International Journal of Advance Research in Science and Engineering
Volume No.07, Special Issue No. (01), January 2018
www.ijarse.com
IJARSE
ISSN: 2319-8354

Facial Expressions and Emotions in Cognitive Science

B.Pandu Ranga Raju[1],K.Arun Kumar[2] , V.Sathyendra Kumar[3]

[1,2]Assistant Professor, Department of IT,

AITS- Rajampet, AP, (India)

[3]Assistant Professor, Department of MCA,

AITS- Rajampet, AP, (India)

ABSTRACT

Facial Expressions has to examine evaluations of surprised faces, which signal that an unexpected and unambiguous event has occurred in the expresser's environment. There are few multimodal fusion systems that integrate limited amount of facial expression, speech and gesture analysis. Specifically expressions are examined by older and younger participants, evaluation of happy, angry and surprised facial expressions. We predicate that merging of mindsets on the basis of age-related changes in the processing of emotional information, such as positive and negative facial actions with positive meaning. In this paper we describe the implementation of a semantic algebra based formal modal that integrates six basic facial expressions, speech phrases and gesture trajectories. This system has capable of real-time interaction.

Keywords:Aging, Emotions, Facial Expressions, Multimodal, Decision level fusion.

I. INTRODUCTION

1.1. Facial Expressions

The face plays an important role in social interaction, both in its static dimensions (structural feature, physiognomy) and in its dynamic dimension (facial expression), being a rich source of information and interactive signals. The face is in fact able to send a lot of information concerning age, gender, social status, etc., and affects impression of personality through the process of interpersonal perception. Facial expression on the other hand is an effective signaling system in interpersonal communication. In combination with other nonverbal signals it has a strong and immediate impact in expressing emotions such as fear, anger, happiness, sadness and in communicating interpersonal attitudes such as cordiality, hostility, dominance, submission and so on; it communicates also other mental activity such as attention, memory, thinking, etc. Moreover the face takes part actively in conversation: the "speaker" accompanies his/her words with facial expression to emphasize or modulate the meaning of verbal communication; the "listener" during conversation provides a constant feedback through facial expression.

1.2. Emotions

Most emotion analysis applications attempt to annotate video information with category labels that relate to emotional states. However, since humans use an overwhelming number of labels to describe emotion, we need to incorporate a higher-level and continuous representation that is closer to our conception of how emotions are

expressed and perceived. Activation-emotion space is a simple representation that is capable of capturing a wide range of significant issues in emotion. It rests on a simplified treatment of two key themes:

1.2.1. Valence

The clearest common element of emotional states is that the person is influenced by feelings that are "valenced", i.e. they are centrally concerned with positive or negative evaluations of people or things or events.

1.2.2. Activation level

Research has recognized that emotional states involve dispositions to act in certain ways. Thus, states can be rated in terms of the associated activation level, i.e. the strength of the person's disposition to take some action rather than none.

1.3. Emotions of Persons of Different Age

People also have beliefs about age and emotionality. Photos of individuals from four different ages groups (18–29; 30–49; 50–69; 70+) and asked them to indicate how likely they thought it that the person shown in the photo would express each of four emotions (happiness, sadness, anger, and fear) in everyday life. The responses differed with regard to both sex and age. Thus, as they get older, men were perceived to be less likely to show anger whereas the reverse was the case for women. Men were also perceived as more likely to show sadness as

