



The Effect of the Number of Key-Frames on the Facial Emotion Recognition Accuracy

Suhaila N. Mohammed ^{a*}, Alia K. Abdul Hassan ^b

^a Department of Computer Science, University of Technology, Baghdad, Iraq
suhailan.mo@sc.uobaghdad.edu.iq

^b Department of Computer Science, University of Technology, Baghdad, Iraq, 110018@uotechnology.edu.iq

*Corresponding author.

Submitted: 20/08/2020

Accepted: 17/11/2020

Published: 25/03/2021

KEY WORDS

Facial expressions
 Fuzzy C-Means
 algorithm
 Graph-based
 Substructure Pattern
 (gSpan) algorithm
 Key-frame selection.

ABSTRACT

Key-frame selection plays an important role in facial expression recognition systems. It helps in selecting the most representative frames that capture the different poses of the face. The effect of the number of selected keyframes has been studied in this paper to find its impact on the final accuracy of the emotion recognition system. Dynamic and static information is employed to select the most effective key-frames of the facial video with a short response time. Firstly, the absolute difference between the successive frames is used to reduce the number of frames and select the candidate ones which then contribute to the clustering process. The static-based information of the reduced sets of frames is then given to the fuzzy C-Means algorithm to select the best C-frames. The selected keyframes are then fed to a graph mining-based facial emotion recognition system to select the most effective sub-graphs in the given set of keyframes. Different experiments have been conducted using Surrey Audio-Visual Expressed Emotion (SAVEE) database and the results show that the proposed method can effectively capture the keyframes that give the best accuracy with a mean response time equals to 2.89s.

How to cite this article: S. N..Mohammed and A. K. Abdul Hassan, "The Effect of the Number of Key-Frames on the Facial Emotion Recognition Accuracy," Engineering and Technology Journal, Vol. 39, Part B, No. 01, pp. 89-100, 2021.

DOI: <https://doi.org/10.30684/etj.v39i1B.1806>

This is an open access article under the CC BY 4.0 license <http://creativecommons.org/licenses/by/4.0>

1. INTRODUCTION

Keyframes can be defined as the most representative frames in the sequence of video frames. Video keyframe selection must capture the main video content in an accurate and fast manner [1, 2].

Human face in video recognition is commonly used in video analytics and video surveillance. Recently, numerous face-based recognition systems such as facial emotion recognition have been proposed by the researchers and very high accuracy is achieved on still-image-based datasets. In real-world environments, for example, the detection of the emotion from a video with different talking

faces poses great challenges on facial emotion recognition [3]. Many reasons behind that, firstly, face detection and emotion recognition require a lot of computation resources, especially when every frame in the video needs to be processed; secondly, different poses of the face while talking make the recognition of the involved emotion a difficult task. To overcome these challenges, keyframe selection becomes a necessary and essential preprocessing step before performing the emotion recognition task [4].

There are several key-frame selection strategies used by the researchers in the field of pattern recognition, such as motion-analysis-based strategy, shot-boundary-based strategy, visual-content-based strategy, and clustering-based strategy. Motion-analysis-based strategy computes the optical flow for each frame in the video to estimate whether or not the change happens. The shot-boundary-based strategy selects the first, the middle, and the last frames of the video as the key-frames. On the other hand, the visual content-based strategy employs more than one criterion (i.e., the combination of shot-based, color features, and motion-based criteria). While in the clustering-based strategy, each frame is allocated to a specific cluster, and the frames which are closest to the center of each cluster are then selected as key-frames [5]. Generally, the clustering methods demonstrate good performance from the other three methods [6].

Key frame selection using clustering technologies has recently attracted a wide range of researchers' attention and some of the proposed key frame extraction methods are Shumin et al.[7] used an improved artificial fish swarm algorithm to obtain the initial clustering centers of the extracted color feature vector. After that, K-Means was conducted to obtain the final clustering result; the center frame of each clustering was selected as the key frames. In the work of Liu and Hao[1], the improved hierarchical clustering algorithm was used to obtain an initial clustering result and K-means is then conducted to optimize the obtained result and produce the final clustering result. Again the center frame of each clustering is extracted as key frame. Lv and Huang [6] used nearest neighbor clustering by merging the smaller clusters into the nearest larger cluster which it is next to. The authors based on the fact that the extremely small clustering may deviate from the video's topic content, but some smaller clustering is also part of video content, which can be considered to be incorporated into a larger cluster.

