A Study on Recognition of Expressionsusing Features of Facial Landmarks

Dr. K.N. Venkata Ratna Kumar[1] K.Srinivasa Rao[2] D.Subba Rao[3]

[1]*Asst. professor-Department of Electronics and Computer Engineering, Wolaita Sodo University, Ethiopia*

[2]*Associate Professor & Principal-Diploma, BRIG Engineering College, Opp. Ramoji Film City, Hyderabad, Telangana, India*

[3]*Associate Professor & Head-Dept. of CSE, A.M. Reddy Memorial College of Engineering & Technology, Narasaraopeta, Andhra Pradesh, India*

Abstract - Facial expressions are one of the most powerful, natural and immediate means for human being to communicate their emotions and intensions. Recognition of facial expressions has many applications including human-computer interaction, cognitive science, human emotion analysis, personality development, etc. In this study, it is aimed to use appearance-based features obtained from the landmarks for instant facial expression recognition. In the study, the local binary pattern (LBP) attributes obtained from the surrounding of the landmarks using active shape models are used. In order to find the most discriminating subset of the obtained attributes, the selection of the attributes has been applied for improving the recognition rate.

Keywords: *Facial expression recognition; Local Binary Patterns; Feature Selection; Cohn- Kanade Dataset; Sequential Forward Feature Selection.*

I. INTRODUCTION

Muchof research interest has been attracted by Artificial intelligence because of its applications in ability to recognize human emotions. Human vision is easily replicated by computers due to the emergence of robust Artificial Intelligence algorithms. Computer learns human vision and performs necessary action to get accurate output. A computer plays key role in human computer interaction by recognizing facial expressions. Psychological characteristics such as heartbeat and blood pressure, speech, hand gestures, body movements, facial expressions are termed to be as emotions of humans. Facial expressions are more effective among all these characteristics. Mehrabian's research indicates that facial expressions convey 55% of a message in a face-to-face communication. A facial expression is a change in position of muscles beneath the facial skin. These movements on face indicates the emotional state of an individual. Facial expressions refers to very powerful non-verbal communication. Anger, disgust, fear, happiness, sadness, and surprise are identified to be as the six basic facial expressions[1]. These six expressions are broadly categorized into *positive* and *negative emotions*. *Surprise* and *happy* emotions are included under positive emotions and *fear, sad, angry, disgust* expressions are included under negative emotions. By observing facial features and facial muscle movements, one can identify whether the individual is in pain or frustrated or happy or in any other emotion. Many applications such as Medicine, E-Learning, marketing research, data-driven animation, interactive games, entertainment, etc. make use of automated facial expression recognition systems. The architecture of facial expression recognition system is shown in Fig.1below:

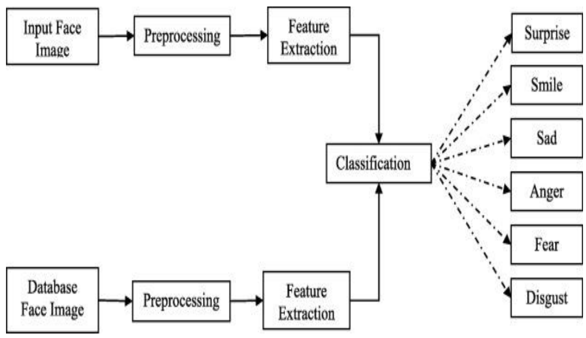

Fig-1. Architecture of Face Expression Recognition System

Convolutional Neural Network (CNN), which is a deep learning method is highly successful for outstanding classification in the area of image classification and recognition. CNN has the ability of automatic feature extraction and translation invariance which makes it feasible neural network for image classification. Each Layer in CNN extracts unique feature from the given input image which crafts it as a more powerful neural network. The objective of the Convolutional neural network is to transform a set of inputs into accurate and meaningful outputs. The Local Binary Pattern Convolutional Neural Network is employed in the research to achieve maximum efficiency. Cohn_Kanade Facial Expression Database (CKFED)[2] is a standard database used as training and testing data for expression recognition.

