Received: 29 July 2020 | *Revised: 20 September 2020* | *Accepted: 21 September 2020*
DOI: 10.37917/ijeee.16.2.4 *Vol. 16| Issue 2| December 2020*

Open Access

Iraqi Journal for Electrical and Electronic Engineering
Original Article

Emotion Recognition Based on Mining Sub-Graphs of Facial Components

Suhaila N. Mohammed *[1,2], Alia K. Abdul Hassan [1]
[1]Department of Computer Science, University of Technology, Baghdad, Iraq
[2] Department of Computer Science, College of Science, University of Baghdad, Baghdad, Iraq

Correspondence
* Suhaila N. Mohammed
Department of Computer Science,
University of Technology, Baghdad, Iraq
Email: suhailan.mo@sc.uobaghdad.edu.iq

Abstract
Facial emotion recognition finds many real applications in the daily life like human robot interaction, eLearning, healthcare, customer services etc. The task of facial emotion recognition is not easy due to the difficulty in determining the effective feature set that can recognize the emotion conveyed within the facial expression accurately. Graph mining techniques are exploited in this paper to solve facial emotion recognition problem. After determining positions of facial landmarks in face region, twelve different graphs are constructed using four facial components to serve as a source for sub-graphs mining stage using gSpan algorithm. In each group, the discriminative set of sub-graphs are selected and fed to Deep Belief Network (DBN) for classification purpose. The results obtained from the different groups are then fused using Naïve Bayes classifier to make the final decision regards the emotion class. Different tests were performed using Surrey Audio-Visual Expressed Emotion (SAVEE) database and the achieved results showed that the system gives the desired accuracy (100%) when fusion decisions of the facial groups. The achieved result outperforms state-of-the-art results on the same database.

KEYWORDS: Deep Belief Network (DBN), emotion recognition, facial components, graph mining, gSpan algorithm, Naïve Bayes.

I. INTRODUCTION

Facial expressions are one of the fundamental data in any face-to-face communication. Thus, it is general that the exploration of facial impressions has gained a lot of interest over many decades ago with diverse applications in effective computing [1]. However, there are many challenges in facial emotion detection caused by illumination variation, subject-dependence, and head pose-changing which widely affect the performance of the facial expression recognition system [2].

Facial emotion recognition is essentially a pattern recognition problem and involves finding the optimal feature set from the facial data being analyzed [3]. Various techniques are followed by the researchers to detect the facial emotion which can be fall basically into four main categories according to the type of the extracted features which are: appearance-based features, geometric-based features, using Ekman Action Units (AUs) and using deep learning techniques [2].

The appearance-based features appear temporarily in the face during any kind of facial expression (for example, the presence of specific facial wrinkles, bulges, forefront and the texture of the facial skin in regions surrounding the mouth and eyes). Cid, Manso, and Nunez [4] suggested a framework for the identification of five different emotional expressions (happy, angry, fear, sad and neutral). The system based on applying Gabor filter to the facial image and then using dynamic Bayesian classifier. The result of this approach on the Surrey Audio-Visual Expressed Emotion (SAVEE) database was 94.6%. However, appearance-based features are not robust against different facial variations, like scaling of the face, appeared face region, and head area orientation, etc., and cannot work well with noisy images [5]. In the geometric method, the location of key facial components such as eyes, eyebrows, mouth and nose are being tracked and the geometric relationship between certain key points on the face (e.g., distances, angles, and shapes) are considered when making the decision [6]. Haq, and Jackson [7] extracted 240 visual features based on the marker locations on the face. They used Principal Component Analysis (PCA) and Linear Discriminate Analysis (LDA) for feature selection. Using SAVEE database as evaluation material, the achieved accuracy was 91 %. However; the extraction of such types of features can be considered expensive from the computation view because it requires reliable and accurate methods for facial feature tracking and detection [8]. AUs simulate facial muscle movement during facial expression. The emotional state can be determined using a combination of facial action units. However; AUs recognition is a challenging problem due to different factors such as illumination changes, pose variations or individual subject differences [9]. With the advances in neuro-based learning algorithms, deep learning algorithms found their wide applications on the facial emotion recognition task. Barros, and Wermter [10] utilized the key point's movement information with a conventional neural network for emotion recognition. The achieved result on the SAVEE database was 93.9%. Despite these advances, the selection of optimal features and parameters in deep learning remain challenging issues [11].

Graphs are becoming extremely valuable for describing complex components such as images, biological networks, chemical substances, web and XML files. Many significant characteristics of graphs can be used in describing these domains. Facial graph contains all the information about the face image which is useful for recognition. Therefore, different algorithms of graph mining can be used to analyze various characteristics of graphs to recognize facial emotions [12]. A novel method for facial emotion recognition has been proposed by Hassan, and Mohammed [13] where a turnover point was introduced by converting the unstructured representation of the facial image into a structured one (i.e., graph). The facial zone is depicted as a graph of vertices and edges. Graph mining concepts are then applied to extract the sub-graphs which are affected during each emotion expression. The final system accuracy was 90.00% using SAVEE database.

Although different works have concentrated on extracting features for the detection of facial emotions, these attributes cannot sufficiently distinguish human emotions. So, the "emotional gap" is defined as "the lack of coincidence between the extracted features, and the expected emotion status". In other words, the emotional gap essentially represents the differences between emotions and the extracted features.

