Research Journal of Recent Sciences ____ ISSN 2277-2502
Vol. **3(2),** 61-68, February **(2014)** *Res.J.Recent Sci.*

Computationally Efficient Invariant Facial Expression Recognition

Muhammad Hussain, Sajid Ali Khan, Nadeem Ullah, Naveed Riaz and Muhammad Nazir
Department of Computer Science, Shaheed Zulfikar Ali Bhutto Institute of Science and Technology, Islamabad, PAKISTAN

Available online at: www.isca.in, www.isca.me
Received 28th June 2013, revised 12th July 2013, accepted 13th August 2013

Abstract

The most important bottleneck for facial expression recognition system is recognizing the expression in uncontrolled environments with minimum computational time consumption. This problem has been addressed by combining the robust local texture descriptors which are invariant to illumination effects. In this work, the illumination effects are eliminated by using Weber Local Descriptor (WLD). Next, Local Ternary Descriptor (LTP) was introduced to preserved discriminatory local information, which is robust to noise. Both types of features are concatenated to produce more discriminatory feature set. The proposed technique is computationally efficient and gives very good results on JAFFE face database.

Keywords: Facial expressions, local ternary pattern, Weber local descriptor, feature level fusion, computationally efficient.

Introduction

Facial expression recognition is one of the major problems occurs in computer vision. For man that's so simple to recognize people weather he/she is upset, angry, sad and so on but is difficult for computer to recognize or differentiate these behavior. Facial expression classification deliberates an vital and major role in face recognition system. A lot of techniques introduced in the past year to classify.

In human-to-human conversation, facial expressions in addition to voice form a communication channel, which carries important information about the mental, emotional and physical state of the human being in conversation. Simplest form of facial expressions differentiates whether a person is happy or angry. In more detailed view, an intended or unintended feedback can be observed from facial expressions from listener to speaker to highlight understanding of, sympathy for or even some information about disbelief/belief toward what the speaker is saying. Recent research has shown that using facial expressions, one can check whether interrogated human being is attempting to mislead their interviewer.

Finding an efficient and discriminative facial emergence descriptor[1,2], which can be robust to noise, illumination, partial occlusions, poses and other changes, is one of the key challenges in face expression recognition[3] and face recognition[4]. Appearance feature-based descriptor and Geometric feature-based descriptor are the two main approaches. In geometric-based descriptors, features are extracted from different facial points while in appearance-based descriptors, features are extracted by using transformation techniques like Discrete Cosine Transform (DCT). Recently, representation based on local features capture much more attention because they capture all small important information. Methods in this group contain Local Binary Pattern (LBP)[5], Gabor wavelets[6] and local directional pattern[7].

The research on facial expression gets started with John Blur biological investigation on facial expression in 1640. Similarly, Darwin[8] and Bell[9] investigate the expressions of animal and human in 19th century. After this study, Darwin wrote a book title as "The expression of emotions in man and animal". Before 1970's mid, the analysis of an expression process gets an attraction of many research groups. A new scheme named as Facial Action Coding Scheme has been introduced by Ekmen and Frison in 1978. According to this scheme, the expression has been identified by the combination of 64 basic action units and also used the combination of AU's represents movements of facial muscles. Automatic facial data extraction is performed before mid 1990's by using facial motion analysis. A survey on automatic analyzing, recognizing human faces and expressions is presented Ashok et al[10]. According to Moses et el [11], mouth shape is the most important feature for identifying expression. An integrated system has been developed in Pantic M. and Rothkrantz L.J.[12] to recognize the expression from dual still view images. This proposed system identified 30 face images different actions and then categories them into six different expressions.

Xiaoye et al extracted facial appearance features by using local binary pattern[13]. They have divided their process into two stages. At first, stage, two candidate expressions have been selected. One of the two expressions is varied as a final expression at the second stage. Wallhoff et al have used self organizing map and holistic approaches for facial expression recognition[14]. They used Macro motion technique for feature extraction and SVM-SFFS for classification. Accuracy rate of 61.67% has been achieved by using FEEFTUM publicly available face database.

In 2009, Shenchuan et al[15] identified expression from video sequence. First median filter is used to reduce the noise. In the next step, mathematical models and cross correlation from the face points are used. In the last step, ELMAN neural is trained and tested to classify the expressions.

