See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/284724383

Affect Representation and Recognition in 3D Continuous Valence–Arousal–Dominance Space

Article *in* Multimedia Tools and Applications · January 2017

DOI: 10.1007/s11042-015-3119-y

CITATIONS
84

READS
3,432

2 authors:

Gyanendra Kumar Verma
National Institute of Technology Raipur

67 PUBLICATIONS **1,451** CITATIONS

SEE PROFILE

Uma Shanker Tiwary
Indian Institute of Information Technology Allahabad

127 PUBLICATIONS **1,295** CITATIONS

SEE PROFILE

Some of the authors of this publication are also working on these related projects:

Psycholinguistic View project

Hindi TTS View project

All content following this page was uploaded by Gyanendra Kumar Verma on 31 May 2018.

The user has requested enhancement of the downloaded file.

Affect representation and recognition in 3D continuous valence–arousal–dominance space

Gyanendra K Verma[1] **· Uma Shanker Tiwary**[2]

Received: 25 February 2015 / Revised: 8 October 2015 / Accepted: 25 November 2015 /
Published online: 8 January 2016

© Springer Science+Business Media New York 2016

Abstract Currently, the focus of research on human affect recognition has shifted from six basic emotions to complex affect recognition in continuous two or three dimensional space due to the following challenges: (i) the difficulty in representing and analyzing large number of emotions in one framework, (ii) the problem of representing complex emotions in the framework, and (iii) the lack of validation of the framework through measured signals, and (iv) the lack of applicability of the selected framework to other aspects of affective computing. This paper presents a Valence – Arousal – Dominance framework to represent emotions. This framework is capable of representing complex emotions on continuous 3D space. To validate the model, an affect recognition technique has been proposed that analyses spontaneous physiological (EEG) and visual cues. The DEAP dataset is a multimodal emotion dataset which contains video and physiological signals as well as Valence, Arousal and Dominance values. This dataset has been used for multimodal analysis and recognition of human emotions. The results prove the correctness and sufficiency of the proposed framework. The model has also been compared with other two dimensional models and the capacity of the model to represent many more complex emotions has been discussed.

Keywords Affect representation · Emotion recognition · Valence · Arousal · Dominance · Physiological signals · EEG · Classification and clustering of emotions

✉ Gyanendra K Verma
gyanendra@nitkkr.ac.in

Uma Shanker Tiwary
ust@iiita.ac.in

[1] Department of Computer Enggineering, National Institute of Technology, Kurukshetra 136119, India

[2] Department of Information Technology, Indian Institute of Information Technology, Allahabad 211012, India

1 Introduction

Affective Computing is a prime research area of Human Computer Interaction (HCI) which combines engineering and computer science with Cognitive Science, Sociology, Physiology, Psychology and many other fields. In the last few decades most of the research was based on spontaneous as well as posed data acquired in the laboratory setting for affect recognition. The different affective states like thinking, embarrassment, depression etc. can be considered complex affective states and expressed via several anatomically possible facial expressions or body gestures [8]. Since the complex affective states cannot be expressed by a single label, therefore researchers have used the dimensional model of affect to express them. Recently, the focus of research on human affect recognition has been shifted from discrete emotions to affect recognition in continuous two or three dimensional space. Researchers started exploring dimensional model of emotion as this model is able to capture action units (AUs) effectively. The complex emotion states can be represented by two/three dimensions emotion primitives. The 2D model is covered by valence and arousal on two axes (in both positive and negative directions), representing all emotions in four quadrants, whereas, a 3D model deals with three emotion primitives i.e. valence, arousal and dominance. Some researchers have used different nomenclature for different emotion dimensions, for example Whissell [27] proposed 2D emotion model by considering evaluation-activation as two dimensions.

Key contributions of this study are as follows:

- It presents the approach for affect prediction in terms of valence, arousal and dominance based on physiological (EEG etc.) and visual cues.
- It also predicts the correlation among emotions and demonstrates significant improvement in emotion recognition performance.

In this paper, we have studied and reviewed the research work on emotion representation in two dimensional space. Then, we have mentioned the limitations and shortcomings of two dimensional model and highlighted the need for a three dimensional emotion model. Moreover, a 3D emotion model and methodology is explained. An emotion graph is generated by representing a large number of emotions in three dimensional space and the findings obtained from emotion graph are discussed and validated. Emotion prediction from multimodal cues is also presented.

We have used DEAP [10], a database for the analysis of spontaneous emotions. The database contains samples of physiological signals along with frontal facial video of participants. Each participant watched different videos and recorded their emotional responses in terms of Arousal, Valence and Dominance [10].

This paper includes three major experiments. Experiment 1 includes emotion representation and anlysis of distribution of emotions in VAD space. Clustering and relative distance measure followed by emotion graph generation is done under experiment 2. Experiment 3 is performed to validate the emotion graph through emotion prediction from multimodal physiological cues. The overall paper is divided into six sections, including this introduction section. The second section deals with state-of-the-art of affect representation and modeling of emotions. The proposed methodology for emotion prediction in 3D space is described in third section. Emotion recognition from multimodal cues is described in section four. Results and discussions are discussed in section five and concluding remarks are given in the last section.

2 State-of-the-art of affect representation and modeling

Recently, researchers started exploring dimensional model of emotion, in which emotions are represented in 2D or 3D space. The 2D model is covered by two dimensions i.e. valence and arousal, whereas 3D model deals with three emotion primitives i.e. valence, arousal and dominance. Many researchers have mapped directly the visual signal onto emotion dimensions [7, 14, 28]. Emotion categorization is critical to know the affective states of human in the application of emotion recognition. Some researchers have used a simple strategy to automatic classification of affect, and that is to simplify the problem of classifying six basic emotions to a two class (positive-negative) or three class (positive, neutral and negative) classification problem. A similar simplification is to reduce the emotion classification problem to a two-class problem—positive vs. negative and active vs. passive classification problem. Few researchers [1] used Pleasure, Arousal and Dominance (PAD) as three emotion primitives. The valence scale ranges from unpleasant to pleasant. S. Kolestra et al. [10] added one more emotion dimension i.e. liking in their work [10]. An emotion can be independently described by each of these three primitives or measures on a continuous valued scale.

Many researchers [3, 6] have reported a four-class classification problem—i.e. each quadrants of 2D AV space. Glowinski et al. [7], for instance, analysed four emotions, each belonging to one quadrant of AV emotion space: high arousal positive valence (joy), high arousal negative valence (anger), low arousal positive valence (relief) and low arousal negative valence (sadness).

2.1 Affect representation in 2D space

Whissell [27] presented a two dimensional emotion model, taking a pair of values: evaluation and activation. They have shown the position of affective words in evaluation-activation space in the range of (−3, +3). The neutral is placed at the origin. The emotion mapping is carried out by considering each of the six basic emotions as 2D weighted points in the evaluation-activation space, where the assigned weights are the affective weights obtained by the facial affect recognizer for each emotion.

In Whissell space (Fig. 1), each emotion has a specific location in evaluation-activation space. It is interesting to note that the emotion ‘joy’ is placed far apart from other emotions, whereas it should be near ‘cheerful’ and ‘pleased’. Moreover, emotion ‘surprise’ and ‘joy’ are in the same quadrant-2. In addition, the Whissell space hardly detects intermediate states between ‘joy’ and rest of the emotions. Therefore, we can say that two dimensional representation is not able to correctly represent the relationship among various emotions. Should we require more dimensions to represent all emotions accurately?

A researcher presented a mood and emotions tracking experiment in his article [2]. In this the average valence and arousal scores for emotion were mapped in the valence-arousal space as depicted in Fig. 2. Although, it is obvious in Fig. 2 that the two dimensional graph represents emotional states relatively well, but as can be seen in the graph, all the positive and intense emotions like ‘pride’, ‘thrill’ and ‘joy’ end up in the upper right quadrant of the space. The negative and intense emotions like ‘irritation’, ‘suffering’ and ‘anxiety’ are in the opposite third quadrant. The less intense emotions are close to the middle of the arousal scale, but in the appropriate upper and lower parts of the plane, according to their values of valence [2].