This paper based on the work that proposed by [13] where graph mining technique is used to mine the facial region. However; instead of mining the graph that is built on the whole face region, different graphs are built to represent the facial components and their relations. The main objective of the proposed work is getting insight in the facial components and finding the micro changes that happen within their graphs during some emotions occur. More accurate sub-graphs will then be discovered from the local regions of the facial image. The achieved results demonstrate the effectiveness of mining facial component graphs

The structure of remaining paper as follows: Section 2 presents the theoretical background for the principles of graph mining and landmark extraction methods. In section 3, the detailed description of the proposed system is presented. Experimental results and discussions are given in Section 4. Finally, the work conclusion and possible future research directions are presented in Section 5.

II. Theoretical Background

A. Th Principles of Graph Mining

A graph is a conceptual structure made up from a collection of vertices joined by edges. Each node in the graph represents a discrete information piece, while the edge represents the relation between two vertices [14]. A database of graphs (D_n) is a set of N_g various graphs, $G_1 = (N_1, A_1) \ldots G_n = (N_n, A_n)$, where the set of nodes in the i^{th} graph is represented by N_i, and the set of edges in the i^{th} graph is represented by A_i. Each node $p \in N_i$ is linked with a label represented by $l(p)$ [15].

Graph matching is the task of determining the degree of resemblance between graphs. To verify if a coincide exists between any two graphs, it is required to create a one to one relationship between the nodes of both graphs, such that a match exists in nodes and edges labels (i.e., the difference between these two graphs should be zero) [16].Frequent sub-graphs are useful in the tasks of graph classification, analysis and clustering. Frequent Sub-graph Mining (FSM) is the task of searching for sub-structures which often exist in the dataset of graphs. FSM indicates that the queried graph occurs more than some a predetermined number of times (minimum support) [15] [16].

Graph-based Substructure PAtterN (gSpan) is one of frequent sub-graphs mining algorithms. GSpan prevents the creation of redundant sub-structures by reducing the number of supernumerary sub-graphs nominee to be tested. It based on depth first search technique, which could be described by the following principles [14]:

1) *gSpan Encoding.* It refers to an arranged set of edges, where the arrangement is defined by the way from which the edges are reached during the first depth traverse. The edge in the encoding of gSpan made up from a five-tuple consisting of the same endpoint indices (1st and 2nd elements), endpoint labels (3rd and 5th elements), and edge labels (4th element).

2) *Rightmost Expansion*. Each edge are extended via rightmost expansion in gSpan encoding, so that a new edge can only be added in the case one of the endpoints is located on the right path.

3) *Lexicographic Order*. The gSpan encodings can be ordered linearly (i.e., from the smallest to the largest). This arrangement is achieved by seeking for edges commonly exist within two encodings of gSpan. This encoding helps in identifying sub-graphs that are similar to each other and in effect prevents gSpan from wasting time to do tasks which are already carried out.

4) *Minimum gSpan Encoding (MgSE).* MgSE is actually the smallest of all gSpan encoding possibilities (i.e., the lowest in lexicographic order"). MgSE is the basis to candidate pruning in gSpan since MgSE allows gSpan to realize when it is about to build a sub-graph already deemed by it. This principle helps in making the gSpan algorithm more effective than original Apriori-based algorithms.

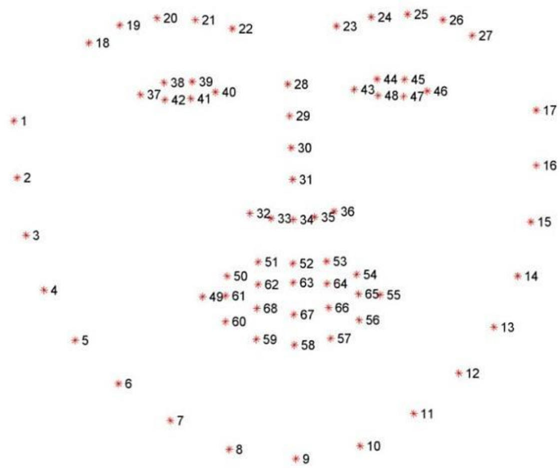
B. Facial Landmark Extraction

Facial landmark extraction is defined as the process of localization of the positions of certain key points on the face region. These key-points have an impact on the subsequent face-based tasks, such as animation, face recognition, gaze detection, face tracking, recognition of expression, gesture perception, etc. Examples of landmarks' points are the points that surround the head, nose, lip or jaw areas [17].

Facial landmark detection algorithms can be divided into three main categories: holistic methods, Constrained Local Model (CLM) methods, and regression-based methods, depending on how facial appearance and facial patterns are modeled. The facial appearance refers to the distinctive pixel intensity patterns around the facial landmarks or throughout the facial area, while facial shape patterns refer to the patterns of the facial shapes as described by the landmarks' locations and their spatial relationships [18].

The holistic methods explicitly model the facial appearance and global patterns of facial shape. Active Appearance Model (AAM) is a classic example of holistic methods [19]. CLM methods infer the location of landmark points based on global facial shape patterns as well as independent local appearance information around each landmark that is easier to capture and more resilient to illumination and occlusion compared to the holistic appearance [18]. Regression-based approaches use holistic or local appearance knowledge and may indirectly incorporate the global facial shape patterns for landmark localization. Multiple patch predictions can be combined to determine by voting the final prediction of the probability or confidence score. Recently, the regression-based approaches show the best performance compared with holistic and CLM methods [20].

Based on regression-based approaches, Kazemi, and Sullivan [21] presented an algorithm to accurately estimate the locations of 68 facial landmarks in a computationally efficient way as shown in Fig. 1. They are based on gradient boosting to learn an ensemble of regression trees which optimizes the sum of squared error loss.