This paper aims at studying the effect of some selected key frames on the accuracy of the facial emotion recognition system. Fuzzy C-means is used for the selection of key-frames and a facial emotion recognition system based on graph mining is then used for the classification task. The rest of the paper is organized as follows: the theoretical background for the methods used in this work is presented in Section 2. The proposed methodology is discussed in Section 3. The achieved results are demonstrated in Section 4. Finally, the work conclusions and future work are shown in Section 5.

2. THEORETICAL BACKGROUND

1. First-order statistics

First-order features can effectively describe the change in the texture involved within the frame image. Some of the first-order features are [8]:

A. Mean (μ_1): the mean value can give an indicator if the colors of the frame pixels have been changed. This feature is calculated using the following equation:

$$\mu_1 = \sum_{i=0}^{255} i * P(i) \quad (1)$$

$$\text{and } P(i) = \frac{\text{Histogram}(i)}{(W*H)} \quad (2)$$

Where W and H are dimensions of the frame under processing.

B. Variance (μ_2): the variance is defined as the histogram width. It is a measure of how many the gray levels differ from the mean, and can be used to describe smoothness.

$$\mu_2 = \sum_{i=0}^{255} (i - \mu_1)^2 * P(i) \quad (3)$$

C. Skewness(μ_3): it is a measure of the degree of histogram symmetry around the mean. Equation (6) has been used to compute skewness values.

$$\mu_3 = \sum_{i=0}^{255} (i - \mu_1)^3 * P(i) \quad (4)$$

D. Kurtosis (μ_4): kurtosis is a measure of the histogram sharpness and can be found using equation (7).

$$\mu_4 = \sum_{i=0}^{255} (i - \mu_1)^4 * P(i) \quad (5)$$

E. Entropy (E): E is a measure of histogram uniformity. The closer to normal distribution the higher the E . it can be defined as:

$$H = -1 * \sum_{i=0}^{255} \log_2(P(i)) * P(i) \quad (6)$$

II. Feature normalization

The normalization helps in selecting the proper value of the parameters used in the proposed system such that they will be within the same range (i.e., [0-1]). The equation used for Min-Max normalization is [9]:

$$V'_i = \left(\frac{V_i - \text{Mino}}{\text{Maxo} - \text{Mino}} \right) * (\text{Maxn} - \text{Minn}) + \text{Minn} \quad (7)$$

Where V_i and V'_i are old and new data values; respectively, Mino and Maxo are the old range confines, and Minn and Maxn are the new range confines.

III. Fuzzy C-Means Algorithm

Fuzzy C-Means algorithm is a method of clustering that uses concepts from the field of fuzzy logic and fuzzy set theory. Fuzzy set theory allows an element to belong to a set with a degree of membership between 0 and 1 [10]. Fuzzy C-Means consists of a collection of C clusters, C_1, C_2, \dots, C_N , and a membership matrix $\mu = \mu_{ij} \in [0,1]$, for $i = 1 \dots K$ and $j = 1 \dots C$, where each element μ_{ij} represents the degree of membership of object i in cluster C_j . It is based on minimization of the following objective function [11]:

$$J_m = \sum_{i=1}^K \sum_{j=1}^C \mu_{ij}^m \|x_i - c_j\|^2, \quad 1 \leq m \leq \infty \quad (8)$$

where K is the number of data elements, C is the number of clusters, m is any real number greater than 1 which represents the intensity of fuzzification, X_i is the i^{th} of the d -dimensional measured data item, C_j is the d -dimension center of the cluster j , μ_{ij} is the degree of membership of X_i in the cluster j , and $\| \ \|$ is any similarity measure between the data item and cluster center. Fuzzy partitioning is carried out through an iterative process. In each iteration, the membership μ_{ij} and the cluster centers C_j are updated as following [11]:

$$\mu_{ij} = \frac{1}{\sum_{z=1}^C \left(\frac{\|x_i - c_j\|}{\|x_i - c_z\|} \right)^{\frac{2}{m-1}}} \quad (9)$$

IV. Emotion Recognition Using Graph Mining

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]. The novel method that has been proposed by Hassan and Mohammed [13] for facial emotion recognition is used for facial emotion recognition. This method can be summarized by the following steps:

1) Step 1: 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 [14] introduced, is utilized to locate the coordinates of facial landmark points. Twenty-two

points located on the eyes, eyebrows, mouth, nose, and jaw regions are selected for the graph building task.

2) Step2: The facial component graph is then built using point indices as node labels and the distance between each two landmark points as edge weight.