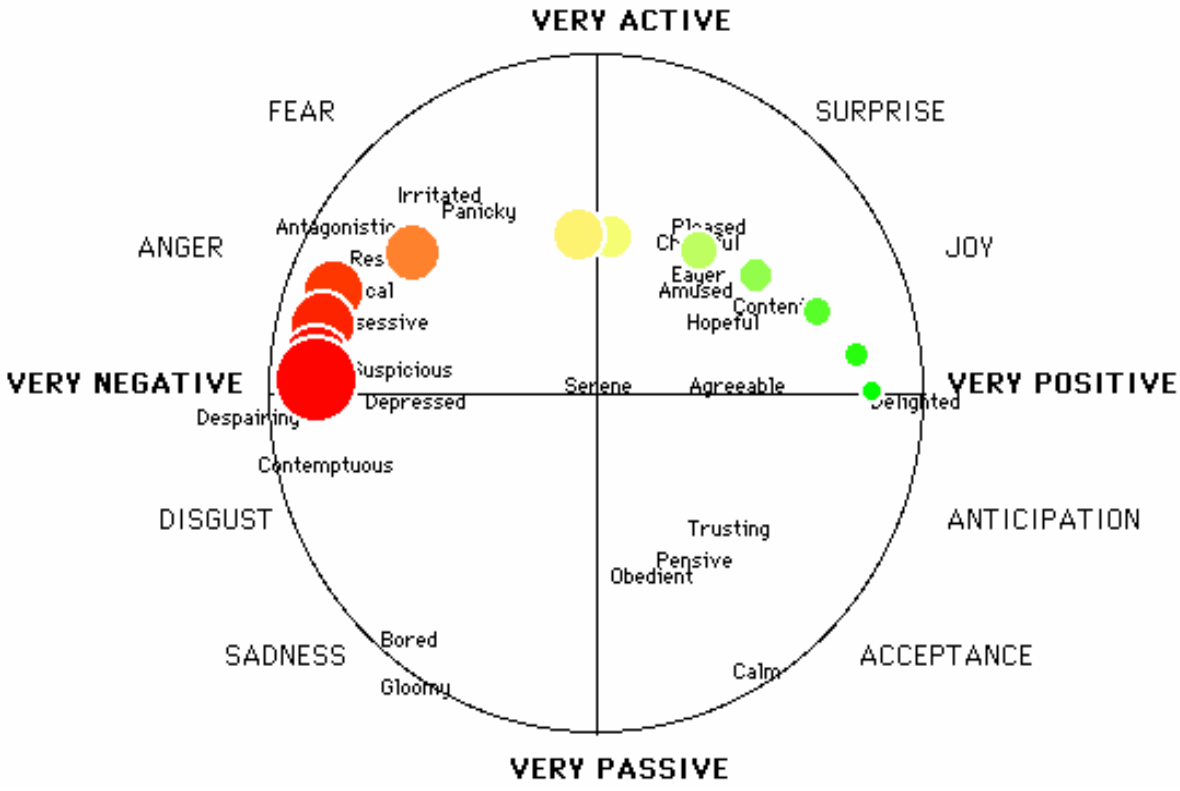

Fig1. The Activation-emotion space

II. BACKGROUND

2.1. Emotion Recognition

There are three popular psychological theories of emotions: *James-Lange theory* , *Cannon-bard theory* and *Schater -singer* theory. James-Lange theory states that the mental state in response to the reactions which caused by external stimuli is emotion. *Cannon-Bard* theory is based upon anticipation rather than as a reaction to specific action. *Schater -Singer* theory states that encountering an emotionrequires both an interpretation of the bodily response as well as specific circumstance at a specific moment. Also, there are three major classes of emotions:

a) Basic emotions,

b) Emotions that having same basic class, but having different intensity,

c) Mixed emotions that are a combination of one or more basic and/or mixed emotions.

Although, there are some disagreements among researchers, and a popular computational theory of Ekman identifies six basic emotions: *happiness, sadness, surprise, disgust, anger* and *fear*. An example set of emotions having same basic class, butdifferent intensities are {relaxed, happy, delighted, and euphoric}. Another set is {upset, anger, rage} etc. An example of mixed emotion is {amazed} that is a combination of {surprise and happiness} or {envy} which is the combination of {sadness and anger} or {despair} which is the combination of {fear and sadness}. In general, Facial Expressions have been done using these types of systems:

a) Facial Action Coding System (FACS) based on the simulation of facial muscle movement,
b) Geometric Features Modeling (GFM) based upon the movement of major feature-points of the face such as dynamic change in location endpoints and curvature of the mouth, eye, lips, forehead furrows and space betweeneyebrows.

