The Effect of Classification Methods on Facial Emotion Recognition Accuracy

Suhaila N. Mohammed[1*], Loay E. George[1] and Hayder A. Dawood[2]

[1]*Department of Computer Science, College of Science, Baghdad University, Baghdad, Iraq.*
[2]*Department of Computer Science, University of Technology, Iraq.*

Authors' contributions

This work was carried out in collaboration between all authors. All authors performed the statistical analysis, wrote the protocol, wrote the first draft of the manuscript, managed literature searches and the analyses of the study. All authors read and approved the final manuscript.

Original Research Article

ABSTRACT

The interests toward developing accurate automatic face emotion recognition methodologies are growing vastly, and it is still one of an ever growing research field in the region of computer vision, artificial intelligent and automation. However, there is a challenge to build an automated system which equals human ability to recognize facial emotion because of the lack of an effective facial feature descriptor and the difficulty of choosing proper classification method. In this paper, a geometric based feature vector has been proposed. For the classification purpose, three different types of classification methods are tested: statistical, artificial neural network (NN) and Support Vector Machine (SVM). A modified K-Means clustering algorithm has been developed for clustering purpose. Mainly, the purpose of using modified K-means clustering technique is to group the similar features into (K) templates in order to simulate the differences in the ways that human express each emotion. To evaluate the proposed system, a subset from Cohen-Kanade (CK) dataset have been used, it consists of 870 facial images samples for the seven basic emotions (angry, disgust, fear, happy, normal, sad, and surprise). The conducted test results indicated that SVM classifier can lead to higher performance in comparison with the results of other proposed methods due

_____

Corresponding author: E-mail: suhaila.najam@yahoo.com;

to its desirable characteristics (such as large-margin separation, good generalization performance, etc.).

1. INTRODUCTION

Facial expression is a visible manifestation of the affective state, cognitive activity, intention, personality, and psychopathology of a person [1]. Since each person has different cultural and surrounding background, it is difficult to precisely determine the number of facial expressions that are used daily in our life. However, researches on analyzing facial impressions have focused on seven basic expressive gestures which are: anger, disgust, fear, happiness, sadness, neutral, and surprise [2]. Till now, the recognition of facial expression still a problem with a big number of challenges. This is mainly due to the following reasons: (i) face detection from the captured image and segmentation process for the purpose of Region of Interest (ROI) extraction is a difficult task [3], (ii) Generating effective descriptors from ROI region (i.e., face region) that reduce the percentage difference between samples per class and increase between classes differences is an important and difficult step for building facial impression recognition system with high accuracy and (iii) difficulty of selection the classifier that classifies the face expression into correct emotional state [4].

In general, there are two main methods for facial emotion feature generation: geometric-based features and appearance-based features. Geometric features always present in the face but may be deformed due to any kind of facial expression. The basic components of the face or the points that represent facial features are extracted from the image spatial domain for purpose of classification can be used to get discriminating geometrical based features [5]. The appearance-based classification methods use features that 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). Transforms filters, such as Haar wavelets and integral image filters, are applied to ROI region, which either forms the overall face area or specific parts of the face region, to extract feature vector [6].

Classification is the procedure for discriminating class data from other data sources in the feature space. There are many methodologies in pattern classification; the most widely used are those based on statistical approach, syntactic approach, template matching, support vector machines, and neural network [7].

Anmar & Ahmed [8] have used Euclidian metric to measure the distance between feature vector that is generated using wavelets transform and Discrete Cosine Transform (DCT) and the feature vectors for the six basic facial expression images. Recognition rate of about 92.2% is achieved using 123 facial image samples taken from CK database for the six basic facial emotions (expect normal class).

Zhang et al. [9] have used ANN to recognize emotion class of the input image using two typical appearance facial features, Local Binary Patterns (LBP) and Gabor wavelets representations. Experiment results on 470 images samples taken from CK database gives accuracy of about 97.14% and 98.09%, for LBP and Gabor wavelets, respectively.