(a)

Facial Component	Point Region
Left jaw line	1-8
Chin	9
Right jaw line	10-17
Left eyebrow	18-22
Right eyebrow	23-27
Bridge of nose	28-31
Bottom of nose	32-36
Left eye	37-42
Right eye	43-48
Outer edge of lips	49-60
Inner edge of lips	61-68

(b)

Fig. 1: Locations of facial landmark points proposed by Kazemi and Sullivan [21]; (a) locations of the points in the face region; (b) indices of the points.

Each regression function in the cascade effectively predicts the shape of landmark points from an initial estimate ($S^{(t)}$) that determined from the mean face pose of all faces in the database. The intensities of a sparse collection of pixels indexed relative to this initial estimate, used as input to the regressor. Each regressor, $r_t(\cdot,\cdot)$, in the cascade predicts an update vector from the image (I) and $S^{(t)}$ and $r_t(I, S^{(t)})$ is added to the current shape estimate ($S^{(t)}$) to improve the estimate as in the following equation [21]:

$$\hat{S}^{(t+1)} = S^{(t)} + r_t(I, S^{(t)} \qquad (1)$$

Where I is the facial image and $\hat{S}^{(t)}$ is the shape that represents the set of landmarks points in regressor r_t.

The regression function (r_t) is learned using gradient tree boosting with a sum of squared error loss. This process is repeated until all the regressors ($r_0, r_1, \ldots, r_{T-1}$) in the cascade of T regressors are learned which when combined give an improved level of accuracy. The gradient boosting algorithm of each regression function (r_t) is based on the regressor tree to fit the residual targets. At each split node in the regression tree, the decision is made based on thresholding the difference between the intensities of two pixels (the landmark point and the current pixel within the collection of pixels) [21].

III. The Proposed Method

As shown in Fig. 2, the proposed facial emotion recognition system involves five important stages which are: landmark extraction, construction of facial component graphs, frequent sub-graphs mining, feature vector generation and emotion classification. Locations of landmark points are determined first in landmark extraction stage and twelve different groups are created to represent different facial components and their relations. After that, a graph is built for each group based on its landmark points. The frequent sub-graphs of each graph are mined using gSpan algorithm. The resulted sub-graphs are filtered in feature generation stage to select the most effective ones which then formed the final feature vector of the given graph. Emotion classification stage includes two levels of classifications: group level decision and fusion level decision. Each of selected group will give its decision regards the emotion class in group level decision and then the decision vector is fed to Naïve Bayes to fuse them and make the final emotion classification.

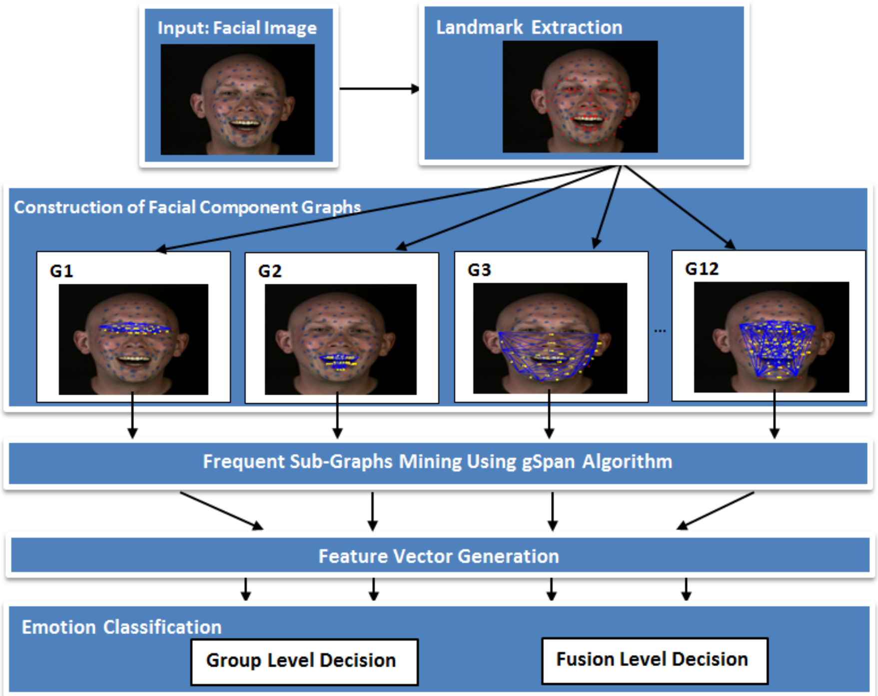


Fig. 2: General overview of the proposed system.

A. Landmark Extraction

Landmark points are identified first to determine the key parts of the face like eye-brows region, eyes region, mouth region, nose region, and jawline region. The dlib library, that Kazemi and Sullivan [21] introduced, is utilized to locate the coordinates of facial landmark points. After that, twelve different groups of landmark points are created to represent the different facial components (eyes and eyebrows, mouth, jaw and nose), in addition to the relations among these facial components as shown in Table 1.

