

Facial Expression Recognition using Hybrid Texture Features based Ensemble Classifier

M. Arfan Jaffar

College of Computer and Information Sciences,
Al Imam Mohammad Ibn Saud Islamic University (IMSIU),
Riyadh, Saudi Arabia

Abstract—Communication is fundamental to humans. In the literature, it has been shown through many scientific research studies that human communication ranges from 54 to 94 percent is non-verbal. Facial expressions are the most of the important part of the non-verbal communication and it is the most promising way for people to communicate their feelings and emotions to represent their intentions. Pervasive computing and ambient intelligence is required to develop human-centered systems that actively react to complex human communication happening naturally. Therefore, Facial Expression Recognition (FER) system is required that can be used for such type of problem. In this paper, FER system has been proposed by using hybrid texture features to predict the expressions of human. Existing FER system has a problem that these systems show discrepancies in different cultures and ethnicities. Proposed systems also solve this type of problem by using hybrid texture features which are invariant to scale as well as rotate. For texture features, Gabor LBP (GLBP) features have been used to classify expressions by using Random Forest Classifier. Experimentation has been performed on different facial databases that demonstrate promising results.

Keywords—Expression classification; ensemble; adaboost; facial; features

I. INTRODUCTION

Communication is fundamental to humans. Many scientific research studies have shown that most of the human communication (55% to 93%) is non-verbal [1], [2]. Along with the movement of the head, facial expressions are the main part of the non-verbal communication [3]. Facial expressions are one of the most compelling, of course, excellent means for people to communicate their feelings and emotions to clarify and to strain their understanding, disagreements, and intentions. The next generation computing, such as, pervasive computing and ambient intelligence, needs to develop human-centered systems that enthusiastically respond to complex human communication happening naturally [4]. Traditional HCI ignore bulk of information communicated through non-verbal ways and just caters for user's intentional input. FER plays a key role in the development of interfaces that can understand the emotional expressions and emotional responses of the user [5], [6]. The effective application areas of FER are: affective computing (autism syndrome diagnosis, analysis of depression, pain detection, stress recognition, drowsiness detection, etc.), smart media (smart home, smart meeting), video conferencing and visual surveillance, lie detection, psychiatry, emotion and paralinguistic communication,

robotics, computer games (e.g. Microsoft Kinect), medicine, expression driven facial animation (e.g. facial movements in the film Avatar, Cartoons for children), security, HCI, facial image fusion for gender conversion and different age groups' fusion [7]-[9], etc. Thus, there is a strong interest for both academia and industry, and it makes the researchers to pursue for the goal to develop robust and efficient FER systems.

To develop a reliable and robust system for the FER is still a difficult and challenging problem [10], [11]. The FER is innately data driven. The challenge of generalization and robustness for various ethnicities and cultural variations can only be reasonably addressed if a representative database of such variations is available. But, there is no database constructed yet by keeping this problem in view. A good number of traditional expression recognition systems, trained over a benchmark data set and tested over the corresponding one, can offer an optimal performance. It is observed that while being in physical/practical environment such systems reflect a substantial decline in their performance due to in-appropriation or insufficiency of training datasets for facial expression patterns which could be expected in feature [12]. An expression can be demonstrated by so many combinations of facial expression patterns. Each of the images in facial expression datasets represents a single common expression i.e. disgust, but a variety of facial expression patterns easily observable. So, it is too difficult to engender such all possible patterns and to use these as training data even when the construction of an expression recognition system is based on a large number of facial images. This is nearly impossible to consider all possible future variations in facial expression patterns ahead-in-time. Thus, based on existing datasets, the highest recognition performance is never expected in practical environments. Although, some efforts have been made to construct naturalistic databases but they still have many limitations along with publically accessibility issues [13]. However, researchers in the FER community are agreed that spontaneous data gathering with real-life environments is very monotonous task. Consequently, the process for data base construction requires new methods those could expedite its creation process and represent the actual real conditions. Furthermore, the construction of robust algorithms with reliability is also required. Under such sort of situation when these databases are not available in absolute, an incremental learning along with dynamically weighted majority voting based solution is proposed. Head movements, pose variations and illumination variations are frequently encountered while dealing with spontaneous expressions. This problem is

addressed by using a feature set that is robust against in-plane image rotations and illumination variations.

II. PROPOSED METHOD

Proposed method has been divided into three different phases. First phase is to extract region of interest (ROI) and in this case is to extract the face portion only that is used for facial expression recognition. Hairs, or head or neck and other parts of bodies has no involvement in facial expression recognition therefore, face extraction is the most important step. In the second phase, the most suitable features that represent the facial expression are important. So in this phase texture and geometrical features has been extracted for the decision of expression. In the third phase, intelligent classifier has been used to decide the expression type by using the features extracted in second phase. Fig. 1 shows proposed method.