Hen et al. [10], have proposed the use of Gabor filter with SVM for face verification purpose. Clearly, when the proposed SVM is compared to the state-of-the-art methods such as PCA-SVM and Eigen face, it has outperformed these two methods with an error rate which equals to 6.19%. Wang et al. [11] have used SVMs to perform classification task on feature vector that is generated using Discriminant Laplacian Embedding (DLE) method. Six emotions from CK database have been used for performance evaluation, the attained test results were 87.4%.

In this paper, the effectiveness of three different classification methods (classical statistical method, neural network method and support vector machine method) on recognition abilities of facial emotion is discussed. The used features are of geometric-type representing various facial expressions.

2. MATERIALS AND METHODS

To build a facial emotion recognition system four main stages are needed: (i) face detection, (ii) Regions of Interest extraction (ROIs) stage, (iii) feature vector extraction stage, and (iv) classification stage.

2.1 Face Detection and ROIs Extraction Stage

To extract a set of good discriminating features from the allocated ROIs; first the face region is allocated using knowledge-based methods for face detection where the human knowledge about face structure and symmetry property is exploited. Face pattern is built and the search for this pattern is done. After detection of face region, the basic facial components (i.e. eyebrows, eyes, nose, and mouth) are extracted. Since the illumination condition and the skin and hair colors are highly different among tested dataset, adaptive contrast enhancement methods are used to extract ROIs regardless of these difficulties (A more detailed description about this stage could be found in the published paper [12]).

2.2 Feature Vector Extraction Stage

In this paper, a set of geometric based features is extracted using two types of geometrical operations; which are: the distance conveys between any two points and the angle that is convey among any three points.

A set of 26 facial points of interest is generated from the basic facial components (i.e., five points for each segment except the nose region where one point is taken to represent it). Fig. 1 shows the locations of these 26 points; the point locations are labeled using capital English alphabetic letters (from A to Z).

The possible distances and angles between these facial points are determined; in total they are 5525 possible features for both types (i.e., distance and angles).

For distance-based features, normalization process is applied to reduce the effects of face size and camera zooming variability. Normalization was performed according to the formula (1):

$$D_{jn}(i) = \frac{D_j(i)}{\mu(j)} \qquad (1)$$

Where, j is the index of the person subject, i is the index of the distance to be normalized, $\mu(j)$ is the mean of all distances determined between any possible two pairs of facial reference points extracted from subject sample ,and Dj(i), Djn(i) is the distance feature of index i for subject j before and after normalization, respectively.
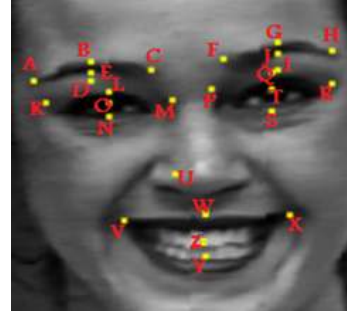

Fig. 1. Points of interest locations for basic facial components

2.3 Feature Vector Clustering

The next step to perform as part of feature extraction stage is feature clustering; this step is needed for the following reasons:

a. The single person can express a given emotion in different ways. Each time with a way that may differ from another time (as shown in Fig. 2a).
b. The way that person expresses a given emotion may differ from the way that another person expresses it (as shown in Fig. 2b).
c. For a given emotion there are different intensities to express it (as shown in Fig. 2c).

To obtain more than one template for each class, data clustering is needed. Data here means best N features that give highest discrimination abilities.

