conditions, and the face of the subject.

The research question was answered by conducting an experiment. We can say that the percentage of face detection in the image and the ability to classify the emotional state is strongly dependent on the rotation of the subject's head. At the beginning of the experiment, we measured the monitor's distance from the webcam and tested the ability to detect the subject's face at different distances. We found that the minimum distance for Face Detection is 20cm, maximum is 7.5m. For the implementation of the test environment, a uniform distance of 1 meter was chosen subsequently. The answer to the research question is that the ability to detect the subject's face by using our Affectiva SDK solution is:

1. Independent of the distance of the scanned object,
2. Ambient lighting conditions,
3. Dependent on facial rotation (if rotation is greater than 15%),
4. Independent from image rotation (face rotation of the object is not realised).

Thus, hypothesis 1 can be partially accepted. The results of the experiment did not need to be statistically verified because the software either detects or doesn't detect the face in the image and statistically calculates the level of emotional expression. Our task, however, was not to determine the percentage of successful classification of the correct emotional expression, but the overall average percentage of success of the detection of the proposed solution using the Affectiva SDK. This is 84.27%.

While conducting the experiment, however, we realised that it would be appropriate to add the solution to the software's ability:

1. Identify the sex of the subject,
2. Determine the age of the subject,
3. Show the duration of individual emotions,
4. Show the duration of detection of at least 1 face of the subject,
5. Show the number of faces successfully detected,
6. Graphically represent the statistical results,
7. To provide the results of the emotional state classification for further processing (save in a suitable format),
8. Detect the face of the subject that is rotated by more than 15 °.

VI. CONCLUSION

Face recognition in real-time using a webcam, face detection itself, as well as the classification of the emotional state of the subject, is currently of great importance in various areas of our lives. Emotions affect our lives, they are a manifestation of how we feel. The emotional status classification can be used in different sectors of our life, from education (determining learner feelings), through industry (employee's

feelings), trade (the area of neuromarketing) through the automotive industry (controlling the aggressiveness of motor vehicle drivers).

In the publication, we pointed out our proposed solution using Affectiva SDK. This solution has been thoroughly tested. The total average detection rate is 84.27%. However, with the front view of the subject (if the subject does not have a rotating head above 15 °), we achieve values around 100%. Looking to the left or right, when the subject's head is over 15% face detection fails as well as the classification of the emotional state. We are aware of this problem and it will be the subject of our further research.

REFERENCES

- [1] P. Augustyniak, et al. Seamless tracing of human behavior using complementary wearable and house-embedded sensors. *Sensors*, 14(5), 2014, p. 7831-7856.
- [2] K. Bahreini, R. Nadolski, & W. Westera. Towards multimodal emotion recognition in e-learning environments. *Interactive Learning Environments*, no. Ahead-of-print, 2014, p. 1-16.
- [3] V. Bettadapura. Face Expression Recognition and Analysis: The State of the Art. arXiv: Tech Report, (4), 2012, p. 1-27.
- [4] Y. Chen, R. Hwang, & C. Wang. Development and evaluation of a web 2.0 annotation system as a learning tool in an e-learning environment, *Computers and Education*, 58(4), 2011, p. 1094-1105.
- [5] C. Darwin. *The Expression of the Emotions in Man and Animals*. Oxford University Press, USA (1998).
- [6] P. Ekman. All Emotions Are Basic. In Ekman & Davidson (Eds.), *The Nature of Emotion*. New York: Oxford University Press, 1992, p. 15-19.
- [7] P. Ekman. Strong evidence for universals in facial expressions: A reply to russell's mistaken critique. *Psychological Bulletin*, 115(2), 1994, p. 268-287.
- [8] P. Ekman, & W. Friesen. *Facial Action Coding System: Investigator's Guide*. Palo Alto, CA: Consulting Psychologists Press, 1978.
- [9] B. Fasel, & J. Luetin. Automatic facial expression analysis: A survey. *Pattern Recognition*, 36(1), 2003, p. 259-275.
- [10] M. Feidakis, T. Daradoumis, & S. Caballe. Emotion Measurement in Intelligent Tutoring Systems: What, When and How to Measure. *Third International Conference on Intelligent Networking and Collaborative Systems*, 2011, p. 807-812.
- [11] G. Giannakakis, M. Padiaditis, D. Manousos, E. Kazantzaki, F. Chiarugi, P.G. Simos, & M. Tsiknakis. Stress and anxiety detection using facial cues from videos. *Biomedical Signal Processing and Control*, 31, 2017, p. 89-101. doi:10.1016/j.bspc.2016.06.020.
- [12] M.X. Huang, J. Li, G. Ngai, & H. V. Leong. StressClick: Sensing stress from gaze-click patterns. Paper presented at the MM 2016 - Proceedings of the 2016 ACM Multimedia Conference, 2016, p. 1395-1404. doi:10.1145/2964284.2964318.
- [13] Y. Juan, S. Dobson, & S. McKeever. Situation identification techniques in pervasive computing: A review. *Pervasive and mobile computing*, 8(1), 2012, p. 36-66.
- [14] D. Keltner. *Born to be good: The science of a meaningful life*. New York: WW Norton & Company, 2009.
- [15] A. Kumar, A. Kumar, S.K. Singh, & R. Kala. Human Activity Recognition in Real-Times Environments using Skeleton Joints. *International Journal of Interactive Multimedia and Artificial Intelligence*, 3(7), 2016, p. 61-69.
- [16] J. Li, G. Ngai, & V. Hong. Multimodal Human Attention Detection for Reading from Facial Expression, Eye Gaze, and Mouse Dynamics. *Applied Computing Review*, 16(3), 2016, p. 37-49.
- [17] Y.F. Li, J. Zhang, & W. Wang. *Active sensor planning for multiview vision tasks*. Vol. 1. Heidelberg: Springer, 2008.
- [18] A.A. Liu, et al. Coupled hidden conditional random fields for RGB-D human action recognition. *Signal Processing*, 2015, p.74-82.
- [19] M. Magdin, M. Turcani, L. Hudec. Evaluating the Emotional State of a User Using a Webcam. *International Journal of Interactive Multimedia and Artificial Intelligence*, 4(1), Special Issue, 2016, p. 61-68.
- [20] D. McDuff, E.R. Kaliouby, T. Senechal, M. Amr, J.F. Cohn, & R. Picard. Affectiva-mit facial expression dataset (AM-FED): Naturalistic and spontaneous facial expressions collected 'in-the-wild'. Paper presented at the IEEE Computer Society Conference on Computer Vision and Pattern