3) Step3: 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. Graph-based Substructure PAtterN (gSpan) [12] 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.

4) Step4: levels-based feature selection is then used to classify the emotion in levels wherein each level; one emotion is classified from the remaining ones. Six levels of classification are used and cat swarm intelligence is used to select the effective set of features in each classification level.

5) Step5: a feed-forward neural network is then used for classification purposes within each level of classification.

6) Step6: To make the final decision regards the emotion class of the selected key frames; Naïve Bayes classifier is used to fuse the outputs resulted from the key frames. Naïve Bayes then gives the final decision regards the emotion class.

3. THE PROPOSED METHOD

The proposed method involves two important phases. The first one is key-frame selection using fuzzy C-means clustering algorithm based on the static features that are extracted from the candidate set of key-frames. The second phase includes the classification of the emotion using the selected set of key frames. Figure 1 shows the general view of the proposed method.

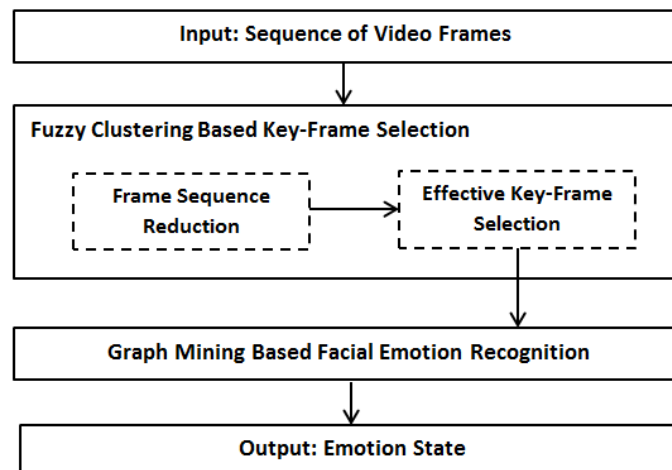


Figure 1: The general view of the proposed method.

I. Fuzzy Clustering Based Key-Frame Selection

A. In this phase, the most effective C key frames are selected to work as a source for emotion recognition system training and testing. Key-frame selection phase involves two main stages which are frame sequence reduction using dynamic information and effective key-frame selection using static information and fuzzy C-means algorithm as shown in Figure 2. Following subsections describe these stages in detail.

1) Frame Sequence Reduction

The clustering process will become very time-consuming when all video frames are involved during key-frame selection. However; in fact, the difference between adjacent frames is very small. Using this characteristic to properly reduce the number of frames involved in the clustering process will dramatically reduce the time of key frame selection. In other words, the number of redundant frames can be reduced to improve the speed of the clustering process. The dynamic information between the adjacent frames must be extracted to achieve this goal. The following steps have been followed to reduce the number of frames in video sequence:

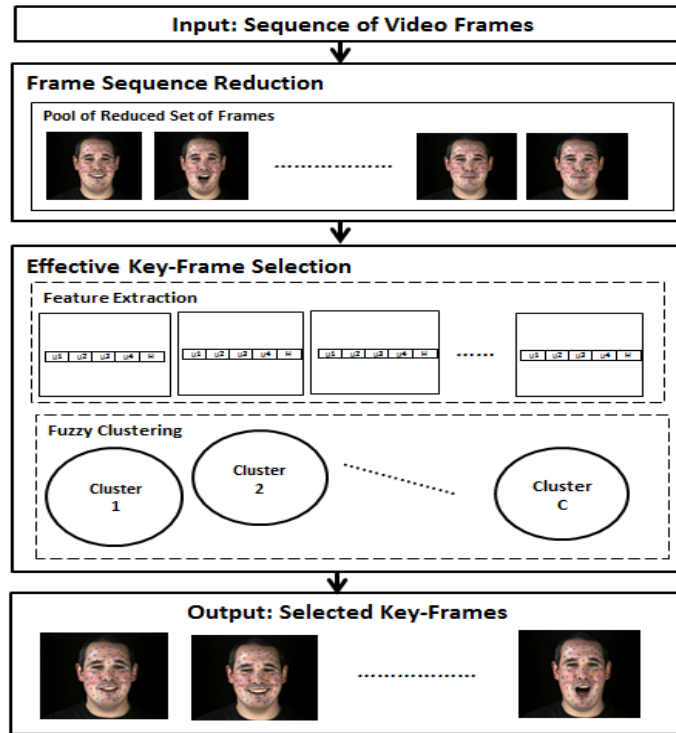


Figure 2: General view of the key frame selection phase.