Different algorithms can be used for clustering purpose; one of them is K-Means clustering algorithm which clusters each class into fixed (K) subclasses number. In this paper modified K-Means clustering algorithm is proposed. The modification is based on the idea of minimizing clustering error in cascade way (i.e., at each step minimization is done by tuning each feature independently, such that the accumulative error over whole features is minimized). In mathematics form, for the feature vector of

sample (S_i) containing N features (f_{1i}, f_{2i}, f_{3i},……., f_{Ni}), the minimized accumulative mean absolute error for the features of a given class (j) from its mean vector (T_j) is calculated as following:

$$MAD(S_i, T_j) = Min(\sum_{k=1}^{N} |S_i(k) - T_j(k)|) \qquad (2)$$

In equation (3), the error is calculated for all features to form one value that represents total error (accumulative error), this mean if we need to minimize the error of clustering we need to minimize the total error that is resulted from all features together. The error can be calculated and minimized for each feature individually as following:

$$MAD(S_i, T_j) = Min(|S_i(1) - T_j(1)|) + Min(|S_i(2) - T_j(2)|) + ... + Min(|S_i(N) - T_j(N)|) \qquad (3)$$

By minimizing the error of each feature in individual fashion, we can effectively reduce the total clustering error for each class. So, we have exploited this propriety to enhance K-Means clustering abilities by reducing final clustering error for each class by swapping templates values for each feature within each class and the total error for that class is recalculated. If the new error is less than the old error then new templates values within that class is adopted. Algorithm 1 illustrates the steps of modified K-Means clustering algorithm.

Algorithm 1: *Modified K-Means Algorithm*

Input:
    FeatVect(): Array of feature vectors for all samples within a given class
    Fc: Feature count
    N: Number of iterations of K-Means
Output:
    tMean1(): Array holds temple1 values for all features
    tMean2(): Array holds temple2 values for all features
Begin
Step1: ///Calculate Initial Mean and Standard ///deviation values for each feature
   For all J **Do** {where 0≤J< Fc }
      Set tMean(J) ← Mean value of feature J within all samples
      Set tStd(J) ← Standard deviation value of feature J within all samples
   End For
Step2: ///Perform K_Means clustering for N times ///using initial two templates (tMean1 & ///tMean2)
   For all J **Do** {where 0≤J < Fc }
      Set tMean1(J)←tMean(J)-tStd(J)
      Set Mean2(J)←tMean(J)+tStd(J)
      Set tDis1←0, tDis2←0
      Set N1←0, N2←0
      For all itr **Do** {where 1<itr≤N}
        Set tDis1←0, tDis2←0
        Set N1←0, N2←0
        For all I **Do** {where 0≤I< FeatVect.Length}
          If Abs(tMean1(J)– FeatVect (I, J))<**Abs** (tMean2(J)–FeatVect(I,J))**Then**
            Set tDis1←tDis1+FeatVect(I,J)
            Set N1←N1+1
          Else
            Set tDis2←tDis2+FeatVect(I,J)
            Set N2←N2+1
          End If
        End For
        Set tMean1(J)←tDis1/N1
        Set tMean2(J)←tDis2/N2
      End For
   End For
Step3:/// Calculate total clustering error

```
   For all I Do {where 0≤I< FeatVect.Length}
      Set Dis1←0, Dis2←0
      For all J Do {where 0≤J<Fc}
         Set Dis1←Dis1+Abs(FeatVect(I,J)– tMean1(J))
         Set Dis2←Dis2 +Abs(FeatVect(I,J)– tMean2(J))
      End For
      If Dis1<Dis2 Then Set Dis←Dis1
                     Else Set Dis←Dis2
      Set OptDis←OptDis+Dis
   End For
Step4:///Minimize total clustering error by
        ///swapping templates values for each
        ///feature individually  and recalculate
        ///total clustering error
   For all J Do {where 0≤J< Fc}
    ///Swap feature templates values
      Set Ff←tMean1(J)
      Set tMean1(J))←tMean2(J)
      Set tMean2(J))←Ff
    ///Calculate total clustering error after
    ///swapping templates values of feature J
      Set Dis1←0, Dis2←0 , TotDis←0
      For all I Do {where 0≤I<FeatVect.Length}
         For all J1 Do {where 0≤J1< Fc}
            Set Dis1←Dis1 +Abs(FeatVect(I,J1)–tMean1(J1))
            Set Dis2←Dis2 +Abs(FeatVect(I,J1)–tMean2(J1))
         End For
         If Dis1<Dis2 Then Set Dis←Dis1
                        Else Set Dis←Dis2
         Set TotDis←TotDis+Dis
      End For
      ///If new total clustering error less than old
      ///total clustering error then save new total
      ///clustering error value
      If OptDis>TotDis Then
         Set OptDis←TotDis
      Else
         ///If new error greater than or equal to  ///the old error re-swap templates values
         Set Ff←tMean1(J)
         Set tMean1(J)←tMean2(J)
         Set tMean2(J)←Ff
      End If
   End For
Step5: End
```