In our day to day communication with other people, facial expression is an essential part in the means of emotion recognition towards others. Human beings demonstrate facial expressions when experiencing emotion. The human spectrum of emotion, having different intensities or levels can be shown in a lot of ways, e.g., verbal and non-verbal communications, however, facial expressions remain to be the main channel. Facial expression is described as the positions of the muscle beneath the skin of the face. Studies pointed out that humans are capable of showing six basic human emotions and a neutral state. The seven basic human emotions are anger, disgust, fear, joy, neutral, sadness and surprise as shown in Fig. 2 :

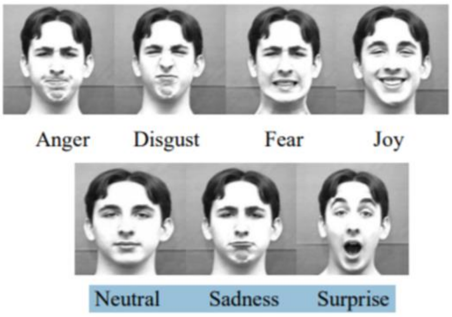

Fig-2.The seven basic human emotions

Emotion Recognition via Facial Expressions (ERFE) studies are increasing from the last two decades since the technology is rapidly advancing along with it. ERFE is mainly being applied in intelligent human-computer interaction, safety, medicine [3], affective computing and is deeply studied due to its non-intrusive method of recognition. Though advancement has been made in both hardware and software, recognizing facial expressions with high accuracy remains challenging as there is much variability and complexity in each face. This is especially true with "in the wild" facial expression recognition. The "in the wild" facial expression recognition is classifying facial images outside the closed-set of classes or the training classes. This is a challenge to facial expression recognition systems which will most likely result in misclassification. This is the objective that many studies have been working on as it is problematic to train every system with all possible facial feature [4]. The goal of this study is to develop frameworks that can be used in facial expression recognition. To achieve this objective, this study determines which feature descriptor will best fit a respective machine learning algorithm to classify facial expressions by (1) using and examining different feature extraction methods and combinations of the methods as feature descriptors, (2) classifying the images using the feature descriptors by utilizing several machine learning algorithms, and to evaluate their respective performances and (3) determining how the models will perform "in the wild" by introducing locally gathered images.

II. LITERATURE SURVEY

Butalia MA et al., (2012) [5] implemented geometry-based features for facial expression recognition. This technique searched for 153 possible distances among 18 critical candidate points/landmarks. Correlation-based feature subset selection (CFS) method was applied to select 16 of these distances having significant contribution to accuracy. This CFS+ANN method has 91.2% accurate classification rate.

Dureha et al., (2014) [6] used automatic fiducial point location algorithm locating 58 fiducial points and calculated the Euclidean distances between the center of gravity coordinate and the fiducial points coordinates of the face. A person's neutral expression and the other seven basic expressions are used to extract geometric deformation difference features. This feature vector acts as input to multiclass SVM classifier which classifies data input for seven basic expressions.

Wu Y et al., (2008) [7] proposed FER using efficient Local Binary Pattern (LBP) for feature extraction and Artificial Neural Network (ANN) for classification. Local Binary Pattern is illumination variant and detects the features using very simple calculation. Extracted features from LBP were given to ANN for classification. This algorithm improves recognition rate for 64x64 window size.

Zhang et al., (2006) [8] Simona et al. designed a model for detection of face, eyes and mouth which uses Haar functions and applied Bezier curves to extract the distances between facial parts. Two layered feed-forward neural network is used along with K-means algorithm for pre-classification resulting in accuracy of 85%.

Hsieh et al., (2010) [9] proposed facial expression recognition method using Hidden Markov Model. The relative displacement of the feature points between the current frame and the neutral frame are extracted as the facial features. Classification entropy threshold and model parameters are found out using iterative algorithm during training process. During testing, an image sequence is assigned an expression category when the entropy of the expression likelihood obtained from early HMMs is below the threshold by gradually increasing sequence length. The overall recognition rates achieve about 82% using about 45% sequence length on CK+ database, and about 52% with about 23% sequence length on MMI database.