- Recognition Workshops, 2013. p. 881-888. doi:10.1109/CVPRW.2013.130
- [21] D. McDuff, A. Mahmoud, M. Mavadati, M. Amr, J. Turcot, & E.R. Kaliouby. AFFDEX SDK: A cross-platform real-time multi-face expression recognition toolkit. Paper presented at the Conference on Human Factors in Computing Systems - Proceedings, 07-12-May-2016, 2016, p. 3723-3726. doi:10.1145/2851581.2890247
- [22] M. Pantic, & L.J. Rothkrantz. Automatic Analysis of Facial Expressions: The State of Art. *IEEE Transactions on Pattern Recognition and Machine Intelligence*, 22(12), 2000, p.1424-1445.
- [23] F. Pasquale, G. Percannella, & M. Vento. Graph matching and learning in pattern recognition in the last 10 years. *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 28, 2014, p. 1450001 (40 pages).
- [24] J.A. Russell. Is there universal recognition of emotion from facial expressions? A review of the cross-cultural studies. *Psychological Bulletin*, 115(1), 1994, p. 102-141.
- [25] T. Senechal, D. McDuff, & E.R. Kaliouby. Facial action unit detection using active learning and an efficient non-linear kernel approximation. Paper presented at the Proceedings of the IEEE International Conference on Computer Vision, 2016, p. 10-18. doi:10.1109/ICCVW.2015.11.
- [26] V. B. Semwal, et al. An optimized feature selection technique based on incremental feature analysis for bio-metric gait data classification. *Multimedia Tools and Applications*, 76 (22), 2016, pp. 24457-24475.
- [27] N. Sharma, & T. Gedeon. Modeling observer stress for typical real environments. *Expert Systems with Applications*, 41(5), 2014, p. 2231-2238. doi:10.1016/j.eswa.2013.09.021.
- [28] N. Sharma, & T. Gedeon. Objective measures, sensors and computational techniques for stress recognition and classification: A survey. *Computer Methods and Programs in Biomedicine*, 108(3), 2012, p. 1287-1301. doi:10.1016/j.cmpb.2012.07.003.
- [29] M. Suk, & B. Prabhakaran. Real-time facial expression recognition on smartphones. Paper presented at the Proceedings - 2015 IEEE Winter Conference on Applications of Computer Vision, WACV 2015, 2015, p. 1054-1059.
- [30] M. Trojahn, F. Arndt, M. Weinmann, & F. Ortmeier. Emotion recognition through keystroke dynamics on touchscreen keyboards. Paper presented at the ICEIS 2013 - Proceedings of the 15th International Conference on Enterprise Information Systems, 3, 2013, p. 31-37.
- [31] P. Viola, & M. Jones. Rapid object detection using a boosted cascade of simple features. Paper presented at the Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001, p. 511-518.
- [32] L. Xia, Ch.Ch. Chen, & J.K. Aggarwal. View invariant human action recognition using histograms of 3d joints. *Computer Vision and Pattern Recognition Workshops*, IEEE Computer Society Conference on IEEE, 2012.
- [33] Z. Zeng, M. Pantic, G.I. Roisman, & T.S. Huang. A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 31(1), 2009, p. 39-58. doi:10.1109/TPAMI.2008.52.
- [34] Z. Zhang. Feature-Based Facial Expression Recognition: Sensitivity Analysis and Experiment with a Multi-Layer Perceptron. *International Journal of Pattern Recognition Artificial Intelligence*, 13 (6), 1999, p. 893-911.
- [35] Nordic Apis. (2017, November 3). Nordic Apis. Retrieved from: <http://nordicapis.com/20-emotion-recognition-apis-that-will-leave-you-impressed-and-concerned/>
- [36] Face Detection Home Page. (2017, November 3). Face Detection and Recognition. Retrieved from <http://facedetection.com>
- [37] Affectiva. (2017, November 3). Developer Portal. Retrieved from <http://developer.affectiva.com>



M. Magdin

M. Magdin works as a professor assistant at the Department of Computer Science. He deals with the theory of teaching informatics subjects, mainly implementation interactivity elements in e-learning courses, face detection and emotion recognition using a webcam. He participates in the projects aimed at the usage of new competencies in teaching and also in the projects dealing with learning in virtual

environment using e-learning courses.



F. Prikler

F. Prikler was student at the Department of Computer Science. He deals with the theory of teaching informatics subjects, mainly face detection and emotion recognition using a webcam. At present he works as programmer in the company Muhlbauer Technologies.