

Unobtrusive emotion sensing and interpretation in smart environment

Oleg Starostenko^{*}, Ximena Cortés, J. Alfredo Sánchez and Vicente Alarcon-Aquino

Department of Computing, Electronics and Mechatronics, Universidad de las Americas Puebla, Cholula, Pue. 72810, Mexico

Abstract. Currently, a particular focus of human centered technology is in expanding traditional contextual sensing and smart processing capabilities of ubiquitous systems exploiting user's affective and emotional states to develop more natural communication between computing artefacts and users. This paper presents a smart environment of Web services that has been developed to integrate and manage different existing and new emotion sensing applications, which working together provide tracking and recognition of human affective state in real time. In addition, two emotion interpreters based on the proposed 6-FACS and Distance models have been developed. Both models operate with encoded facial deformations described either in terms of Ekman's Action Units or Facial Animation Parameters of MPEG-4 standards. Fuzzy inference system based on reasoning model implemented in a knowledge base has been used for quantitative measurement and recognition of three-level intensity of basic and non-prototypical facial expressions. Designed frameworks integrated to smart environment have been tested in order to evaluate capability of the proposed models to extract and classify facial expressions providing precision of interpretation of basic emotions in range of 65–96% and non-prototypical emotions in range of 55–65%. The conducted tests confirm that such basic as non-prototypical expressions may be composed by other basic emotions establishing in this way the concordance between existing psychological models of emotions and Ekman's model traditionally used by affective computing applications.

Keywords: Affective computing applications, sensing basic and non-prototypical emotions, facial expression recognition

1. Emotion sensing in affective computing

The detection and interpretation of emotions is helpful in areas such as marketing studies and consumption, distance education, affective computing and usability testing, monitoring and training people for more effective interpersonal communication, etc. The subject of emotion analysis has been widely studied by psychologists to determine, which facial features and their intensity precisely describe the emotional state of human [2,7,11].

As usual, six basic consistently recognized emotions that all the people express in the similar manner are analyzed. They are *happiness, sadness, disgust, surprise, anger, and fear*. However, for the classifica-

tion of states related to moods, feelings and attitudes, the complex non-prototypical emotions must be taken into account. These specific unusual emotions such as *serenity, acceptance, trust, admiration, distraction, boredom, disapproval, awe, optimism, aggressiveness* and others may be considered as combination of basic emotions and their recognition depends on the context, mood, culture, personality, breed, gender or may be also the result of incomplete stimuli presentation or variation in the strength of emotion expression [23].

Quite acceptable model of non-prototypical emotions has been proposed by Plutchik, who describes the relationships between emotional concepts as it is shown in Fig. 1 [27]. The vertical dimension represents intensity of emotion while the circle represents the degree of similarity between them. The eight sections correspond to the eight dimensions of the primary emotions defined by the model, which are arranged

^{*}Corresponding author. E-mail: oleg.starostenko@udlap.mx.

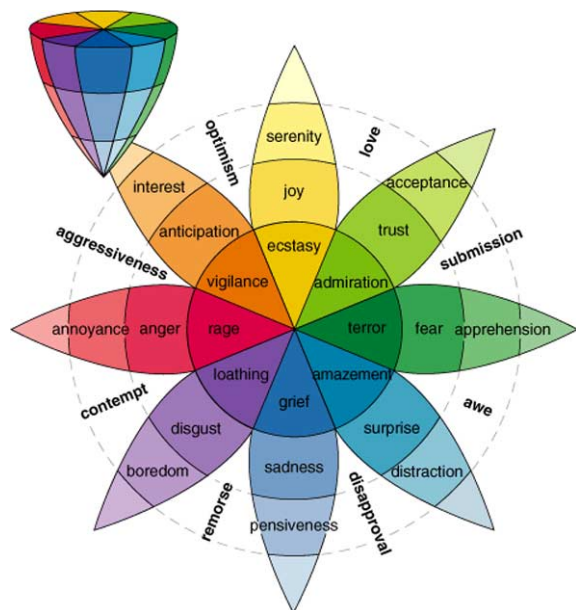


Fig. 1. Plutchik's emotion model. Reprints from J. American Scientist [27].

into four pairs of opposing emotions. The emotions of the blank spaces are called primary dyads, which are the combination of two primary emotions.

Recently, a novel model introducing compound expressions is widely discussed by researchers [6]. The fundamental concept used in last decades, when only six basic emotions: *happiness*, *surprise*, *sadness*, *anger*, *fear*, and *disgust* may be consistently described by facial expressions, now is extended to 21 emotions, which are expressed in the same way by everyone.

It means that the number of emotions recognized by observers or cognitive and affective systems is larger than previously thought. Compound emotions are created combining two basic component categories of six traditional emotions or neutral one. These computational models of the perception of facial expressions of emotion will be successfully exploited to achieve great recognition results in computer vision applications [6,22].

Usually, automatic emotion interpretation process consists of three steps: head finding and face detection in a scene (image segmentation and face localization), extracting facial features (described by pixel position, colors, shape deformations, regions, texture dynamics, etc.) and finally, emotion classification into some categories useful for interpretation.

Numerous techniques and many high performance approaches have been proposed for facial expression

recognition and emotion interpretation however, they have no particular emphasis to real world applications in context of human activity analysis providing contextual sensing and smart processing capabilities of modern ubiquitous systems [2,3,17,21,36]. It is important to define particular specifications and features required for development of cost-sensitive approaches. Among them the most useful features are high processing speed (design of real time applications), acceptable precision (correct interpretation not only basic but also non-prototypical emotions), low cost (accessible for any user), portability (simple migration to any platform), low complexity (running on mobile devices with restricted capabilities), compatibility (simple integration to existing smart environments), reliability (assurance of confident results using standard or non-standard databases or digital collections) and others.

Several authors propose to develop systems for measuring non-prototypical emotions such as *enjoyment*, *hope*, *pride*, *relief*, *anxiety*, *shame*, *hopelessness*, *boredom*, etc. providing more natural communication between computing artefacts and users [21,23,32,36,37,41]. This problem may be solved by development of novel models for ambient aware systems that take into account emotional state of users [2,3,10,40].

Therefore, an extensible software platform is introduced. It provides developers and users with uniform interfaces and services so their applications can access the results and resources from existing or newly implemented emotion sensing tools in real time. In addition, two original 6-FACS and Distance models for emotion interpretation based on recognition of facial expressions and measurement of their intensity using either Ekman's AUs (Action Units) or FAPs (Facial Animation Parameter) and FDPs (Facial Definition Parameter) of MPEG-4 standard are proposed and implemented.

The paper is structured as follows. First, an overview of background and the related works for facial expression recognition are presented. Then, the proposed platform that integrates existing and novel emotion interpreters is described. In the next section two models for measuring facial deformations and corresponding emotion interpretation are presented and evaluated. Then, the evaluation of the proposed models is provided discussing used standard databases, designed system functionality, performance and ability to recognize complex non-prototypical expressions. Finally, the critical discussion of obtained results, contribution and future works are presented in conclusion.

Table 1
Basic emotions encoded by AUs

Emotion	Primary Visual Cues				Auxiliary Visual Cues					
	AU	AU	AU	AU	AU	AU	AU	AU	AU	AU
Happiness	6	12				25	26	16		
Sadness	1	15	17			4	7	25	26	
Disgust	9	10				17	25	26		
Surprise	5	26	27	1 + 2						
Anger	2	4	7	23	24	17	25	26	16	
Fear	20	1 + 5	5 + 7			4	5	7	25	26

2. Facial feature extraction and classification

The facial actions expressed by visually detected displacement of fiducial points or landmarks are defined either by the *Action Units* (AUs) of Ekman's *Facial Action Coding System* FACS or by *Facial Animation Parameters* (FAPs) and *Facial Definition Parameters* (FDPs) of MPEG4 standard used for coding of audio-visual objects [5,7,9]. Action units represent muscular activity that corresponds to unique facial changes.

FAPs are sets of parameters used in animating MPEG-4 model that defines reproduction of emotions from facial expressions. Each parameter set is closely related to muscle actions. The definition of FAPs is based on fiducial points defined by manual or automatic tool for extraction of face features FDPs [16]. The change of facial features (displacement of facial fiducial points) are classified and encoded by forty four AUs. Each AU represents the simplest visible facial motion, which cannot be broken down into more simple units.

Table 1 defines a well-known model for recognizing six basic emotions in terms of the AUs [7]. FDPs are the references used in MPEG-4 standard either to customize facial model or to express emotions for example, on an animated robot [4,5]. Thus, the relative motion of face gestures may be quantified for recognizing particular expression by processing AUs or FAPs through FDPs using well-known methods of computer vision. Figure 2 shows FDPs used in MPEG-4 standard for description of FAPs.

In Table 2 some examples of facial actions defined by FAPs encoding two basic emotions (*anger* and *sadness*) are shown. *MNS* (Mouth-Nose Separation) or *MW* (Mouth Width) means the measuring unit for particular FAP. For example, in the second row of Table 2 the FAP4 (*lower_t_midlip*) for vertical top middle inner lip displacement is defined by *MNS* distance that has bi-directional (*B*) displacement.

Its value is decreased (*down*) with respect to reference point FDP 2.2 presented in Fig. 2 [35]. In the last column of Table 2 the encoding *anger* and *sadness* emotions by FAPs is presented. For instance, *squeeze_l_eyebrow(+)* means the contraction of the left eyebrow, when the distance from reference point used for this FDP is increased (+). The particular FAP is defined by position of fiducial points (FDPs in MPEG-4) that may be easily detected and measured during their displacement.

There are several techniques that have been used for facial feature extraction. Some of them are based on Gabor filter that significantly improves accuracy of expression recognition [17], the active appearance and geometric models, which allow detecting more accurately the dimensional facial changes under large variation [19,20,42], the principal components analysis and hierarchical radial basis function network to provide better feature discrimination of similar emotions [18], the optical flow and deformable templates for flexible pose- and texture-independent approaches that exploit head pose independent temporal facial action parameters [5,14,34], the multilevel Hidden Markov Model [16], the dynamic Bayesian belief networks to perform Bayesian inference based on statistical feature models (SFM) and Gibbs-Boltzmann distribution [44], the models based on SIFT features resilient to object scaling, rotation and noise in image [1,12,43] and others.

The common disadvantages of mentioned approaches are the presence of errors during spatial sampling, restrictions for input visual queries, which must have small number of well-defined and separated faces without occlusion, sensitivity to scaling or rotation of analyzed regions, low precision of recognition if features in image have weak borders or complex background. The analysis of factors like tolerance to deformation, robustness against noise, feasibility of indexing of facial expression, significant amount of required memory are other factors that must be taken into ac-

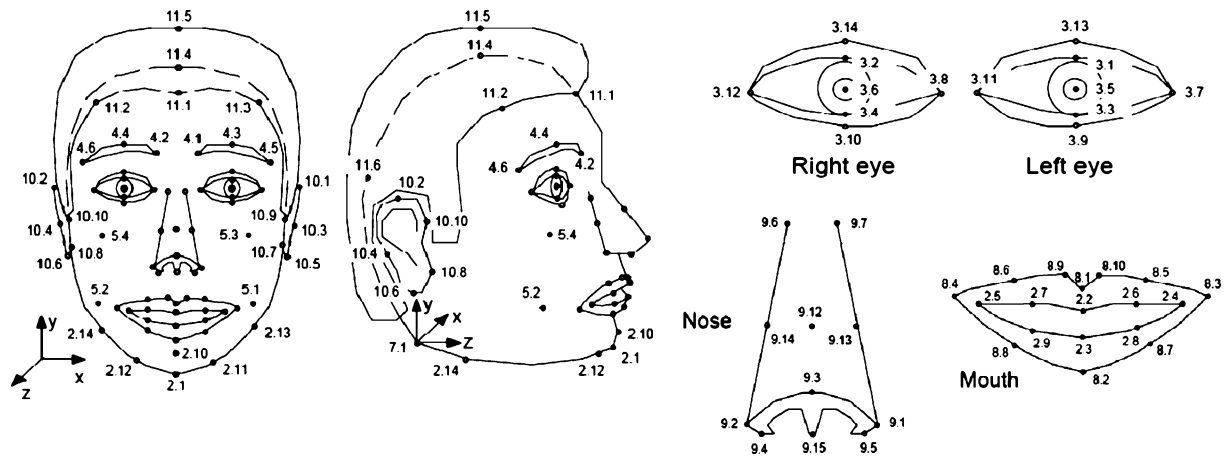


Fig. 2. FDPs used in MPEG-4 standard for description of FAPs [13].