TABLE 1
GROUPS OF FACIAL COMPONENT POINTS

Group ID	Facial Components	Number of points	Landmark points' indices
G_1	Eyes and eyebrows region	8	18, 27, 22, 23, 37, 46, 40, 43
G_2	Mouth region	8	49, 55, 51, 59, 52, 58, 53, 57
G_3	Jaw region	8	1, 17, 3, 15, 6, 13, 8, 11
G_4	Eyes, Eyebrows and nose regions	9	18, 27, 22, 23, 37, 46, 40, 43, 34
G_5	Mouth and nose regions	9	49, 55, 51, 59, 52, 58, 53, 57, 34
G_6	Jaw and nose regions	9	1, 17, 3, 15, 6, 13, 8, 11, 34
G_7	Mouth and eyes and eyebrows regions	12	18,22,23,27,37,40,43,46,49,52,55,58
G_8	Jaw and eyes and eyebrows regions	12	18,22,23,27,37,40,43,46,4,1,7,11
G_9	Jaw and mouth regions	8	49,52,55,58, 4,14,7,11
G_{10}	Mouth, nose and eyes and eyebrows regions	13	18,22,23,27,37,40,43,46,49,52,55,58,34
G_{11}	Mouth, nose, Jaw regions	9	49,52,55,58,34, 4,14,7,11
G_{12}	Eyes and eyebrows, mouth and Jaw regions	14	18,22,23,27,37,40,43,46,49,52,55,58, 7,11

B. Construction of Facial Component Graphs

After creating twelve different groups of facial landmark points, a facial component graph is then built for each group using the same steps proposed by [13] which can be summarized with the following points:

1) The indices of landmark points are used as vertices' labels.
2) For every two points ($p_1(x_1, y_1)$ and $p_2(x_2, y_2)$), the distance confined between p_1 and p_2 coordinates are computed.

3) To prevent the effect of the differences in camera zooming and face scaling among different human faces, the normalization process is performed by dividing each measured distance by the highest measured distance in that facial component. After performing normalization step, the distances will be within the range [0-1].

4) To be able of applying the gSpan algorithm, the continues values of the normalized distances in each graph must be categorized. Categorization process is performed by dividing the range of distances (i.e. [0-1]) into sub-ranges, each sub-range with length equals to 0.05. After that each sub-range is given a label from 1 to 20 to accommodate all potential distances. Fig. 3 shows examples for the constructed graphs within the different facial component groups using an example image taken from the used database.

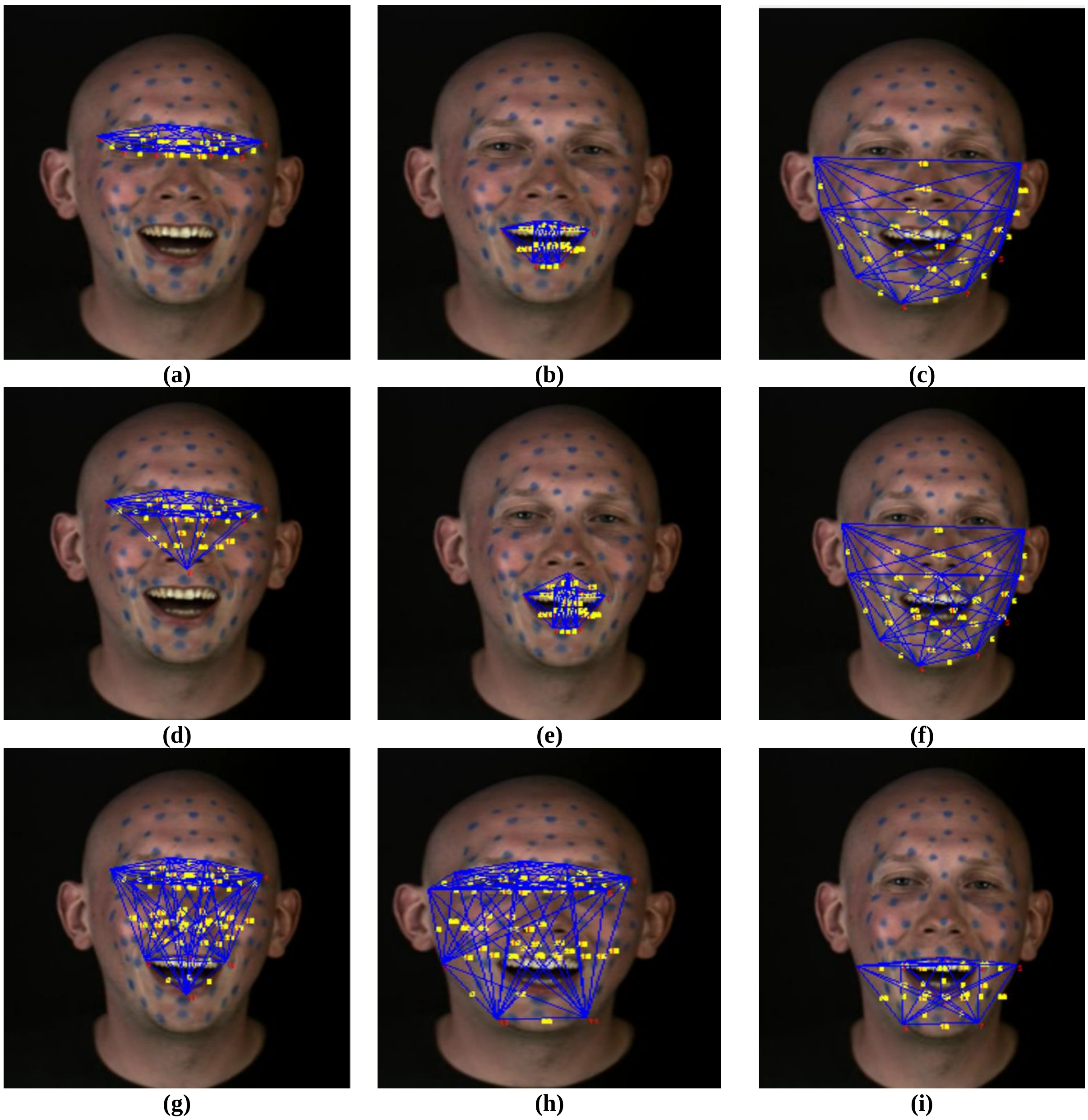


Fig. 3: Examples for the constructed graphs; (a) G_1 graph; (b) G2 graph; (c) G_3 graph; (d) G_4 graph; (e) G_5 graph; (f) G_6 graph; (g) G_7 graph; (h) G_8 graph; (i) G_9 graph; (j) G_{10} graph; (k) G_{11} graph; (l) G_{12} graph.