Step1: Firstly, the first frame of the video is added to the pool of the reduced set of frames ($Frame_R$).
 Step2: The difference between the frame which recently added to $Frame_R$ and the frame which is under consideration is computed. This difference is measured by finding the accumulated absolute difference between the corresponding pixels of the two frames as shown in Eq. (10).

$$D = \sum_{i=0}^W \sum_{j=0}^H f(F_1(i, j), F_2(i, j)) \quad (10)$$

Where F_1 and F_2 are the two video frames, W and H are the width and height of the frame, respectively and f is a different function that can be computed using the following equation:

$$f(F_1(i, j), F_2(i, j)) = \begin{cases} 1 & \text{if } |f(F_1(i, j), F_2(i, j))| > 0 \\ 0 & \text{otherwise} \end{cases} \quad (11)$$

Step3: The resulted difference (D) is then divided by the total number of pixels in the frame (i.e., $W*H$) to find the percentage of the change ($change_{pct}$) concerning the size of the frame.

Step4: Finally, the value of $change_{pct}$ is compared with a predefined threshold ($change_T$). The threshold $change_T$ is used to decide whether the current frame will be considered further in the clustering process. If $change_{pct} \geq change_T$ then the current frame will be added to $Frame_R$, otherwise it will be discarded.

Step5: Go to step 2 until all the frames of the video sequence are tested.

2) Effective Key-Frame Selection

The main purpose of this stage is selection the C representative key-frames from the pool of the reduced set of video frames ($Frame_R$). A clustering strategy is utilized in this paper to select the key-frames. However; if the clustering algorithm is applied using the color of the pixels in the frame as input, then the time of the clustering process will be high due to the large dimensionality of the used features. Thus, a discrimination set of features must be extracted from each frame in $Frame_R$ before applying the clustering algorithm.

B. Feature extraction step

Since the response time is a very important factor in the facial emotion recognition system, the extracted features must be effective in representing the content of the video frames and at the same time, the computation time of these features must be small. Five first-order statistics features have

been generated from each frame after converting it to the greyscale version. These features are selected because they can effectively describe the change in the texture involved within the frame image. Equations (1) to (6) are used as representative features for the given frame. The extracted features are then normalized to be within the range [0-1] using the Min-Max normalization method.

C. Fuzzy clustering step

The outputs of applying the modified fuzzy C-Means on the extracted features of the reduced set of frames are: (1) the clusters' centers and (2) the membership degree of each frame within each cluster. To select the C best key-frames that represent the different contents of the video, the frame that achieves the highest membership degree in each cluster is selected to be the key frame of that cluster. Algorithm 1 illustrates the steps followed to implement the proposed key-frame selection method.

Algorithm 1: Key-frame Selection Using Fuzzy C-Means	
Input	FrameList: a sequence of video frames, change _{pct} : percentage of the change, m: the intensity of fuzzification [1, +∞], C: number of clusters, IterNum: number of iterations.
Output	SelectedFrames: List of selected key-frames
Steps	<pre> 1. Set w← Width of video frames, h← Height of video frames // Frames reduction using dynamic information 2. Set Frame_R()← null // Frame_R represents the pool of reduced set of frames 3. Add FrameList(0) to Frame_R 4. Set oldGrey←Grey-level representation of the first frame in FrameList 5. For i = 1 to FrameList.Length Step 1 6. Begin 7. Set Grey←Grey-level representation of the frame with index (i) in FrameList 8. Set count←0 9. For x = 0 to w-1 Step 1 10. Begin 11. For y = 0 to h-1 Step 1 12. Begin 13. If (Abs(Grey(x, y)–oldGrey(x, y)) > 0) Then Set count←count + 1 14. End For 15. End For 16. Set change_v←count / (w × h) 17. If (change_v >= change_{pct}) Then 18. Add FramList(i) to Frame_R 19. Set oldGrey←Grey 20. End If 21. End For //Static information extraction (feature extraction step) 22. Set FeatureVector(.)←0 23. For i = 0 to Frame_R.Length Step 1 24. Begin 25. Set Grey← Grey-level of Frame_R 26. Set Histogram() ←0 27. For i = 0 to w Step 1 28. Begin 29. For j = 0 to h Step 1 30. Begin 31. Set Histogram(Grey (i, j)) ← Histogram(Grey (i, j))+1 32. End For 33. End For // Find the probability of each color distribution 34. Set prob () ←0 35. For x = 0 to 255 Step 1 36. Begin 37. Set prob(x) ←Histogram(x) / (w × h) 38. End For // Compute features using prob() 39. Set Mean ←0, Std ←0,Skewness ←0, Kurtosis←0, Entropy←0 </pre>