This clustering algorithm is applied for each feature of 5525 possible features which are previously extracted and each feature is clustered into two templates within each class. More than two templates can be generated for each feature but in our work we have confined the list of templates to two templates per class to handle the clustering task.

2.4 Feature Dimensional Reduction

Dimensionality reduction means the transformation of high dimensional data into a meaningful representation of reduced dimensionality. Ideally, the reduced representation should have a dimensionality that corresponds to the intrinsic dimensionality of the data with minimum number of parameters needed to account for the observed properties of the data [13]. Fisher discriminant analysis (LDA) is an approach used for supervised dimensionality reduction and classification; it is based on classical statistical methods. LDA works on reduction feature dimensionality by simultaneously minimizing the within-class distance among each
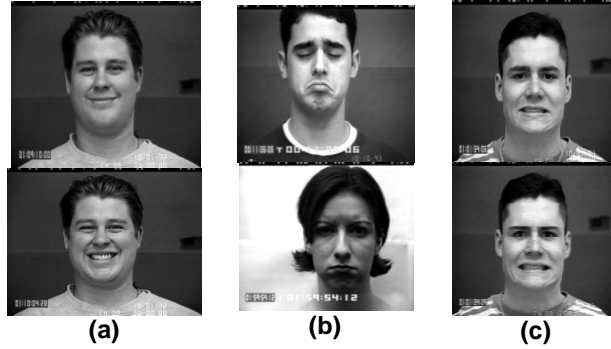
Fig. 2. Different cases in facial emotions expression: (a) two different ways to express "happy" emotion for the same subject, (b) two different ways to express "sad" emotion for two different subjects, (c) two different intensities to express "fear" emotion for the same subject

class samples and maximizing the between-classes distances among all classes, thus achieving maximum class discrimination by combining features with each other (with assumption there is two templates for each feature) and choosing feature combination that give best recognition ability [14].

LDA algorithm indicated that the features listed in Table 1 are the best ones from 5525 possible features (of distances and angles types).

Table 1. The proposed geometric features for representing the seven considered facial expressions

$F_{No.}$	Description	$F_{No.}$	Description
f_1	$\angle NJX$	f_{21}	$\angle WNX$
f_2	$\angle TVY$	f_{22}	$\angle IXW$
f_3	$\angle YAX$	f_{23}	$\angle YKV$
f_4	D_{IU}	f_{24}	D_{EU}
f_5	D_{WZ}	f_{25}	$\angle WFX$
f_6	$\angle VKT$	f_{26}	D_{MI}
f_7	D_{OV}	f_{27}	$\angle DXW$
f_8	D_{AN}	f_{28}	$\angle YPV$
f_9	$\angle VCY$	f_{29}	D_{GS}
f_{10}	D_{MO}	f_{30}	$\angle DBX$
f_{11}	$\angle AYZ$	f_{31}	D_{BU}
f_{12}	$\angle BTX$	f_{32}	D_{LF}
f_{13}	$\angle YLV$	f_{33}	D_{GP}
f_{14}	D_{GU}	f_{34}	D_{HI}
f_{15}	$\angle TSV$	f_{35}	D_{LX}
f_{16}	$\angle KTY$	f_{36}	$\angle MXW$
f_{17}	D_{AD}	f_{37}	$\angle QAV$
f_{18}	D_{DF}	f_{38}	D_{CK}
f_{19}	D_{HP}	f_{39}	D_{GT}
f_{20}	D_{IS}		

2.5 Classification Stage

In this stage three different classification methods have been analyzed to evaluate the recognition ability of the proposed feature vector.