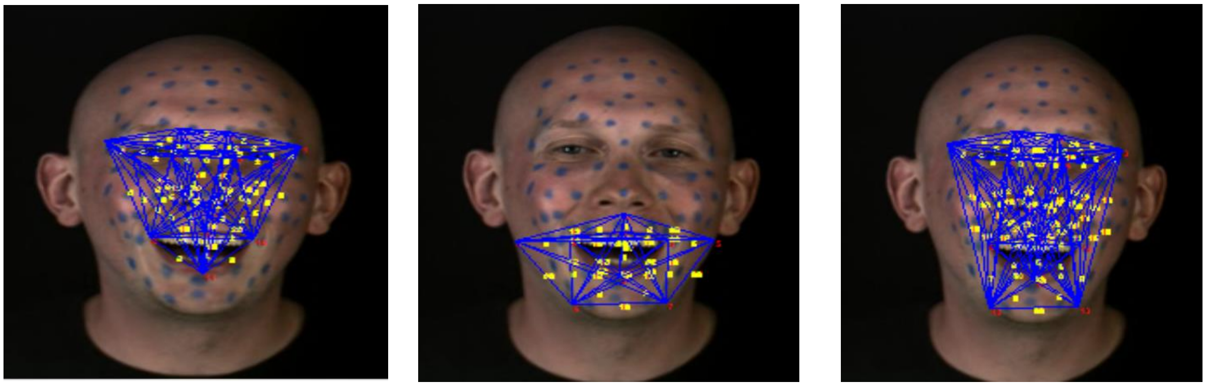


Fig. 3: Continued.

C. Frequent Sub-Graphs Mining

To extract discriminating attributes that reflect the frequent changes in the facial graph of each facial component group, the frequent sub-structures within each emotional class must be mined first. gSpan is utilized for this task. The algorithm takes as input a dataset of graphs for each emotion within the group and the value of minimum support, and returns as output the frequent sub-structures for the given dataset of graphs. gSpan algorithm is applied for number of times equals to 84(Number of groups (12) × Number of emotions (7)= 84). However; the mining time is very small because the size of constructed graphs is smaller than the size of the graph that represents the whole facial region.

D. Feature Vector Generation

Twelve different feature vectors are generated; each one represents a facial component group. Following steps are followed during feature vector generation stage:

1) For each emotion, the overlap ratio that each sub-structure makes with remaining emotional graphs is computed.

2) Then, the sub-graphs are ordered in ascending order according to the overlap ratio.

3) The best two sub-graphs that give the smallest overlap ratio are then selected.

4) This procedure is followed by picking the best three sub-graphs with the least overlap ratio and the best four sub-graphs and so on till the overlap ratio value becomes equal to zero or it does not improve.

5) The set of selected sub-graphs for all emotions are then merged together to form the final feature vector of the given group.

6) In order to be able of applying the algorithms of optimization and machine learning, the final vector of selected sub-structures should be transformed to multi-dimensional feature vector. Binary encoding is utilized for this task where every facial graph within the graph dataset is checked for a correspondence with the predefined selected sub-structure. If the match occurs, the sub-graph is represented in multidimensional vector with value 1 otherwise it is represented with 0.

E. Emotion Classification

To make the final decision about the emotion class for the given facial image, each group must give its decision using the generated feature vector of sub-graphs (i.e., group level decision). The decisions taken by each group are then fused in fusion level decision to make the final emotion decision. Fig. 4 demonstrates the main idea of emotion classification stage.

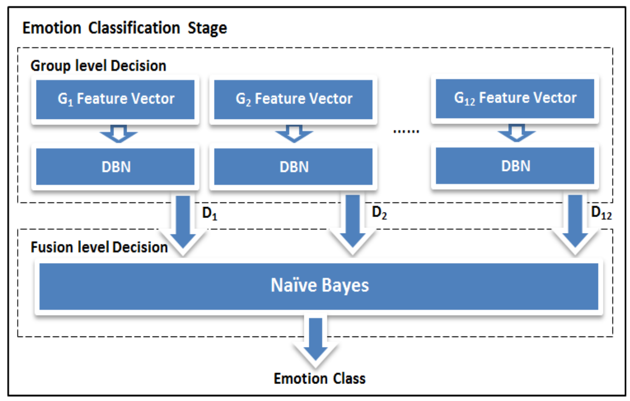


Fig. 4: Emotion classification stage.

1) Group level decision

The Deep Belief Network (DBN) classifier is utilized for the task of emotion classification within each facial group due to its strong classification ability. DBN consists of stacked Restricted Boltzmann Machines (RBMs) that firstly performs the unsupervised learning to avoid vanishing problem that occurs when training the network from scratch weights (i.e., random weights). The weights of the network are then fine-tuned using supervised learning (i.e., regular back propagation training). RBM is a Boltzmann machine which has multiple hidden layers and one visible layer [22].The classification stage involves two steps which are training step and testing step. In training step, the DBN will be trained with the training part of the database to find the optimal DBN architecture and adjust the network weights. The resulted weights are then fed to testing step for guessing the emotion of the given group feature vector.