Chin S and Kim KY (2009) [10] proposed a deep Convolutional neural network (CNN) for facial expression recognition system which is used for deeper feature representation of facial expression to achieve automatic recognition. The proposed system results 76.7442% and 80.303% accuracy in the JAFFE and CK+, respectively.

Fasel B and Luettin J (2003) [11]used Action Units for Facial Expression Recognition and analysis by using random forest classifier in a video. First random frog will detect action units and these detected AUs are classified by second random forest which detects expressions. On first frame, Facial Landmarks are generated by active Appearance Model (AAM) landmarks are tracked throughout the sequence of frames in a video optical flow tracker. A displacement vector is created between natural and peak expression. First Random forest detects Action units from DNNP features and these AUs are sent to 2nd Random Forest as an input that then process these AUs into Facial Expressions. The proposed methods of facial expression recognition achieve the accuracy rate of 89.37% for the two-fold Random Forest classifier. It can achieve accuracy rate of 96.38%. The results have been achieved by randomly selecting training and testing sets from the database for 9 times.

Ekman P and Friesen WV (2003) [12] Carcagnì et. al. proposed a system which implements Histograms of Oriented Gradients (HOG) on FER system. HOG was dense feature extraction method for single image. It extracts all regions of interests from image through gradients. This technique was pretty fast. This study describes about how to set perimeters of HOG so it could distinguish the facial expression traits to its best. Algorithmic pipeline pattern splits the system in 3 phases. In 1st phase, input frontal face in system which then performs registration of face, after which HOG is applied on face. Support Vector Machine (SVM) technique is applied for classification. Phase 2 applies HOG perimeters which are then tested on datasets; sequence of input faces starts with neutral face and ends with expressive face. Phase 3 validates system in real world. This system gave performance for edge and shape molding up to 95.8% accuracy. Strength of applied technique lies in choice of parameters plus it gives performance 95.9%, a precision 98%, and accuracy of 98.9%.

III. PROPOSED METHOD

In this study, we have proposed a method to combine LBP features and gray values to classify the facial expressions. The gray values and the LBP features are all extracted from the active patches. Active patches are those face regions which undergo a major change during different expressions. The active patches usually lie around eyes, nose and mouth. So by choosing these active parts we could gain the key information of different expressions. In order to gain the active patches we use the automated facial landmark detection method to mark the faces. Using these landmarks we could gain patches. By the application of PCA we could reduce the dimensions of the features. The algorithm we have applied needs less memory and computational cost, and gains a good grade which is close the state-of-the-art recognition accuracy which uses the deep learning framework. The proposed framework in our project has 7 stages which are shown in Fig. 3. In the next sections of this study, we will refer to these seven stages in details.