2.5.1 Statistical classification method

In statistical classification approaches, the underlying probability model is exploited, which provides a probability of being in each class rather than simply a classification [8]. To compare patterns, this approach uses measures (such as Euclidian distance measure) by observing distances between points in this statistical space [7]. In this approach, patterns to be classified are represented by a set of features defining a specific multidimensional vector. By doing so, each pattern is represented by a point in the multidimensional feature space [15].

To perform classification task using this method, the distance of the feature vector for the input image from the templates of each emotions class is calculated using Euclidian distance measure and the class that shows lowest distance will be considered as final classification result. Taking in account, there are two templates for final feature vector within each class (total number of templates for all classes will be 14), so the Euclidian distance is applied for all possible templates of feature vectors as shown in Fig. 3.

2.5.2 ANN classification method

An ANN is an information processing paradigm inspired from the way that the biological nervous systems (e.g., the brain) process information. An ANN is configured for a specific application (for example, pattern recognition, function approximation, prediction, clustering and many others) [16].
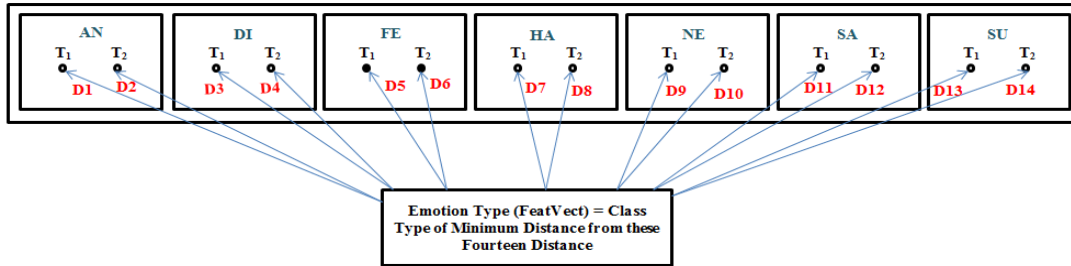
Fig. 3. Statistical classification method

Feed forward neural network with following structure is adopted (see Fig. 4):

No. of input layer nodes = 39
No. of hidden layer nodes = 20
No. of output Layer nodes = 3

The network is learned using supervised learning to change weights values of the feature vector for achieving the best representation of emotions classes. A set of feature vectors (for training set) is given to ANN during training phase and the suitable number of clusters (No. of hidden nodes in hidden layer) is chosen based on data structure, in our system the suitable number of clusters that give best accuracy was 20 nodes in hidden layer.
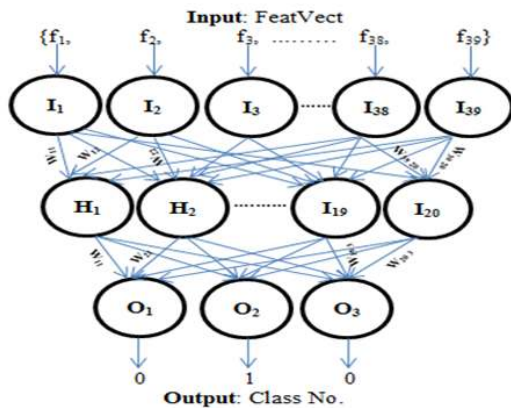

Fig. 4. ANN method

2.5.3 SVM classification method

Recently, many machine learning algorithms have been used in many applications of pattern recognition. In fact, Support Vector Machine (SVM) is the main choice to be used as a classification method in the last decade. The SVM classifier was presented by Cortes and Vapnik, [10], as a supervised machine learning algorithm. This algorithm shows that it has outperformed the state-of-the-art methods that were used for the same purpose [17]. This algorithm was proposed for binary classification which has two classes of classification, but this algorithm shows that it can be easily extended to multi-classes classification. The main point in this algorithm is to find the optimal hyper-plane that separates the classes that have positive +1 from the classes that have negative -1. Basically, when determining the boundary of the data, it can construct the hyper-plane (as shown in Fig. 5).