Emotional speech has multiple features such as phonemes, emotional phrases, amplitude, syllable envelope, pitch, rhythm, quantile and silence. Phonemes are the basic units of speech. During emotional interaction, pitch, amplitude, syllable envelope, duration of silence and utterances change significantly; act as parameters for the recognition of interactive emotions. Gesture is a nonverbal communication using perceptible bodily actions such as body-postures and body-part movements, including movements of the head, torso, hands, face and eyes. Different components of the emotions are measured using different sensors. Facial-Expression uses image analysis techniques to identify the movement of facial feature points; speech analysis uses wavelet analysis, FFT analysis, morphology analysis, text-to-speech conversion for phoneme detection and dictionary lookup to identify phrases. Gesture recognition requires image analysis to derive postures and video-frame analysis to derive motion of various body parts such as head, arm, eyes, hand, palm, fingers. The posture and motion are modeled as fuzzy values to reduce the computational space. The motion of the body parts can also be derived using skeletal and depth analysis used in Kinect.

The unit of FACS is an Action Unit (AU) that involves a segment of a muscle in facial expression. There are 17 major AUs involved in basic facial expressions. Examples of AUs involved in facial expressions are: inner brow

raiser, outer brow raiser, brow lowered and drawn together, upper eye-lid raised, cheek raised, upper lip raised, lip corners pulled down, etc. The major geometric feature points, involved in facial expression analysis are given in Figure 2 which these features-points include:

a) 3 eyebrow points in each of the eyebrows: $b_1^L, b_2^L, b_3^L, b_1^R, b_2^R, b_3^R$

b) 2 endpoints of eyes in each of the eyes: $e_1^L, e_2^L, e_1^R, e_2^R$

c) Middle points eye-lid in each of the eyes: el_L and el_R

d) 2 endpoints of nose: n^T *and* n^B

e) 2 endpoints of mouth: m^L *and* m^R

f) 2 middle points of the mouth based on top and bottom lips: m^T *and* m

g) Chin-point denoted as: *ch.*

The points shaded in dark black- $e_1^L, e_2^L, e_1^R, e_2^R, n^T \& n^B$ do not move, and act as reference-points. Remaining spotted-points move with emotions, and their displacement is used to derive the facial expression.

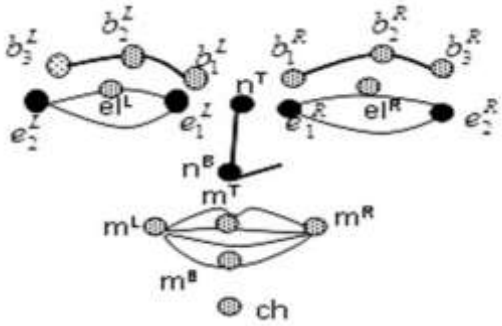

Fig 2. Major feature-points on the face

2.2. Mathematical concepts

The Fuzzy values map a large value-space to a smaller finite space. The major advantages of the use of fuzzy values are:

a) Reduction of the computational complexity

b) Nearness to human perception and

c) Tolerance from the sensor noise.

We use two types of fuzzy sets:

a) Discrete fuzzy set, and

b) Ordered fuzzy sets.

A discrete fuzzy set has values that have no relationship that shows transitivity. For example, a head posture can be {rotated-left, rotated-right, normal, tilted-left, tilted-right, looking-down, looking-up}. An ordered fuzzy set shows transitive relationship between the values, and is used to model motion intensity in gesture analysis for

better classification of emotion. For example, the speed of a head-motion can be modeled as {still, slow, normal, fast, very fast}. The values in the fuzzy set can be mapped onto the ordinals 0… 4:

Still→0, Slow→1, Normal→2, Fast→3 and Very Fast→4 (1)

The use of this mapping allows the use of comparison operators on ordered fuzzy sets. Cartesian product of the N sets returns a set of N-tuples such i^{th}-field of an element is a member of the i^{th} set as shown:

$$X_1 \times \ldots \times X_n = \{(x_1, \ldots, x_n) \mid x_i \in X_i \; \forall \; i = 1, \ldots, n\} \quad (2)$$

Two domains can be joined using:

a) Product-domain that uses the Cartesian product $^{A \times B}$, or

b) A sum - domain that uses disjoint-union $^{A+B}$, or

Function Domain mapping on lifted domains f: A⊥B.Where ⊥ is the bottom symbol used to catch all ill-defined mappings.