In section 2 of this article, we have presented a detailed literature review of some well-known facial expressions techniques. In section 3, we present our proposed algorithm and discuss its layout. In section 4 we present the results of the empirical evaluation and compare these with some of the well-known techniques. We conclude our work in Section 5 and also discuss the possible future work.

Literature Review

Human being interacts and interchanges their feeling and emotion using facial expression. This expression plays an important role in our daily life. In order to play with features some basic steps are to be taken and for these steps special procedures or techniques are used[16]. In first step input is taken in the form of image. After feeding these images to the system another step is to remove noise from images. This paper mainly sorts out the selecting of salience feature for expression recognition. Here we apply fixed filter and non linear 2-D filter. Fixed filter are use to extract main and simple feature while adaptive filter are use to extract complex type of features. The main is on feature selection investigation, and to reduce the mutual information between selected features. They use JAFFE dataset for testing purpose. Using the above technique they got 96.7% classification rate (CR).

The author[17] of this paper presents paper good idea about facial expression is based on static images which is combination of both fixed filter and adaptive 2-D filters as hierarchical position. Both filter has own goal in this context. Main feature are used to gain and extract main features. After apply fixed filter adaptive filters are trained on dataset to extract and get more complex features. After apply the above filters than support vector machine is applied to classify the extracted feature. The proposed approach is checked on the JAFFE database. These dataset have seven types of different facial expressions: anger, fear, disgust, happiness sadness, neutral and surprise. Over all flow o this paper is as below.

AAM is use to extract the facial feature vector by passing images to system, than convert facial feature vector into its related neutral facial expression as a vector either using direct transformation or indirect facial expression transformation. At last, distance-based matching techniques is used to check nearest neighbor classifier along with combination linear discriminate analysis (LDA). A face is located at head used to get contour points using its related motion information. Eyes are use as features to get face size. Visual features are use as modeled for support vector machine (SVM) to gain facial expression recognition. The following technique is use here in this paper. Support vector machine, fixed and adaptive 2-D filters.

In Thiyagarajan R. and Arulselvi S.[18] presents his idea 3D images rather than 2D images. Recently most of the work is done on 2D images which were not successful result on variant pose, illumination and expression. This paper presents a high performance face recognition system for database images. Geometrics techniques has use with two types of neural networks; multilayer perceptron and probabilistic. The recognition is performed by 3D facial recognition. Geometrics technique is also used for analyzing and mapping the face and to relate each selected human face point in space with 3D coordinates (X, Y, and Z). Photo Modeler is used to perform the coordinate of specific points that used to specified geometrics parameters for face recognition. The neural network is used for making the required recognition. There is some technique which is used in this paper Neural network, Geometrics techniques, Photo Modeler, Image processing, Biometric recognition, Photogrammetric technique.

Raheja and Kumar[19] mainly focus on average half face rather than using full image of face. Overall system accuracy of system is fine and well in result as compare to full face images in place of half face image. Whenever they pick half face of an image than computation and saving of resultant image point will be fast and computation will perform fast due to lest data and feature point which use in computation. The information stored in average-half-face may be more clear and easy to get specific portion face identification, which is well especially in 3D databases. Average half face recognition based on the extraction of facial fudicial points such as head, nose and ear and measuring the Euclidean distance between these features using Elastic bunch Graph matching algorithm. In this facial fudicial features on the face are head, nose and ear which are described by set of wavelet (jets) components. Image graph is a bunch graph, which is constructed between the jets. Recognition is based on the Euclidean distance measurement using bunch graph. The distance is considered as a unique factor for the specific features for each person.

Proposed Algorithm: Figure-1 depicts presents an overview of proposed system architecture. Two types of face features are extracted using Weber Local Descriptor (WLD) and Local Ternary Pattern (LTP). More powerful feature's sets are generated by fusing both, WLD and LTP features.

Weber Local Descriptor: One of the newly built-up, strong and authoritative confined descriptor that is rapidly gaining popularity nowadays is WLD[20]. Gradient orientation and differential excitation are the two main components of WLD. A psychosomatic decree called "Weber's Law" stimulated WLD. It restrains numerous advantages like, being vigorous against elucidation alteration, ability to unvaryingly haul out the boundaries of an image even in the presence of profound noise Weber's law states that any alteration in stimulus (like lighting, sound) is proportional to the original stimulus. If the change is lesser than the constant ratio, then a human being would identify it as surroundings noise instead of a valid signal.