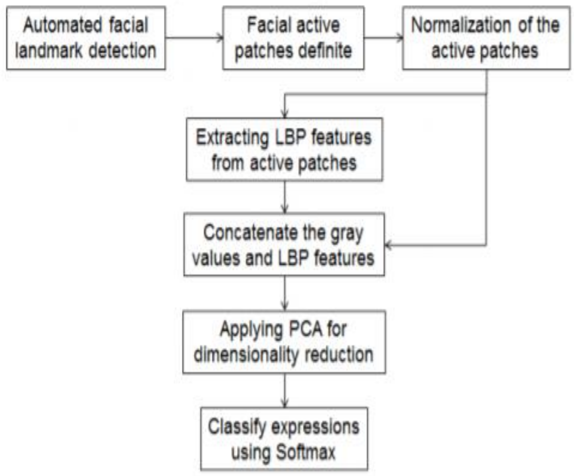

Fig-3. The proposed algorithm framework

A. *Automated Facial Landmark Location:*

Feature extraction is one important step for expression recognition, which could contribute to fast and accurate expression classification. For the expression recognition, extract features from the whole faces may collect enormous useless data which could cost much time for calculating, also these useless data may make it hard to classify expressions. For these reasons, we chose to use the active patches to extract the key information. Automated facial landmark detection is the first step in our method. Facial landmark detection is an important foundation for facial expression classification. There are many methods to detect landmarks from the faces. It has proposed a Fast-SIC method for fitting AAMs to detect marks on the faces. This study includes the use of a model based on mixture of trees with shared pool marks of 68 landmarks on the face. For our method, more landmarks are better for the patches' location, because more landmarks mean more accurate location. Comparing with the faces marked with fewer landmarks, we choose to use the method which locates 68 landmarks on the face. These landmarks mark the shape of eyebrows, eyes, nose, mouth and the whole face, we could use the landmark to cut the active patches. The image which is marked 68 marks is shown in Fig. 4.

Fig-4. 68 landmarks on the face

B. Facial Active Patches Definition and Normalization:

In the second step, we define the active facial patches. In this study, we used multitask sparse learning to confirm the common part of a face. Therefore, in our paper we choose three parts in the face for classification. The obtained result is illustrated in the Fig. 5, and we call the patch in the red, green, blue rectangles : *forehead patch, cheek patch and mouth patch*. As these faces have different sizes, the patches get from the faces have various dimensions, it is important to apply normalization on the patches.

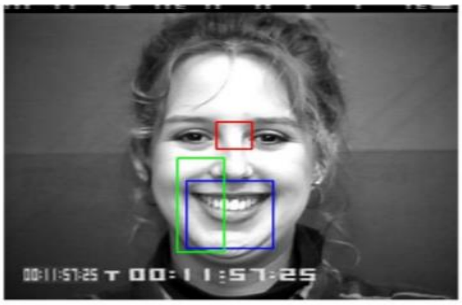

Fig-5. Facial active patches definite

In this study, we used the landmarks which are found from the landmarks detection method to find the patches, then using the proportion of the lateral edge with the normal size to normalize the other edge. Normalization is an important step to get the information from the images, as after the normalization we could draw features in the same dimensions. Also by applying the normalization method, we could gain more information from the images, which helped to gain better grade using the small benchmarks.

C. Local Binary Patterns (LBP) Features:

Texture information is an important descriptor for the pattern analysis of image. LBP was presented to get the texture information from the images. The Fig. 6 shows the calculation progress of the LBP value. The feature vector is processed using the softmax regression classifier. A useful extension to the original operator is the so-called uniform pattern, which can be used to reduce the length of the feature vector and implement a simple rotation invariant descriptor.

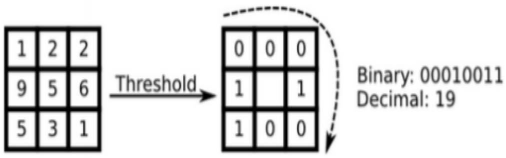

Fig-6. Calculation of LBP value

In our study, we used the uniform pattern LBP to gain features from the patches, the patches are all separated to small patches. Using uniform patterns, the length of the feature vector for a single cell reduces from 256 to 59. For example, the size of the mouth patch is 40*60 and the small patches' size is 10*15, so the mouth patch is divided into 16 patches. The uniform LBP features are extracted from each small patch and mapped to a 59-dimensional histogram. In the Fig. 7 we separated the mouth to 16 small patches.

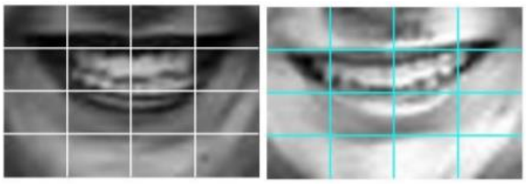

Fig-7. Calculation of LBP value

We use extended Cohn-Kanade (CK+) database to evaluate the proposed framework. CK+ database is a standard database which consists of 100 university students aged from 18 to 30 years old, of which 65% were female, 15% were African-American and 3% were Asian or Latino. The CK+ database contains 327 expression-labeled image sequences, each of which has one of 7 expressions, i.e., anger, contempt, disgust, fear, happiness, sadness, and surprise activated.