Facial Expressions	Major Feature-points displacements
Anger	$(e_1 \leftarrow \; el\uparrow) + [e_2\uparrow] + [m^T\uparrow \; m^B\uparrow]$
Disgusted	$(m^T\uparrow ch\uparrow) + [\{m^L, \; m^R\}\downarrow] + [m^B\uparrow]$
Fear	$(e_1\uparrow, \; m^L\downarrow \; m^R\downarrow) + [m^T\downarrow] + [e1\leftarrow]$
Happiness	$(m^L\nearrow m^R\nearrow, \; M^T\uparrow m^B\downarrow \; ch\downarrow m^L\leftrightarrow m^R\leftrightarrow)$
Sadness	$(el\downarrow \; m^L\leftrightarrow m^R\leftrightarrow) + [ch\downarrow]$
Surprised	$(e^1\uparrow e^2\uparrow e^3\uparrow el\uparrow ch\downarrow) + [m^T\uparrow m^B\downarrow]$

TABLE1. Feature Point displacements

Analysis by using facial symmetry and invariance under head Motion. There are 13 moving-points (11 active points and 2Passive points) and 6 references-points .FACS system analysis has been used to derive the features-points that are significant during the expression of a specific facial expression. For example, for a surprise the all eyebrow points are uniformly raised; for happiness mouth corners arestretched, the eye-lid point gets lowered; for anger distance between eyebrows becomes smaller, inner eyebrow points get lowered. These FAUs have been translated to the corresponding feature-point movements as given in Table 1. We denote vertical-up motion by ↑, vertical-down motion by ↓, horizontally stretched outwards by '↔', horizontally compressed inwards by '↔', oblique-stretched downwards by '↘', oblique-stretched upwards by '↗'. If the emotion is symmetric, then the subscripts L and R have been omitted. If the movement is optional or shows higher intensity increase then it has been placed within the square brackets. Conjunction has been shown using concatenation Essential feature-point have been within parenthesis () separated by ','. At least one of the essential feature point motion has to be present for the emotion to occur. Scores are associated with the presence of each feature-point motion for each feature point; we measure the displacement distance and the direction of the displacement. Thus the derivable facial expressions are mapped to a vector of (displacement-distance ratio, direction).Direction is a discrete-fuzzy set with six possible values:

a) Vertical-up,
b) Vertical-down,
c) Horizontal-compressed-inwards,
d) Horizontal-stretched-outwards,
e) Oblique-stretched-upwards, and
f) Oblique-stretched-downwards.

III. METHOD

3.1. Facial-Expressions Agreement

To assess the value of obtained judgments we first used the majority vote method to determine the label of each image. The confusion matrix is presented in table. This matrix illustrates the judgments confusion among the nine alternatives: all six basic expressions (neutral, angry, contempt, disgust, fear, happy, sad and surprise), the composed expression contempt an ambiguous and a noisy capture (not a face). The matrix is organized in a judgments verses label fashion: each column represents the actual label obtained by majority vote and each row indicates how the label of each column is confused with the other ones (for example, images with a facial expressions of fear, are identified as Disgust 10.74% of the times). The diagonal of the confusion matrix illustrates how the majority of expressions are clearly separable from the others. The facial expressions happy and surprise were the most consensual among all annotators with an agreement of over 0.8. Many facial expressions are confused with neutral. The most dubious facial expression is contempt which is often confused with neutral, once more due to the intensity of expression.