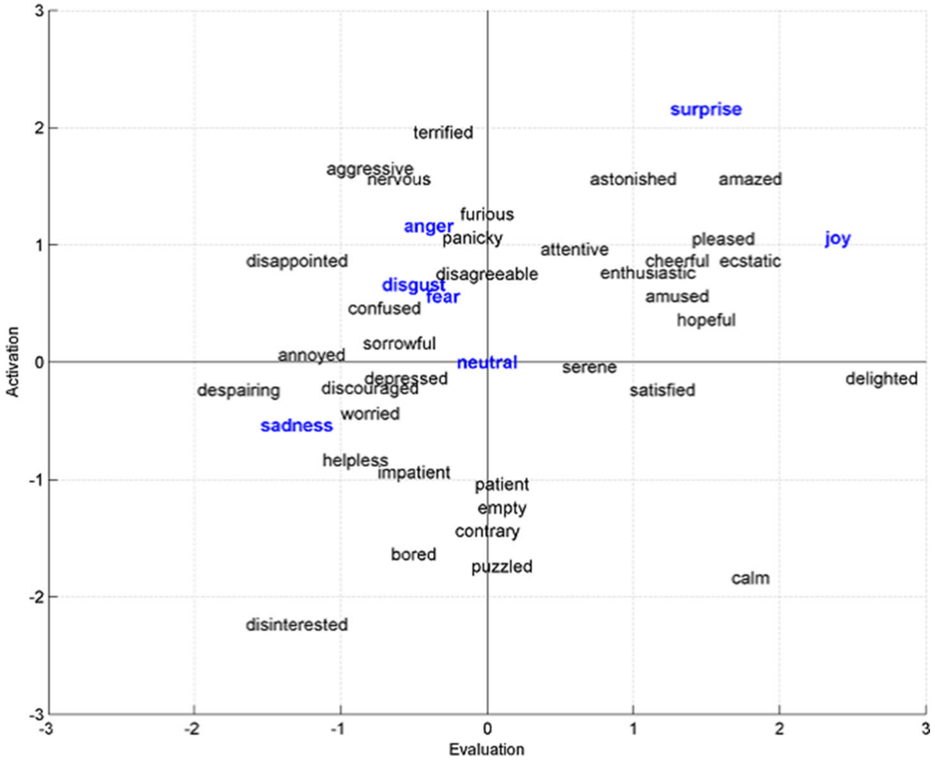


Fig. 1 Whissell's evaluation-activation space

In the valence-arousal space, the three emotions, 'contentment', 'affection' and 'sadness' are on the valence axes itself, which is not possible. Emotion 'affection' belongs to happy group, so it should be nearby 'joy'. In addition, only limited number of emotions are represented in this model. Hence, this valence-arousal model is insufficient to represent emotions accurately. So again the question arises here, should we require more dimensions to accurately represent the emotions?

The answer is YES. From the above two studies, it is clear that the representations of emotions in two dimensional space are not sufficient. Therefore, we have decided to explore the three dimensional representation of emotions.

According to Schachter and Singer [18], many psychologists believe that physiological signals do not distinguish more than arousal level. R. Picard [15] claimed that these distinctions were not limited to arousal, but also included discrimination of emotions having similar arousal and varying positive or negative (valence) characteristics. They developed pattern recognition algorithms that attained 81 % classification accuracy instead of the predicted 12.5 % of a random classifier. They found that emotions could be distinguished at levels significantly higher than chance. Furthermore, they concluded that recognizable physiological differentiation does appear with the eight emotions they investigated.

2.2 Affect representation in 3D space

According to Schuller B. [20] estimating emotion on a continuous valued scale is an important alternative to emotion categories for computational community to describe human's affective

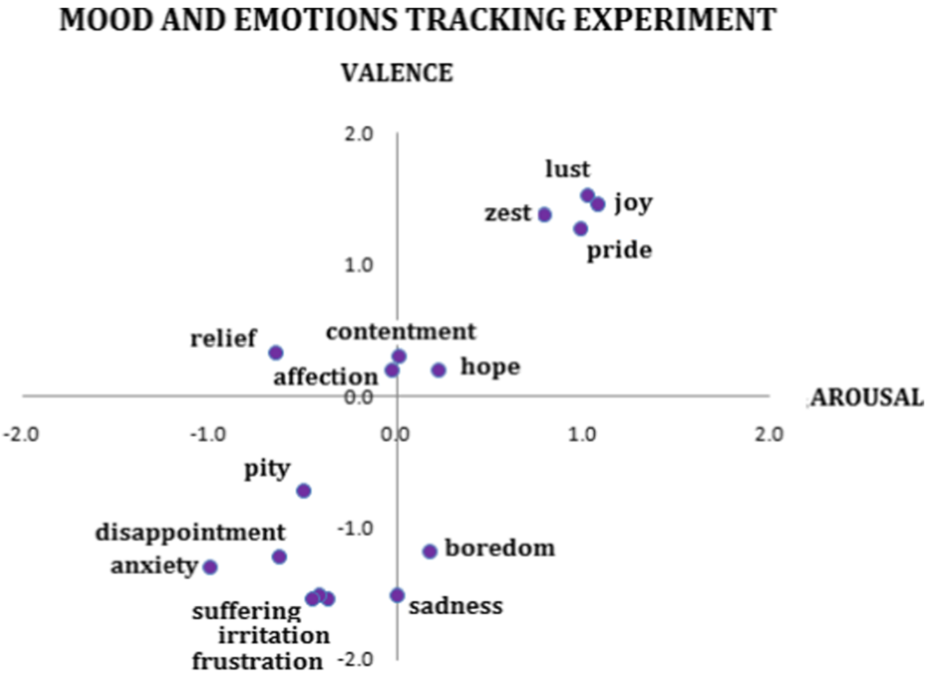


Fig. 2 Mapping emotional states in valence-arousal mood space

states because it is able to describe the intensity of emotion, which can be used for recognizing dynamics and allows for adaptation to individual moods and personalities. The 3D emotion model consists of three emotion dimensions i.e. Valence, Arousal and Dominance (in short VAD).

- Valence: The valence scale ranges from unhappy or sad like emotions (negative emotions) on one end to happy or joyful like emotion (positive emotions) at the other end.
- Arousal: The arousal scale ranges from 0 to 9 signifying calm emotions to stimulated or excited emotions.
- Dominance: The dominance scale ranges from submissive (or “without control”) on one end to dominant (or “in control” or “empowered”) at the other end.
- Other possible candidates for dimensions may be liking and familiarity. Liking inquires about the participant’s tastes, not their feelings. Also familiarity varies from subject to subject for each emotion. Hence, liking and familiarty cannot be considered as dimensions for representation of emotion independent of subjects.

DEAP database has been successfully used for emotion recognition from EEG and peripheral cues in [9, 11, 24]. This work is a novel effort to correlate video and other multimodal cues (i.e. EEG and other physiological signals) with Valence, Arousal and Dominance values. The EEG signals of 32 subjects (in DEAP dataset) were recorded while participants watched one-minute fragments from 40 music video clips. The VAD values were obtained using Self Assessment Manikins (SAM) [13], a method of self-assessment emotional states. Valence, arousal, dominance and liking were rated directly after each trial on a continuous 9-point scale using a standard mouse. For liking (i.e. how much did one like the video?), thumbs-up and thumbs-down icons were used as shown in Fig. 3. The intensity scales of emotional reactions of excitation, arousal, were represented using graphic pictures that express nine levels from (1–9) to indicate their emotional state. Similarly for valence, for dominance and liking, SAM method, as depicted in Fig. 3 was used.

It is worth noting that dimensional representation has mostly been used for emotion recognition from physiological signals. In Table 1, we briefly summarised automated systems that attempt model and recognize affect in continuous dimensional space. Table 2 summarizes representative systems with classification methods and results for dimensional affect recognition. S. Koelstra et al. [10] proposed an emotion recognition system by using a fusion of facial expressions and EEG signals. They utilized methods for facial expression and EEG signal analysis to investigate the possibilities for multi-modal fusion in affect recognition and implicit tagging. Mamalis A. Nikolaou [14] presented emotion recognition system based on multimodal cues such as facial expression, shoulder gesture and audio cues and fusion thereof. They mapped the multimodal cues of continuous emotions in two dimensional space of valence and arousal. They used Support Vector Regression (SVR) and Long Short Term Memory Neural Network (BLSTM-NNs) machine learning algorithms to compare the performance of the system. An output associative fusion framework was also proposed by them for the feature and model level fusion of multiple cues.

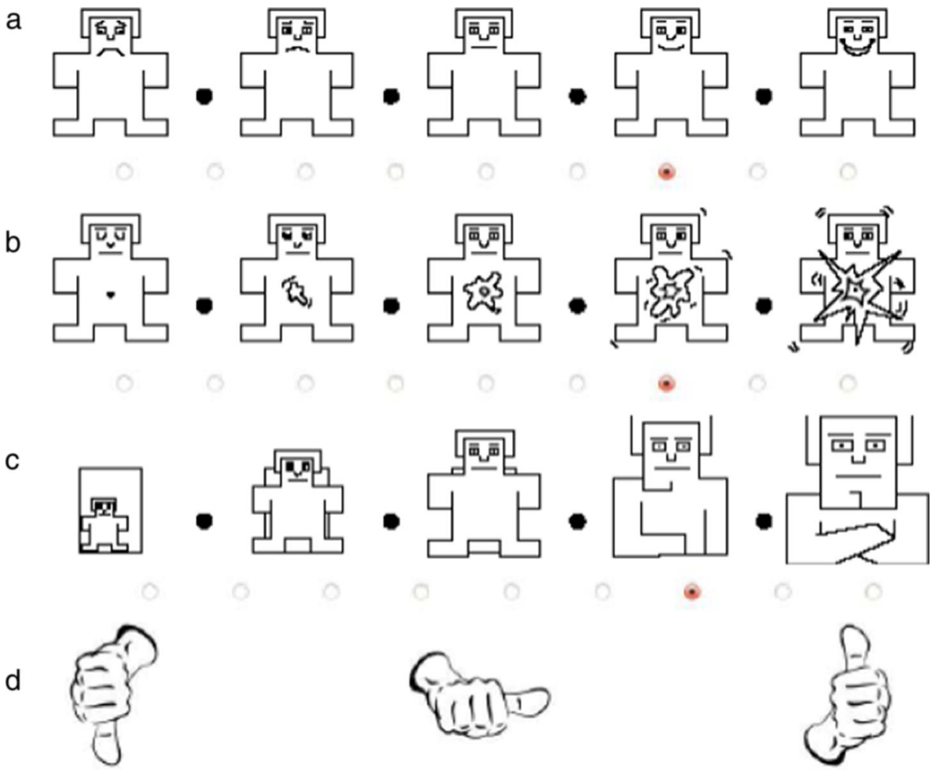


Fig. 3 Self-Assessment Manikins (SAM) method to record emotion states, **a** Valence **b** Arousal **c** Dominance and **d** Liking [10]