<p>40. For x = 0 to 255 Step1 41. Begin 42. Set Mean ← Mean + (prob (x)× x) 43. End For 44. For x = 0 to 255 Step1 45. Begin 46. Set Std ← Std + ((x - Mean)²×prob (x)) 47. Set Skewness ← Skewness + ((x - Mean)³ × prob(x)) 48. Set Kurtosis ← Kurtosis + ((x - Mean)⁴ × prob(x)) 49. Set Entropy←Entropy + prob(x) × Log₂(prob(x) + 0.0000001) 50. End For 51. Set FeatureVector(i, 0)←Mean 52. Set FeatureVector(i, 1)←Std 53. Set FeatureVector(i, 2)← Skewness 54. Set FeatureVector(i, 3)← Kurtosis 55. Set FeatureVector(i, 4)← Entropy×-1 56. End For 57. Normalize the extracted features using Min-Max normalization method (Equation (7)) 58. Apply FuzzyC_Means on FeatureVector with number of clusters equal to C and number of iterations equal to ItrNum <i>//Select the best C frames</i> 59. For i = 0 to C-1 Step1 60. Begin 61. Set SelectedFrames ← The frame that gains the highest membership degree with respect to cluster (i) 62. End For 63. Return SelectedFrames</p>
--

II. Graph Mining Based Facial Emotion Recognition

The novel method that has been proposed by Hassan and Mohammed for facial emotion recognition is utilized in this work for facial emotion recognition. The input is the dataset of graphs for each emotion class while the output is the frequent sub-graphs within each emotion. The final feature vector is then used with five different neural networks to classify the queried facial image in levels.

4. EXPERIMENTAL RESULTS

The proposed method is tested using samples taken from Surrey Audio-Visual Expressed Emotion (SAVEE) database [15]. SAVEE includes recordings of seven different emotions for four male actors, a total of 480 samples. The sentences are in British English utterances and chosen from the standard corpus of Texas Instruments/Massachusetts Institute of Technology (TIMIT). The data were recorded in a visual media lab with high-quality audio-visual equipment, processed, and labeled. To check the quality of performance, the recordings were evaluated by ten subjects under audio, visual, and audio-visual conditions. The length of the videos in the dataset lasts from 1 to 7 seconds.

Table I shows the effect of change_T value on the number of a reduced set of frames when applied to an example video chosen from the database. As shown in Table I, as the change_T value increases the number of the frames in Frame_R will decrease till only the first frame will remain in the pool of the reduced set of frames. However; to make a tradeoff between the response time of key-frame selection process and the number of frames which will then contribute in the clustering process, the value 0.25 is obtained as a parameter setting for change_T.

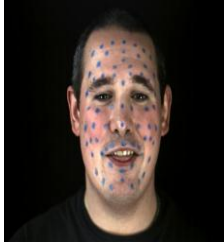
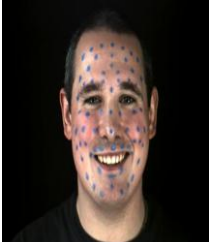
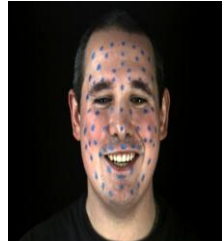
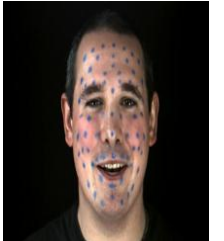
TABLE I: The effect of changeT parameter on the number of reduced set of frames in FrameR

change _T	0.10	0.15	0.20	0.25	0.30	0.35	0.40	0.45	0.50	0.55	0.60
Number of frames in Frame_R	207	200	183	137	95	50	26	8	2	2	1

Table II shows the values of the extracted feature vector for four different frames selected randomly from a sample video taken from the SAVEE database. Table II clearly showed that the

extracted features are varied with respect to the change in the facial expression and this reflects the effectiveness of the used features.