2) Fusion level decision

Naïve Bayes is utilized for decision level fusion. Naive Bayes is a supervised learning algorithm which is relied on Bayes theorem. It simplifies the learning process by assuming the features are independent on the given class. Bayesian classifiers are very simple, requiring only a single scan of the data, thereby providing high accuracy and speed for large databases.

IV. Experimental Results and Analysis

A. Database

The proposed system is evaluated using samples taken from the Surrey Audio-Visual Expressed Emotion (SAVEE) database [23]. The SAVEE database was established at the University of Surrey in the United Kingdom. It includes recordings for six basic emotions, i.e. anger, disgust, fear, happiness, sadness, and surprise as well as natural state. The database contains 480 native British utterances formed of 60 samples for each of the emotional states except for natural emotion which includes 120 samples. Emotional and text prompts were displayed in front of the subject on a monitor while each video is recoding. In each prompt, a video clip and three images were included such that the text was divided into three groups for each emotion, in order to avoid fatigue. The samples were captured with a video camera sampled at 60 frames per second (fps) with resolution of about 81,920 pixels and mono voice signals in a 16-bit format sampled at 44,000 Hz.

For system evaluation purpose, the frame that positioned in the middle of each video is picked up because it represents the peak state of the emotion. To avoid over fitting during training process, 80% of samples within each emotion are used for system training and 20 % for testing purpose.

B. Results

Different experiments are carried out to find the best configuration of the proposed system parameters. Following subsection demonstrates the achieved results within sub-graphs mining, feature vector generation and classification stages.

1) Sub-Graphs Mining Results

Table 2 shows the number of frequent sub-graphs that are generated using gSpan algorithm when applied on the database of graphs for each emotion within each group using minimum support value=70. As shown in the table, Group 5 generates the least number of sub-graphs due to the unsteady behavior of the mouth facial region because the samples in the database are taken for subjects while talking.

TABLE 2
NUMBER OF GENERATED SUB-GRAPHS WITHIN EACH EMOTION IN THE DIFFERENT GROUPS

Group ID	Anger	Disgust	Fear	Happy	Natural	Sad	Surprisr
G_1	380	489	2146	2078	1277	772	924
G_2	4	4	4	4	2	3	5
G_3	1561	153	329	185	224	73	359
G_4	432	500	2679	2609	1277	777	999
G_5	5	4	4	4	2	3	5
G_6	2376	412	2728	180	1258	935	550
G_7	120	71	15	94	423	38	12
G_8	315	31	87	86	251	119	102
G_9	11	20	23	12	26	37	14
G_{10}	4007	4006	432	4005	4025	4008	1581
G_{11}	16	21	25	13	27	39	16
G_{12}	190	23	95	108	432	388	235

2) Feature Vector Generation Results

In feature vector generation stage, the best frequent sub-graphs within each group are selected. Table 3 demonstrates the number of selected sub-graphs within each emotion for the different groups. The total number of selected sub-graphs which represents the length of feature vector for the given group is also shown in the table. The resulted overlap from each combination of sub-graphs is demonstrated in Table 4. The highest overlap is yielded from groups that contain mouth region that has the unstable behavior. The combinations those include eyes or jaw region have small overlap which reflect the ability of these regions in reflecting the emotions without effecting by the change that occur in the face while talking.

TABLE 3
NUMBER OF SELECTED SUB-GRAPHS WITHIN EACH EMOTION IN THE DIFFERENT GROUPS

Group ID	Anger	Disgust	Fear	Happy	Natural	Sad	Surprise	Total number of Sub-graphs
G_1	4	4	4	7	7	4	3	33
G_2	3	3	3	3	2	2	2	18
G_3	6	9	7	8	7	8	7	52
G_4	4	4	5	8	7	5	4	37
G_5	5	4	4	4	2	3	3	25
G_6	6	6	6	8	7	8	8	49
G_7	7	4	9	7	6	8	3	44
G_8	7	9	8	9	9	8	5	55
G_9	4	4	4	4	6	7	5	34
G_{10}	7	6	9	7	4	3	9	45
G_{11}	7	5	4	5	7	8	5	41
G_{12}	9	9	9	7	8	8	8	58

TABLE 4
OVERLAP RATIO RESULTED FROM THE SELECTED SUB-GRAPHS WITHIN EACH EMOTION IN THE DIFFERENT GROUPS

Group number	Anger	Disgust	Fear	Happy	Natural	Sad	Surprise	Total overlap
G_1	0.0083	0.2625	0.1146	0.0167	0.0262	0.1667	0.1813	0.7763
G_2	0.2958	0.2979	0.3417	0.3396	0.5396	0.3958	0.3313	2.5417
G_3	0.0604	0.0979	0.1000	0.0729	0.0896	0.1083	0.0583	0.5874
G_4	0.0291	0.1708	0.0520	0.0104	0.0229	0.1188	0.1188	0.5228
G_5	0.1708	0.2813	0.3333	0.3292	0.5146	0.3708	0.3271	2.3271
G_6	0.0095	0.0769	0.0833	0.0563	0.0625	0.0688	0.0458	0.4875
G_7	0.0188	0.2750	0.0854	0.0333	0.0063	0.0292	0.2896	0.7376
G_8	0.0271	0.0646	0.0188	0.0417	0.0271	0.0250	0.0125	0.2168
G_9	0.1542	0.2042	0.2646	0.2021	0.0458	0.0625	0.2771	1.2105
G_{10}	0.0604	0.1396	0.1146	0.0688	0.0313	0.1021	0.0208	0.5376
G_{11}	0.0479	0.0958	0.1813	0.1250	0.0208	0.0417	0.1917	0.7042
G_{12}	0.0208	0.0438	0.0542	0.0021	0.0063	0.0146	0.0292	0.1710

3) Emotion Classification Results

To take the decision about the emotional class of the given facial image, two experiments are conducted. In the first one, the DBN classifier is trained using different set of feature vectors of the facial component groups. In the second one, the Naïve Bayes classifier is trained with the decisions that are obtained from the DBNs of the different groups. The measure that is used in system evaluation is the classification accuracy that can be calculated as follows [24]:

$$Accuracy = \frac{correct}{Total} \times 100 \qquad (2)$$

Where correct represents the number of correctly classified speech signals and Total is the number of speech signals in the database.