In our study, we only used six basic expressions to classify. In order to get a good result, we chose the last two peak images which contribute to 618 images. To be accurate, these expressions have different number with each other, in order to use all the images in the database we chose to use leave-one-out validation method to validate the classification effect.

D. Use the Gray Value for Classification:

For the purpose of this study, and to know the influence of the part patch such as the mouth patch and the cheek patch, we design two experiments. One is using only the data of mouth patch for classification, and the other uses the gray value of all the three patches. The progress is shown in Fig. 8.

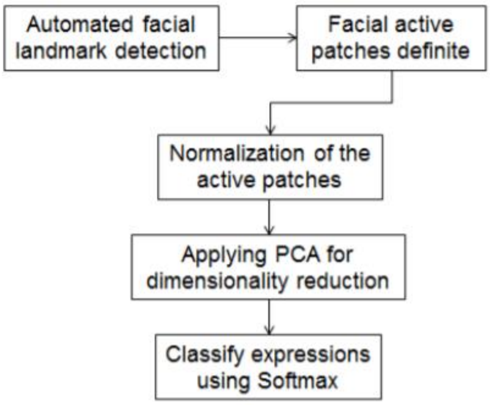

Fig-8. Process of using only gray value

This idea of using the active part in this study, we are using multi-task learning algorithm to distinguish active and inactive patches. This study also shows the active patches in the image. Also, the study distinguishes active patches to common and specific patches. The common patches in the study are those face regions that change mostly during different expressions. The specific patches are those patches that different expressions have different display.

Different from the method, we chose only the mouth patches to classify the different expressions to do the first experiment and we apply all the active patches to recognize the expressions as the second experiment. This is different from these methods which use the common patches to distinguish all the six expressions and apply the specific patches to sort the special expressions which have different performs in these regions. The results are shown in Fig. 9 and Fig.10

Fig-9. Confusion matrix of using mouth patch's gray value

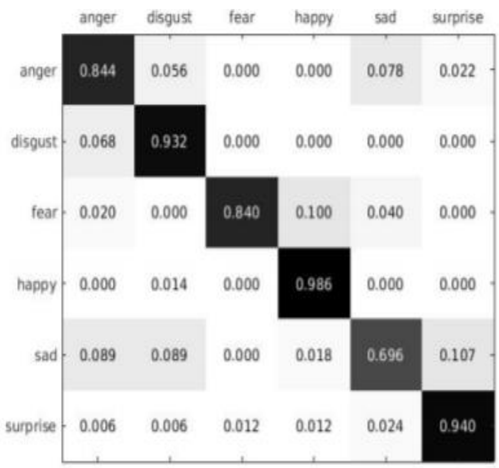

Fig-10. Confusion matrix of using three patches' gray value

Learning from the results of the experiments we could draw a conclusion that more effective data means higher accuracy. Besides, our framework also could gain a good result with these data.

E. Apply the LBP Feature for Classification:

In this study, we also designed one experiment to evaluate the LBP feature's effectiveness. We extracted the LBP features from the mouth patch, cheek patch and forehead patch. In the second section, we divided the mouth patch into 16 blocks, then we gain 59 dimensions feature from every small block. So only from the mouth patch we extracted 59*16=944 dimensions feature.

In the third experiment, we used the LBP features of the all patches and obtained a better result than the gray value. The result is shown in the Fig. 10. In the next part, we compared all the results from the experiments.

F. Use the Concatenation of the Patches' Gray Value and LBP Features for Classification

In the second part of the paper, we proposed a method to concatenate the gray value and the LBP features. We do as the fourth part in the second section, normalize all the gray value, then fusion the gray value of all the patches and concatenate the LBP features extracted from all the patches and use the PCA to reduce the dimensions.