Ambiguous expressions achieved a surprising agreement of 0.47 because it was confused with neutral 10% of the time. This means that when a user is not performing one of the other facial expressions, some annotators assign the neutral label while others assign the ambiguous label and the remaining annotators try to choose a label. So, we can conclude that annotators follow different decision criterion when they are faced with ambiguous expressions. One of the most relevant contributions of this dataset concerns the judgments quality at such a large-scale for a facial expression dataset. An image with agreement of 1.0, means that all 5 votes were on same label, this happens on 39.7% of our dataset and 62.4% has at least 0.8 of agreement, which means 4 to 5 votes were on the same label. There were a total of 20.9% images with an agreement of 60%. Thus .this is a very high agreement for such a large dataset:approximately 25,000 images (62.4% of the dataset) have an agreement of 80% and 34,000 images (83.3% of the dataset) have an agreement of 60% or more. On the other hand, 15.5% of dataset has an agreement of 40% and 1.2% of the images has an agreement of 0.2, in other words, all the votes were on different labels. Although these images have a low agreement (16.5% of images have an agreement of 40% or less), these results allows drawing an important conclusion. We observed that annotators avoided the ambiguous label (0.47%) and tried to make a decision, however, vote statistics show that 16.5% of the faces were actually ambiguous.

3.2. Labels Quality

To compare the crowd sourcing labels to expert ground-truth, we repeated the previous process with the CK+ dataset and corrected them with the statistical consensus methods. We used the implementation provided by Sheshadri and Lease, who conducted an extensive evaluation of such methods in natural language datasets and two image datasets. Both image datasets concerned binary annotations (annotation of a face smile and discrimination between two types of birds). Besides the importance of analyzing our data with different methods, it is also important to note that in contrast to the image datasets used in the Nova Emotions dataset contains uncleaned data, has multiple labels and data contain far more ambiguity. The results of table show the accuracy of each crowd sourcing method for facial expression. The last line presents the accuracy of each crowd sourcing method. Note that, we did not take into account the facial expression contempt because we had very few examples. The best accuracy was 91.45% and that result was achieved by RY. IT is interesting to notice that RY only support binary labels but achieved better results than DS and ZC, which actually support multiclass. On the other hand, CUBAM achieved only 82.28%. We identified two causes of this: (1) our approach to use CUBAM in multiclass problem did not work and (2) CUBAM cannot handle a data set with different class proportions. These results show that crowd sourcing labels are less than 9% different from expert labels.

3.3 Classifiers

3.3.1. k-Nearest Neighbors

In a k-NN classifier, the model is the entire training set, where each training sample corresponds to a multi dimensional feature vector and a facial-expression label. To determine the label of a test image, we can simply calculate the majority of the labels on the set of the k nearest elements, sorted by the Euclidean distance.

3.3.2. Weighted k-Nearest Neighbors

Since the k-NN classifier is majority voting method, it means that all k-neighbors contribute equally to the classification of a test image. An extension of the k-NN is to weight the vote of each nearest neighbor by the inverse of the distance (1/d) to the test image. We will refer to this modified nearest neighbor classifier as weighted k-NN.

3.3.3. Kernel Density Estimation

The KDE is an approach to estimate the true probability density distribution from the training data. Unlike the k-NN method that uses only the nearest k elements, the KDE method uses the entire training set to compute a smoothed estimate of the true probability density function. The contribution of each element also differs from the weighted k-NN: instead of using the Euclidean distance to weight each neighbor, it applies a kernel function to every point of training set to compute the contribution of every training sample. This kernel function is usually a standard probability distribution function. For a matter of convenience, we will use Gaussian Kernel as follows

$$K(z) = 1/\sqrt{2\pi}\ (e\text{-}1/2)\ z^2 \quad (1)$$

To estimate the density function on a given test point x^3, the aggregate contributions of all training samples correspond:

$$f\hat{}(x) = 1/n \sum_{i=1}^{n} K_h(x\text{-}x_i) \quad (2)$$

Due to the fact that f(x) is dependent on the distance of point x to the training samples, we need to compute this sum for every test image that we need to classify.Formally, we have one function $f\hat{}l_j$ (x) for each label l_j of our problem, where j=1,…, L. In our case, we will have a function $f\hat{}l_j$ (x) for each facial expression. Thus each training sample contributes exclusively to the density function of its own label. This leads us to the following formalization:

$$nf\hat{}(x) = 1/n\sum_{i=1} K_h(x\text{-}x_i) * 1l_j(x_i) \quad (3)$$

Where $1l_j(x_i)$ is an indicator function, taking the value 1 if the sample x_i belongs to the label l_j and 0 otherwise. It is now straightforward to address the multiclass nature of facial expressions. Using Bayes' Theorem we can merge all individual density estimates $f\hat{}l_j$ (x) with j=1,…..,L:

$$P(l = l_j|X = x0) = \pi l_j\ f\hat{}l_j\ (x0\ ) \ /\ \sum^{L}_{i=0} \pi l_j\ f\hat{}l_j\ (x0\ ) \quad (4)$$

Where πicorresponds to the label i prior. This definition allows computing the probability of one image x0 belonging to certain label lj. To classify test images, we only need to find the label l_j that maximizes the above expression.