Table 2
Description of FAPs used for codification of *anger* and *sadness* emotions

FAP	Action	Unit	Direct	Changes	FDP	Emotion	FAPs for Emotion
open-jaw	Vertical jaw displacement not affect mouth opening	MNS	U	down	2.1	Anger	squeeze_l_eyebrow(+) lower_t_midlip(-)
lower_t_midlip	Vertical top middle inner lip displacement	MNS	B	down	2.2		raise_l_i_eyebrow(+) close_t_r_eyelid(-) close_b_r_eyelid(-)
raise_b_midlip	Vertical bottom middle inner lip displacement	MNS	B	up	2.3		raise_l_i_eyebrow(+) close_t_l_eyelid(+)
stretch_l_cornerlip	Horizontal displacement of left inner lip corner	MW	B	left	2.4	Sadness	raise_l_m_eyebrow(+) raise_l_o_eyebrow(-) close_b_l_eyelid(+)

count during development of models for emotion interpretation.

It is important to mention some relevant systems for emotion sensing and interpretation, which features are resumed in Table 3. Each of existing approaches exhibits advantages and disadvantages, depending on the contexts, in which they are used.

Also there are several systems that help developers to incorporate specific approaches for emotion sensing into their applications. However, programming libraries and development kits typically focus on the use of one particular approach, which relies on its own tool, representation model and data collection method. This makes it difficult for programmers to combine emotion detection techniques and develop applications with higher performance. To take advantage of wide variety of systems and approaches we propose extensible software platform that may be considered as a smart environment, which combines embedded emo-

tion sensing computing devices and applications to access through uniform interfaces services and resources in real time.

Analyzing relevant systems used for emotion interpretation from facial expressions, it is important to mention that they as usual, do not compute the intensity of emotions and interpret only six basic emotions without analysis of non-prototypical ones. Frequently, researchers propose models based on non-standard facial features and report high recognition results obtained from tests on prepared ad-hoc face data bases making them useless for implementation and reproduction in practice. Even though these approaches have high recognition precision, there are not reports about how they may link standard facial actions with particular formal models or rules for automatic emotion interpretation. Finally, used algorithms are quite complex and slow to operate in real-time applications or run on mobile devices with limited capabilities.

Table 3
Comparison of relevant systems used for emotion interpretation

System	Recognized Emotions	RES	RER	OS	DMP	API	CA	Eo/c	Mo/c	POF	DDL	Classification Method, Recognition Precision
Compound Expressions [6]	21 expressions discriminated as basic	Y	Y	N	N	N	Y	Y	Y	N	N	SVM, 6 basic expressions: 96.9% accuracy, 15 compound expressions: 76.9%
eMotion [33]	Neutral +6 basic, no intensity	Y	N	N	N	N	N	N	N	Y	N	Units of Movement, Bayesian net and SVM classifiers
Expert System [26]	5 basic, no intensity	N	N	N	N	N	N	Y	Y	N	N	AUs multi-detector, 86%
FaceReader [25]	Neutral +6 basic	Y	Y	N	N	Y	Y	Y	Y	N	Y	Active Appearance Model, FACS model, 85%
Fuzzy Facial Expression Recognition [19]	6 basic, no intensity with 3 level intensity	Y	N	N	N	N	Y	Y	Y	Y	Y	KNN classification: 75.2% fuzzy inference system (FIS): 91.3%, GA: 93.96%
Fuzzy Classifier [8]	6 basic with 3 level intensity	Y	N	N	N	N	N	Y	Y	N	Y	Distance model with angle analysis, 72%
RealEyes [32]	Boredom, anxiety, anger, tension, enjoyment	N	N	Y	N	N	N					machine Learning
SHORE [30]	Neutral +4 basic, (joy, anger, sadness, surprise)	N	N	N	Y	Y	N	Y	Y	Y	N	Facial: Machine Learning (AdaBoost) using lookup tables, 95.3 % (for happy)

Used abbreviations: RES – Ranking emotional state, RER – Record of emotional reports, OS – open source, DMP – Detection of multiple persons, CA – Calibration Ability, Eo/c – Eyes opened/closed, Mo/c – Mouth opened/closed, POF – Partially occluded face, DDL – Detecting direct look.

3. Extensible platform for integration of emotion interpreters

There exist several systems that automatically interpret emotions using different approaches and classification techniques. For greater accuracy of analysis of emotional state the behavioral and physiological features are suggested to take into account such as facial expressions, voice, gesture, vital signs and others. As usual, the physiological measures are obtrusive. In contrast, voice analysis and computer vision measures are unobtrusive but noise sensitive [29]. For processing only unobtrusive computer vision measures a Web-based extensible platform is proposed. It facilitates integration and management of new and existing applications for emotions sensing and analysis. In the future, a Web platform may be expanded with services for analyzing other behavioral and physiological features providing better and more robust results.

The platform consists of Web services, which are responsible for keeping information detected by integrated emotion interpreters. This platform will be useful to demonstrate the importance of emotion sensing applications working together and tracking in real time facial expressions recognizing them and displaying detected basic and non-prototypical emotions although integrated to platform automatic emotion interpreters have not been designed precisely for analysis of user emotional state in particular context of virtual educa-

tion, pilot training, interpersonal communication, market evaluation and others.

The proposed platform presented in Fig. 3 operates as an intermediate agent between users (clients) and integrated emotion sensing applications. A client has access to all records with results of emotion interpretation by n integrated emotion sensing applications via *Client Web Application*. This Web application has been developed in HTML5 so that, all the information may be visualized on different devices such as computers, tablets or smart phones.

Integrated to platform computer applications are communicated to the *Server* to access the content (recorded processed videos, digital collections, reports of expression recognition sessions, user content, etc.) that may be displayed to users and stored to database. The proposed platform was developed in ASP.NET 4.5 using as initial template ASP.NET MVC 4 Web application in C # with MS SQL database.

The server of Web application is based on the MVC (*Model-View-Controller*) design pattern. The objects of *Model* are classes that represent relationship between emotions detected by different emotion interpreters connected to platform. Designed *Controllers* are responsible for processing requests from *Emotion Sensing Applications* and work with the *Model* and the *View*.

API Controller represents service layer to manage emotion detection applications and to control sessions, user content, files, emotion description, etc. The *Views*

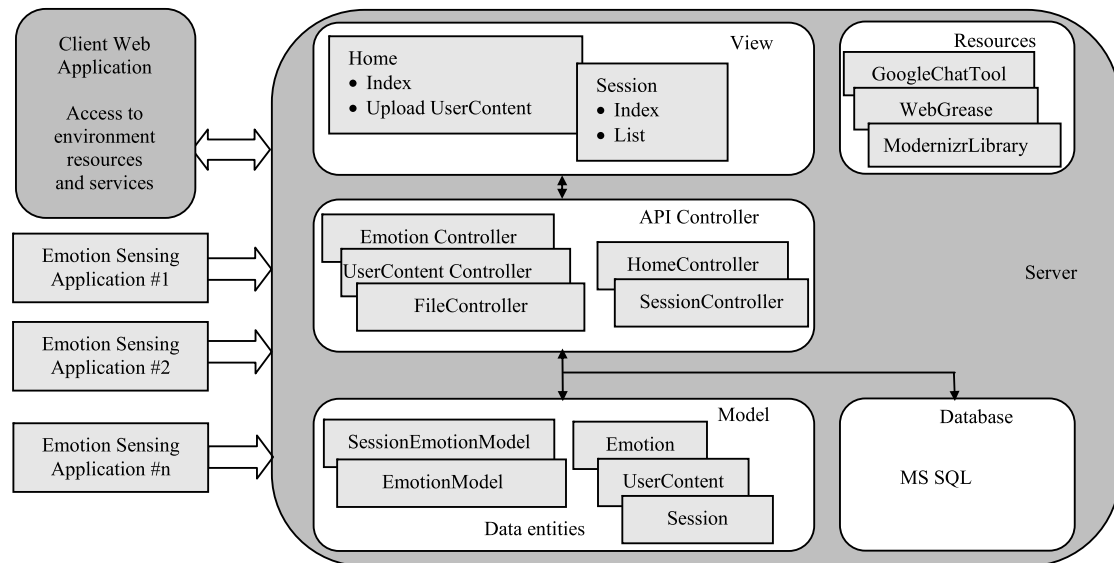


Fig. 3. Block diagram of the proposed extensible Web-based platform for integration and management of emotion sensing applications.

are pages, where users will see the content of sessions, in which emotions are interpreted.

To get the data and display graphs of interpreted emotions, JQuery JavaScript libraries and Google Chart Tools are used. JQuery is open library for multiple browsers that simplifies HTML coding client side, event handling and automating Ajax usage. Google Chart Tools is a library that provides various types of plotted graphs using HTML5/SVG technology, which like JQuery supports multiple browsers.

Provided by server REST services are the central part of this architecture. They are represented by the controllers and can be accessed by applications and monitoring services. Communication is based on stateless HTTP requests, which result in very low network requirements, given the simple attributes defined for information exchange. The main entities in the model comprise users, emotions and sessions.

In order to facilitate interoperability among applications and emotion analysis components, the data model is intentionally kept very simple. This allows for any technique to map its results to the available set of emotions and the instants, in which emotions are detected as timestamps.

4. Proposed facial models for emotion analysis

4.1. 6-FACS model for facial expression recognition

FACS describes expressions in terms of the configuration and strength of action units. For estimation

and recognition of AU strengths various types of local 2D and 3D shape indicators have been considered such as mean curvature, Gaussian curvature, Gabor moments, shape index and curvedness, as well as their fusion [1,14,30,39].

However, in the literature various fiducial points have been selected as relevant ones that better define particular facial expression. For example, eye opening, mouth opening, mouth corner angles and eyebrow constriction have been chosen as the most relevant features [10,19,36].

Taking into account these relevant features additionally some requirements must be considered in development of facial model. For instance, facial recognizer must be fast enough to be used in real time applications, the complexity of used algorithms for emotion interpretation must be low nevertheless it must provide acceptable precision of recognition and finally, emotion sensing tool must be designed on known standards for its simple integration to smart environments.

After various experiments with existing applications for facial expression recognition, only six relevant features, which ensure the correct description of basic and non-prototypical emotions, have been adopted in the proposed facial model [25,30,32]. They are presented in Table 4. For 6-FACS model it was decided to implement emotion interpreter based on AUs of Ekman's FACS. Due to specific requirements defined for facial features and in order to explore new technologies, the Microsoft's Kinect sensor has been used. It consists of a standard CMOS image sensor that cap-

Table 4
Facial expression features adopted in the proposed 6-FACS model

Facial Action Parameter	Action Unit Description	
	Kinect	FACS
Raised Upper Lip	AU0 – Upper Lip Raiser	AU10 – Upper Lip Raiser
Opened Mouth	AU1 – Jaw Lowerer	AU26 – Jaw Drop
Stretched Lips	AU2 – Lip Stretcher	AU27 – Mouth Stretch
Lowered Eyebrow	AU3 – Brow Lowerer	AU4 – Brow Lowerer
Squeezed Corners of Lips	AU4 – Lip Corner Depressor	AU12 – Lip Corner Puller
Raised Outer Eyebrow	AU5 – Outer Brow Raiser	AU2 – Outer Brow Raiser

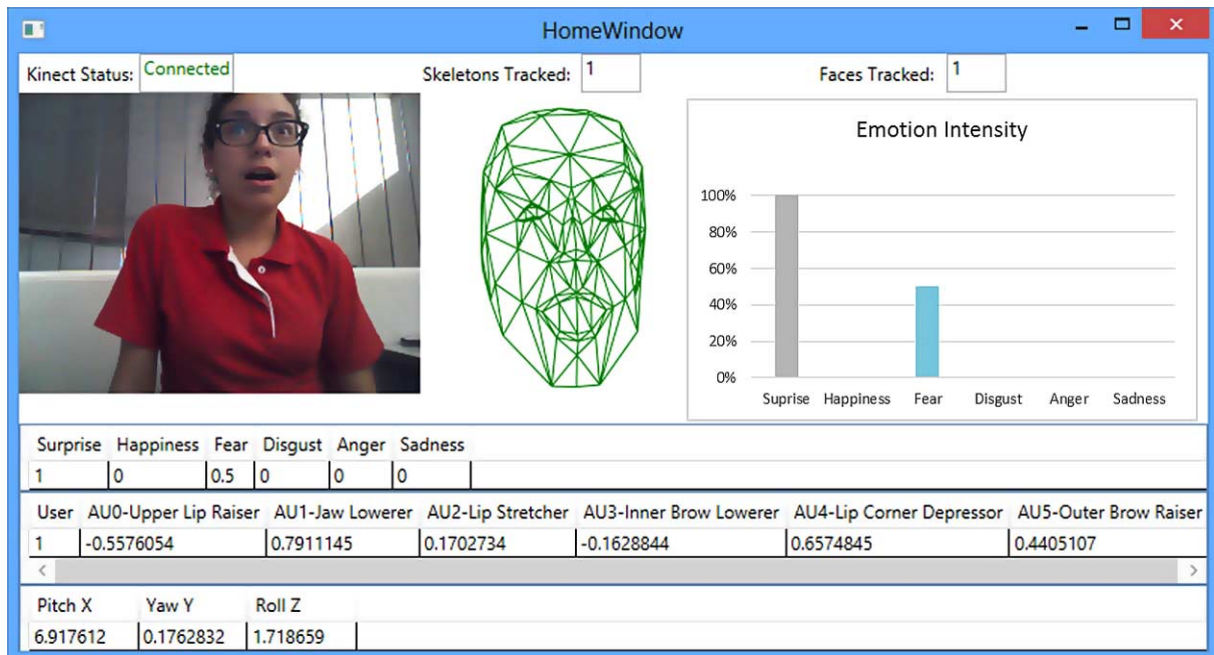


Fig. 4. Kinect session with detected AUs and interpreted emotions: *surprise* (1.0) and *fear* (0.5).

tures the reflected light from the laser (this technology allows to build 3D depth maps in great detail); an integrated RGB camera with a maximum resolution of 1600×1200 (UXGA) and an integrated auditory microphone [40].