Differential Excitation: Initially, WLD calculates its differential excitation component, by manipulating the percentage sandwiched between the summation of strength variation of the center pixel next to its adjoining pixels and the intensity of the center pixel. Equation 1 represents it mathematically;

$$\xi(x_c) = \arctan\left[\sum_{i=0}^{p-1}\left(\frac{x_i - x_c}{x_c}\right)\right] \tag{1}$$

In equation 1, x_c denotes the center pixels, and $\xi(x_c)$ is the differential excitation of center pixel. x_i, where i = 0, 1,2,3…. P-1; is the *ith* a neighbor of x_c and p is the number of neighbors.

Gradient Orientation: The following steps are used to calculate the gradient orientation for the center pixel[20];

$$\theta = \arctan\left[\frac{v_{10}}{v_{11}}\right] \tag{2}$$

Where $v_{11} = x_7 - x_3$ and $v_{10} = x_5 - x_1$

$f : \theta \longrightarrow \theta'$ (Mapping)

$$\theta' = \arctan 2(v_s^{11}, v_s^{10}) + \pi \tag{3}$$

$$\arctan 2(v_s^{11}, v_s^{10}) = \begin{cases} \theta & v_s^{11} > 0 \text{ and } v_s^{10} > 0 \\ \pi + \theta & v_s^{11} > 0 \text{ and } v_s^{10} < 0 \\ \theta - \pi & v_s^{11} < 0 \text{ and } v_s^{10} < 0 \\ \theta & v_s^{11} < 0 \text{ and } v_s^{10} < 0 \end{cases}$$

Where $\theta \in \left[\frac{-\pi}{2}, \frac{\pi}{2}\right]$ *and* $\theta' \in [0, 2\pi]$

In this step, below function is used to quantize the θ'.

$$\Phi_t = f_q(\theta') = \frac{2t}{T}\pi \quad \text{Where } t = \mathrm{mod}\left[\left(\frac{\theta}{2\pi/T} + \frac{1}{2}\right), T\right] \tag{4}$$

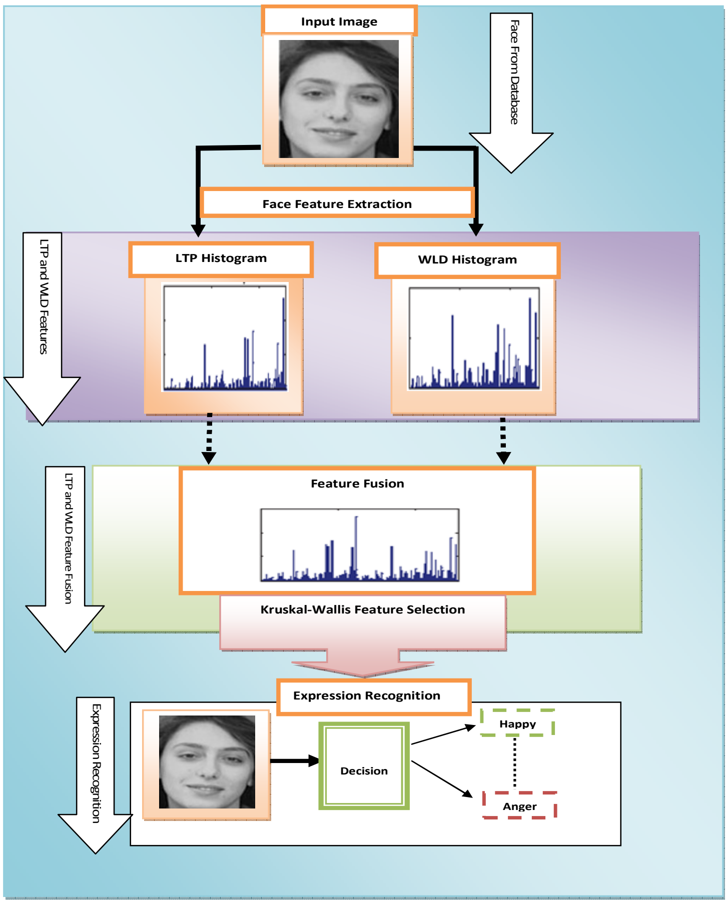

Figure-1
Proposed technique flow diagram

WLD Histogram: In this step, Differential excitation and gradient orientation are used to compute the histogram.