IV. RESULT

The table below gives the recognition result of all the four experiments. Comparing these details in the Table-I, we could find that the accuracy of the expressions all have promotion by using the all-patches' features. On one hand, this shows that more information will result higher accuracy in our algorithm. On the other hand, by using the fusion features we could gain better grades of all the expressions. This means that our algorithm could improve the recognition accuracy of facial expression. As for the computational cost, 8 days to complete the overall training for 6 expressions in an 8-fold experimental setup on a 6-core 2.4GHz PC using MATLAB implementation. For our algorithm, we take about 5 seconds to train for 6 expressions in leave-one-out validation method. The experiment is running on a 4-core 3.2 GHz PC using MATLAB. Though we use less time than the algorithm in the previous studies, we could get a grade similar to their result.

Table-1. Result of Contrast Experiments

Facial expressions	Classification Result			
	Exp. one	Exp. two	Exp. three	Exp. four
Anger	0.744	0.844	0.856	0.956
Disgust	0.839	0.932	0.949	0.966
Fear	0.780	0.840	0.800	0.920
Happy	0.986	0.986	0.993	1.000
Sad	0.732	0.696	0.875	0.857
Surprise	0.934	0.940	0.970	0.982
Average	0.869	0.905	0.932	0.963

V. CONCLUSION & FUTURE WORK

In this study, we have proposed an automated and simply equipped facial expression recognition framework which uses combined LBP features and gray values to classify the facial expressions. Besides, we used the active patches of the face instead of the whole face. By using these active patches, we could use the key information of the different expressions. The combination of the LBP features and gray values has better classification result than only one part, so that we could merge these features to do other recognition work. In this study, we applied PCA method to reduce dimensions which could reduce computational cost and memory cost. Our algorithm could gain better results from the static images, but recognizing the basic expressions from the videos would still be a big challenge for the researchers. In the future, we will attempt to apply our algorithm on the videos.

VI. REFERENCES

[1] Darwin C, Ekman P, and Prodger P (1998). The expression of the emotions in man and animals. Oxford University Press, Oxford, UK.

[2] Mehrabian A (1968). Some referents and measures of nonverbal behavior. Behavior Research Methods and Instrumentation, 1(6): 203-207.

[3] Ekman P and Friesen WV (1971). Constants across cultures in the face and emotion. Journal of Personality and Social Psychology, 17(2): 124-129.

[4] Mandal MK, Pandey R, and Prasad AB (1998). Facial expressions of emotions and schizophrenia: A review. Schizophrenia Bulletin, 24(3): 399-412.

[5] Butalia MA, Ingle M, and Kulkarni P (2012). Facial expression recognition for security. International Journal of Modern Engineering Research (IJMER), 2(4): 1449- 1453.

[6] Dureha A (2014). An accurate algorithm for generating a music playlist based on facial expressions. International Journal of Computer Applications, 100(9): 33-39.

[7] Wu Y, Liu W, and Wang J (2008). Application of emotional recognition in intelligent tutoring system. In the First International Workshop on Knowledge Discovery and Data Mining, WKDD 2008, IEEE, Adelaide, Australia: 449-452.

[8] Zhang Z and Zhang J (2006). A new real-time eye tracking for driver fatigue detection. In the 6th International Conference on ITS Telecommunications Proceedings, IEEE, Chengdu, China: 8-11.

[9] Hsieh CK, Lai SH, and Chen YC (2010). An optical flow based approach to robust face recognition under expression variations. IEEE Transactions on Image Processing, 19(1): 233-240.

[10] Chin S and Kim KY (2009). Emotional intensity-based facial expression cloning for low polygonal applications. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews), 39(3): 315-330.

[11] Fasel B and Luettin J (2003). Automatic facial expression analysis: a survey. Pattern Recognition, 36(1): 259-275.

[12] Ekman P and Friesen WV (2003). Unmasking the face: A guide to recognizing emotions from facial clues. Institute for the Study of Human Knowledge (ISHK), California, USA.