1. Group level Results

Table 5 shows the accuracy that is achieved in each group of facial components using DBN classifier. Different configurations for DBN parameters are tested which include Number of hidden number (H), Learn Rate (LR) and Momentum (Mom). The best configuration that gave the highest accuracy is then obtained. As shown in Table (5), the highest accuracy (97.86%) is reached when training DBN with G_{12} feature vectors while G5 is the worst group with accuracy equals to 23.33%.

2. Fusion level decision results

Six different experiments are conducted in fusion level decision based on the accuracy of each group classifier to assess the effect of group's accuracies on Naïve Bayes results. First, the decisions of all groups are given to Naïve Bayes as a vector of attributes to train and test the classifier. Second experiment is conducted using groups with DBN accuracy more than 50%. Third experiment is based on the outputs of groups with DBN accuracy >60 and so on until DBN accuracy >90. Table 6 shows results which are obtained for the six different cases. As clearly shown in the table, Naïve Bayes provides the best accuracy when obtaining all groups' decisions during making the final decision which refers to the fact that groups with lowest accuracy can help in improving the decision by working as weak attributes that support the strong ones.

TABLE 5
RESULTS OBTAINED WHEN TRAINING DBN WITH DIFFERENT COMPONENT GROUPS

Group	Accuracy	DBN Configuration
G_1	75.95%	H=17, LR= 0.2, Mom= 0.3
G_2	25.24%	H=16, LR= 0.1, Mom= 0.9
G_3	42.38%	H=16, LR= 0.4, Mom= 0.2
G_4	86.19%	H=16, LR= 0.1, Mom= 0.1
G_5	23.33%	H=16, LR= 0.1, Mom= 0.4
G_6	60.48%	H=16, LR= 0.2, Mom= 0.1
G_7	90.48%	H=15, LR= 0.1, Mom= 0.1
G_8	94.29%	H=17, LR= 0.6, Mom= 0.2
G_9	50.71%	H=16, LR= 0.1, Mom= 0.4
G_{10}	77.62%	H=15, LR= 0.1, Mom= 0.2
G_{11}	60.95%	H=15, LR= 0.1, Mom= 0.3
G_{12}	97.86%	H=16, LR= 0.4, Mom= 0.2

TABLE 6
RESULTS OF THE DIFFERENT SIX EXPERIMENTS CONDUCTED ON THE OUTPUTS OF FACIAL COMPONENT GROUPS

Case	Groups	Number of Attributes	Naïve Bayes Accuracy
All groups	All groups	12	100.00%
DBN Accuracy > 50	G1+G4+G6+G7+G8+G9+G10+G11+G12	9	99.77%
DBN Accuracy > 60	G1+G4+G6+G7+G8+G10+G11+G12	8	99.77%
DBN Accuracy > 70	G1+G4+G7+G8+G10+G12	6	99.77%
DBN Accuracy > 80	G4+G7+G8+G12	4	99.77%
DBN Accuracy > 90	G7+G8+G12	3	99.77%

C. Comparison with Other Works

Table 7 shows a comparison made between the proposed facial emotion recognition system and other works which utilized the same database (i.e., SAVEE database). The table also shows the used methods and achieved accuracy by each work. As it is clearly shown in the table, the accuracy that are achieved by the proposed system outperforms those are achieved by other works with an improvement equals to 5.4%. The table also reflects the ability of the extracted sub-graphs features on determines the facial emotion class regardless of illumination, contrast, rotation, and scaling and shift variant effects that exist in other feature extraction methods such as Gabor filter.

TABLE 7
COMPARISON BETWEEN THE PROPOSED SYSTEM AND OTHER WORKS USING SAVEE DATABASE

Authors	Method	Database	Number of emotions	Achieved accuracy
Haq, and Jackson (2009) [7]	240 visual features based on the marker locations on the face, PCA and Linear Discriminate Analysis (LDA) for feature selection	SAVEE	Five different emotional expressions (happy, angry, fear, sad and neutral)	91.00%
Cid, Manso, and Nunez (2015) (2015) [4]	Gabor filter and dynamic Bayesian classifier	SAVEE	Seven basic emotions	94.60%
Barros, and Wermter (2016) [10]	Key point's movement information with a conventional neural network	SAVEE	Seven basic emotions	93.90%
Hassan, and Mohammed (2019) [13]	Graph mining for the whole facial region	SAVEE	Seven basic emotions	90.00%
The proposed system	Mining sub-graphs of different facial components	SAVEE	Seven basic emotions	100%

V. CONCLUSIONS

A facial emotion recognition system has been presented in this paper based on mining sub-graphs of facial components of human face image. However; instead of directly applying the gSpan algorithm on the facial graph of the whole face image, the graphs that represent the facial components and their relations are used as inputs to graph mining task. Mining graphs of facial components helps in focusing the mining process on the micro changes in the graphs of different facial components and in effect leads to more effective sub-graphs. Deep belief network also gives the proposed system the power in understanding the hidden patterns in the selected sub-graphs from each emotion within each group. Experimental results showed that the desired accuracy (100%) was achieved on SAVEE database when fusion the decisions of the twelve groups of facial components using Naïve Byes classifier. As a future work, deep learning technique can be employed through the construction of a new type of deep network that is specialized in graph recognition. The represented sub-graphs can be mined in more than one layer with different techniques of mining for each layer. Similarly, the generated sub-graphs can be reduced using different techniques in more than one layer. Finally, feed forward neural network can be used in the classification layer.