The line between the two dash lines is called the hyper-plane. The distance between these two lines is called margin and the point lying on the boundaries is called support vector. Apart from that, the boundary implementation is based on the complexity of the production function and a small subset of training vector is trained on the margin. Several researches have used SVM for classification purposes such as fingerprint, face detection, speech recognition, image retrieval system and hand writing. This is due to the fact that SVM has a lot of desirable characteristics which makes it very popular and used commonly.

Basically, different kernels have been proposed for SVM, namely Linear, Polynomials, Radial Basis Function (RBF) and Sigmoid. In fact, SVM is mainly based on two parameters namely Cost (C) and Gamma (γ), the best values will be achieved by using Grid search algorithm with search space $\{2^{-5}, 2^{-3}, ..., 2^{-15}\}$ and $\{2^{-15}, 2^{-13}, ..., 2^{3}\}$ for C and γ respectively on the training data. In addition, in this research one versus all classification schema have been used. Lastly, this research paper has implemented the SVM classifier based on the LIBSVM with its four kernels and grid search algorithm to find optimal SVM parameters [18]. The algorithm below illustrates the LIBSVM implementation.

3. RESULTS AND DISCUSSION

To analysis the performance of the proposed system using different classification methods, a subset consists of 870 facial images samples is taken from Cohen-Kanade (CK) database (from the last three frames). This set of samples holds the seven basic facial expressions which are neutral (NE), happy (HA), angry (AN), disgust (DI), fear (FE), sad (SA) and surprise (SU). The images are gray images (in png format), the size of each image used is either 640x490 or 640x480 pixels.

When Euclidian distance used for classification purpose with two templates for each emotion class, the recognition rate was reached to 83.22. Table 2 shows confusion matrix resulted when statistical classifier is used.

To evaluate ANN and SVMs classifier methods performance, the used dataset was divided into two subsets: training and testing. The training subset consists of about 75% of each class samples, taken, randomly. The remaining samples have been assembled as test samples.

After training ANN with training set and determining the weights of each feature in the final feature vector, the recognition ability of the proposed system is checked. The system classification ability is checked with different random training and testing sets, Table (3) shows accuracy results of best 10 runs. Table (4) shows ANN parameters (i.e. Learning Rate LR, No. of nodes in hidden layer HL, Momentum MOM, and Number of Epochs EP).
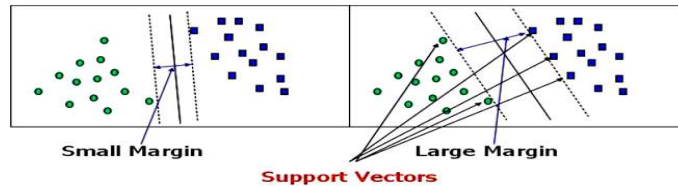

Fig. 5. SVM classifier used to separate two classes

Algorithm 2 : Steps in Applying SVM Classifier

Input:
    Train and test data
Output:
    Classification decision
Begin

- Make scaling to the training and testing data using the following equation $X = \dfrac{2*(X - Min)}{(Max - Min)} - 1$

- Apply Grid search method on the Scaled Training Data with K-fold cross validation =10 to find the optimal SVM parameters C and γ.
- Select one of the SVM Kernels to train the SVM based on the optimal C and γ and build the training model.
- Test the SVM model using testing data.

End

Table 2. Confusion matrix for seven emotion classes' classification using statistical classification

Class	AN	DI	FE	HA	NE	SA	SU
AN	47	8	6	0	7	5	0
DI	0	72	4	0	1	0	0
FE	1	1	106	3	2	6	0
HA	0	1	8	151	3	0	1
NE	5	6	13	0	130	9	0
SA	5	7	13	0	17	67	4
SU	3	0	0	0	2	5	151

Table 3. Facial emotion recognition accuracy based ANN with best 10 different random runs

Run	Accuracy %
1	91.72
2	91.61
3	90.8
4	90.34
5	90.22
6	89.89
7	89.77
8	89.66
9	89.31
10	89.31
Average	90.26 %
STD	0.82