Table 1 An overview of the systems for emotion recognition in terms of modalities used, database employed, feature extracted and dimensional recognized

System	Modality/cue	Database	Number of samples	Features	Dimensions
S. Koelstra and I. Patras (2013)	Visual and physiological signals	MAHNOB HCI dataset	24 subjects out of 30	AUs for visual features/Power spectral density (PSD) in the Theta, slow alpha, alpha, beta and gamma bands for EEG Features	Arousal, valence, control
M. A. Nicolaou et al. (2011)	audiovisual	SAL database	2 male and 2 female subjects	MFCC for audio, tracking 20 facial feature point (FFP) and 4 shoulder point for video	Arousal-valence
Y. Wang et al. (2012)	audiovisual	RML/eNTERFACE	400 video clips for RML (total from 8 subjects)/10 subjects out of 44 for eNTERFACE database	Kernel canonical correlation analysis (KCCA)/Kernel cross-modal factor analysis (KCFA)	6 emotions (Anger, disgust, fear, happiness, sadness and surprise)
M. Paleari (2010)	audiovisual	eNTERFACE	40 subjects for training and 4 subjects for testing	Various video features for video/mfcc for audio	6 emotion
M. Mansoorizadeh (2010)	audiovisual	Multi-databases	Various no. for different databases	Linear discrimination analysis(LDA)	6 emotion
D.Datcu et al. (2009)	audiovisual	Multi-databases	Not available	FACs/mfcc	6 emotion
B. Schuller et al. (2009)	audio	Multi-databases	Various no. for different databases	MFCC	Arousal-valence

Table 2 Systems shown in Table 1 with classification methods, fusion and results

System	Classification	Explicit fusion	Results
S. Koelstra and I. Patras	binary classification on the arousal, valence and control ratings	Feature level and decision level fusion	For arousal, valence, and control, video tag classification rates of 80.0 %, 80 % and 86.7 % are attained respectively when aggregating across all 24 participants.
M. A. Nicolaou et al. (2011)	The bidirectional Long Short-Term Memory neural networks (BLSTM-NNs), and Support Vector Machines for Regression (SVR)	Feature level, model level, and output associative fusion	In terms of RMSE (0.141) and correlation (0.84), the inter-coder SAGR (0.86)
Y. Wang et al. (2012)	HMM	Feature level and score level fusion	RML dataset result 82.22 % and eNTERFACE dataset result 72.47 %
M.Paleari (2010)	Neural Network	Feature level fusion	73 %
M. Mansoorizadeh (2010)	SVM (10-foldscross validation, that is, 90 % of the data is used for training and the remaining 10%is used for testing).	Feature level, Decision level, Optimal decision level, Optimal feature level fusions	77 % for TMO-EMODB and 71 % for eNTERFACE
D. Datcu et al. (2009)	2-fold Cross validation method for testing the performance of the models.	High level data fusion	80.02 % on still pictures and 85 % on a sequence of frames.
B. Schuller et al. (2009)	SVM HMM/GMM	Supra-segmental modelling	80.2 % for SVM and 80.5 % for HMM/GMM

Y. Wang et al. [26] investigated multimodal information extraction and analysis based on kernel method. They utilized Kernel cross-modal factor analysis for modeling the nonlinear relationship between two multidimensional variables. They have also introduced an approach to identify optimal transformations to represent patterns. M. Mansoorizadeh [12] proposed an asynchronous feature level fusion approach to create unified hybrid feature space for clustering and classification of basic affective states from speech and facial expressions. They claimed that the proposed fusion approach performs better than unimodal face and speech based system. They have also provided comparative results based on synchronous feature level and decision level fusion approaches. B. Schuller et al. [19] compared the performance of nine standard corpora using modeling on a frame-level by means of Hidden Markov Models (HMM) and supra-segmental modeling by systematic feature brute-forcing for emotion recognition. They cluster each database's emotion into binary valence and arousal to provide better comparability among different datasets. They claimed that supra-segmental modeling performed better.

3 Proposed 3D emotion model in VAD space

We propose a 3D emotion model based on three continuous emotion dimensions (valence, arousal and dominance) by considering a large number of emotions (fun, happy, joy, cheerful, melancholy, depressing, terrible, exiting, love, lovely, sentimental, sad, mellow, shock, hate etc.). The DEAP database used for this work provides the values of valence, arousal and dominance for large number of emotions. There is no prior knowledge about their relative locations in three dimensional VAD space. In order to evaluate the model, it is necessary to establish the region in VAD space where each emotion can be considered to be correctly located and their relative positions and distances. For this purpose, a total of 1280 (40 trials for each 32 subjects) VAD values are represented in 3D space as shown in Fig. 4. Each emotion was placed in the valence-arousal-dominance coordinate system.

3.1 Experiment 1: Emotion representation in VAD space

To validate our model we have plotted all 1280 emotion-points in VAD space as shown in Fig. 4. All the values of valence, arousal and dominance are continuous and on a continuous scale of 0–9. In Fig. 4, it can be seen that the most of the 'valence' values are distributed from 2.0 onwards. However, the values of dominance are distributed throughout the range from 1 to 9. The average standard deviation (for all emotions) for valence, arousal and dominance are 1.5511, 1.8367 and 1.8558 respectively. Mean and standard deviation for different emotions are given in Table 3.

As evident from Fig. 4, all emotions are continuously distributed in VAD space and although the standard deviation (see Table 3) of any particular emotion is not more than 2.51, values of V, A and D for various instances of any one emotion can vary a lot. One reason for this can be attributed to the presence of noise in each measurement. Therefore, in experiment 2, we have analyzed the centroids of each emotion, which indicates a definite pattern. But, such a big and continuous spread of every emotion cannot be only due to noisy measurement. It definitely proves the continuous distribution of each emotion in VAD space. Any emotion (say Happy) can have different values of Valence, Arousal and Dominance depending on different stimulation effects. Hence, at different instances, a subject can be Happy in one instance and can be more Happy at the other instance. Our representation model can accommodate all these variations as all the three axes are continuous.

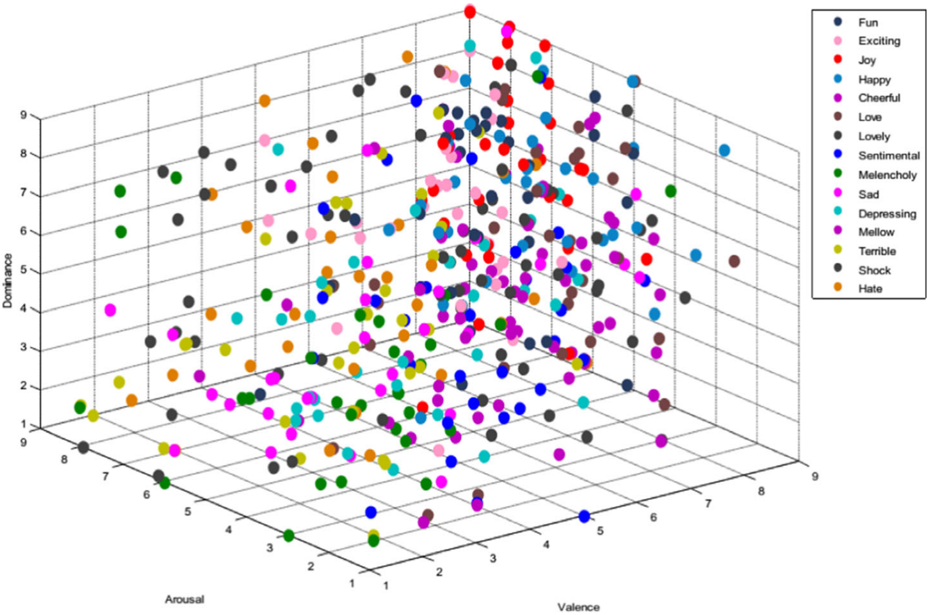


Fig. 4 Emotion distribution in 3D space

Table 3 Mean and standard deviations (SD) of different emotions on VAD dimensions