3.4. Datasets

The comparative evaluation will use two facial expressions datasets: the Cohn-Kanade Extended (CK+) dataset and the Nova Emotions dataset. These datasets provide an adequate setting for comparison, as the first was annotated by experts and second by crowd sourcing.

3.4.1. Cohn-Kanade

The CK+ dataset contains 593 sequences of video-frames from 123 subjects, where each image illustrates a facial expression at its maximum intensity. The dataset is composed by a sequence of images where each sequence represents a facial expression. The first and last images of each sequence are annotated by experts.

3.4.2. Crowd sourcing labels

The CK+ dataset already has expert labels; we collect the crowd sourcing labels for the CK+ dataset. The datasets have crowd sourcing labels but only the CK+ has expert labels.

TABLE 2. Confusion matrix for each facial expression. The agreement is computed assuming that the most voted expression is the correct one. The correct labels are in the columns and each row of a column indicates the distribution of votes across

	Neutral	Angry	Contempt	Disgust	Fear	Happy	Sad	Surprise	Ambig.	NAF
Neutral	66.74	11.30	19.21	5.25	5.21	4.09	7.55	2.35	10.05	13.91
Angry	2.27	49.91	4.16	4.30	1.98	0.20	2.06	0.53	2.34	1.30
Contempt	6.10	8.17	42.18	4.70	2.53	1.08	3.17	0.67	6.03	3.48
Disgust	2.64	10.65	8.08	60.45	10.74	1.16	4.94	1.92	8.82	3.91
Fear	1.02	1.57	1.76	3.34	52.03	0.38	1.71	2.98	1.98	2.17
Happy	8.51	3.91	6.66	5.63	3.73	88.23	2.43	5.07	9.68	10.00
Sad	5.28	6.91	8.00	6.61	4.61	0.55	73.41	0.53	3.94	2.17
Surprise	2.42	2.39	3.25	3.53	14.70	2.48	0.92	82.97	7.97	6.09
Ambiguous	4.46	4.87	6.35	5.74	4.33	1.66	3.56	2.67	47.09	12.17
Not a face	0.55	0.30	0.35	0.45	0.14	0.17	0.24	0.31	2.11	44.78

all labels. NAF stands for Not a Face.

TABLE 3. Crowd sourcing labels comparison to expert labels on the CK+ dataset. The statistical consensus methods are: Majority Vote (MV), CUBAM, Dawid and Skene (DS), Generative model of Labels, Abilities and Difficulties (GLAD), Raykar (RY) and ZenCrowd (ZC).

Facial expression	MV	CUBAM	DS	GLAD	RY	ZC
Angry	80.00	80.00	78.75	76.25	80.00	81.25
Disgust	96.81	72.34	98.94	95.74	96.81	94.68
Fear	97.22	30.56	94.44	97.22	97.22	77.78
Happy	91.89	88.51	91.22	91.22	91.22	93.24
Neutral	88.16	89.02	84.73	87.14	90.05	90.05
Sad	93.88	69.39	89.80	91.84	89.80	91.84
Surprise	100.00	74.13	100.00	100.00	100.00	98.60
Accuracy	90.74	82.28	88.71	89.68	91.45	91.01

Fig 3. Example images from the CK+ and the Nova Emotions datasets

IV. CONCLUSION AND FUTURE WORKS

In this paper, we have described a detailed methodology and an initial prototype implementation of real-time multimodal fusion to derive interactive emotion for interaction with social-robots and intelligent machines with limited emotional phrase based interaction. The proposed integrated system has many novelties such as: an abstract model of fusion based upon a semantic algebra that the maps Cartesian product of different components

to derivable emotions, the use of invariant displacement of geometric feature-points to identify facial-expressions, and Gestures based upon head-trajectory and fuzzy values of other upper body parts to reduce the search space. Currently, the gesture based system is limited to, image analysis of feature-points in the head and hand to derive posture. We are looking into Kinect based analysis to integrate skeleton based body posture, motion and depth analysis for better accuracy.