Table 4. ANN parameters for the best 10 different random runs

Run	HL	LR	Mom	EP
1	20	0.1	0.4	4000
2	20	0.1	0.3	4000
3	20	0.1	0.1	4000
4	21	0.1	0.7	4000
5	20	0.1	0.1	4000
6	20	0.1	0.1	4000
7	20	0.1	0.7	4000
8	20	0.1	0.1	4000
9	20	0.1	0.1	4000
10	20	0.1	0.1	4000

As shown in Table 3 the best system accuracy with NN method from 10 random runs was reached to 91.72%. Table 5 shows confusion matrix for this run.

As it mentioned earlier, the SVM kernels are mainly based on two parameters namely Cost and Gamma, to find the optimal parameters, the grid search has been used for this purpose. In the experiments, the best parameters that achieved are $C = 32$ and $\gamma = 0.5$, with k-fold cross validation= 10. The table below illustrates the classification result of the SVM classifier with its four kernels.

Table 5. Confusion matrix for seven emotion classes' classification ANN classifier is used

Class	AN	DI	FE	HA	NE	SA	SU
AN	67	2	1	1	2	0	0
DI	1	70	2	0	0	4	0
FE	6	9	95	1	2	0	6
HA	1	0	0	161	2	0	0
NE	2	0	0	3	153	2	3
SA	0	2	1	3	7	95	5
SU	0	0	2	0	1	1	157

Table 6. Facial emotion recognition accuracy based different SVM kernels

SVM Kernel	C	γ	Accuracy %
Linear	32	0.5	78.281
Polynomial With 3rd Degree	32	0.5	88.235
Radial Basis Function (RBF)	32	0.5	92.308
Sigmoid	32	0.5	52.489

As it shown in Table 6, the Radial Basis Function kernel based SVM classifier gives higher level of accuracy when it compare with the other kernels. In addition, to verify the stability of the proposed system, 10 different random training and testing sets are tested with Radial Basis Function kernel based SVM classifier. Table (7), illustrates the achievement result of 10 different random runs.

Table 7. Facial emotion recognition accuracy based Radial Basis Function (RBF) with 10 different random runs

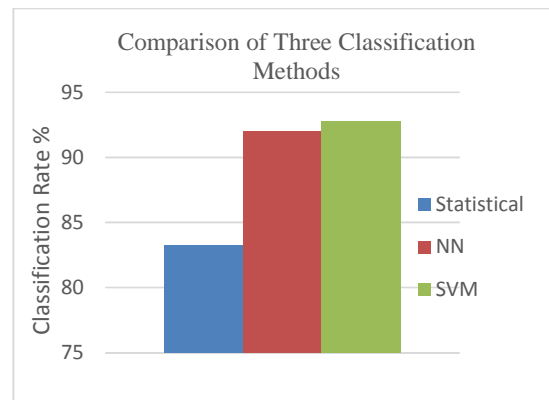| Run	Accuracy %
1	92.760
2	90.950
3	91.855
4	90.045
5	91.855
6	91.403
7	90.498
8	91.855
9	90.498
10	91.855
Average	91.855 %
STD	0.7954

Fig. 6. Comparison of the three used classification methods results

From Table 7, the average of 10 different random runs gives higher level of classification rate when it compare with the other state of the art algorithms. Besides, the standard deviation (STD) gives an indication that the proposed system is much stable and perfume very well under different random runs.

Fig. 6 gives a classification comparison of three proposed classification methods namely statistical, NN and SVM. Basically, the artificial methods denoted by SVM and NN outperform the Statistical method denoted by Euclidian Distance. On the other hand, the NN classifier gives higher level of classification rate due to its clustering ability but the SVM classifier outperform the benchmark methods with a stable results this is because of, SVM works well with high dimensionality data and its classification decision based on the located support vectors.