Available software for interpreting information from the Kinect sensor as OpenKinect, CL NUI and OpenNI could not provide facial expression recognition [38]. Therefore, for emotion interpreter that captured facial gestures described by Kinect-Codification Facial Stocks (see Table 4), the.NET-based application has been designed using C#.

Due to limited number of action units and taking advantage of the existing works, the initial method of FACSaid program conducted by Friesen [10] has been exploited, in which the decision-making process

to categorize emotions is made considering the assessment of the value of essential FACS [40]. Connected to designed platform that has been introduced in previous section, the Kinect based emotion interpreter uses six predefined AUs that may be appreciated in the interface shown in Fig. 4. Kinect measures AUs in the interval $[-1, 1]$. If all the AUs have intensity equal to 0, the face is considered as neutral.

Using interval provided by Kinect and directions of AU changes, the simplified version of description of expression intensity is provided by following levels: absence of expression is 0.0, medium is 0.5 and maximum is 1.0. At the middle of interface the head position is found and in the bottom line the values of detected rotations in three axes are presented. Possible values of head rotation are from -90° to 90° although

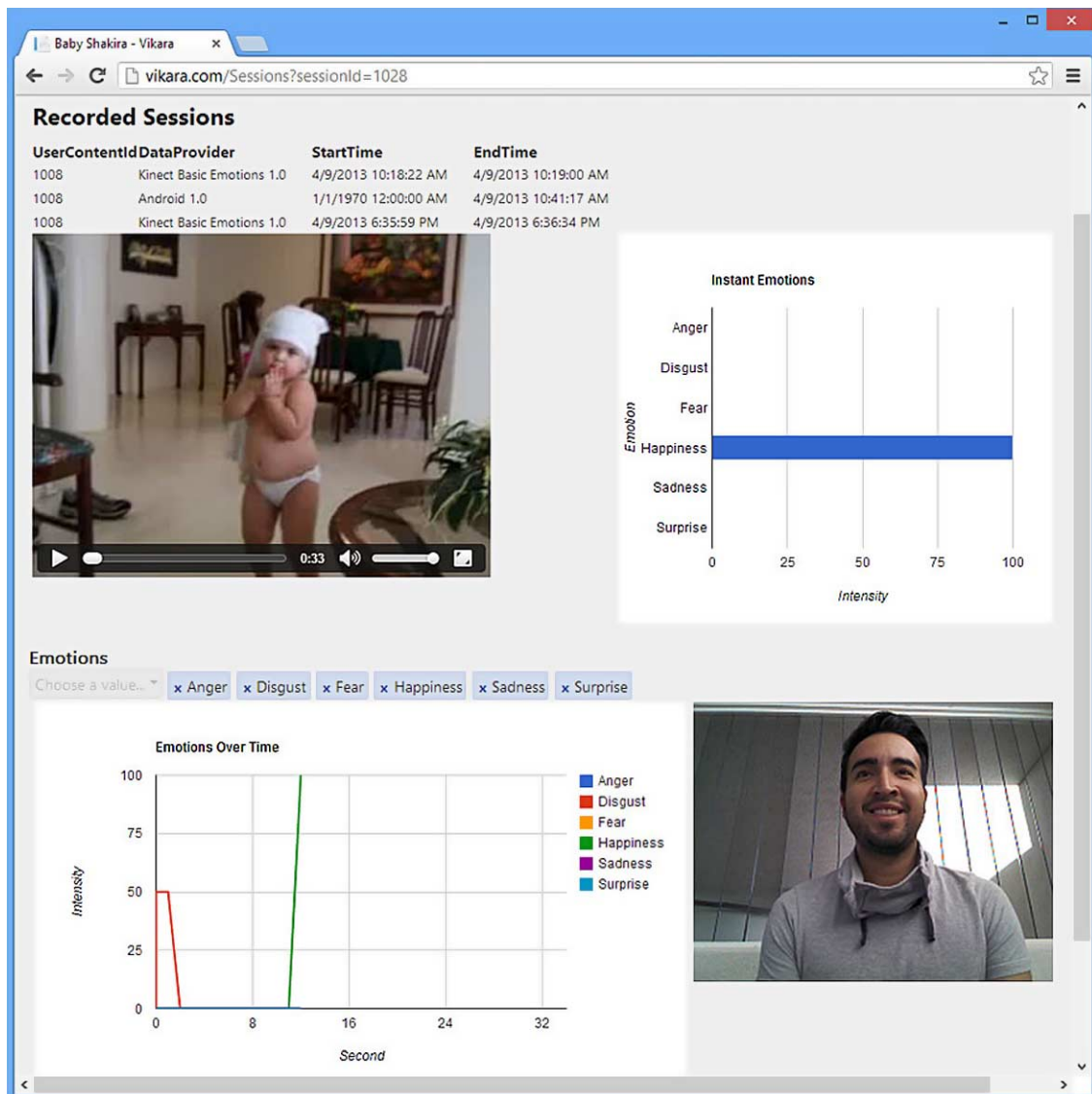


Fig. 5. Emotions detected by the integrated to Web platform framework designed using 6-FACS model.

in practice, when rotation is around of 75° , the sensor ceases to follow a face.

In Fig. 5 one particular session on Web page of a client of previously introduced platform is presented, where a user laughing (bottom right) while watching the video of dancing baby (upper left corner). In platform the visualization client listens to changes reported by the monitoring services and produces a chart that represents graphically the emotions occurring at a given instant. The graph of emotions (top right) indicates the intensity of emotions that were interpreted at the time, when a user was watching. The graph (bottom left) indicates *disgust* emotion reduction in 2 seconds

and increasing *happiness* from 11th second to maximum intensity.

4.2. Performance of emotion interpreter based on 6-FACS-model

The principal goal of tests is to evaluate performance of the designed framework based on 6-FACS model in controlled environments. In conducted tests still images from standard Kanade's and Pantic's databases have been used. Kanade's CMU-Pittsburgh AU-Coded Facial Expression Database consists of 2105 digitized image sequences (640×490 or

Table 5
Confusion matrix of facial expression recognition using 6-FACS model

	Happiness (%)	Anger (%)	Fear (%)	Disgust (%)	Surprise (%)	Sadness (%)	Neutral (%)
Happiness	43.7%	9.5%	0.00	12.7%	22.2%	2.4%	9.5%
Anger	11.1%	42.6%	16.7	3.7%	20.3%	5.6%	0.00
Fear	0.00	0.00	48.6%	0.00	31.9%	2.8%	16.7%
Disgust	13.3%	13.3%	15.5%	20%	15.5%	2.4%	20%
Surprise	0.00	0.00	33.3%	4.2%	62.5%	0.00	0.00
Sadness	0.00	6%	61.6%	0.00	1.4%	19.9%	11.1%
Neutral	0.00	11.1%	36.1%	0.00	0.00	2.8%	50%

Table 6
Definition of metrics for evaluating performance of emotion interpreter

	Metrics and Equations	Interpretation
A	$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$	Percentage of correct interpretations
P	$Precision = \frac{TP}{TP + FP}$	Percentage of correct positive predictions
R	$Recall = \frac{TP}{TP + FN}$	Percentage of correctly identified positive instances
S	$Specificity = \frac{TN}{TN + FP}$	Percentage of correctly identified negative instances

640×480 pixel arrays with 8-bit gray-scale or 24-bit color values) from 182 adult between the ages of 18 and 50 years of varying ethnicity, performing multiple tokens of most primary AUs that have been FACS coded [15].

The Pantic's MMI Facial Expression Database actually includes more than 2900 videos and still images (24-bit true color images of 720×576 pixels) of 75 male and female subjects of varying ethnic background in frontal and in profile view displaying wide range of facial expressions of emotion [26]. Particularly, about 400 images with six basic expressions previously labeled in Kanade's and Pantic's databases have been selected. Images of twenty nine subjects (18 women and 11 men) were chosen from Kanade's collection and images of eighteen subjects (five women and thirteen men) were selected from Pantic's database.

It was decided to use a confusion matrix that is a typical tool in supervised learning to display the results obtained in tests (see Table 5). The rows represent real emotions in tests (how they have been labeled in standard databases) and columns show the percentage of correct recognition of expressions by system.

Presented in Table 6 some performance metrics such as accuracy, precision, recall and specificity have been additionally evaluated. The used variables are: *TP* (true positive) is a case, when a system interprets emotion correctly or *FN* (false negative) incorrectly. A case, in which the emotion was identified as something that

was not, is interpreted as *FP* (false positive) and a case, when real emotion was not interpreted as emotion, is *TN* (true negative).

Figure 6 shows how each emotion has been evaluated by metrics mentioned in Table 6.

As in other systems and as it happens to human, the emotions were not recognized with the same accuracy. During experiments with existing emotion sensing systems connected to the platform it was detected that in FaceReader *disgust* and *fear* are the emotions, which had fewer correct response and in Shore system *disgust* emotion is not included [25,30]. The best recognition as usual, has *surprise*, *fear* and *happiness* due to wide range of changes of corresponding AUs.

In the conducted tests *fear* is the emotion that generally has more frequent incorrect interpretation accuracy of 69% and precision of 21%. The achieved average accuracy of recognition by the system is around 82%. The low accuracy of emotion interpretation is due to small range of changes of AUs that define these emotions as well as limited number of AUs used in the proposed model.

The designed system provides quite acceptable expression recognition similar to reports about relevant emotion sensing systems. For example, the recognition rates of some systems are: 72% in Esau's emotion fuzzy classifier [8], 85% in FaceReader [25], 86% in Pantic's expert system [26], 91.6% in eMotion system [33], 92.4% in SHORE framework [30],

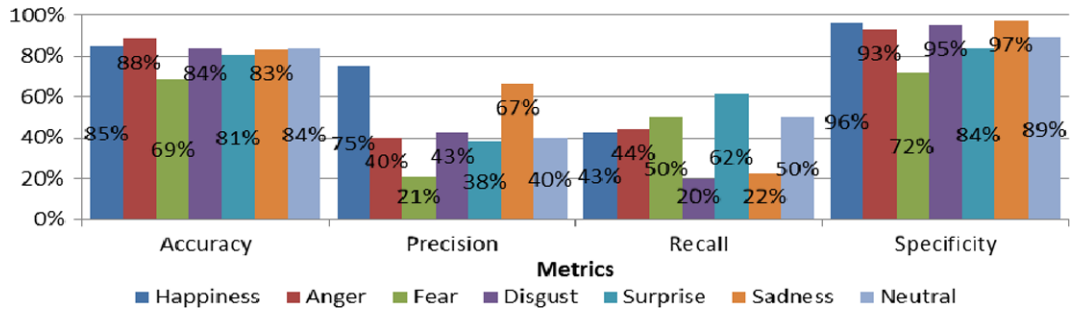


Fig. 6. Performance analysis of framework designed using 6-FACS model.

and 96.9% in Compound expression approach by Martinez [6].

After validation and estimating performance of the proposed system, it was concluded that recognition rate will be similar on other collections of images. In order to analyze the system performance, the designed framework has been integrated to the platform for real-time expression recognition. Forty video records about 30–40 seconds each with resolution of 1280×720 pixels with different affective content have been stored in database of the platform.

These records have been used to cause real emotional responses of ten persons, which expressions were evaluated by system displaying every second the instant detected expression (only ten frames per second are processed) and plotting the emotion variation over time as it is shown in Fig. 5. Although the used records have not been previously labeled, the recognition rate in real time was similar to results presented in Fig. 6. Basically, it is because the processed frames of video records have higher resolution than images in Kanade's and Pantic's databases.