Emotions	Valence		Arousal		Dominance	
	Mean	SD	Mean	SD	Mean	SD
Fun	6.8571	1.3015	5.8571	2.1993	6.0	1.5584
Exciting	5.9286	2.0516	6.9286	1.9808	5.5	2.4128
Happy	7.1429	1.1867	4.8571	1.4569	5.2143	1.319
Joy	6.9333	2.3228	6.4667	1.9276	5.8	2.0067
Cheerful	5.9286	1.7914	3.3571	1.342	4.9286	1.7914
Love	6.5714	1.3997	4.2143	2.5122	5.4286	2.0603
Lovely	6.4667	1.3597	4.0	1.7889	4.9333	1.9137
Sentimental	4.2	1.4236	3.7333	1.8062	3.9333	1.9482
Melancholy	3.3333	1.1926	4.4667	1.9956	3.2	1.376
Sad	3.3333	1.3499	2.9333	1.6918	4.6667	2.1499
Depressing	4.2	1.6411	3.6	1.2	4.6	1.8184
Mellow	4.2	1.7963	3.0	1.5055	3.3333	1.7764
Terrible	3.6667	1.4907	5.4667	2.0613	4.6	1.7436
Shock	4.6667	1.4907	6.4	1.9253	4.9333	1.5691
Hate	3.9333	2.0483	6.1333	2.0934	5.5333	1.8927

3.2 Experiment 2: Clustering and the relative distance pattern of the centroids of emotions

After representing emotions in VAD space, we calculated the emotion centroid of each emotion (considering all instances except very few outliers) and calculated the Euclidean distances of each emotion from the other as reported in Table 4. Although, there are several ways to compute the cluster of a set of emotions, we chose K-means clustering (K = 1) for this purpose. We have removed some of the outliers; a point is considered an outlier if its coordinate values are too high or too low to fall in any cluster. Table 4 shows the maximum distance of 5.463 between Joy and Melancholy and the minimum distance of 0.356 between Mellow and Sentimental. Further, we have plotted a graph of 15 emotions as nodes and the distances as edges in Fig. 6. Firstly, we have only shown the distances up to 25 % of the maximum value as solid lines. This emotion network (or graph) interestingly groups 15 emotions in 5 groups, which are exactly equal to the number of clusters as shown in Fig. 5. Interestingly these five clusters are: C1: Happy, Joy, Fun, Exciting (Happy Group); C2: Love, Cheerful, Lovely (Love Group); C3: Depressing, Sentimental, Mellows (Sentimental Group); C4: Sad, Melancholy, Terrible (Sad Group); and C5: Shock, Hate (Hate Group).

In Fig. 5, it can be seen that the cluster1 of happy group of emotions is associated with high valence and high dominance. This is in line with some studies, e.g. [17], where happy group of emotions are classified in positive valence and positive dominance space. The emotions within a cluster are highly associated with each other. The cluster 2, interestingly related to low dominance and high valence emotions, consists of Love, Lovely and Cheerful emotions. There are debates on whether to include love in the list of emotions or not, but we have considered whatever was provided in DEAP dataset. Generally Cheerful can be included in happy group (Cluster 1), but here it is with love group (Cluster 2), which could also be because of the

Table 4 Euclidean distances among different emotions

	Fun	Exciting	Joy	Happy	Cheerful	Love	Lovely	Sentimental	Melancholy	Sad	Depressing	Mellow	Terrible	Shock	Hate
Fun	0	0.7191	1.0142	0.7742	1.8702	1.3909	2.1563	2.9577	4.4610	3.2831	2.8988	3.1423	3.5375	3.0608	2.9264
Exciting	0.7191	0	1.2609	1.4385	2.3689	1.9499	2.5948	3.1865	4.3755	3.1563	2.9210	3.14164	3.2796	2.5661	2.7324
Joy	1.0142	1.2609	0	0.8810	2.7356	2.2068	3.0495	3.9483	5.4634	4.2711	3.9115	4.1179	4.4736	3.7932	3.8685
Happy	0.7742	1.4385	0.8810	0	1.9523	1.4321	2.292	3.2908	5.0170	3.8882	3.4161	3.4211	4.2083	3.8155	3.5738
Cheerful	1.8702	2.3689	2.7356	1.9523	0	0.5701	0.3848	1.4779	3.5073	2.6247	1.9720	1.5481	3.1630	3.5444	2.6463
Love	1.3909	1.9499	2.2068	1.4321	0.5701	0	0.8650	1.9489	3.8735	2.8736	2.2932	2.0722	3.3294	3.5224	2.8273
Lovely	2.1563	2.5948	3.0495	2.292	3.3848	0.8650	0	1.1830	3.2477	2.4260	1.7745	1.2720	2.9815	3.5581	2.5830
Sentimental	2.9577	3.1865	3.9483	3.2908	1.4779	1.9489	1.1830	0	2.1201	1.5678	0.9113	0.3569	2.2210	3.2038	1.9650
Melancholy	4.4610	4.3755	5.4634	5.0170	3.5073	3.8735	3.2477	2.1201	0	1.2332	1.6089	2.2361	1.4507	2.9397	1.9078
Sad	3.2831	3.1563	4.2711	3.8882	2.6247	2.8736	2.4260	1.5678	1.2332	0	0.7256	1.8290	0.6870	2.0379	1.0018
Depressing	2.8988	2.9210	3.9115	3.4161	1.9720	2.2932	1.7745	0.9113	1.6089	0.7256	0	1.1674	1.3879	2.3567	1.1147
Mellow	3.1423	3.14164	4.1179	3.4211	1.5481	2.0722	1.2720	0.3569	2.2361	1.8290	1.1674	0	2.5016	3.4479	2.1616
Terrible	3.5375	3.2796	4.4736	4.2083	3.1630	3.3294	2.9815	2.2210	1.4507	0.6870	1.3879	2.5016	0	1.7853	1.2508
Shock	3.0608	2.5661	3.7932	3.8155	3.5444	3.5224	3.5581	3.2038	2.9397	2.0379	2.3567	3.4479	1.7853	0	1.3747
Hate	2.9264	2.7324	3.8685	3.5738	2.6463	2.8273	2.5830	1.9650	1.9078	1.0018	1.1147	2.1616	1.2508	1.3747	0

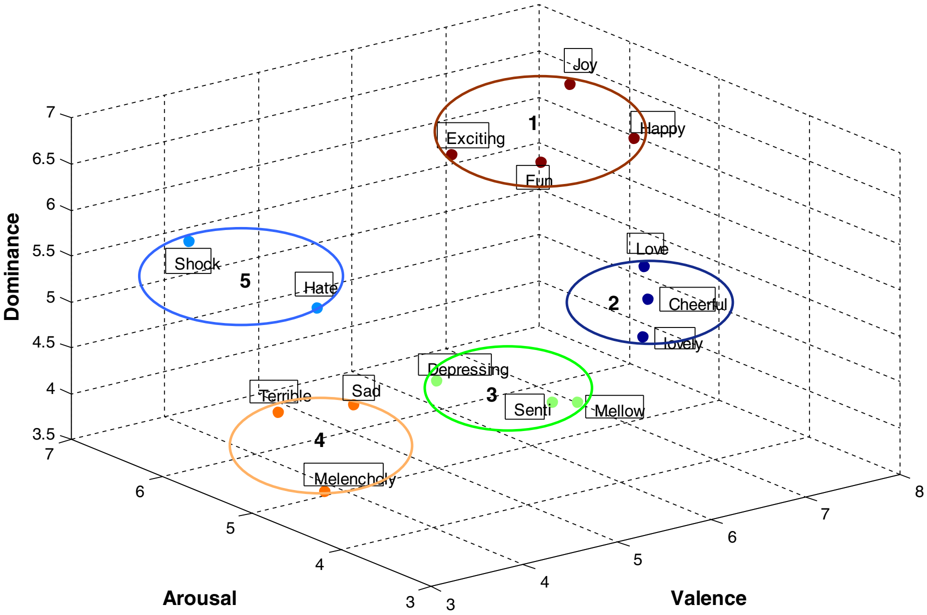


Fig. 5 Clusters of emotions in 3D space

variations in labeling of emotions in various cultures. But this also indicates the problem of discrete linguistic labelling. More Cheerful (having more dominance) may change its cluster to happy group. Similar comments can be made about odds in cluster C3, C4 and C5, where Depressing should be transferred from C3 to C4 and Terrible from C4 to C5. To clarify this and visualize the relative distances of each emotion from the other, we have calculated Euclidean distances among centroids of each emotion (Table 4) and shown the 'Emotion-Graph' in Fig. 6 in the proposed VAD space. In 'Emotion-Graph' each centroid of emotion is shown as a node (15 nodes for 15 emotion centroids) and the nearest distance node is connected through bold lines and next two nearest nodes are connected through dotted and dashed lines respectively. As can be seen in Table 4 and Fig. 6, Depressing is nearest to Sad and hence must be included in C4 (Sad group). Although Terrible is nearest to Sad but the second nearest to Hate and hence with some compensation in error it can be transferred from C4 (Sad group) to C5 (Hate group). This new grouping was verified linguistically in terms of synonyms and literary use of emotion words. Hence, the new grouping of above 15 emotions should be:

- C1: Happy, Joy, Fun, Exciting, Cheerful (Happy Group);
- C2: Love, Lovely (Love Group);
- C3: Sentimental, Mellow (Sentimental Group);
- C4: Sad, Depressing, Melancholy, (Sad Group) and
- C5: Shock, Hate, Terrible (Hate Group).