4. CONCLUSIONS

A new facial emotion recognition system is presented using geometric based features. LDA algorithm was used for dimensional reduction of feature space. To cluster each feature into K templates, modified K-Means clustering algorithm is presented to minimize the final clustering error. Three different classification methods are tested in the classification stage. A subset consists of 870 facial images samples from CK database is used to evaluate these methods. The experimental results show that the SVM classifier with Radial Basis Function outperform the benchmark classifier methods, this is due to the fact that the SVM has the ability to deal with the high dimensionality feature vectors and constructed a hyper-plan based on some points of the training data (Support Vector) not all the training data. As a future work more than two templates can be used in modified K-Means clustering and the results of classification methods are tested based on these feature clusters where more accurate results could be achieved especially for statistical analysis and SVM.

DISCLAIMER

The images used in this research belong to the public data base Cohn-Kanade (Ck) which can be downloaded from web site: The Affect Analysis Group At Pittsburgh (http://www.pitt.edu/~emotion/ck-spread.htm)

COMPETING INTERESTS

Authors have declared that no competing interests exist.

REFERENCES

1. Murty G, Kiran J, Kumar V. Facial expression recognition based on features derived from the distinct LBP and GLCM. Image, Graphics and Signal Processing. 2014;2:68-77.
2. Iatraki G. Emotional facial expression recognition & classification. Department of Media and Knowledge Engineering. Delft University of Technology; 2009.
3. Youssif AA, Asker WA. Automatic facial expression recognition system based on geometric and appearance features. Computer and Information Science. 2011;4(2):115-124.
4. Shan C, Gong S, McOwan PW. Facial expression recognition based on local binary patterns: A comprehensive study. Image and Vision Computing. 2009; 27(6):803-816.
5. Verma V, Sharma LK. A comprehensive survey on human facial expression detection. International Journal of Image Processing (IJIP). 2013;7(2):171-182.
6. Dhall S, Sethi P. Geometric and appearance feature analysis for facial expression recognition. International Journal of Advanced Engineering Technology. 2014;7(111):01-11.
7. Parasher M, Sharma S, Sharma AK, Gupta JP. Anatomy on pattern recognition. Indian Journal of Computer Science and Engineering (IJCSE). 2011;2(3):371-378.
8. Razzak AA, Hashem AR. Facial expression recognition using hybrid transform. International Journal of Computer Applications. 2015;119(15):12-18.
9. Zhang S, Zhao X, Lei B. Facial expression recognition using sparse representation. WSEAS Transactions on Systems. 2012;11(8):440-452.
10. Hen YW, Khalid M, Yusof R. Face verification with Gabor representation and support vector machines. First Asia International Conference on Modeling & Simulation (AMS '07); 2007.
11. Wang H, Huang H, Makedon F. Emotion detection via discriminant Laplacian embedding. Springer. 2013;440-452.

12. Mohammed SN, George, LE. Illumination invariant facial components extraction using adaptive contrast enhancement methods. British Journal of Applied Science & Technology. 2016;12(3). (In Press).

13. Nishana SS, Surendran S. Graph embedding and dimensionality reduction-A survey. International Journal of Computer Science & Engineering Technology (IJCSET). 2013;4(1):29-34.

14. Reif M, Shafait F. Efficient feature size reduction via predictive forward selection. Pattern Recognition. 2014;47: 1664–1673.

15. Michie D, Spiegelhalter DJ, Taylor CC. Machine learning, neural and statistical classification. ACM SIGART Bulletin; 1994.

16. Maind SB, Wankar P. Research paper on basic of artificial neural network. International Journal on Recent and Innovation Trends in Computing and Communication (IJRITCC). 2014; 2(1):96-100.

17. Ayad H, Abdulameer MH, George LE, Hassan NF. Descriptor trends in texture classification for material recognition. Research Journal of Applied Sciences, Engineering and Technology. 2015;10(10): 1089-1101.

18. Chang CC, Lin CJ. LIBSVM: A library for support vector machines ACM Transactions on Intelligent Systems and Technology; 2011. Available:http://www.csie.ntu.edu.tw/ ~cjlin/libsvm

_____