Histogram is $\{\Phi_t, \xi(x_c)\}$, Where t= 0, 1, 2,… T-1, c = 0, 1, 2… N-1. T represents number of dominant orientations and N represents the dominant differential excitations. WLD histogram is converted to row vector of size $W = T \; X \; N$.

Local Ternary Pattern (LTP): LBP is consistent with gray-level transformation due to which it resists to lighting effects. LBP is sensitive to noise because it threshold to center value.

Local Ternary Pattern (LTP) [21] is used to extract the face features. LTP is the extended form of LBP. The gray-level values around i_c and in the area of width $\pm t$ are converted to zero, once above these are converted to +1 and below are converted to -1.

$$s'(u, i_c, t) = \begin{cases} 1, & u \geq i_c + t \\ 0, & |u - i_c| < t \\ -1, & u \leq i_c - t \end{cases}$$

In equation 5, t represents the threshold value which makes the LTP resilient to noise. Figure-2, depict the LTP encoding procedure in which the threshold value is set to 5.

To make it more simplify the Ternary code is further divided into negative and positive parts, which are illustrated by the figure. After division into negative and positive portion, they will act like LBP separate channel descriptor for which separate metrics and histogram will compute and will combine at the end.

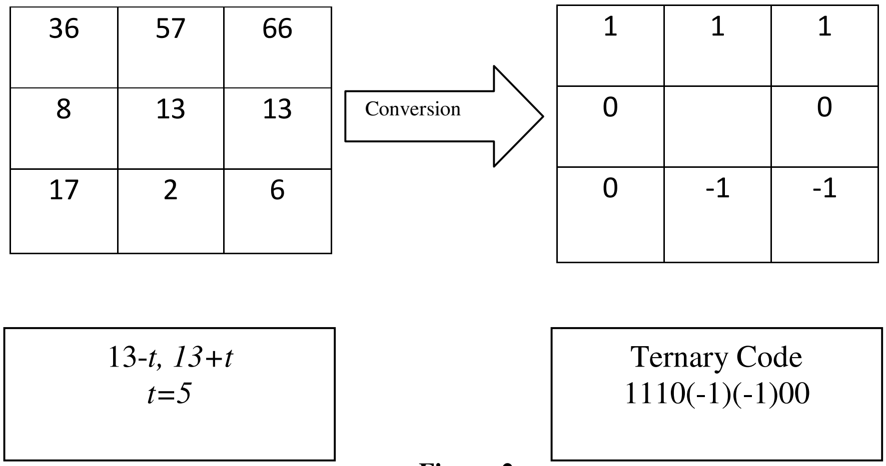

Figure-2
Proposed technique flow diagram

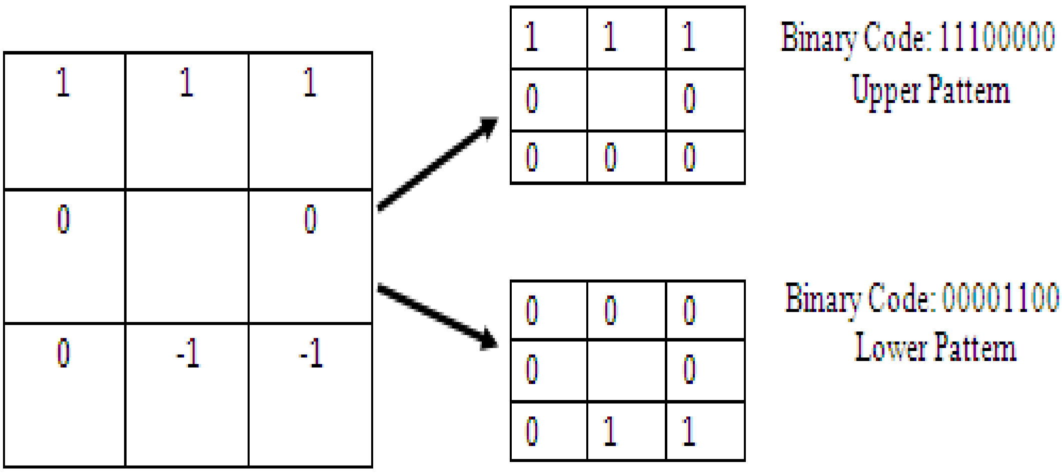

Figure-3
LTP upper and lower patterns

In figure-3, the Local Ternary Pattern is converted into upper and lower level pattern. In the next stage histogram is generated using LTP-based extracted features.