4.3. Distance model for face description

In order to improve the introduced in previous section emotion interpreter, a Distance model based on analysis of fifteen distances ($Distance(fdp1, fdp2)$) between nineteen FDPs has been proposed. This model describes all the necessary facial actions defining either basic or non-prototypical emotions.

Figure 7 shows the selected FDPs with corresponding number of associated FAPs that better describe facial expressions. The variable $Distance(fdp1, fdp2)$ quantifies facial changes in terms of units defined by MPEG-4 standard. Table 7 shows the fifteen instances (Dd) chosen for our model, the geometrical definitions of these distances (Dif of FDPs), the measuring units,

Table 7

Description of instances in the proposed Distance model

Dd	Dif of FDPs	Units	FDP	Action Description
D1	d{3.11, 4.1}	ENS	31	raise l i eyebrow
D2	d{3.8, 4.2}	ENS	32	raise r i eyebrow
D3	d{3.7, 4.3}	ENS	33	raise l m eyebrow
D4	d{3.12, 4.4}	ENS	34	raise r m eyebrow
D5	d{3.7, 4.5}	ENS	35	raise l o eyebrow
D6	d{3.12, 4.6}	ENS	36	raise r o eyebrow
D7	d{4.1, 4.2}	ES	–	squeeze l/r eyebrow
D8	d{3.3, 3.1}	IRISD	21–19	close t/b l eyelid
D9	d{3.4, 3.2}	IRISD	22–20	close t/b r eyelid
D10	d{8.3, 8.4}	MW	53–54	stretch l/r cornerlip
D11	d{3.11, 8.3}	ENS	59	raise l cornerlip o
D12	d{3.8, 8.4}	ENS	60	raise r cornerlip o
D13	d{9.15, 8.1}	MNS	–	lower t midlip
D14	d{9.15, 8.2}	MNS	–	raise b midlip
D15	d{8.1, 8.2}	MNS	51–52	raise b/t midlip

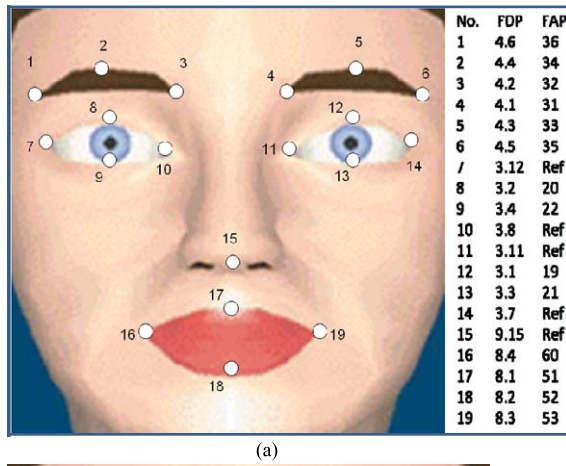
the relation between FAPs and the corresponding actions, which they describe.

Some reports suggest a geometrical model of face, which includes not only distances but also angles between the lines connecting standard FDPs. However, this approach does not contribute significant precision and makes the processing too complex [8,10,36].

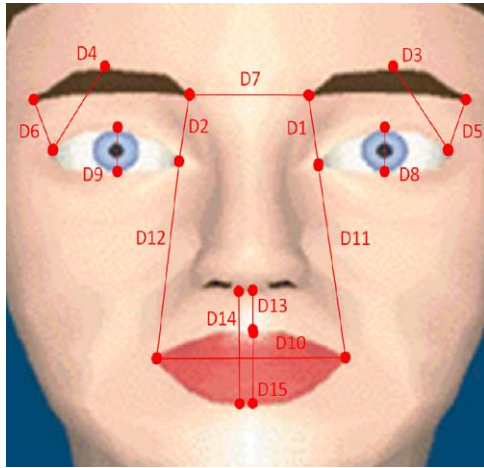
5. Fuzzy emotion classifier for designed framework based on Distance model

5.1. Fuzzyfication-defuzzyfication processes

For recognition of facial expressions and interpretation of basic and non-prototypical emotions the rule-based fuzzy classifier has been used, because the equilibrium between simplicity and precision is required. Fuzzy logic based systems provide a type of reasoning, where logical statements are not only true or false but



(a)



(b)

Fig. 7. Nineteen FDPs used for recognizing facial expressions and definition of fifteen distance instances.

can also lie in range from almost certain to very unlikely. A software based on fuzzy logic allows computers to mimic human reasoning more precisely so that, decisions can be taken with incomplete or uncertain data.

The fuzzy approach and its combination with other approaches for pattern recognition and interpretation have widely used [8,35,42]. In the area of facial expression recognition the application of fuzzy reasoning remains marginal despite that some researchers have successfully used classifying systems, which emulate how the humans identify prototypical expression [5,8,18,19]. Some well-known systems use other types of classifiers based on the multiple adaptive neuro-fuzzy inference approach, support vector machine, hidden Markov model, evolutionary algorithm, genetic algorithm, etc. Even though these approaches detect and recognize facial features, the precision of

recognition sometimes is low and only basic emotions are interpreted without quantitative analysis of intensity of facial expression [1,16,20,39,44].

A fundamental process of fuzzy reasoning is fuzzyfication of input variables and definition of the corresponding membership function used for indexing facial deformations. The input variables of classifier are FAPs, which represent variation of distances between fiducial points selected in the proposed Distance model. The membership function is used to associate a grade to each linguistic term. The membership grade for all the members defines a fuzzy set. Given a collection of objects U , a fuzzy set A in U is defined as a set of ordered pairs presented in Eq. (1)

$$A \equiv \{(x, \mu_A(x)) | x \in U\} \quad (1)$$

where $\mu_A(x)$ is called the membership function for the set of all objects x in U . According to particular requirements we can use different membership functions. The most used are piecewise linear function, Gaussian distribution function, sigmoid curve, and quadratic or cubic polynomial curve.

From a mathematical point of view, a fuzzy rule-based system *FRBS* presented in Eq. (2) may be introduced as a set of membership functions μ_{ab} , a fuzzy rule base R , the t -norm for fuzzy aggregation T (i.e. operations within one rule), the s -norm for fuzzy composition S (i.e. operations among rules) and the defuzzification method *DEF*:

$$FRBS = (\mu_{ab}, R, T, S, DEF). \quad (2)$$

Perhaps, the most popular defuzzification method is the centroid calculation, which returns the center of area under the curve. Other methods are the bisector, middle of maximum (the average of the maximum value of the output set), the largest of maximum and the smallest of maximum. The proposed fuzzyfication-defuzzification processes provided by inference engine are shown in Fig. 8.

5.2. Knowledge base for the proposed fuzzy classifier

The proposed fuzzy classifier for emotion interpretation consists of two principal modules. The first one is a knowledge base (KDB) that is used for modeling and indexing facial deformations by FAPs and AUs developed according to well-known standards [7,13,35]. The second module is used for recognizing facial ex-

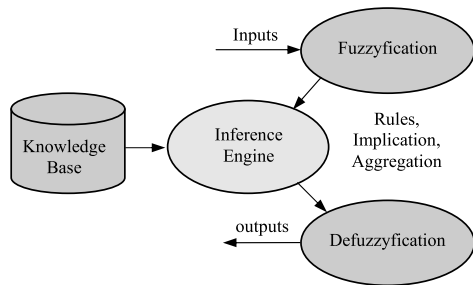


Fig. 8. Fuzzyfication-defuzzyfication processes for facial expression recognition.

pressions by defuzzyfication providing interpretation of emotion and their intensity.

The quantification of emotion intensity is handled by measuring the range of geometrical displacement of selected FDPs. To reduce relative subjectivity and lack of psychological meaning of emotional intensity levels, the statistical analysis of facial actions in preprocessed Cohn-Kanade's and Pantic's image databases has been carried out [15,26].

Additionally, for evaluating performance of facial expression tool the data base with more than 5000 manually indexed images with definition of 94 fiducial points in each one also may be used [6]. Presented in Fig. 9 the proposed KDB allows measuring facial deformations in terms of distances between FDPs modeled by FAPs and AUs represented by rule-based descriptors used then in the process of fuzzyfication.

For FDPs MPEG-4 standard provides the automatic normalization of measured facial deformations making them invariant to scale of input images. Figure 9 shows the structure of KDB that is a basis of our fuzzy reasoning system. For KDB four classes based on AUs, FAPs, FDPs, and Distance have been created. The *Emotion_Model* class provides creation of the rule-based models for emotion indexing using classes of the *Face_Model*. The *Face_Model* class defines different approaches for representation of face features. Particularly, the instances of *Face_Model* class contain the basic facial actions (AUs, FAPs) that include action number, its name, description, direction of motion, involved facial muscles and part of a face, where action is occurred [35].

The proposed approach is able to detect and measure any type of facial expression however, it has been tested using as six basic expressions (*happiness, sadness, disgust, surprise, anger, and fear*) as some combinations of them interpreting in this way non-prototypical expressions. Some rules that define the re-

lationship between measured facial deformations and their mathematical description by the corresponding AUs and FAPs have been exploited. The KDB has been implemented using ontology editor Protégé that provides extensible, flexible, and plug-and-play environment, which allows fast prototyping and application development [28,36]. The advantage of the proposed KDB is that, the classes and instances with attributes represent knowledge about facial expressions. The parameters of any model may be automatically converted to parameters of each other. For example, if input feature vector corresponding to particular emotion has been created on base of non-standard *Distance(fdp1, fdp2)* model, these parameters may be immediately represented by standard AUs or FAPs attributes and vice versa [35].

5.3. Emotion interpretation by the framework designed using Distance model

Recognition and quantization of facial expressions is provided in the proposed classifier in two steps as it is shown in Fig. 10. The first one is a fuzzyfication process, when Fuzzy Inference Engine (FIS) identifies and measures intensity of AUs using as inputs the *Distance* variables defined in the proposed model. The first stage classifier has two outputs: numerical value in range from 0% to 100% representing intensity of action and qualitative linguistic description of AUs in terms of their *low, medium or high* (L-M-H) intensity. In order to associate selected *Distance(fdp1, fdp2)* with standard action units of FACS, twelve AUs have been chosen for classifier. For example, in Table 8 for recognizing the action unit AU1, which indicates that the inner portion of eyebrows are raised, the *distances D1* and *D2* are measured in images with neutral face and with one of particular expression. If these two variables were increased approximately in the same proportion, then the action expressed by face is AU1.

Thus, the first stage of classification provides detection of particular AUs by means of Distance variables according to the rules defined in Table 8. For each AU corresponding fuzzy inference system FIS is designed. A set of 12 individual FISs constitutes the first classifier. The advantages of this kind of system are simple calibration and easy modification processes.

The second step uses numerical values obtained in the first stage as inputs for defuzzyfication process that provides estimation of emotions and their intensity in three levels: low, medium and high. Table 9 presents how the selected in our model AUs identify six emo-

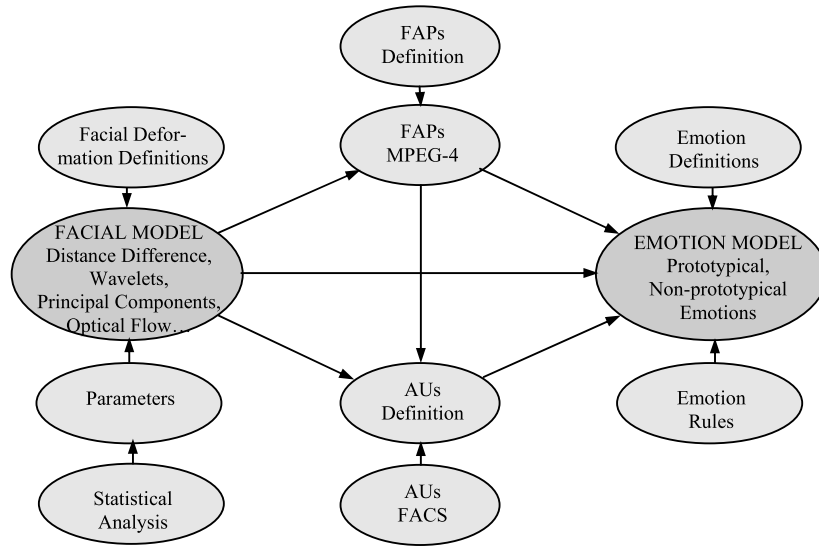


Fig. 9. Structure of KDB based on AUs, FAPs, FDPs, and Distance differences.