TABLE II: The extracted features for four different frames taken from a sample video

Example Frame	Extracted feature values					Example Frame	Extracted feature values				
	μ_1	μ_2	μ_3	μ_4	E		μ_1	μ_2	μ_3	μ_4	E
	0.255548047827502	0.294317462927024	0.321226534897292	0.301315701702757	0.356626629932555		0.613237756690863	0.619030853901325	0.512955562699044	0.505661265904333	0.500155354691561
	0.64742207467906	0.766159665029087	0.958487796715885	0.969948560938145	0.885683875966219		0.933499809042575	0.987224578092588	0.899975939004221	0.931898995152563	0.658343022601403

The degree of overlap ratio that is resulted during mining graphs of each emotion is also affected by the number of selected key frames. Tables III, IV and V shows the first ten smallest overlap ratio within each emotional class using minimum support equals to 70% and some key frames equal to 1, 5 and 10 respectively. As it is clearly shown in Tables III, IV, and V, the smallest overlap has resulted when C=10 which reflects the fact that more characteristics of each emotion class are captured and this leads to more stable sub-graphs.

TABLE III: The first ten smallest overlap ratio achieved within each class of emotion using graph mining when C=1

#	Angry		Disgust		Fear		Happy	
	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio
1	85	3.13%	464	31.67%	2863	18.13%	2060	16.88%
2	86	3.13%	289	37.29%	2864	18.13%	2061	16.88%
3	1256	5.83%	288	37.29%	2870	18.13%	2062	16.88%
4	1257	5.83%	436	37.50%	2871	18.13%	2063	16.88%
5	273	6.04%	435	37.50%	2878	18.13%	2524	16.88%
6	1244	6.46%	580	37.50%	2879	18.13%	2525	16.88%
7	1245	6.46%	592	37.5%	2940	18.54%	2528	16.88%
8	81	6.67%	463	39.17%	2941	18.54%	2529	16.88%
9	274	6.67%	207	39.58%	2947	18.54%	2530	16.88%
10	82	6.67%	462	39.58%	2948	18.54%	2531	16.88%

#	Natural		Sad		Surprise	
	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio
1	516	7.92%	643	11.67%	1569	18.13%
2	1073	7.92%	644	11.67%	1570	18.13%
3	1748	7.92%	654	11.67%	1632	18.13%
4	315	8.13%	655	11.67%	1633	18.13%
5	1457	8.13%	3231	12.08%	2995	18.54%
6	872	8.13%	3232	12.08%	2996	18.54%
7	1493	8.54%	1119	12.29%	3058	18.54%
8	1495	8.54%	1120	12.29%	3059	18.54%
9	1496	8.54%	1108	12.29%	1858	19.38%

10	1492	8.54%	1109	12.29%	2601	19.38%
----	------	-------	------	--------	------	--------

TABLE IV: The first ten smallest overlap ratio achieved within each class of emotion using graph mining when C=5

#	Angry		Disgust		Fear		Happy	
	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio
1	567	12.67	121	43.29	184	43.79	681	13.88
2	568	12.67	120	43.92	168	49.88	742	16.46
3	569	13.29	123	45.00	183	50.42	1001	18.83
4	570	13.29	122	45.13	132	50.42	713	23.29
5	46	14.00	42	47.54	324	50.88	714	23.29
6	47	14.00	43	47.54	100	50.92	712	23.46
7	580	15.29	45	47.92	11	51.13	715	23.46
8	581	15.29	46	47.92	271	51.46	687	23.63
9	591	15.38	40	48.17	184	43.79	688	23.63
10	592	15.38	41	48.17	168	49.88	697	24.58

#	Natural		Sad		Surprise	
	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio
1	116	20.71	181	36.00	57	17.96
2	117	20.71	186	36.46	58	17.96
3	123	20.75	185	36.54	59	17.96
4	124	20.75	189	36.54	60	17.96
5	119	20.83	182	36.63	260	19.54
6	120	20.83	180	36.71	261	19.54
7	122	20.88	817	37.00	262	19.54
8	125	20.88	818	37.00	263	19.54
9	152	20.96	187	37.17	158	19.71
10	153	20.96	184	37.25	159	19.71

TABLE V: The first ten smallest overlap ratio achieved within each class of emotion using graph mining when C=10

#	Angry		Disgust		Fear		Happy	
	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio
1	409	0.1	50	0.24	3	0.32	682	0.4
2	411	0.1	49	0.24	4	0.32	688	0.4
3	410	0.1	203	0.24	8	0.32	781	0.4
4	408	0.1	204	0.24	9	0.32	787	0.4
5	406	0.1	82	0.24	2	0.33	699	0.4
6	407	0.1	208	0.25	5	0.33	705	0.4
7	187	0.11	209	0.25	7	0.33	816	0.4
8	188	0.11	234	0.29	10	0.33	822	0.4
9	190	0.11	233	0.29	6	0.36	725	0.4
10	191	0.11	254	0.31	1	0.36	731	0.4