Fusion of LTP and WLD features: The features extracted using LTP and WLD do not encode the same information. After performing different experiments, the 15% error rate has been observed for both LTP and WLD. This specifies that fusion of LTP and WLD features might lead to high recognition rate. Feature level fusion strategy is used by concatenating the LTP and WLD histograms. The figure-4, represents the WLD and LTP histograms, before and after fusion.

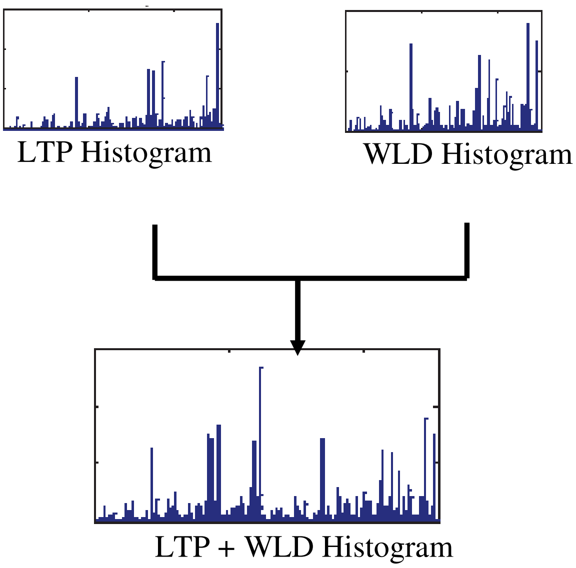

Figure-4
LTP and WLD histogram concatenation

Discriminative Feature Selection: Some of the features would possibly not contain adequate discriminative data that is why they can be ranked in lower recognition performance. Furthermore, handling many characteristics degrade the categorization procedure. Solution for this issue is to select the optimize features. Different feature selection techniques have been proposed by Huan L. Hiroshi S. Rudy S. and Zheng Z.[22], but these are usually complex in nature and are computationally expensive. In this work, we have used Kruskal-Wallis method[23] to select the most prominent features. The Kruskal-Wallis method is simple in nature and has a low computational cost. It has the ANOVA (analysis of variance) functionality that can be applied to two or more classes. According to Kruskal-Wallis's hypothesis, it checks whether two or more classes have equal median and return the value of p. Features with discriminative information are selected if for a certain feature, the value of p is near to zero. If the value of p is far from zero, then the feature will be of low quality and will not be useful for recognition.

LTP and WLD fused features are processed using Kruskal-Wallis's technique. Those features are selected for the next recognition steps which gave a value of p less than the threshold. The selected features are of more importance and contribute more for face recognition.

Facial Expression Classification: After feature extraction, the subsequent most vital task is to have a proper classifier that is fast and robust to any particular problem. This section describes the architecture of classifier. There are several classifiers that can be utilized for multi-classification problems but there is a demand of such a classifier or a combination of classifiers that can effectually categorize the facial expression alongside elevated accuracy.

It has been noticed that performance is significantly improved after combining different binary classifiers as compared to single classifier. In this work, bank of neural network is design to improve the accuracy rate. First, results from different classifiers have been obtained to decide that which classifier would be the best one. It is similar to a daily life decision where we search different option before making any decision[24]. Maximum rule is used to combine the response of different classifier (i.e. every single classifier contains information related to facial expression). The combination of different classifier can be efficiently used for facial expression recognition. Figure-5 illustrated the design of a bank of neural network. Figure 5, depict the Neural Network training phase phenomena.

Results and Discussion

In this paper, two different feature extraction techniques have been implemented (i.e. LTP and WLD). Neural Network is used for training and testing purpose.

JAFFE database has been used for all experiments which contain 213 images of female with different facial expression of size 256x256. Each image corresponds to the following seven categories of expressions. (i.e. Fear, happiness, neutral, sadness, disgust, anger and surprise). All the images are frontal and of grayscale in a tiff format.

The arrangement utilized to attain the pictures in the database consisted of a camera climbed on a table and encircled in a box. The user-facing side of the box had a semi-reflective plastic sheet. Every single subject seized a picture as looking towards the camera and looking at the reflective sheet. Tungsten lights were positioned in order to craft an even illumination result on the face. The actual terms of the subjects are not exposed but they are denoted alongside their initials: KA, KL, KM, KR, MK, NA, NM, TM, UY and YM. Each picture in the database was rated by 91 people for degree of every single of the six frank expressions present in the image. The semantic locale of the pictures displayed that the error for recognizing the fear expression was higher than that of any other expression. This displays that even humans cannot guarantee the 100% correct credit of expressions.