The blank space in the graph shows the possibility of many other emotions to be accommodated in emotion graph. The nearby emotions (in clusters) are related to each other and their relative positions in VAD space can be viewed as some measure of relatedness with each other.

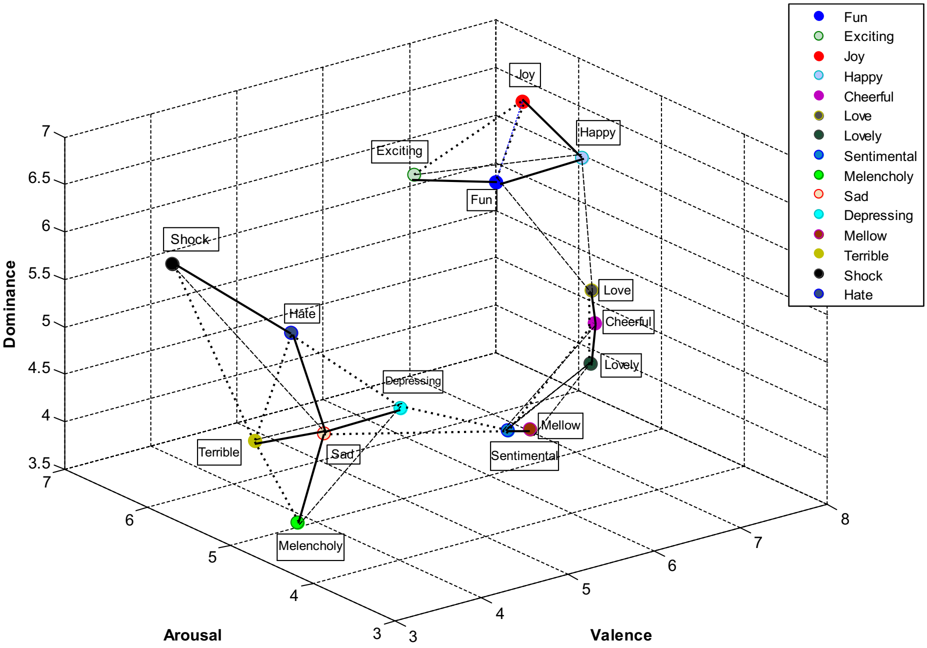


Fig. 6 Emotion graph in VAD space

The following are the findings from the Emotion Graph as depicted in Fig. 6.

- It can be concluded that the emotions are completely represented in three dimensional space and each emotion is a combination of the quantities represented in three dimensions i.e. valence, arousal and dominance (rather than a single value as in the case of discrete emotion model).
- The valence and arousal are relatively high in happy group of emotions (Joy, Exciting, Happy and Fun), but, valence is low in sad group (Sad, Depressing and Melancholy). This is what represents positive and negative emotions. If we have a perpendicular plane at valence 6.5, all emotions on the left side are called negative emotions while on the right side are called positive emotions.
- This emotion graph validates the existing emotion theory where happy group of emotions is far away from sad group of emotions.

The above findings also demonstrate to some extent the sufficiency of the three continuous dimensions, namely Valence, Arousal and Dominance (Although in DEAP database the values of Liking and Familiarity are also given, but we find that they do not provide any useful information about emotion). As per the cognitive appraisal theory, emotions are responses to the cognitive appraisal of the abnormal situations. Smith C.A. et al. [21] has pointed out eight dimensions of cognitive appraisal. These dimensions include Pleasantness, Attentional Activity, Control, Certainty, Goal-Path Obstacle, Legitimacy, Responsibility and Anticipated Effort. Out of these, Goal Path Obstacle and Legitimacy has not been considered by many other researchers. Although, no one–to–one mapping can be done from remaining six dimensions to VAD space, but we are of the view that the dimension of Pleasantness and Anticipated

Effort are in some way included in Arousal, while Control and Responsibility gets included in Dominance and other dimensions like Attentional Activity and Certainty are represented by Valence. Hence, we think that Valence, Arousal and Dominance continuous dimensions are sufficient to uniquely represent an emotion.

4 Model validation through emotion prediction and multimodal emotion recognition

In this section, we have validated the proposed 3D emotion model through emotion prediction and multimodal emotion recognition. The multimodal cues used in this study are EEG and visual cues.

Electroencephalogram (EEG) measures voltage fluctuations resulting from ionic current flows within the neurons of the brain. EEG records the brain's spontaneous electrical activity over a short period of time, from multiple electrodes placed on the scalp. Whereas, visual cue includes video frames obtained from different subjects of DEAP database. We have considered only basic emotions (Ekman's emotion) due to the limitation in the emotion content of videos in DEAP database.

4.1 Experiment 3: multimodal affect-group recognition framework

We have proposed a multimodal affect-group recognition framework in this section which was used to validate emotion grouping in VAD space. Multimodal signals used in experiments are video and EEG signals. The visual signals include facial video frames of different subjects, extracted from DEAP database. In DEAP database, the spontaneous responses of participants, while watching music videos, were recorded with resolution of 720 × 576 pixels at 50 frames per seconds. We have extracted the video frames after selecting and grouping the video into three emotion categories i.e. happy, sad and surprise. We have considered only three basic emotions (Ekman's emotion) due to the limited emotion content of videos in DEAP database. Although 32 subjects participated in the experiments, the face videos of 22 subjects are available in the database. Hence, we have used the facial expressions of 22 participants only in this study.

The affect-group recognition framework consists of the three major steps: 1) VAD data processing 2) Multimodal data processing and 3) Affect-group prediction. First the Valence, Arousal and Dominance values were processed to create ground truth data in VAD space. Segmentation was performed on VAD data to segment V-A-D values into different

Table 5 Valence, arousal and dominance class with range

Valence class (range)	Arousal class (range)	Dominance class (range)
Low Valence (1–4.5)	Low Arousal(1–4.5)	Low Dominance (1–4.5)
Medium Valence (4.5–5.5)	Medium Arousal (4.5–5.5)	Medium Dominance (4.5–5.5)
High Valence (5.5–9)	High Arousal (5.5–9)	High Dominance (5.5–9)

class as shown in Table 5. Multimodal data processing includes pre-processing, feature extraction and feature normalization of multimodal cues (visual and physiological). Different pre-processing approaches have been employed for each modality, as described in detail under section of pre-processing of EEG and video cues. An architecture of affect-group recognition framework is illustrated in Fig. 7.

4.1.1 Emotion Grouping in VAD Space

Valence, Arousal and Dominance (VAD) data processing includes annotation of emotions in three categories based on low, medium and high Valence, Arousal and Dominance values separately as given in DEAP database. The ground truth data of DEAP database is the ratings given by each participant in the continuous range of [0, 9] individually in terms of V, A and D. The ratings for each of these scales are threshold into three classes (low, medium and high) as shown in Table 5. This procedure segments multimodal data in terms of low, medium and high categories.

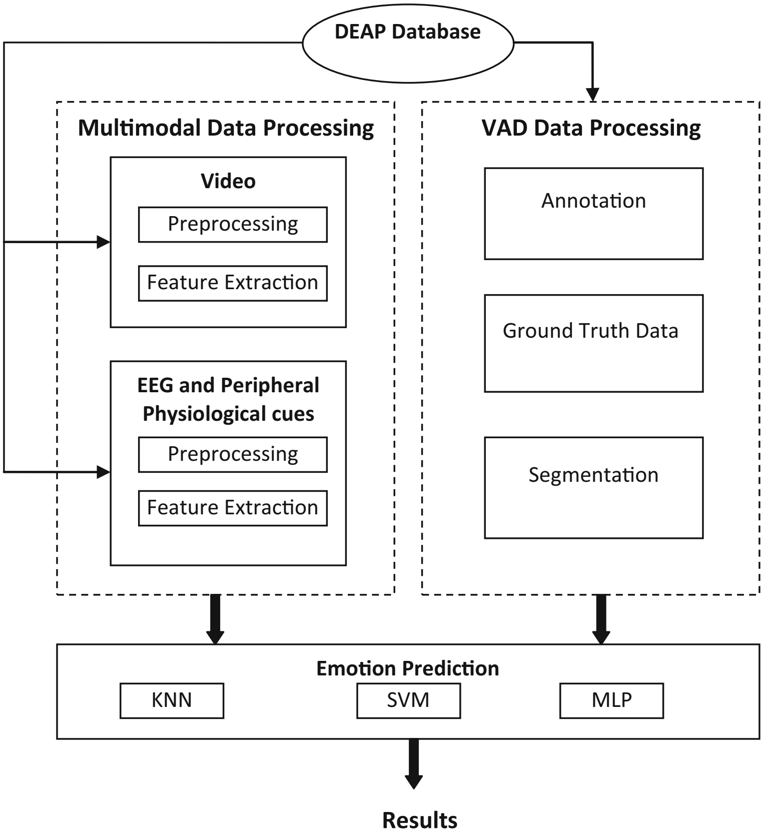


Fig. 7 Proposed affect prediction framework

4.1.2 Multimodal data processing and Feature Extraction

Multimodal data processing involves processing of video, EEG and peripheral physiological signals. For the video cue, the pre-processing involves to detect the frontal face from a video frame. We have used face detection utilities of Viola Jones [25] which is based on cascade classifier. To extract the features from frontal face, Discrete Wavelet Transform (DWT) was applied with "daubechies" wavelet family to extract wavelet features. Similarly, to capture the spontaneous emotion from EEG and peripheral physiological cues, we have also applied DWT. Followed by feature extraction, min- max normalization was performed to normalize features obtained from visual, EEG and peripheral physiological signals.