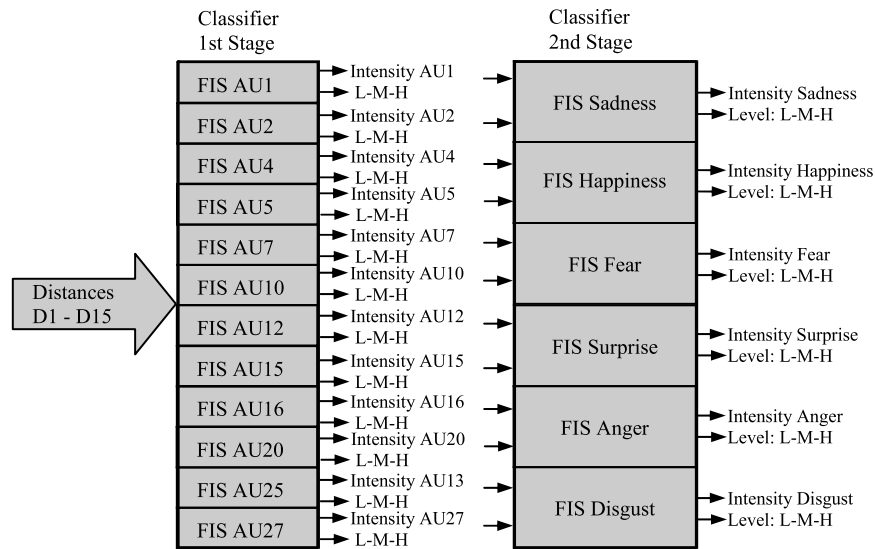


Fig. 10. Emotion interpretation steps using fuzzyfication and defuzzyfication in the proposed classifier.

tions according to recognition rules established in previously developed KDB. These variables and rules were obtained from psychological and physiological studies of human behavior [7,11].

The detailed description of fuzzy classifier is provided, because it explains how basic and then non-prototypical emotions may be detected and interpreted using three-level intensity. The blocks diagram of the Fuzzy Inference System (FIS) for measuring and recognizing AUs (particularly, for AU12) is presented in Fig. 11. According to Table 8, AU12 is defined by three

distances D10, D11 and D12, which are used as input of corresponding FIS AU12. Additionally, variable called *asymmetry* is applied to input specifying difference between D11 and D12. This variable is being used to detect an asymmetry in facial expression. For example, if *asymmetry* variable becomes true, then we do not have a smile but a grimace.

FIS AU12 has internally two fuzzy systems named *Detect AU12* (for evaluation of AU12 intensity in linguistic terms as *low* L, *medium* M or *high* H) and *Detect AU12int* (for quantification of intensity level

Table 8

Description of used distances and rules for definition of chosen AUs

Code	Description	Distances	Recognition Rules
AU1	Inner Brow Raiser	D1, D2	Both increase in same proportion
AU2	Outer Brow Raiser	D5, D6	Both increase in same proportion
AU4	Brow Lowerer	D3, D4, D7	D3&D4 increase, D7 decrease
AU5	Upper Lid Raiser	D8, D9	Both increase in same proportion
AU7	Lid Tightener	D8, D9	Both decrease in same proportion
AU10	Upper Lip Raiser	D13	D13 decrease
AU12	Lip Corner Puller	D10, D11, D12	D10 increase D11&D12 decrease
AU15	Lip Corner Depressor	D11, D12	Both increase in same proportion
AU16	Lower Lip Depressor	D14	D14 increase
AU20	Lip stretcher	D10, D11, D12	D10, D11&D12 increase
AU25	Lips part	D15	D15 increase
AU27	Mouth Stretch	D10, D15	D10 decrease, D15 increase

Table 9

Established rules and used AUs for Distance model-based framework

Emotion	Selected AUs	Recognition Rules
Sadness	AU1, AU4, AU15	Increasing 3 actions increase expression intensity
Happiness	AU12, AU7	Presence of AU12&AU7 but not AU7 (blinking). Increasing values increase expression intensity
Fear	AU1, AU2, AU4, AU5, AU20, AU27	Presence of the 6 actions but not AU7 (blinking). Increasing values increase expression intensity
Surprise	AU1, AU2 AU5, AU27	Presence of the 4th action but not AU5 (blinking). Increasing values increase expression intensity
Anger	AU4, AU7	Presence of AU4&AU7 but not AU7 (blinking). Increasing values increase expression intensity
Disgust	A10, AU25, AU27	The infraorbital triangle and center of the upper lid are pulled upward

of AU12 in the (0–100)% range). Multiplexor *Mux* is used to send these values to *Detectors* one by one for computing how significant the changes of analyzed distance are with respect to neutral face. In order to evaluate a distance variation in linguistic terms, the distance range of each variable is divided into three sections L-M-H, where the center and width of medium section are the mean and deviation of particular distance. This process has been carried out by statistical analysis of images in used databases [15,26].

For example, the range that was chosen for the variable distance D10 is defined analyzing horizontal displacement of the left inner lip corner in more than 500 images used for tests. So, the range 0–600 MW (Mouth Width) has been selected for D10. Figure 12 shows three plots of Gaussian membership function for D10 with specification of low, medium and high sections of intensity. For example, in Fig. 11 distance D10 = 348 MW, this corresponds to linguistic expression “the distance D10 is high”, because in Fig. 12 on the horizon-

tal axis this value belongs to high section (black curve “high”).

For detection of AUs in the fuzzyfication process a Gaussian function has been chosen, because it provides smooth transition between linguistic ranges of variable intensity. The trapezoidal function has been chosen as auxiliary in a case, when the higher processing speed is required. The description of tested membership functions (mf) is presented in Table 10. The same evaluation process is applied to D11 and D12 finally, generating the linguistic value of AU12 according to the rules presented in Fig. 13.

Detect AU12int FIS provides generation of the quantitative value of AU.

Additionally, *Gain* and *K* parameters (see Fig. 11) are used for adjustment and calibration of *Detect AU12int* output to range from 0% to 100%, because D10, D11 and D12 have different measurement units and ranges of variation. In more details the fuzzy logical reasoning process provided by *Detect AU12* is

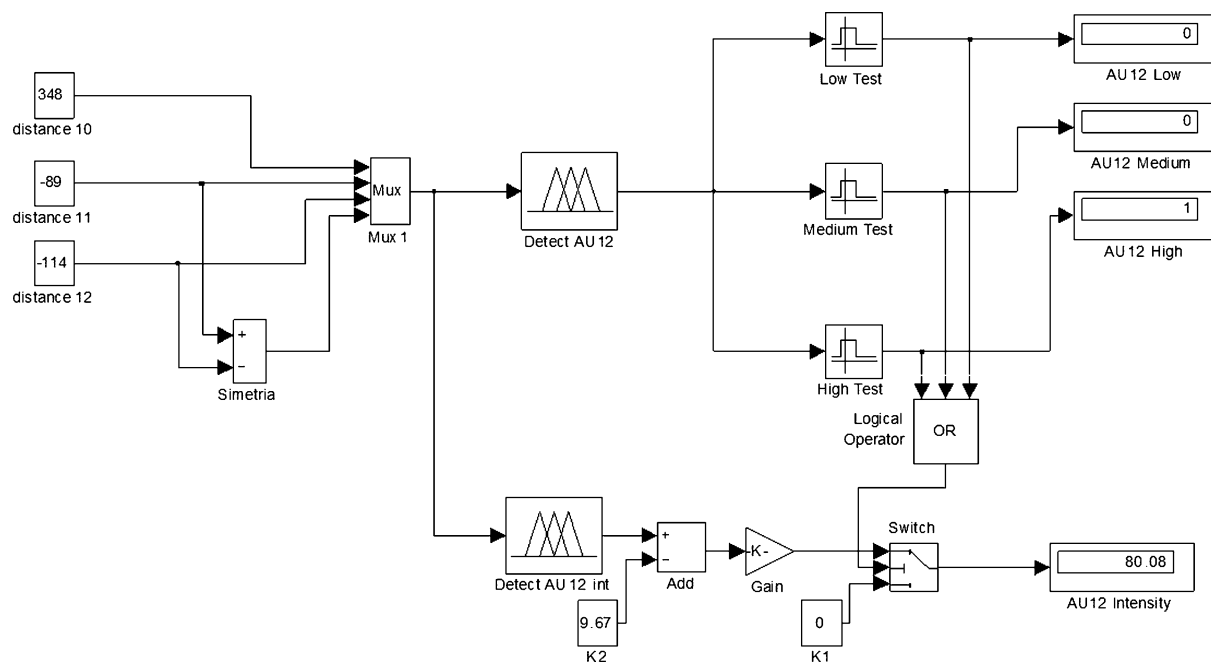


Fig. 11. Block diagram of Fuzzy Inference System to detect AUs (particular case of AU12 recognition).

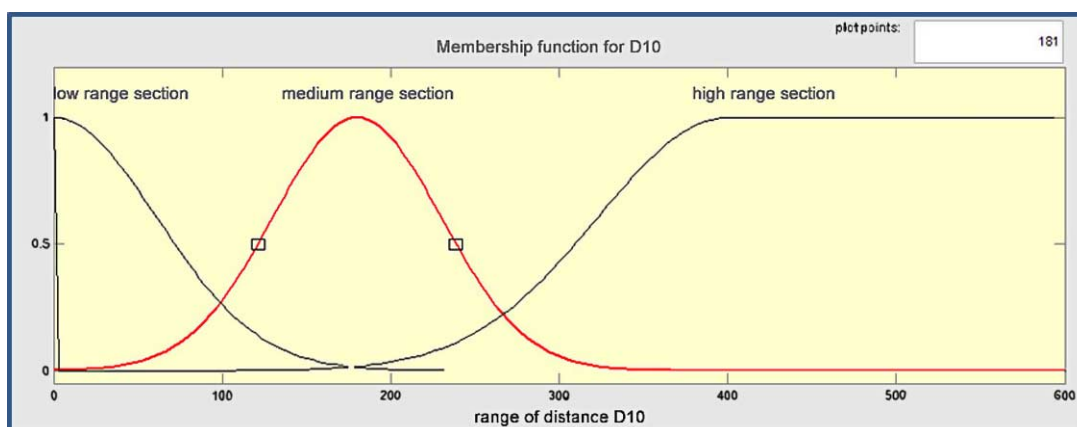


Fig. 12. Plots of Gaussian membership functions for partition of D10 range into low, medium and high sections.

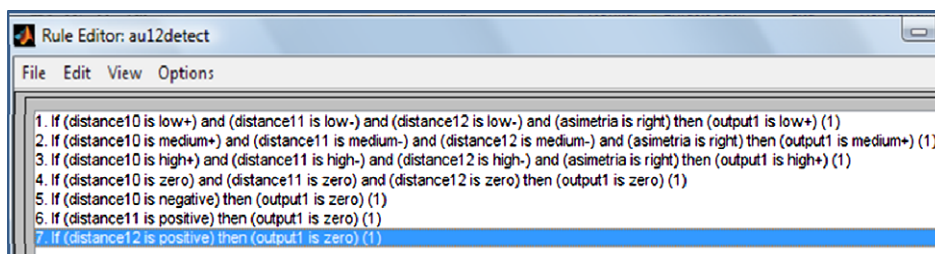


Fig. 13. Fuzzy reasoning in AU12detect” FIS.