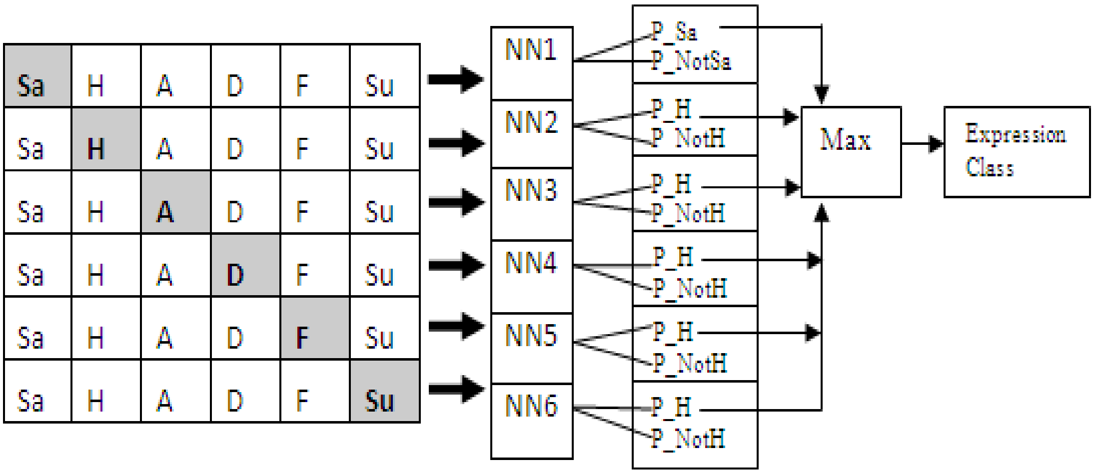

Figure-5
Neural Networks training

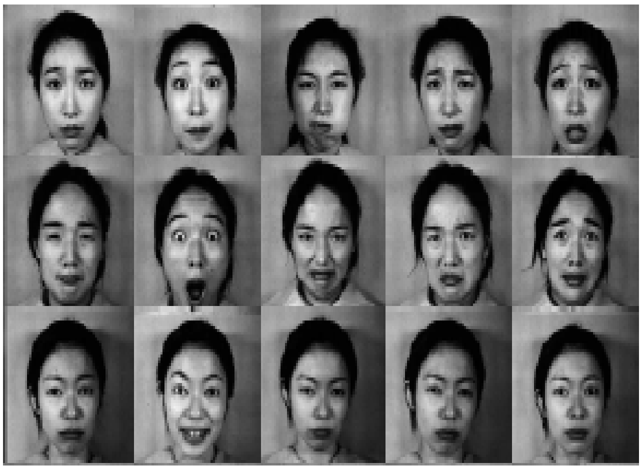

Figure-6
JAFFE face database images

Six expressions were considered for experiments from the database.

In, WLD[21] the three parameter that greatly affect the technique performance are optimized. These parameters include the number of differential excitations (N) and the number of dominant orientations (T). We found out that there is a chance of finding more discriminatory features if the size of these parameters is large. Features are more reliable if the parameter size is small. In our experiment, different values (T = 5 to 8 and N = 4 to 6) of T and N were used. Feature sets of distinct sizes (FS-1 30, FS-2 40 and FS-3 50) were generated using different values of T and N.

In, LTP the t parameter value is set to 0.1 to 0.2. The same size of feature set was generated in the LTP case. Figure-7, depict individual WLD and LTP accuracy rate using different number of feature sets. On average, the WLD features are 75% correctly classified while 72% accuracy rate has been achieved in the LTP case.

Next, Feature level fusion is performed to concatenate both types of local features. After fusion process the size of the feature vector increases (i.e. FS-1 60, FS-2 80, FS-3 100.

After fusion, the dimensions of the feature set become larger that degrades the recognition performance in terms of computational complexity. Kruskal-Wallis feature selection technique is used to select the important features which contribute more in a face recognition task. In Kruskal-Wallis's algorithm the value of p is changed to eliminate the redundant features and find the optimum threshold. The variation [0.01 ~ 0.19] and [2.0 of 4.0] of p was followed in the experiments. The dimensions were decreased and the number of FS-1 40, FS-1 60, FS-3 80 optimal features were selected using Kruskal-Wallis's algorithm. These features contained more discriminative information about-face and produced high recognition rate.