CONFLICT OF INTEREST

The authors have no conflict of relevant interest to this article.

REFERENCES

[1] M. Dubey and L. Singh, "Automatic Emotion Recognition Using Facial Expression: A Review", International *Research Journal of Engineering and Technology (IRJET)*, vol. 3, no. 2, pp. 488-492, 2016.
[2] S.N. Mohammed, and A.K. Abdul Hassan, “A Survey on Emotion Recognition for Human Robot Interaction”, *Journal of Computing and Information Technology*, in press.

[3] C. Wu, J. Lin and W. Wei, "Survey on Audiovisual Emotion Recognition: Databases, Features, and Data Fusion Strategies", *APSIPA Transactions on Signal and Information Processing*, vol. 3, pp. 1-18, 2014.

[4] F. Cid, L.J. Manso and P. Nunez, "A Novel Multimodal Emotion Recognition Approach for Affective Human Robot Interaction", presented at the IEEE/RSJ International Conference on Intelligent Robots and Systems, 2015.

[5] Z. Liu, M. Wu, W. Cao, L. Chen, J. Xu, R. Zhang, M. Zhou and J. Mao "A Facial Expression Emotion Recognition Based Human-Robot Interaction System", *IEEE/CAA Journal of Automatica Sinica*, vol. 4, no. 4, pp. 668-676, 2017.

[6] V. Verma and L.K. Sharma, "A Comprehensive Survey on Human Facial Expression Detection", *International Journal of Image Processing (IJIP)*, vol. 7, no. 2, pp. 171-182, 2013.

[7] S. Haq and P.J.B. Jackson, "Speaker-Dependent Audio-Visual Emotion Recognition", presented at the Int. Conf. on Auditory-Visual Speech Processing, 2009.

[8] A. Meghdari, M. Alemi, A. G. Pour and A. Taheri, "Spontaneous Human-Robot Emotional Interaction through Facial Expressions", presented at the International Conference on Social Robotics, 2016.

[9] S. Dutta and V.B. Bari, "Review of Facial Expression Recognition System and Used Datasets", *International Journal of Research in Engineering and Technology (IJRET)*, vol. 2, no. 12, pp.641-645, 2013.

[10] P. Barros and S. Wermter, "Developing Crossmodal Expression Recognition Based on a Deep Neural Model", *Adapt Behav*, vol. 24, pp. 373–396, 2016.

[11] W.G. Hatcher and W. Yu, "A Survey of Deep Learning: Platforms, Applications and Emerging Research Trends", *Human-Centered Smart Systems and Technologies*, vol. 6, pp. 24411- 24432, 2018.

[12] C.C. Aggarwal and H. Wang, *Managing and Mining Graph Data*, Springer, 2010.

[13] A.K. Hassan and S. N. Mohammed, “A Novel Facial Emotion Recognition Scheme Based on Graph Mining”, *Defence Technology*, vol. 16, no. 5, , pp.1062-1072, 2020.

[14] N.F. Samatova, W. Hendrix, J. Jenkins, K. Jenkins and A. Chakraborty, *Practical Graph Mining With R*, CRC Press, 2014.

[15] C. C. Aggarwal, *Data Mining the Textbook*, Springer, 2015.

[16] D. J. Cook and L.B. Holder, *Mining Graph Data*, John Wiley & Sons, 2007.

[17] B. Johnston and P. D. Chazal, "A Review of Image-Based Automatic Facial Landmark Identification Techniques," *EURASIP Journal on Image and Video Processing*, vol. 2018, no. 86, pp. 1-23, 2018.

[18] Y. Wu and Q. Ji, "Facial Landmark Detection: A Literature Survey," *International Journal of Computer Vision*, vol. 127, pp. 115–142, 2019.

[19] N. Wang, X. Gao, D. Tao, and X. Li, "Facial Feature Point Detection: A Comprehensive Survey," Neurocomputing, vol. 275, pp. 50-65, 2018.

[20] O. Çeliktutan, S. Ulukaya, and B. Sankur, "A Comparative Study of Face Landmarking Techniques," *EURASIP Journal on Image and Video Processing*, vol. 2013, pp. 1- 27, 2013.

[21] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees", presented at the 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014.

[22] A. Bondarenko and A. Borisov, “Research on the Classification Ability of Deep Belief Networks on Small and Medium Datasets”, *Information Technology and Management Science*, vol. 16, pp. 60-65, 2013.

[23] Surrey Audio-Visual Expressed Emotion (SAVEE) Database Website, http://kahlan.eps.surrey.ac.uk/savee/

[24] S. Mohammed, F. Alkinani, and Y. Hassan, “Automatic Computer Aided Diagnostic for COVID-19 Based on Chest X-Ray Image and Particle Swarm Intelligence”, *International Journal of Intelligent Engineering and Systems*, vol.13, no. 5, pp. 63-73, 2020.