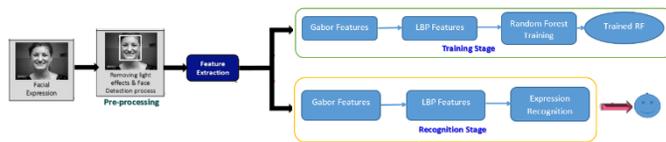


Fig. 1. Proposed method.

A. Preprocessing to extract Face part

First step in proposed method is to extract face from whole image. Most of the time during photography or image acquisition, it is not common to take only face image. Always whole body image or some time head with neck is also taken in the images. Facial expressions are only available on the faces. Therefore, it is important to extract the face part from the images. Proposed method used one of the standard method available in the literature to extract face part by using voila and Jones method. Proposed system did not require any other step as preprocessing except this face part extraction. Result of voila jones method has been shown in Fig. 2 that has been taken from MUG database.



Fig. 2. Face extraction using Viola and Jones method.

B. Feature Extraction

Feature represents the characteristics of an object to distinguish from other objects. Facial image also contains some characteristics, six based upon it can be distinguished and can be identified the expression as well. So some impressive types

of features are required that can be used for all type of images as well as for all ethnicities and cultural people so there should be no discrepancies in different cultures and ethnicities. Most of the times, frontal faces are not available, some posed either left or right faces available. Similarly, it is not necessary to take picture from same distance from camera. Sometimes, it has been taken from long distances and some time from very short distances so scaling is also a factor to consider during features extraction. So such type of features is required that are rotation invariant as well as scale in variant. Texture features is special type of features extraction that can be used to handle such type of variations. So hybrid texture features have been used by extracting Gabor based features and local binary patterns (LBP) features.

C. LBP Features using Gabor Filter

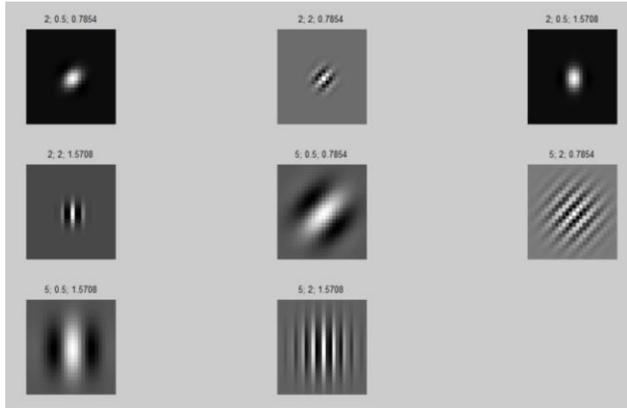
Gabor filter can be used to extract texture information. Texture shows a specific pattern and facial images have some specific pattern that represents a specific expression. Like laughing has specific pattern that always remain same for all faces similarly most the time sad also has some specific pattern, smile also has a specific pattern. Therefore, texture is the most suitable for features extraction in the case of facial expression. So the characteristics of texture can be represented by a spatial frequencies and it can also be represented by their orientations. There are different types of Gabor filter that can be applied on images to extract texture features [14]-[18]. But in facial expression 2-D Gabor filter is most suitable due to nature of images that are in 2-D form. Gabor filter is a Gaussian kernel function and that can be modulated by a sinusoidal wave of precise frequencies and orientation [19]-[21]. To represent the 2-D Gabor filter, following equations can be used:

$$g(x, y) = \frac{1}{2\pi\sigma_x\sigma_y} e \left[-\frac{1}{2} \left(\frac{\bar{x}^2}{\sigma_x^2} + \frac{\bar{y}^2}{\sigma_y^2} \right) \right] e(2\pi jW\bar{x})$$

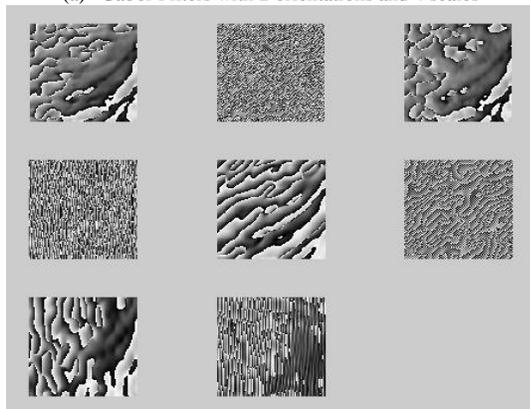
$$\bar{x} = x \cdot \cos \theta + y \cdot \sin \theta \text{ and } \bar{y} = x \cdot \sin \theta + y \cdot \cos \theta$$