Table 10
Membership functions used in fuzzyfication process

mf	Equation	Comments
Symmetric Gaussian mf	$f(x, \sigma, c) = e^{-\frac{(x-c)^2}{2\sigma^2}}$	c defines the position of peak and σ controls the width of the bell shaped Gaussian curve
Combination of Two Gaussian mfs	$f(x, \sigma_{1,2}, c_{1,2}) = e^{-\frac{(x-c)^2}{2\sigma^2}}$	Combination of two Gaussian functions with two sets of parameters σ and c
Trapezoidal Shaped mf	$f(x, a, b, c, d) = \begin{cases} 0, & x \leq a \\ \frac{x-a}{b-a}, & a \leq x \leq b \\ \frac{d-x}{d-c}, & c \leq x \leq d \\ 0, & d \leq x \end{cases}$	Trapezoidal curve is a function of x and depends on four scalar parameters a, b, c and d . The parameters a and d locate the “feet” of the trapezoid and the parameter b and c define the “shoulders”

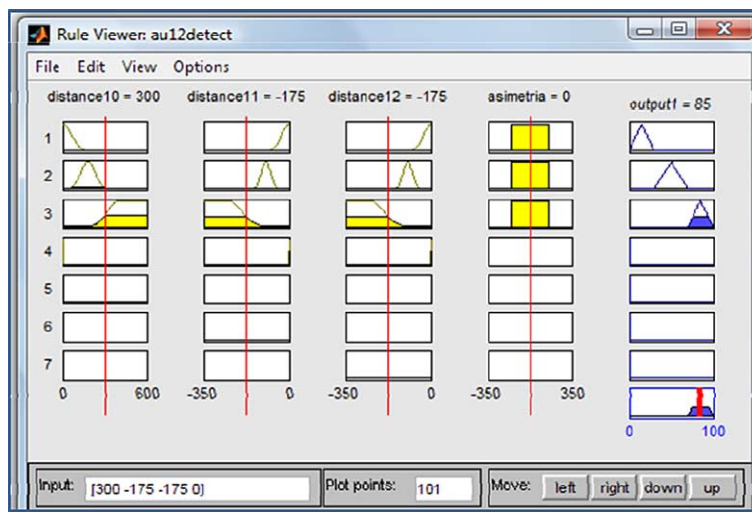


Fig. 14. Fuzzy reasoning for detection of low, medium or high intensity of AU12.

shown in Fig. 14. For instance, input parameters D10, D11, D12 and asymmetry in Fig. 11 have particular values 300, -175 , -175 , 0, respectively. They lie mostly in the sections of high intensity. Therefore, the detector of high intensity in line 3 of Fig. 14 produces the output equal 85%. That means “the facial action AU12 is high”. The same reasoning is used by *Detect AU12int* FIS (see Fig. 15). If input values are, for instance, 250, -121 , -119 , 2, the numerical output value of AU12 intensity is 53.7%. The input to the defuzzification process is a fuzzy set obtained from the previously described step. The outputs of the process are numerical values and a linguistic term, which define intensity of AU. Defuzzification method used in FIS *Detect AU12* is “mom” that means, middle of maximum (computing average of the maximum values of the output set), and the method for FIS *Detect AU12int* consists in centroid calculation, which returns the center of area under the curve. With this

last step of defuzzification the classification process based on fuzzy inference for the facial action AU12 is finished.

The same procedure must be used for other action units. The second stage of our facial expression classifier is interpretation of emotions based on AUs detected in the first phase of classification. This second stage is provided by six FISs, one for each emotion. The recognizing rules for FISs are shown in Fig. 16.

The same methodology presented in Figs 11, 14 and 15 for recognition of AUs is applied for interpretation of emotions.

For instance, in Fig. 17 the final step of computing intensity level of *happiness* emotion (represented by AU7 and AU12) is shown, where in this particular case the inputs to the *Happiness*FIS are AU12 = 64 and AU7 = 14. These values generate an output value equal to 50 corresponding to the linguistic value “medium”.

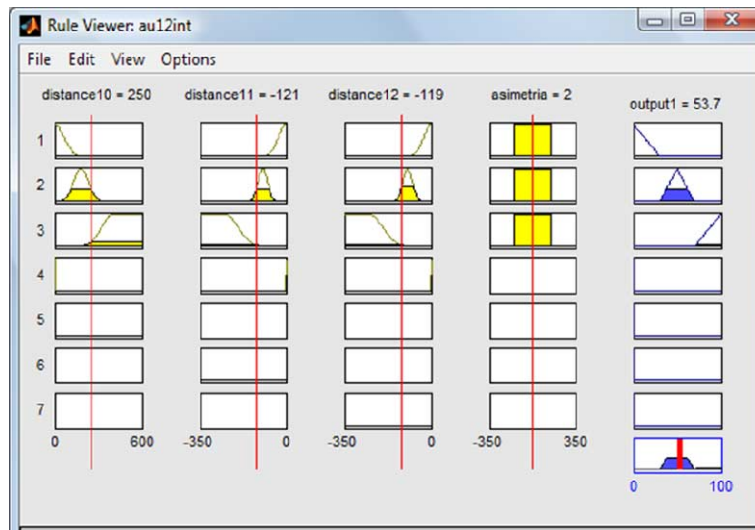


Fig. 15. Fuzzy reasoning for detection of numerical value of AU12 intensity.

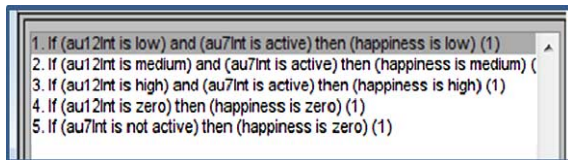


Fig. 16. Applied rules for interpreting emotion of *happiness* as combination of AU7 and AU12.

The used two-stage classifier may be substituted by others frequently used for expression analysis [6,19,31,33,34]. However, fuzzy classifier has some advantages:

- It has an extendible platform that may be easily modified for more functionality for instance, to provide also the detection of non-prototypical emotions;
- Fuzzy logic is more intuitive approach than others without far reaching complexity and required training;
- FIS can be mixed with neuronal networks in order to use adaptive techniques like adaptive neuro-fuzzy inference systems, etc.

5.4. Evaluation of Distance model-based framework

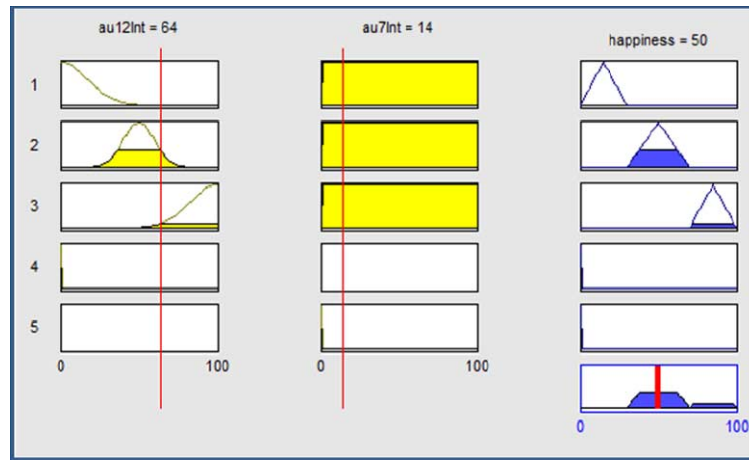
The evaluation of system performance and efficiency of fuzzy classifier have been done in tests with standard Kanade's and Pantic's image databases described in Section 4.2. The principal goal of the first set of tests is to evaluate performance of the designed framework based on Distance model in controlled environments. So, about 400 images with par-

ticular previously labeled expressions from Kanade's and Pantic's databases have been selected. Twenty nine subjects (18 women and 11 men) were chosen from Kanade's collection. From Pantic's database the images of eighteen subjects (five women and thirteen men) demonstrating one of six basic facial expressions of variable intensity were selected.

Tables 11 and 12 show the confusion matrices obtained for six basic expressions in case of medium (low intensity is quite similar to medium) and high intensity. In Figs 18a and 19a two example of images with facial expressions of *happiness* and *sadness* from Kanade's database are presented. The recognition degrees reported by the proposed system for corresponding facial expressions are shown in Figs 18b and 19b respectively.

After evaluation of obtained results in the first set of tests, it was concluded that emotions expressed in test images are correctly identified by the classifier. Expressions with high intensity are recognized more precisely than with low and medium intensities.

For example, the average recognition rate for expressions with high intensity lies in range from 72.2% (for *disgust*) to 97.6% (for *happiness*). The average recognition rate for expressions with low and medium is between 60.1% (for *disgust*) and 96% (for *happiness* and *surprise*). The average percentage of correct recognition for low, medium and high intensities of each corresponding expression is rated as it follows: *happiness* – 96.5%, *anger* – 93.6%, *fear* – 88.7%, *disgust* – 64.1%, *surprise* – 95.5% and *sadness* – 87.3%.

Fig. 17. Fuzzy reasoning for interpreting intensity of *happiness*.Table 11
Confusion matrix of expressions of medium intensity

Emotion	Happiness	Anger	Fear	Disgust	Surprise	Sadness
Happiness	96%	1.9%	0.9%	1.0%	0.2%	0.0
Anger	0.10%	92%	3.4%	2.2%	2.3%	0.0
Fear	0.0	0.0	88.2%	1.3%	5.8%	4.7%
Disgust	6.2%	12.2%	16.0%	60.1%	4.4%	1.1%
Surprise	0.0	0.0	3.7%	0.3%	96%	0.0
Sadness	0.0	0.0	9.5%	0.0	9.5%	81%

Table 12
Confusion matrix of expressions of high intensity

Emotion	Happiness	Anger	Fear	Disgust	Surprise	Sadness
Happiness	97.6%	0.0	1.2%	1.2%	0.0	0.0
Anger	0.0	96.7%	1.6%	0.8%	0.9%	0.0
Fear	0.0	0.0	89.6%	0.0	5.7%	4.7%
Disgust	8.5%	6.5%	9.8%	72.2%	3.0%	0.0
Surprise	1.0%	0.0	2.4%	2.0%	94.4%	0.2%
Sadness	0.0	0.0	8%	0.0	8%	84%

The *happiness* and *surprise* expressions are easier to recognize and the most difficult is *disgust*. *Disgust* and sometimes *sadness* have low recognition rate, because it is easy to confuse them in some instances with other expressions. To solve this problem is possible by incorporating to the proposed models additional facial actions that are more representative for given expression for example, to include processing wrinkles.

Finally, the average recognition rate of the system based on Distance model is 87.6% that is better than of the system based on 6-FACS model. Computing average recognition rate without *disgust* as it has been done in some reports [26,30,32], it achieves 92.3%.

The designed systems compared with relevant emotion sensing applications presented in Section 4.2 have similar acceptable precision providing additionally advantageous measurement of intensity of basic emotions.

The second test consists in comparison of expression intensity recognized by classifier with the intensity values labeled by evaluation committee in set of images selected from Kanade's and Pantic's databases. The evaluation committee was integrated by 32 persons. 19 are women, 22 of them are young university students no more than 23 years old and the rest were older people with ages between 40 and 50 years.

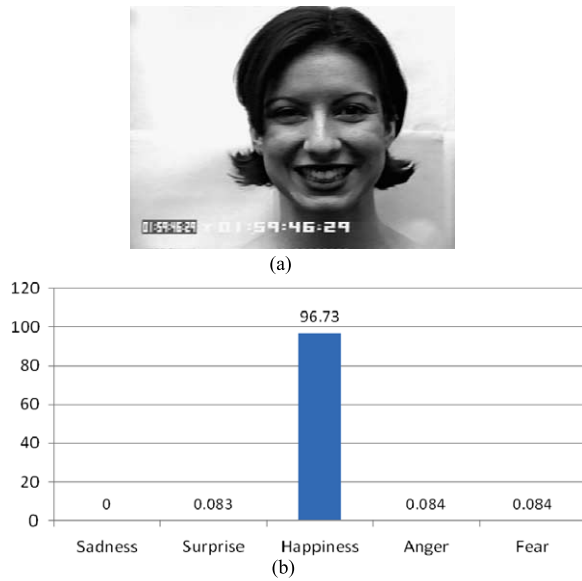


Fig. 18. a) Facial expressions of *happiness* and b) its intensity reported by framework designed using Distance model.

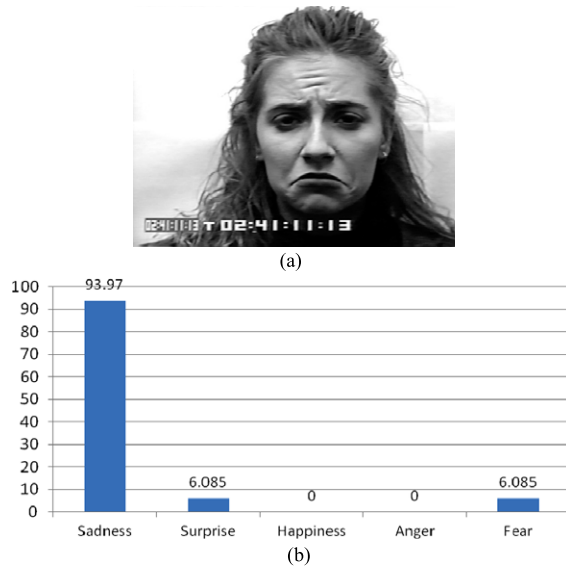


Fig. 19. a) Facial expressions of *sadness* and b) its intensity degree reported by framework designed using Distance model.