REFERENCES

[1] R. Adolphs. "Recognizing emotion from facial expressions: psychologicaland neurological mechanisms," *Behav. Cogn. Neurosci. Rev., Vol. 1,2002, pp. 21-62.*

[2] V. Bevilacqua, D. Barone, F. Cipriani, G. D'Onghia et al., "A new tool for gestural action recognition to support decisions in emotional framework", *Proceedings of the IEEE Symposium of Innovations inIntelligent Systems and Applications (INISTA), 2014, pp. 184-191.*

[3] G. Caridakis, G. Castellano, L. Kessous, A. Raouzaiou, L. Malatesta, et al. "Multimodal emotion recognition from expressive faces, body gestures and speech"; *Artificial Intelligence and Innovations: FromTheory to Applications, Springer Berlin Heidelberg, 2007, pp. 375-388.*

[4] P. Ekman and W. Friesen, "Facial Action Coding System: A Techniquefor the Measurement of Facial Movement," *Consulting PsychologistsPress, Palo Alto, 1978.*

[5] M. Ghayoumi and A. K. Bansal, "Unifying Geometric Features and FacialAction Units for Improved Performance of Facial Expression Analysis,"Proceedings *of the International Conference on Circuits, Systems, SignalProcessing, Communications and Computers (CSSCC 15), pp. 259-266.*

[6] Cowie, R, Douglas-Cowie, E "Emotion recognition in human-computer interaction." *IEEE Signal Processing Magazine. pp. 33-80, 2001.*

[7] M. Ghayoumi, A. Bansal, "An Integrated Approach for Efficient Analysis*of Facial Expressions", SIGMAP 2014.*

[8] P. Ekman, Facial expression and emotion. *American, 8 (4): 384-392, 199.*

[9] M. Ghayoumi, A. Bansal, "An Integrated Approach for Efficient Analysisof Facial Expressions", *SIGMAP2014.*

[10] A, Walter. "The James-Lange Theory of Emotions: A Critical Examination and an Alternative Theory." *The American Journal of Psychology 39: 106-124.*

[11] Aashish Sheshadri and Mathew Lease, Square: A benchmark for research on computing crowd consensus. *In First AAAI Conference on Human Computation and Crowding 2013.*

[12] Alexander Philip Dawid and Allam M Skene Maximum likelihood estimation of observer error-rates using the em algorithm. *Applied Statistics pages 20-28, 1979.*

[13] Takeo Kanade, Jeffrey F Cohn and Yingli Tian. Comprehensive database for facial expression analysis. In Automatic Face and Guesture Recognition, *2000. Proceedings. Fourth IEEE International Conference on, pages 46-53. IEEE, 2000.*

[14] Andre Mourao and Joao Magalhaes. Competitive affective gaming: Winning with a smile. *In Proceedings of the 21st ACM International Conference on Multimedia, pages 83-92.ACM,2013*

[15] Peter Welinder, Steve Branson, Serge Belongie,Perona and San Diego. *The Multidimensional wisdom of Crowds, pages 1-9.*

[16] Andre Mourao, Pedro Borges, Nuno Correia and Joao Magalhaes. Facial expression recognition by sparse reconstruction with robust features. *In Image Analysis and Recognition, pages 107-115. Springer, 2013.*

[17] M. Pantic, and M. S. Bartlett, "Machine Analysis of Facial Expressions," *Face Recognition, K. Delac and M. Grgic, eds., pp. 377-416, Vienna, Austria: I-Tech Education and Publishing, 2007.*

[18] M. Pantic, M. F. Valstar, R. Rademaker *et al.*, "Web-based database for facial expression analysis."