#	Natural		Sad		Surprise	
	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio	Sub-graph No.	Overlap Ratio
1	41	0.02	66	0.04	49	0.04
2	42	0.02	67	0.04	50	0.04
3	43	0.02	68	0.04	51	0.04
4	44	0.02	69	0.04	52	0.04
5	45	0.02	70	0.04	53	0.04
6	46	0.02	71	0.04	54	0.04
7	47	0.02	72	0.04	55	0.04
8	48	0.02	73	0.04	56	0.04
9	49	0.02	74	0.04	57	0.04
10	50	0.02	75	0.04	58	0.04

The final number of selected sub-graphs along with the final resulted overlap ratio for the different values of key frames is illustrated in Tables VI, VII, and VIII. Again, when C=10 a smaller set of sub-graphs are adopted (43 sub-graphs) with a total overlap ratio equals 0.19%. When C=1 the smallest total overlap is achieved but with the number of selected sub-graphs equal to 57 which is more than when C=10.

TABLE VI: The number of finally selected sub-graphs and the overlap ratio within each emotion when C=1

Emotion	Number of Selected Sub-graphs	Overlap Ratio
Anger	5	0.00%
Disgust	7	0.00%
Fear	11	0.01%
Happy	11	0.01%
Natural	7	0.04%
Sad	8	0.03%
Surprise	8	0.00%
Total	57	0.09%

TABLE VII: The number of finally selected sub-graphs and the overlap ratio within each emotion when C=5

Emotion	Number of Selected Sub-graphs	Overlap Ratio
Anger	10	0.05%
Disgust	12	0.05%
Fear	12	0.33%
Happy	12	0.24%
Natural	9	6.89%
Sad	12	0.14%
Surprise	12	3.86%
Total	79	11.56%

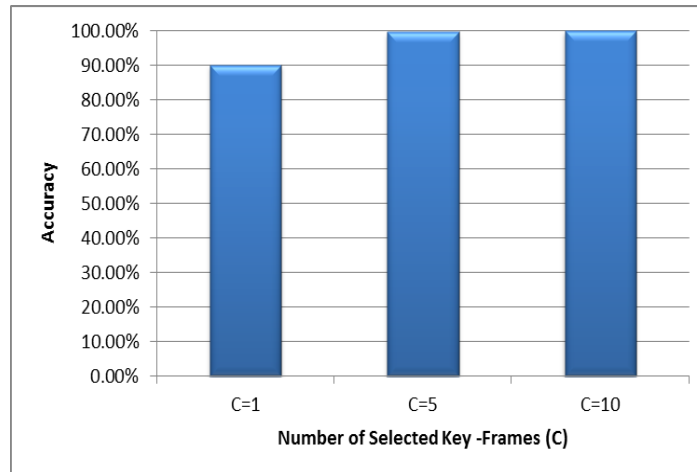
TABLE VIII: The number of finally selected sub-graphs and the overlap ratio within each emotion when C=10

Emotion	Number of Selected Sub-graphs	Overlap Ratio
Anger	7	0.00%
Disgust	9	0.00%
Fear	9	0.00%
Happy	8	0.09%
Natural	3	0.02%
Sad	3	0.04%
Surprise	4	0.04%
Total	43	0.19%

Levels concept is employed to predict the emotion class in levels. The number of selected features within each level for the different values of C is shown in Table IX. In addition, the final achieved accuracy after fusion the results of the different selected key frames are demonstrated in Table IX. The desired accuracy equals 100% achieved when C=10 while accuracy of 99.58% is reached when C=5. So, if the time is not critical, it is better to use the number of key frames equals to 10 because using more frames for predicting the final emotion class, more time is needed during the recognition task of each frame image. On the other hand, if the time is a critical point, it is recommended to use the number of key frames equals 5 with an error equals to 0.42 % which is considered nothing against the time. The graphical representation of the effect of the number of key-frames on facial emotion recognition accuracy is shown in Figure 3.