Pre-processing of visual cues As the emotion content in videos of DEAP database is low, we have considered only three emotion category (i.e. happy, sad and terrible) for affect recognition. The facial expression samples are collected from videos of different subjects having high content of facial emotion. The videos selection is based on the content of emotion manually observed by us. As the emotion frames contains various background objects, which are not significant in emotion recognition, therefore it is essential to extract the frontal face from video frames. To detect the face from the video frames, we have applied face detection utilities of Viola Jones [25] which is based on cascade classifier. The DEAP dataset contains the frontal facial image of different subjects However; there is some other background objects appearing in the video. Therefore, we have used face detection utilities of Viola Jones to extract the frontal face. The size of the extracted frame is 720 × 576.

Pre-processing of EEG signal Electroencephalogram signals in DEAP dataset were recorded with a 1024 Hz sampling rate and later down sampled to 256 Hz to reduce the memory and processing costs. EEG signals were recorded using active AgCl electrodes placed according to the international 10–20 system. The unwanted artefacts, trend and noise were reduced prior to extracting the features from EEG data by pre-processing the signals. Drift and noise reduction were done by applying a 4-45 Hz band-pass filter and eye artefacts were removed with a blind source separation technique [10].

Multimodal signal analysis Most of the emotion theory reveals that physiological features are important for emotion. P. Ekman et al. [5] demonstrated that physiological pattern associated with a particular emotion. Multiresolution methods such as wavelet transforms are rated as potent for feature extraction during the near past in various applications [16, 23, 29]. Hence, we have used multiresolution analysis (MRA) to analyze visual and physiological signals in this study. EEG can be described by frequency and amplitude. The following frequency bands are includes in EEG signal [22].

- Delta: 1–4 Hz.
- Theta: 4–8 Hz.
- Alpha: 8–12.5 Hz.
- Beta: 12.5–28 Hz.
- Gamma: 30–40 Hz.

Feature extraction Discrete wavelet transform used with different wavelet families to extract the approximation and detail coefficients from physiological signals**.** Power spectral features, logarithms of the spectral power from all EEG bands were extracted from approximation and detail coefficient as shown in Table 6**.**

4.1.3 Matching and Affect Prediction

In the final stage, we predict emotions by classification of multimodal (EEG and Video) features using three different well known classifiers namely Multilayer Perceptron (MLP), Support Vector Machine (SVM) and K-Nearest Neighbor (K-NN). A brief description of above classifiers is as under.

A **Multilayer Perceptron (MLP)** is a feed forward artificial neural network model that maps sets of input data onto a set of appropriate outputs. MLP utilizes a supervised learning technique called back propagation for training the network. Classification is starts by assigning input nodes with extracted EEG outcome, $\{x_1, x_2, \ldots x_n\}$ from the proposed technique which then propagated in a forward direction through the perceptron until the output nodes. The network is trained with the back propagation learning algorithm. The learning algorithm adapts the weights; w_n and v_n based on minimizing the error between given output and desired output. The two main activation functions used in current applications are both sigmoid, and are described by Eq. 1.

$$y(v_i) = tanh(v_i) \text{ and } y(v_i) = (1 + e^{-v_i})^{-1}) \tag{1}$$

In Eq. 1, the former function is a hyperbolic tangent which ranges from −1 to 1, and the latter, the logistic function, is similar in shape but ranges from 0 to 1. Here y_i is the output of the i^{th}node (neuron) and v_i is the weighted sum of the input synapses.

***k*- Nearest Neighbor algorithm** assumes that all the data are in feature space and each training data has a feature vector and class label associated with it. In K-NN algorithm the number k decides how many neighbors will influence the classification. If k = 1 then the K-NN algorithm is called the nearest neighbor algorithm. It simply assign the data point to class which has nearest distance from the class by using any distance function like Euclidean distance, Manhattan distance, Minkowski distance or Cityblock distance. Let an arbitrary n-dimensional feature vector=$[x_1, x_2, \ldots x_n]$. Then

Table 6 Video, EEG signal and extracted features

Signal	Extracted features
Video	Standard deviation, mean, entropy of each level and ratio between them.
EEG	Relative Power Energy (RPE):Four band of Delta, Theta, Alpha, Beta and Gamma Logarithmic Relative Power Energy (LRPE) : Four band of Delta, Theta, Alpha, Beta and Gamma Absolute Logarithmic Relative Power energy (ALRPE) : Four band of Delta, Theta, Alpha, Beta and Gamma, Standard deviation all levels of detail coefficients and highest level approximation coefficient Entropy: all levels of detail coefficients and highest level approximation coefficient.

the distance between two feature vectors $X=[x_1,x_2,\ldots x_n]$ and $Y=[y_1,y_2,\ldots y_n]$ can be defined in terms of Eucledean distance given in Eq. 2.

$$dis(X,Y)=\sqrt{\sum_{i=1}^{n}(x_i-y_i)^2} \tag{2}$$

Support Vector Machine (SVM) is a method of classification that uses a supervised learning algorithm to classify data into different classes. The main goal of support vector machine is to design a hyper plane that classifies all training vectors into two classes. A hyper plane is a decision boundary that is of n-1 dimension if the data points are of n dimension. There are many hyper planes for classifying the data, but the best one is that which has a maximum margin from the both classes of training vector. The data points that are close to the hyper plane are called support vector. The movement of support vector causes movement of decision boundary. The equation of the support vector machine classifier is given as Eq. 3.

$$g(x)=w^Tx+b \tag{3}$$

$$\mathrm{g(x)}\geq 1,\forall x\epsilon Class1$$

$$\mathrm{g(x)}\leq 1,\forall x\epsilon Class2$$

Where w is a vector normal to the hyper plane, b is the bias and $g(x)$ is classifier.

5 Results and discussion

All the experiments were carried out on a 64-bit Intel i5 processor (2.40 GHz) with 4 GB RAM. The DEAP database is being used in all the experiments. 40 channel EEG signals with 8056 data with 40 trials each for 32 participants were used in experiments. As the sampling rate is 128 Hz. We have decomposed the physiological signals up to five levels in order to extract detail information from physiological signals. Daubechies wavelet transform (‘db6’) was used to extract the approximation and detail coefficients from physiological signals. Min-max normalization has been applied to normalize feature vectors to the range [0, 1] prior to classification.

A leave-one-out 10 fold cross validation approach has been employed to validate the user independent classification performance. At each step of cross validation, the samples of one participant were taken out as test set and the classifier was trained on the samples from the rest of the participants. In order to get the best performance, various parameters of machine learning methods are configured. For SVM, we got the best accuracy for C = 200 and ε = 0.01 with RBF kernel. Similarly, the MLP network has two hidden layer and learning rate is 0.5. For each video from the dataset, the ground truth data was prepared by the ratings (in terms of valence, arousal and dominance) given by each participants individually. The results are given in terms of classification rate, F1 measure and Receiver Operating Characteristics (ROC) curve as given in Eqs. 4, 5 and 6 respectively.

$$\text{Classification Rate (CR)} = (\text{TP} \times 100)/\text{N} \tag{4}$$

Where TP = number of correct classified instances and N = Total number of instances.

$$\text{F1} = 2 \times (\text{precision.recall})/(\text{precision} + \text{recall}) \tag{5}$$

$$\text{ROC} = \text{False Positive}/\text{True Positive} \tag{6}$$

The Confusion Matrix for 7200 instances of various emotion groups with low, high and mid V, A and D and the predicted emotion groups from EEG and Video Signals are shown in Table 7. Table 7 validates the continuous V, A and D representation model of emotions through EEG signals of the same subjects. Features extracted from EEG signals as described in Table 6, were classified (using K-NN classifier) into same emotion groups (low, high and mid) V, A and D as shown in Table 5. Table 7 shows that the prediction is correct for low and high V (51 % and 63 %), A (62 % and 65 %) and D (57 % and 61 %) emotion-groups considering only one dimension (V, A or D) at a time. However, the results for mid (range 4.5–5.5) V, A or D valued emotion group is between 34 % to 44 %. This is quite obvious as the mid range is only 1/9 of the total scale and it can be debated that they represent no emotion or neutral emotion. In any case the confusion is very high in the neutral cases (can be mistakenly considered as either towards lower or higher side). It is interesting to note that these predictions of emotion groups are based on either Valence or Arousal or Dominance only. Even then the matching through EEG signal is of the order of more than 60 % (except) in case of low V (51 %) and low D (57 %). This also proves the significance of each axis of representation, i.e. V, A and D.