The evaluation committee was trained to recognize action units using FACS and consequently classify more precisely facial expressions. In these tests it was taken into account that recognition of emotions is a deeply human activity. Without the corresponding training the committee members could discern no differences between expressions of *fear*, *anger* and *surprise* of low intensity. The evaluation commit-

Table 13

Comparative results of recognizing expression intensity by classifier and labeled by evaluation committee for *surprise*

Image#	Classifier: Intensity Recognition Rate (%)	Committee Appreciation	Status
1	6.1	low	OK
2	47.43	medium	OK
3	62.18	medium	OK
4	46.07	medium	OK
5	49.34	medium	OK
6	50.31	medium	OK
7	50.71	medium	OK
8	94.11	medium	FAIL
9	49.8	medium	OK
10	65.74	medium	OK
11	49.77	low	FAIL
12	51.0	high	FAIL
13	48.3	medium	OK
14	47.92	medium	FAIL
15	49.74	low	FAIL
16	93.94	low	OK
17	53.49	high	FAIL
18	50.3	high	OK
19	53.74	medium	OK
20	57.4	medium	OK

tee describes the intensity of emotions by terms: *low*, *medium* and *high*, which ranges are (0–30)%, (30–70)% and (70–100)%, respectively. In this set of tests we have evaluated images with six basic expressions.

Twenty images for each expression but with different randomly selected intensities have been tested by classifier and by each member of committee. The detected by classifier intensity, the average of appreciation of intensity by evaluation committee and the inter-rater concordance status are presented in Tables 13 and 14 for the best (*surprise*) and for the worst (*sadness*) cases, respectively. After comparison of expression intensity levels measured by classifier and reported by evaluation committee (average of appreciation of 32 persons participated in usability tests), the inter-rater concordance for the best case of *surprise* expression is about 90%.

The worst case was for *sadness* giving about 70% of inter-rater concordance (in twenty images there are six inconsistencies in agreement between classifier and committee). As it has been mentioned in the first set of tests, for *sadness* additional and more representative action units are required.

The results of the second set of tests for other expressions show the following inter-rater concordance:

Table 14

Comparative results of recognizing expression intensity by classifier and labeled by evaluation committee for *sadness*

Image#	Classifier: Intensity Recognition Rate (%)	Committee Appreciation	Status
1	6.81	low	OK
2	50.33	medium	OK
3	51.05	low	FAIL
4	48.59	medium	OK
5	49.85	medium	OK
6	94.08	high	OK
7	69.97	high	OK
8	51.46	medium	OK
9	93.93	high	OK
10	94.92	high	OK
11	51.03	medium	OK
12	47.7	medium	OK
13	6.68	low	OK
14	50.2	medium	OK
15	17.95	medium	FAIL
16	95.12	high	OK
17	94.05	high	OK
18	49.29	medium	OK
19	93.21	high	OK
20	93.41	high	OK

happiness – 85%, anger – 80%, fear – 85%, disgust – 75%. The average of inter-rater concordance for six basic expressions results about 81%.

In any case, the evaluation by members of committee are quite subjective and depends on various factors such as high person-specific variability in perception of expressions, influences on decision making by personal feelings, tastes or opinions, spontaneous interactions with actor and referee, presence of simulation of particular emotions by actor, training of referee to detect lying and falsification of emotional state, selection of rating ranges for low, medium and high intensity used by committee, etc. The obtained results suggest more specific researches in this still open problem of facial expression recognition.

6. Recognition of non-prototypical emotions

It is a fact, that emotions and social interaction have significant impacts in our lives. Therefore, evaluation and regulation of affective state, prediction and control of emotional crisis get more attentions in recent researches. In order to take effective actions during a stress, it is important to recognize wide range of spe-

cific expressions, which usually are non-prototypical. For example, during critical situations the most people portray *grim* and *stoic* expressions or demonstrate *anxiety*, *nervousness*, *disapproval*, etc., which are not recognized using traditional Ekman's model of six basic emotions [24,41]. Thus, the emotion sensing applications should be able to detect complex state of feeling quantifying non-prototypical facial expressions by combination of known basic emotions according to existing psychological models.

There are some contradictions between two widely used Ekman's model of six basic emotions and Plutchik's model of eight primary emotions. Ekman's model has been developed principally for expression recognition by computing systems, because they are consistently distinguishable from other expressions described by action units. Plutchik's model has been created to provide the basis for an explanation of psychological mechanisms by emotional responses and not precisely for automatic recognition tool. Used in Plutchik's model eight core bipolar emotions such as *joy – sadness*, *anger – fear*; *trust – disgust* and *surprise – anticipation* can be expressed at different intensities and can be mixed to form specific unusual emotions.

In order to describe relationship between non-prototypical emotions from psychological Plutchik's model and Ekman's six basic emotions expressed by AUs, some considerations have been taking into account. Only six basic emotions *happiness (joy)*, *sadness*, *disgust*, *surprise*, *anger* and *fear* are used. *Trust* and *anticipation* from Plutchik's model are not used. The particular recognition rules defined according to Plutchik's model have been implemented in classifier. Table 15 shows how non-prototypical emotions may be formed according to three levels of intensity (low, medium and high with ranges (0–30)%, (30–70)% and (70–100)%, respectively) of basic emotions computed by fuzzy inference systems described in Section 5.3. Particularly, the relationship between *anger – joy*, *anger – disgust* and *sadness – surprise* with corresponding non-prototypical emotions taken from Plutchik's model are presented. Other combinations *joy – fear*, *disgust – sadness*, *surprise – fear* form corresponding non-prototypical emotions in the similar way.

The similar methodology used in the second test of Section 5.4 has been applied in this test to detect non-prototypical emotions. Non-prototypical emotions demonstrated by the same 29 subjects in randomly selected images from databases should be rec-

Table 15

Relationship between *anger – joy*, *anger – disgust* and *sadness – surprise* with low, medium, high intensities and corresponding non-prototypical compound emotions

Basic Emotion 1	Basic Emotion 2	Compound Emotion
<i>anger</i>	<i>joy</i>	
equal intensity	equal intensity	<i>anticipation</i>
low (medium)	medium (high)	<i>optimism</i>
medium (high)	low (medium)	<i>aggressiveness</i>
low	high	<i>ecstasy</i>
high	low	<i>vigilance</i>
<i>anger</i>	<i>disgust</i>	
equal intensity	equal intensity	<i>contempt</i>
low (medium)	medium (high)	<i>boredom</i>
medium (high)	low (medium)	<i>annoyance</i>
low	high	<i>loathing</i>
high	low	<i>rage</i>
<i>sadness</i>	<i>surprise</i>	
equal intensity	equal intensity	<i>disapproval</i>
low (medium)	medium (high)	<i>distraction</i>
medium (high)	low (medium)	<i>pensiveness</i>
low	high	<i>amazement</i>
high	low	<i>grief</i>

ognized by evaluation committee integrated by the same 32 persons.

For this test we selected a set of 300 images from Kanade's and Pantic's databases with previously labeled non-prototypical expressions from Plutchik's model presented in Fig. 1 (15 images for each of 20 chosen expressions).

The detected by classifier non-prototypical expressions have been compared with averaged results obtained by evaluation committee. As it was expected, the inter-rater concordance in correct recognition was low lying in the range of 50–65%. This low level of recognition is because of limited number of AUs used for representation of complex emotions, simplicity of Plutchik's model and subjectivity of facial expression perception by each person. In case of non-prototypical emotion interpretation the subjectivity plays important role, because complex emotions strongly depend on race, sex, age, culture, state of mood and other factors.

Some tests for recognition of non-prototypical expressions by system based on Distance model have been conducted using developed platform. In the similar way as it has been done with the first designed system in Section 4.2, the integrated to platform Distance-based framework was used for interpretation of emotions of ten persons, which observed forty video records with different affective content that cause real

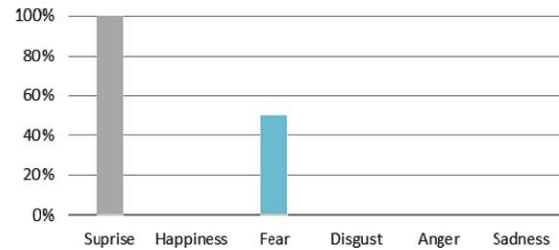


Fig. 20. Reported by system non-prototypical expression of *awe* composed by *surprise* and *fear*.

emotional response. For instance, Fig. 20 shows how the proposed emotion interpreter detects complex expression with presence of *surprise* and *fear*. This emotion according to Plutchik's model presented in Fig. 1 is interpreted as *awe*.

In Fig. 21 another example of detecting compound emotion of *contempt* according to Plutchik's model is presented. However, there also exists relationship between non-adjacent emotions in Plutchik's model. The following combinations of the basic non-contiguous emotions have been also implemented in classifier such as *joy – disgust*, *anger – sadness*, *disgust – surprise*, *sadness – fear*, *surprise – joy*, *fear – anger*, and *surprise – anger*. If classifier detects the presence for example, *anger* and *sadness* the corresponding compound expression is *disgust* although *disgust* as basic emotion has not been detected as such! Similarly, the compound emotion formed by *surprise* and *anger* results another basic emotion of *sadness* although *sadness* has not been detected by classifier. This unexpected surprising finding that has been detected in experiments with Plutchik's model conducts us to the following conclusion: such basic as non-prototypical emotions may be composed by other basic emotions. In Figs 22 and 23 the recognized basic expressions as combination of other basic ones are presented.

It is an interesting fact that facial action units, which define for example, *disgust* in Fig. 22, are not detected by classifier. However, the action units that represent *surprise* and *anger* are recognized. Anyway, according

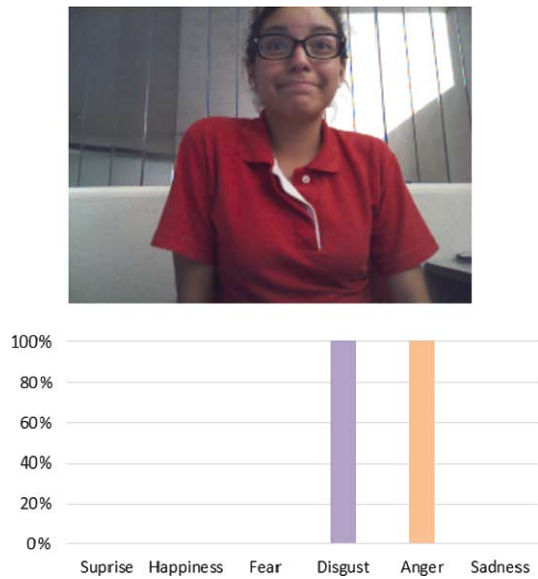


Fig. 21. Reported by system non-prototypical expression of *contempt* formed by basic emotions *disgust* and *anger*.



Fig. 23. Recognized by system facial basic expression of *sadness* formed by two other basic emotions of *surprise* and *anger*.

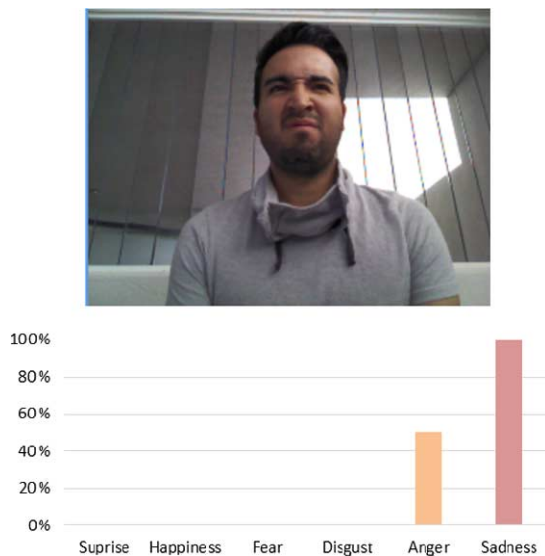


Fig. 22. Reported by system facial basic expression of *disgust* formed by two other basic emotions of *anger* and *sadness*.

to Plutchik's model the composed expression is *disgust* [27]. In recently appeared model of Du and Martinez [6] the archived recognition rate of 15 compound expressions is about 76% processing 94 fiducial points on a face.

However, the compound expressions proposed in [6] such as *happily disgusted* or *happily surprised* in essence are not particular non-prototypical expression

corresponding to existing psychological models. The names of these expressions simply reflect that emotion sensing application detects *disgust* or *surprise* of high intensity with presence of *happiness* of medium or low intensity.

As it has been mentioned in Section 5.4, the recognition rate of the proposed Distance-based system for similar expressions is about 87.6%. Nevertheless, in future researches the idea of extension of traditional concept of Ekman's basic emotions is very attractive. This will provide theoretical basis for automatic emotion sensing that is a still open problem. For example, in Fig. 24 a complex expression is presented however, there is not well suited model that may provide its description by combination of more than two basic emotions with significant intensities.