In the last classification step, One-against-all approach is used for all the experiments. As the number of expressions is six, so six binary NN classifiers are trained using fused features. Output of these classifiers is in the form of probabilities, which represents that to which extent the particular expression is belonged. All the NN outputs are combined using maximum rule.

Table-1 represents each facial expression accuracy rate after fusion in the form of confusion matrices. On average 83%, accuracy has been obtained. It has been noted that the accuracy rate is an increase significantly 8% after feature fusion.

Normally face local feature extraction techniques are computationally expensive but the techniques used in this paper are taking less time to compute the features. Computational time (Sec) of the proposed technique is presented in table-2.

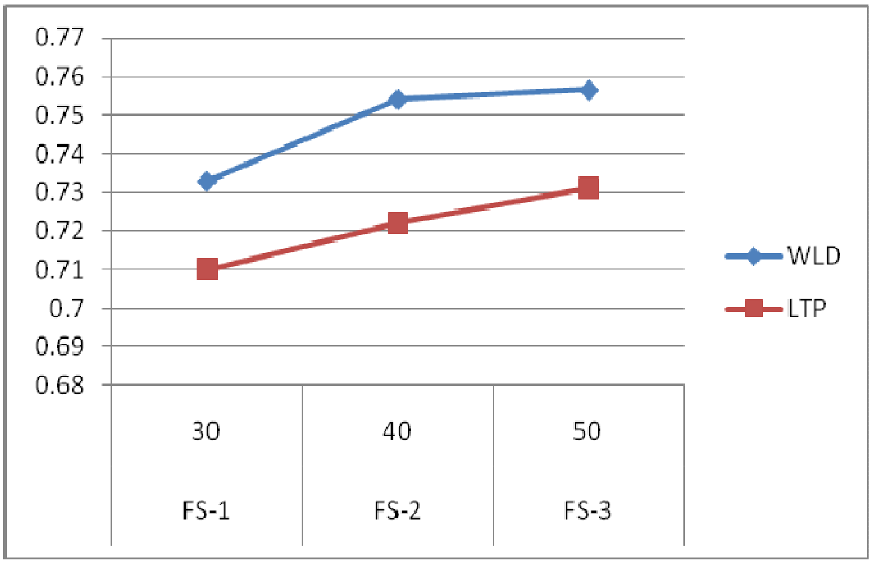

Figure-7
WLD and LTP accuracy rate

Table-1
Expressions classification accuracy rate

	Sa	H	A	D	F	Su
Sa	**86.3**	5.2	1.3	4.2	2	1
H	3.6	**76.7**	3.7	4.6	9.2	2.2
A	4.4	3.3	**84.2**	5.2	0	2.9
D	6.7	2.3	1.2	**83.4**	6.4	0
F	3.4	7.5	4.1	4.4	**73.4**	7.2
Su	3.1	0	1.2	1.2	0.3	**94.2**

Table-2
Proposed modules time complexity using WLD and LTP features

Proposed Modules	**WLD**	**LTP**
Feature Extraction	5.4	7.2
Expression Classification	4.4	4.7
Overall process	9.8	11.9

Conclusion

In this paper, two local feature extraction descriptors are used to extract the local face features. Further most, the following conclusions have been drawn after experiments; Weber Local Descriptor contains important information related to face. LTP features are robust to noise and different facial variations. Overall accuracy rate increases after performing feature level fusion. Feature dimensions are dramatically decreases after applying Kruskal-Wallis algorithm. Kruskal-Wallis feature selection algorithm is computationally en-expensive as compared to other searching algorithms. Combination NN outputs enhance the accuracy rate as compare to single classifier. We are planning to investigate more accurate feature extraction techniques. Those classifiers which training time is minimum will also tested in future.