The results given in Table 8 are based on evaluation matrices: CR (Classification rate), Average F1-measure and ROC area as defined above. A ROC curve for valence, arousal and dominance is given in Fig. 8. The highest classification rate (CR) achieved for EEG data applying various classification techniques (MLP, SVM, K-NN) are 67.5 %, 69.6 % and 65.1 % for V, A and D respectively as shown in Table 8 and Fig. 9 (for MLP). Corresponding ROC curve is also mentioned in Fig. 8. The very high accuracy matching from video data is not a correct representation as only three emotions (happy, sad and terrible) could be selected from video data of emotions in DEAP dataset and hence have little ambiguity. Also, the video frames are extracted from manually selected videos with rich emotion content.

It is observed that our multiresolution analysis is efficient in discrimination of emotion groups based on low, mid (medium) and high values of Valence, Arousal and Dominance separately. We have evaluated our model of three dimensional continuous VAD (valence, arousal and dominance) space of emotion representation through measured EEG data of the subjects and the experiments and results validate our claims of sufficiency of three continuous dimensions in representing complex emotions to some extent.

Experiment 1 clearly establishes that emotions can be represented on a continuous scale, rather than a discrete one. The large value of standard deviation for each emotion indicates the variation in each instance which can be accommodated on continuously varying axes only. Interestingly, the average standard deviation of Valence, Arousal and Dominance for the 15

Table 7 Confusion matrices of classification of EEG and VIDEO (column: classified label; row: ground truth)

Total no. of instances =7200 (low = 2624, Mid = 1120, High = 3456)				
Valence	No. of Instances (%)	Low	Mid	High
	Low	1345 (51.26)	327 (12.46)	952 (36.28)
	Mid	397 (35.45)	376 (33.57)	347 (30.98)
	High	958 (27.72)	327 (9.46)	2171 (62.82)
	(a) EEG			
	%	Low	Mid	High
	Low	99.234	0.316	0.125
	Mid	0.212	98.235	0.0377
	High	0.553	1.449	99.837
	(b) Video			
Total no. of instances =7200 (low = 3232, Mid = 736, High = 3232)				
Arousal	No. of Instances (%)	Low	Mid	High
	Low	2011 (62.22)	209 (6.47)	1012 (30.31)
	Mid	248 (33.70)	251 (34.10)	237 (32.20)
	High	925 (28.62)	193 (5.97)	2114 (65.41)
	(c) EEG			
	%	Low	Mid	High
	Low	98.487	0.391	0.061
	Mid	0.352	98.195	0.091
	High	1.160	1.414	99.848
	(d) Video			
Total no. of instances =7200 (low = 2784, Mid = 1536, High = 2880)				
Dominace	No. of Instances (%)	Low	Mid	High
	Low	1580 (56.75)	409 (14.69)	795 (28.56)
	Mid	462 (30.08)	674 (43.88)	400 (26.04)
	High	738 (25.63)	372 (12.92)	1770 (61.45)
	(e) EEG			
	%	Low	Mid	High
	Low	99.170	0.309	0.092
	Mid	0.331	99.358	0.134
	High	0.497	0.332	99.775
	(f) Video			

Table 8 Single cue prediction results for Valence, Arousal and Dominance Dimensions for different modalities. (CR-Classification Rate, F1- Average F1 measure, ROC- ROC Area)

		MLP			SVM			K-NN		
		CR	F1	ROC	CR	F1	ROC	CR	F1	ROC
Valence	EEG	63.47	0.62	0.739	56.34	0.536	0.605	67.51	0.675	0.725
	Video	98.43	0.984	0.997	96.20	0.962	0.965	98.46	0.985	0.985
Arousal	EEG	69.62	0.696	0.696	52.79	0.528	0.454	68.55	0.685	0.736
	Video	99.43	0.994	0.999	99.63	0.996	0.997	99.30	0.993	0.994
Dominance	EEG	63.57	0.627	0.730	57.71	0.557	0.635	65.10	0.651	0.675
	Video	98.53	0.985	0.999	97.06	0.971	0.975	98.36	0.984	0.986

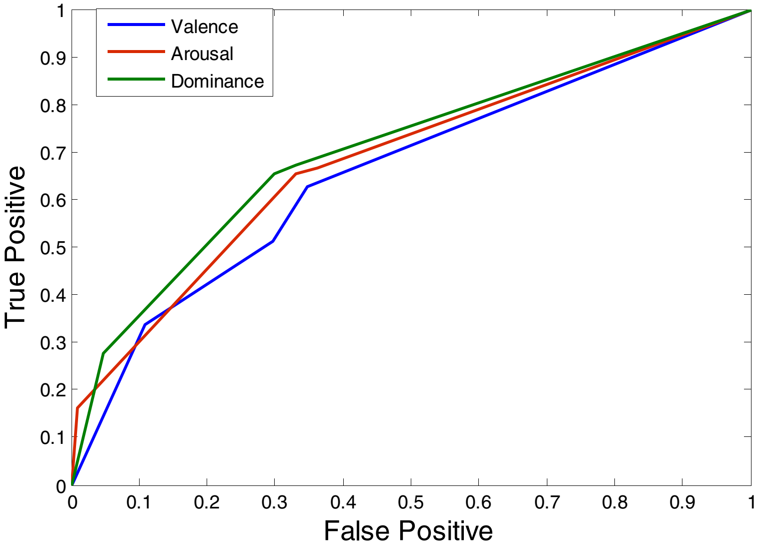


Fig. 8 ROC curve for valence, arousal and dominance

emotions of DEAP dataset is approximately 1.5, which is 1/6th of the total scale (0–9). This indicates the linguistic labelling difficulty for any instance of stimulated emotion, whereas the measured values of physiological signals (e.g. EEG) are generally continuous. Experiment 2 represents the relative distances of emotions from each other in VAD space, where the 'nearness' (Euclidean distance in VAD space) is equated with 'relatedness' property. Within the limited scope of DEAP dataset where sufficient VAD and data for only 15 emotions are available, the relatedness property is established except few exceptions. This also highlights the composition of complex emotions, which can be considered as the mixture of nearby emotions. Experiment 3 validates the VAD model through measured EEG dataset. The emotion group predicted by VAD model more or less, matches with that predicted by measured EEG data of subjects.

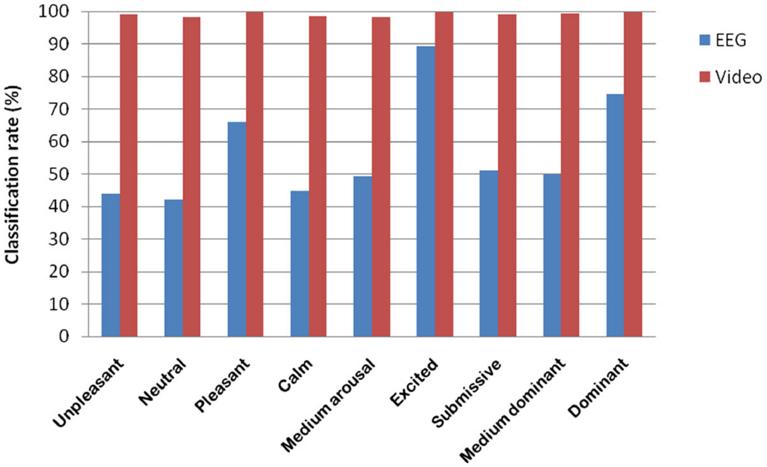


Fig. 9 Classification results (using MLP) of each class for different modalities (EEG Vs. VIDEO)

Table 9 Accuracy comparison with other studies in terms of VAD (for DEAP database)

Author(s)	Year	Approach	Classification Accuracy (%)
Koelstra et al. [10]	2012	Power spectral features	57.6 (V) 62.0 (A)
Chung and et al. [4]	2012	power spectral features and Bays classifier	66.6 (V) for two class and 53.4 (V) for three class
Yoon H. J. and Chung S. Y. [30]	2013	FFT and classifier	70.9 (V) 70.1(A) for two class, 55.4(V), 55.2 (A) for three class
Our method	2015	Multiresolution analysis and MLP	63.47 (V), 69.62 (A), 63.57 (D) for three class

Previous studies [3, 10, 30] using DEAP database are shown in Table 9. Koelstra [10] proposed a system based on power spectral features from EEG signals, Fisher criterion for feature selection and Naive Bayes classifier for classification. They achieved the average accuracies of 57.6 % and 62 % for two classes of valence and arousal using DEAP database. In another study using DEAP database, Yoon and Chung [30] designed an emotion recognition system based on Fast Fourier transform and Pearson correlation coefficient for feature selection and Bayes classifier. They obtained the average accuracies of 70.9 % and 70.1 % for two classes of valence and arousal, and 55.4 % and 55.2 % for three classes. Chung and Yoon [4] proposed an emotion recognition method using Bayes classifier based on a weighed-log-posterior probability function and power spectral features and the best accuracies obtained are 66.6 % and 53.4 % for two and three classes of valence dimension respectively. Among all the studies, feature extraction is based on power spectral features. Whereas, our method is based on MRA and we have obtained improved classification accuracy in terms of three classes of valence, arousal and dominance by at least 8 % compared to others' best classification accuracies as shown in Table 9. Furthermore, in comparison with other new studies, our proposed method has representational capacity including the possibility of complex emotion representation.