TABLE IX: The number of selected features for the different levels of classification for different values of C

C	Level1	Level2	Level3	Level4	Level5	Level6	Final Accuracy
C=1	31	27	27	30	30	28	90.00%
C=5	33	37	36	40	34	43	99.58%
C=10	23	20	24	25	20	20	100.00%

**Figure 3: The effect of the number of key-frames on facial emotion recognition accuracy.**

The response time is a very important factor in any face-based recognition system. Table X demonstrates the time consumed by key-frame selection (in seconds) for seven different videos selected from the used database with lengths ranging from one second to seven seconds. As shown in Table X, the proposed key-frame selection method can choose the best key-frames in one to two seconds depending on the length of the video. The estimated time clearly shows the possibility of using the proposed method as a preprocessing step in any face-based recognition system such as face recognition and facial emotion recognition systems.

TABLE X: The time estimated by the proposed method

Video Length (in seconds)	Total Number of Frames	Estimated Time (in seconds)
01	104	0.7822631
02	131	0.8812016
03	209	1.4513076
04	290	1.8860086
05	321	1.9732318
06	366	2.1990235
07	420	2.3879900

5. CONCLUSIONS

The effect of the number of key-frames on facial emotion recognition has been studied in this paper. The best C key-frames were selected by utilizing a fuzzy clustering-based approach. The dynamic information helps in speeding up the clustering process by focusing the work on a reduced set of frames. Also, the combination of frame reduction and clustering stages effectively assists in the selection of the most represented frames with a mean response time equals 2.89s. The best accuracy was achieved when the number of selected frames equals 10. As future work, other types of dynamic information can be used such as the correlation between the successive frames to further speed up key frame selection time.

References

- [1] H. Liu, "Key Frame Extraction Based on Improved Hierarchical Clustering Algorithm," 11th International Conference on Fuzzy Systems and Knowledge Discovery, Xiamen, China, 2014, pp. 793-797.

- [2] J. M. D., Rodriguez and W. Wan, "Selection of Key Frames through the Analysis and Calculation of the Absolute Difference of Histograms," International Conference on Audio, Language and Image Processing, Shanghai, China, 2018, pp. 423-429.
- [3] A. Fred and S. Wilson, "An Efficient Key frame Extraction Method in Video Based Face Recognition," International Journal of Computer Science (IJCS), Vol. 6, No. 2, pp. 27-33, 2018.
- [4] X. Qi, and S. Schuckers, "CNN Based Key Frame Extraction for Face in Video Recognition," 4th International Conference on Identity, Security, and Behavior Analysis, Singapore, 2018, pp. 1-8.
- [5] F. Noroozi, "Audio-Visual Emotion Recognition in Video Clips," IEEE Transactions on Affective Computing, Vol. 10, No. 1, pp. 60-75, 2019.
- [6] C. Lv, and Y. Huang, "Effective Keyframe Extraction from Personal Video by Using Nearest Neighbor Clustering," 11th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics, Beijing, China, 2018, pp. 1-4.
- [7] S. Shumin, Z. Jianming and L. Haiyan, "Key Frame Extraction Based on Artificial Fish Swarm Algorithm and K-means," International Conference on Transportation, Mechanical, and Electrical Engineering, Changchun, China, 2011, pp. 1650-1653.
- [8] A. A. Goshtasby, "Image Registration Principles, Tools and Methods," Springer-Verlag, London, 2012.
- [9] I. H. Witten and E. Frank, "Data mining: Practical Machine Learning Tools and Techniques," Elsevier, 2005.
- [10] C. Li and J. Yu, "A Novel Fuzzy C-Means Clustering Algorithm," International Conference on Rough Sets and Knowledge Technology, Lecture Notes in Computer Science, vol. 4062, Springer, Berlin, Heidelberg, 2006, pp. 510-515.
- [11] V. K. Malhotra, H. Kaur and M. A. Alam, "An Analysis of Fuzzy Clustering Methods," International Journal of Computer Applications, Vol. 94, No. 19, pp. 9-12, 2014.
- [12] N. F. Samatova, W. Hendrix, J. Jenkins, K. Jenkins and A. Chakraborty, "Practical Graph Mining With R," CRC Press, 2014.
- [13] A. K. Hassan and S. N. Mohammed, "A Novel Facial Emotion Recognition Scheme Based on Graph Mining," Defence Technology, DOI: 10.1016/j.dt.2019.12.006, 2019.
- [14] V. Kazemi and J. Sullivan, "One Millisecond Face Alignment with an Ensemble of Regression Trees," 2014 IEEE Conference on Computer Vision and Pattern Recognition, Columbus, USA, 2014.
- [15] Surrey Audio-Visual Expressed Emotion (SAVEE) Database Home Page, Available: <http://kahlan.eps.surrey.ac.uk/savee/>.