References

1. Zahedi M. and Zahra M., A Fully Automatic and Haar like Feature Extraction-Based Method for Lip Contour Detection, *Res. J. Recent Sci,* **2(1)**, 17-20 **(2013)**

2. Hameed M., Shahrif M., Raza M. and Iqbal M., Framework for the Comparison of Classifiers for Medical Image Segmentation with Transform and Moment based features, *Res. J. Recent Sci,* **2(6)**, 1-10**(2013)**

3. Zhao W. Chellappa R. and Phillips P.J., Face recognition: A literature survey, *ACM Computing Survey*, **34**, 399-487 **(2003)**

4. Jozef B. Matej F. Milos O. and Jarmila P., Non-conventional Approaches to Feature extraction for face recognition, *Acta Polytechnica Hungarica*, **8**, 75-90 **(2011)**

5. Wencheng W. Faliang C. Jianguo Z. and Zhenxue C., Automatic facial expression recognition using local binary pattern, Proceedings of the International conference on Intelligent Control and Automation, 1056-1064 **(2010)**

6. Shishir B. and Ganesh R., Recognition of facial expressions using Gabor wavelets and learning vector quantization, *Journal of Engineering Applications of Artificial Intelligence*, **21**, 1056-1064 **(2003)**

7. Wencheng W. Faliang C. Jianguo Z. and Zhenxue C., Facial expression recognition using Local Directional Pattern (LDP). Proceedings of the IEEE International conference on Image Processing, 1605-1608 **(2010)**

8. Darwin C., The expression of the emotions in man and animal, *J.Murray*, London **(1872)**

9. Bell C., Essays on the anatomy of expression in painting, Longman, Reese, Hurst & Orme, London, third edition **(1806)**

10. Ashok S. and Prasana A.I., Automatic recognition and analysis of human faces and facial expressions, A survey, *Journal of Pattern Recognition*, **25**, 65-77 **(1991)**

11. Moses Y. and Raynard D., Blake: Determining facial expressions in real time. Proceedings of Fifth International Conference on Computer Vision, 296 **(1995)**

12. Pantic M. and Rothkrantz L.J., An expert system for multiple emotional classification of facial expressions. Proceedings of 11th IEEE International Conference on Tools with Artificial Intelligence, 113 **(1999)**

13. Xiaoye F., Facial expression recognition based on local binary patterns and coarse-to-fine classification, Proceedings of Fourth International Conference on Computer and Information Technology, 178-183 **(1999)**

14. Wallhoff F. Schuller B. Hawellek M. and Rigoll G., Efficient recognition of authentic dynamic facial expressions on FEEDTUM database, Proceedings of Conference on Multimedia and Expo, 1026-1033 **(2009)**

15. Shenchuan t. and Hungfu H., Facial expression recognition in video sequences, Proceedings of 6th International Symposium on Neural Networks, 16-30 **(2009)**

16. Khandait S.P. Thool R.C. and Khandait P.D., Automatic Facial Feature Extraction and Expression Recognition based on Neural Network, *International Journal of Advanced Computer Science and Applications*, **2(1)** 113-118 **(2011)**

17. Li P. Phung S.L. Bouzerdoum A. and Tivive F.H.C., Improved facial expression recognition with trainable 2-D filters and support vector machines, Proceedings of 6th International Conference on Pattern Recognition, 3732-3735 **(2010)**

18. Thiyagarajan R. and Arulselvi S., Design a Facial Recognition System Using Multilayer perceptron and Probabilistic Neural Networks Based Geometrics 3D Facial, *International Journal of Image Processing*, **4(6)**, **(2008)**

19. Raheja J.L and Kumar U., Average Half Face recognition by elastic bunch graph matching based distance measurement, *International Journal of computer science and information technology*, **2(1) (2010)**

20. Tan X. and Triggs B., Enhanced local texture feature sets for face recognition under difficult lighting conditions. IEEE Transactions on image processing, **19(6)**, 1635-1720 **(2010)**

21. Chen S.J. Shan S. He C. Zhao G. Pietikainen, M. and Gao W., A robust local image descriptor. IEEE Transactions on *Pattern Analysis and Machine Intelligence*, **32**, 1705-1720 **(2010)**

22. Huan L. Hiroshi S. Rudy S. and Zheng Z., Feature selection: An ever evolving frontier in data mining. Proceedings of Fourth Workshop on Feature Selection in Data Mining, 4-13 **(2010)**

23. Yvan S. Inaki I. and Pedro L., A review of feature selection techniques in bioinformatics, *Bioinformatics*, **23**, 2507-2517 **(2010)**

24. Polikar R., Ensemble based systems in decision making, *IEEE Circuits and Systems Megazines*, **6(3)**, 21-45 **(2006)**