7. Conclusion

This paper presents some approaches for expanding traditional affective computing systems to context sensitive interpretation of emotional state of users. The conceptual contribution of this research consists in development of two face models for description of facial deformations encoded by Ekman's Action Units or Facial Animation Parameters of MPEG-4 standards. Particularly, 6-FACS and Distance models have been proposed and tested. 6-FACS model is used for recog-

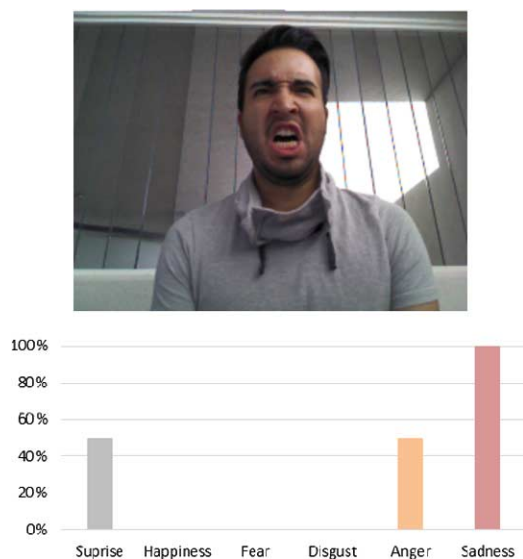


Fig. 24. Complex expression formed by three basic emotions of *surprise*, *anger* and *sadness*.

inition of limited number of AUs with the proposal to design fast and simple emotion interpreter. Due to its low precision of facial expression recognition (about of 82%), the facial model based on distances between specially selected FDPs have been introduced. Additionally, two-stage fuzzy reasoning classifier using Gaussian functions has been developed providing interpreting as basic as non-prototypical emotions with three levels of intensity. The achieved average recognition rate of basic emotions with quantitative measurement of their intensities in Distance model is 87.6%. Finally, interesting results in experiments with Plutchik's model have been obtained, which conduct us to conclusion that such basic as non-prototypical emotions may be composed by other basic emotions! In this way we establish relationship between non-prototypical emotions from psychological Plutchik's model and Ekman's six basic emotions expressed by AUs!

From practical point of view the paper contributes two designed systems for automatic recognition of facial expressions based on proposed facial models that have been implemented using new technological advances of Microsoft's Kinect sensor. Additionally, to take advantage of wide variety of systems and approaches, the extensible Web-based platform that provides developers with uniform interfaces and services has been proposed and designed so, applications can access and manage the existing or newly implemented systems for emotion sensing within smart environment.

It is important to mention that the proposed models and designed systems have some disadvantages that limit their performance. Still low level of recognition is because of the limited number of AUs used for representation of complex emotions (six AUs in 6-FACS model and nine AUs in Distance-based model), the simplicity of Plutchik's model and the subjectivity of perception of facial expression by human. Used fuzzy classifier as usual, has less recognition rate than Multilevel HMM, radial basis function network or SVM-based classifiers.

Our research is still in progress and only partially supports the idea of recognition of non-prototypical expressions. The future works are needed to extend the proposed smart environment integrating more emotion sensing systems that may process both behavioral and physiological features such as facial expressions, speech, gesture, vital signs and others. Improvement of facial action processing models and development of high efficient algorithms are required to ensure the precision and high speed of recognition and classification. In addition, the used emotion sensing models should be reconciled and conformed to existing psychological theories discovering better and more precise way to evaluate human affective state. The evaluation of novel proposal of consistently recognizable twenty one against six traditional basic emotions is not considered in this paper. This approach should be analyzed in further researches for its complete approval.

Acknowledgment

This research is partially supported by European Grant CASES – Advisory Sustainable Manufacturing Services and by Mexican CONACYT project (No. 198881)

References

- [1] S. Berretti, B.B. Amor, M. Daoudi and A. Bimbo, 3D facial expression recognition using SIFT descriptors of automatically detected keypoints, *Visual Computer* **27**(11) (2011), 1021–1036.
- [2] V. Broek, M.H. Schut, J. Westerink and K. Tuinenbreijer, Unobtrusive Sensing of Emotions (USE), *Journal of Ambient Intelligence and Smart Environments* **1**(3) (2009), 287–299.
- [3] C.M. Chen and H.P. Wang, Using emotion recognition technology to assess the effects of different multimedia materials on learning emotion and performance, *Library & Information Science Research* **33**(3) (2011), 244–255.

- [4] S. Chong-Ho, C. Gu-Min and J.N. Kwa, Pixel selection based on discriminant features with application to face recognition, *Pattern Recognition Letters* **33**(9) (2012), 1083–1092.
- [5] F. Dornaika, A. Moujahid and B. Raducanu, Facial expression recognition in videos using tracked facial actions: classifier performance analysis, *Engineering Applications of Artificial Intelligence* **26** (2013), 467–477.
- [6] S. Du, Y. Tao and A.M. Martinez, Compound facial expressions of emotion, *Proc. of the National Academy of Sciences* **111**(15) (2014), E1454–E1462, <http://www.pnas.org/content/suppl/2014/03/26/1322355111.DCSupplemental/pnas.201322355SI.pdf>.
- [7] P. Ekman and W.V. Friesen, *Facial Action Coding System (FACS)*, Consulting Psychologists Press, CA, USA, 1978.
- [8] N. Esau and E. Wetzal, Real-time facial expression recognition using a fuzzy emotion model, in: *Proc. IEEE International Conference on Fuzzy Systems*, 2007, pp. 1–6, http://pdf.aminer.org/000/330/955/exploring_the_time_course_of_facial_expressions_with_a_fuzzy.pdf.
- [9] FACS/AID User's Guide, Java applet for accessing FACS/AID, Retrieved Dec., 2013 from: <http://face-and-emotion.com/dataface/facsaid/facsaidocs.html#intro>.
- [10] H. Gunes and M. Pantic, Automatic, dimensional and continuous emotion recognition, *International Journal of Synthetic Emotions* **1**(1) (2010), 68–99.
- [11] J. Hamm, R. Kohler and V. Gur, Automated Facial Action Coding System for dynamic analysis of facial expressions in neuropsychiatric disorders, *Journal of Neuroscience Methods* **200**(2) (2011), 237–256.
- [12] G. Hermosilla, J. Ruiz-del-Solar, R. Verschae and M. Correa, A comparative study of thermal face recognition methods in unconstrained environments, *Pattern Recognition* **45**(7) (2012), 2445–2459.
- [13] ISO/IEC14496-2:2001(E), International Standard, Information technology – coding of audio-visual objects – Part 2: Visual, 2nd edn, 2001.
- [14] Y. Ji and K. Idrissi, Automatic facial expression recognition based on spatiotemporal descriptors, *Pattern Recognition Letters* **33**(10) (2012), 1373–1380.
- [15] T. Kanade and J.F. Cohn, Comprehensive database for facial expression analysis, in: *Proc. of 4th IEEE Conference on Automatic Face and Gesture Recognition*, France, 2000, pp. 46–53, <http://www.cs.cmu.edu/~face/Papers/FG3.pdf>.
- [16] G.U. Kharat and S.V. Dudul, Neural network classifier for human emotion recognition, in: *Proc. of 1-st Conference on Emerging Trends in Engineering and Technology*, Iran, 2008, pp. 1–6.
- [17] M.O. Kusserow and G. Amft, Modeling arousal phases in daily living using wearable sensors, *IEEE Transactions on Affective Computing* **4**(1) (2013), 93–105.
- [18] D.T. Lin, Facial expression classification using pca and hierarchical radial basis function network, *Journal of Information Science and Engineering* **22** (2006), 1033–1046.
- [19] I. Mahdi and H. Shah-Hosseini, A novel fuzzy facial expression recognition system based on facial feature extraction from color face images, *Engineering Applications of Artificial Intelligence* **25**(1) (2012), 130–146.
- [20] A. Majumder, L. Behera and V.K. Subramanian, Emotion recognition from geometric facial features using self-organizing map, *Pattern Recognition* **47**(3) (2014), 1282–1293.
- [21] G.C. Marchand and A.P. Gutierrez, The role of emotion in the learning process: comparisons between online and face-to-face learning settings, *Internet and Higher Education* **15**(3) (2012), 150–160.
- [22] A. Martinez and S. Du, A model of the perception of facial expressions of emotion by humans: research overview and perspectives, *Journal of Machine Learning Research* **13** (2012), 1589–1608.
- [23] E. Mower et al., Interpreting ambiguous emotional expressions, in: *Proc. of International Conference on Affective Computing and Intelligent Interaction*, 2009, pp. 1–8, http://www.academia.edu/560924/Interpreting_ambiguous_emotional_expressions.
- [24] K.N. Patel, *Style – Dynamic Delivery Techniques*, IEEE-USA E-Books, 2010.
- [25] Noldus Information Technology, FaceReader, Retrieved February 20, 2014, from: <http://www.noldus.com/human-behavior-research/products/facereader>.
- [26] M. Pantic, M.F. Valstar and R. Rademaker, Web-based database for facial expression analysis, in: *Proc. of IEEE Conference on Multimedia*, Netherlands, 2005, pp. 1–6, <http://ibug.doc.ic.ac.uk/media/uploads/documents/PanticEtAl-ICME2005-final.pdf>.
- [27] R. Plutchik, The nature of emotions, *Journal American Scientist* **89** (2001), 344.
- [28] Protégé, Ontology editor, 2009, <http://protege.stanford.edu>.
- [29] C. Ramirez, C. Concha and B. Valdes, Non-invasive technology on a classroom chair for detection of emotions used for the personalization of learning resources, *World Academy of Science, Engineering and Technology* **4** (2010), 1433–1439, <http://waset.org/publications/12756/non-invasive-technology-on-a-classroom-chair-for-detection-of-emotions-used-for-the-personalization-of-learning-resources>.
- [30] T. Ruf, A. Ernst and C. Küblbeck, Face detection with the sophisticated high-speed object recognition engine (SHORE), in: *Microelectronic Systems*, Springer, Berlin, 2011, pp. 243–252.
- [31] A. Savran, B. Sankur and M.T. Bilge, Regression-based intensity estimation of facial action units, *Image and Vision Computing* **30**(10) (2012), 774–784.
- [32] R. Schultz, M. Blech, J. Voskamp and B. Urban, Towards detecting cognitive load and emotions in usability studies using the RealEYES framework, in: LNCS, Vol. 4559, 2007, pp. 412–421.
- [33] N. Sebe, M.S. Lew, Y. Sun, I. Cohen, T. Gevers and T.S. Huang, Authentic facial expression analysis, *Image and Vision Computing* **25**(12) (2007), 1856–1863, <http://disi.unitn.it/sebe/publications/ivc-authentic07.pdf>.
- [34] G. Singh and B. Singh, Feature based method for human facial emotion detection using optical flow based analysis, *An International Journal of Engineering Sciences* **4** (2011), 363–372.
- [35] O. Starostenko, R. Contreras, V. Alarcon-Aquino and L. Flores-Pulido, Facial feature model for emotion recognition using fuzzy reasoning, in: LNCS, Vol. 6256, 2010, pp. 11–21.
- [36] O. Starostenko et al., A fuzzy reasoning model for recognition of facial expressions, *Computación y Sistemas* **15**(2) (2011), 163–180.
- [37] J. Treur and A. Wissen, Conceptual and computational analysis of the role of emotions and social influence in learning, *Journal Procedia Social and Behavioral Sciences* **93** (2013), 449–467.
- [38] N. Villaroman, D. Rowe and B. Swan, Teaching natural user interaction using OpenNI and the Microsoft Kinect sensor, in:

- Proc. of Information Technology Education Conference*, NY, USA, 2011, pp. 227–232.
- [39] W. Xiang, Y.V. Venkatesh, D. Huang and H. Lin, Facial expression recognition using radial encoding of local Gabor features and classifier synthesis, *Pattern Recognition* **45**(1) (2012), 80–91.
- [40] P. Ekman and E.L. Rosenberg, eds, *What the Face Reveals: Basic and Applied Studies of Spontaneous Expression Using the Facial Action Coding System (FACS)*, 2nd edn, Oxford University Press, USA, 2005.
- [41] Z. Yang and L.J. Rothkrant, Emotion sensing for context sensitive interpretation of crisis reports, in: *Proc. of Conference on Information Systems for Crisis Response and Management*, Brussels, 2007, pp. 507–514, http://www.iscramlive.org/dmdocuments/ISCRAM2007/Proceedings/Pages_507_514_55EMOT_05_A_Emotion.pdf.
- [42] A. Yu, C. Elder and J. Yeh, Facial recognition using Eigenfaces, 2009, <http://cnx.org/content/m33180/latest/>.
- [43] L. Zhang, S. Chen, T. Wang and Z. Liu, Automatic facial expression recognition based on hybrid features, *Energy Procedia* **17** (2012), 1817–1823.
- [44] X. Zhao, E. Dellandréa, J. Zou and L. Chen, A unified probabilistic framework for automatic 3D facial expression analysis based on a Bayesian belief inference and statistical feature models, *Image and Vision Computing* **31**(3) (2013), 231–245.