6 Conclusion and future work

The emotion recognition field has recently shifted from six basic (Ekman's emotion visible through facial expressions) discrete emotions to dimensional emotion as complex emotions cannot be measured from facial expressions only. The work presented in this study focused on

1) 3D emotion framework to represent large number of emotions in three dimensional space (Valence, Arousal, Dominance) followed by validation of the proposed model through emotion prediction and recognition from visual and physiological cues. The VAD values are further classified in high, medium (neutral) and low for ground truth data generation. The neutral emotion (no emotion) is ambiguous as with small change in VAD values it can change their class.
2) The proposed model validated through experiments conducted on a benchmark DEAP database.

3) The proposed affect model represents a large number of emotions compared to the previous studies which are based on limited number of emotions.
4) We have shown that two dimensional spaces are insufficient to represent emotions as it has many shortcomings discussed in state-of-the-art section. On the other hand, we have validated a three dimensional model which is sufficient to represent large number of emotions.

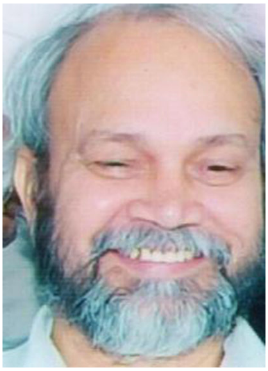
Overall, we conclude that three dimensional framework is sufficient and accurate to represent all emotions compared to two dimensional frameworks. As a future task, the proposed model remains to be evaluated with extensive dataset with large number of simple and complex emotions (and with richer emotional expressions).

References

1. Arifin S, Cheung PYK (2008) Affective level video segmentation by utilizing the pleasure-arousal-dominance information. IEEE Trans Multimed 10(7):1325–1341
2. Emotion Article: www.measuredme.com visited on 20 July, 2014.
3. Caridakis, G., Malatesta, L., Kessous, L., Amir, N., Paouzaiou, A., &Karpouzis, K. (2006, November). Modelling naturalistic affective states via facial and vocal expression recognition. In Proceedings 8th ACM International Conference on Multimodal Interfaces (ICMI'06), Banff, Alberta, Canada (pp. 146–154). ACM Publishing.
4. Chung S. Y., Yoon H. J. (2012) Affective classification using Bayesian classifier and supervised learning. 12th Int Conf Control, Autom Syst (ICCAS) Island: pp. 1768–1771.
5. Ekman P, Friesen WV, O'Sullivan M, Chan A, Diacoyanni-Tarlatzis I, Heider K, Krause R, LeCompte WA, Pitcairn T, Ricci-Bitti PE, Scherer K, Tomita M, Tzavaras A (1987) Universals and cultural differences in the judgments of facial expressions of emotion. J Pers Soc Psychol 53:12–717
6. Fragopanagos F, Taylor JG (2005) Emotion recognition in human-computer interaction. Neural Netw 18: 389–405. doi:10.1016/j.neunet.2005.03.006
7. Glowinski, D., Camurri, A., Volpe, G., Dael, N. and Scherer K (2008) Technique for automatic emotion recognition by body gesture analysis. Proc. IEEE CS Conf. Computer vision and pattern recognition workshops, pp. 1–6.
8. Gunes H. and Pantic M. (2010) Automatic measurement of affect in dimensional and continuous spaces: why, what, and how? Proc Seventh Int',l Conf Methods Tech Behav Res, pp. 122–126.
9. Gunes H, Schuller B (2012) Categorical and dimensional affect analysis in continuous input: current trends and future directions. Image Vis Comput 31(2):120–135
10. Koelstra S, Muhl C, Soleymani M, Lee JS, Yazdani A, Ebrahimi T, Pun T, Nijholt A, Patras I (2012) DEAP: a database for emotion analysis; using physiological signals. IEEE Trans Affect Comput 3(1):18–31
11. Liu Y, Sourina O (2013) Real-time fractal-based valence level recognition from EEG. Trans Comput Sci XVIII Lect Notes Comput Sci 7848:101–120
12. Mansoorizadeh M, Charkari NM (2010) Multimodal information fusion application to human emotion recognition from face and speech. Multimed Tools Appl 49(2):277–297
13. Morris JD (1995) SAM: the self-assessment manikin. An efficient cross-cultural measurement of emotion response.Jounal of. Advert Res 35(8):63–68
14. Nicolaou M, GunesHand PM (2011) Continuous prediction of spontaneous affect from multiple cues and modalities in valence-arousal space. IEEE Trans Affect Comput 2(2):92–105
15. Picard R (2003) Affective computing: challenges. Inter J Hum Comput Stud 59(1–2):55–64
16. Saha, A., Jonathan, Q. M. (2010) Facial Expression Recognition using Curvelet based local binary patterns. IEEE Int Conf Acoust Speech Signal Proc (ICASSP), pp. 2470–2473.
17. Christopher P. Said, James V. Haxby, Alexander Todorov.(2011) Brain systems for assessing the affective value of faces. Phil Trans R Soc London B Biol Sci. 2011 Jun 12:366(1571):1660–1670. doi: 10.1098/rstb.2010.0351.
18. Schachter S, Singer JE (1962) Cognitive, social and physiological determinants of emotional state. Psychol Rev 69:379–399

19. Schuller B (2009) Acoustic emotion recognition: a benchmark comparison of performances. Proc, IEEE ASRU
20. Schuller B (2011) Recognizing affect from linguistic information in 3D continuous space. IEEE Trans Affect Comput 2(4):192–205
21. Smith CA, Ellsworth PC (1985) Patterns of cognitive appraisal in emotion. J Pers Soc Psychol 48(4):813–838
22. Stickel C, Fink J, Holzinger A (2007) Enhancing universal access – EEG based learnability assessment. Lect Notes Comput Sci 4556:813–822
23. Sumana I, Islam M, Zhang DS, Lu G (2008) Content based image retrieval using curvelet transform. Proc. of IEEE International Workshop on Multimedia Signal Processing, Cairns, Queensland, Australia, pp. 11–16
24. Verma G. K. and Tiwary U. S (2014) Multimodal fusion framework: a multiresolution approach for emotion classification and recognition from physiological signals.Vol. 102, Part 1, Pages 162–172 NeuroImage. doi:10.1016/j.neuroimage.2013.11.007.
25. Viola PA, Jones MJ (2001) Rapid object detection using a boosted cascade of simple features. CVPR, Issue 1:511–518
26. Wang Y, Guan L, Venetsanopoulos A (2012) Kernel cross-modal factor analysis for information fusion with application to bimodal emotion recognition. Multimed, IEEE Trans on 14(3):597–607
27. Whissell CM (1989) The dictionary of affect in language, emotion: theory, research and experience, vol 4. Academic Press, New York
28. Wollmer M, Schuller B, Eyben F, Rigoll G (2010) Combining long short-term memory and dynamic Bayesian networks for incremental emotion-sensitive artificial listening. IEEE J Sel Top Signal Proc 4(5): 867–881
29. Wu, X., Zhao, J. (2010) Curvelet feature extraction for face recognition and facial expression recognition. Sixth Int Conf Nat Comput (ICNC), pp. 1212–1216.
30. Yoon HJ, Chung SY (2013) Eeg-based emotion estimation using bayesian weighted-log-posterior function and perceptron convergence algorithm. Comput Biol Med 43:2230–2237

Gyanendra K. Verma is currently Assistant Professor at National Institute of Technology Kurukshetra, India. He has completed his B. Tech. from Harcourt Batlar Technological Institute Kanpur, India and M. Tech. & Ph.D. from Indian Institute of Information Technology (IIITA) Allahabad, India. His all degrees are in Information Technology. He has teaching and research experience of more than 5 years in the area of Computer Science and Information Technology with special interest in Image Processing, Speech and Language Processing, Human Computer Interaction. His research work on the application of Wavelet Transform in Medical Imaging and Computer Vision problems has been cited extensively. He is the member of various professional bodies like IEEE, IAENG & IACSIT.



Uma Shanker Tiwary is currently professor at Indian Institute of Information Technology, Allahabad, India. He has completed his B. Tech. and Ph.D. in Electronics Engineering from Institute of Technology, B.H.U., Varanasi, India in 1983 and 1991 respectively. He has experience of teaching and research of more than 23 years in the area of Computer Science and Information Technology with special interest in Computer Vision, Image Processing, Speech and Language Processing, Human Computer Interaction and Information Extraction and Retrieval. He has co-authored a book on 'Natural Language Processing and Information Retrieval' (Oxford University Press, 2008) and has edited several Proceedings of the International Conferences on 'Intelligent Human Computer Interaction (Springer, 2009 and 2010)' and was publication Chair of 'Wireless Communication and Sensor Networks (IEEE Xplore, 2006, 2007 and 2008)'. His research work on the application of Wavelet Transform in Medical Imaging and Computer Vision problems and Information Retrieval has been cited extensively. He was associated with the research work in the Mechatronics Dept. of Gwangju Institute of Science and Technology, Gwangju, South Korea and with "Anglabharti' project at Dept. of Computer Science and Engg. IIT Kanpur, India. He has delivered lectures, chaired many sessions at IEEE International Conferences and visited many labs in India and abroad, including U.S., South Korea, South Africa, China, Singapore, Thailand. He is the Fellow of IETE and Senior Member of IEEE.

View publication stats