Where, variables x , and y are the spatial variables, σ_x and σ_y represents are the scaling parameters of the filter, and W is the central frequency of the complex wave. Gabor filter bank is a combination of different Gabor filters applied at different scales, frequencies and orientation. It is possible to generate different filter banks with different orientation and scales. In this paper, Gabor filter bank has been created by applying two frequencies, two scales, and two orientations. For this purpose, following values has been used for generation of filter bank. After calculating these filters, convolution is required to apply on the original images. So these eight filters are convolved with the original images so it returns eight new convolved images. After applying Gabor filter bank, there are magnitude values of the Gabor transform. These magnitudes represent changes very slowly with displacement. Thus there is required a process to encode these magnitude values. LBP can be used to encode these magnitude values. Basic advantage of LBP encoding over magnitude images is to improve and enhance the information in the Gabor filtered images. Applying local binary pattern on Gabor is actually a representation approach based on multi-resolution spatial histogram combining local intensity distribution with the spatial information, therefore, it is robust

to noise and local image transformations. Additionally, instead of directly using the intensity to compute the spatial histogram, multi-scale and multi-orientation Gabor filters are used for the decomposition of an image, followed by the local binary patterns (LBP) operator. LBP operator on each pixel of the image to get LBP coded image and then represented as a histogram of that code. The Gabor and LBP combination further improves the representational power of the spatial histogram greatly (Fig. 3).



(a) Gabor Filters with 2 orientations and 4 scales



(b) Implementation of LBP on Gabor filter

Fig. 3. Gabor Filters and Implementation on LBP.

III. CLASSIFICATION USING ADABOOST

Classification is the process to differentiate into classes by using some characteristics. In the literature, many different classifiers are available that can be classified individually. Ensemble classification used different weak classifiers and combine intelligently to combine those classifiers to improve the performance of classification. One of the most important ensemble classifier is AdaBoost that is also known as adaptive boosting. This AdaBoost was proposed by [14] and it improves the simple classifier by using the iterative procedure. In this iterative procedure, during each iteration, there is a process to improve the misclassified samples. This procedure increased weights of misclassified patterns and decrease the weights of correctly classified samples during each iteration. In this way, weak classifiers given more preferences and these weak classifiers are forced to learn more by using difficult samples [14]. In this way, classification performance improves during

this iterative weight adjustment procedures [27]. These adaptive weights can be used for the classification of new samples. In this way, algorithm supposed that the training set contains m samples and these samples are labelled as -1 and $+1$. In this way, classification of the new sample can be find out by using voting for all classifiers M_t with weights α_t . Mathematically, it can be written as:

$$H(x) = \text{sign}\left(\sum_{t=1}^T \alpha_t \cdot M_t(x)\right)$$

Pseudocode of the AdaBoost is given in Fig. 4.

```

Require:  $I$  (a weak inducer),  $T$  (the number of iterations),  $S$  (training set)
Ensure:  $M_t, \alpha_t; t = 1, \dots, T$ 
1:  $t \leftarrow 1$ 
2:  $D_1(i) \leftarrow 1/m; i = 1, \dots, m$ 
3: repeat
4:   Build Classifier  $M_t$  using  $I$  and distribution  $D_t$ 
5:    $\varepsilon_t \leftarrow \sum_{i: M_t(x_i) \neq y_i} D_t(i)$ 
6:   if  $\varepsilon_t > 0.5$  then
7:      $T \leftarrow t - 1$ 
8:     exit Loop.
9:   end if
10:   $\alpha_t \leftarrow \frac{1}{2} \ln\left(\frac{1-\varepsilon_t}{\varepsilon_t}\right)$ 
11:   $D_{t+1}(i) = D_t(i) \cdot e^{-\alpha_t y_t M_t(x_i)}$ 
12:  Normalize  $D_{t+1}$  to be a proper distribution.
13:   $t++$ 
14: until  $t > T$ 
    
```

Fig. 4. Pseudo code of Adaboost Classifier.

IV. RESULTS AND DISCUSSION

For performance evaluation two different datasets has been used to evaluate proposed method. To validate results accuracy has been used as quantitative measure and results has been presented in the form of confusion matrix. After face detection each face image was resized to $[150 \times 150]$ before the extraction of feature vector. These datasets contain different types of expressions like some are spontaneous and some are posed so it is a mixture of different expressions. Some sample images have been shown in Fig. 5. JAFFE database which consists of 213 facial expression images of Japanese females and MUG database which consists of both posed and induced expressions are used in the experimentation. The MUG database consists of 86 subject's image sequences. Out of which 51 subjects are males and 35 subjects are females.

Results has been shown in the form confusion matrix where ANG, DIS, FEA, HAP, SAD, SUR and NEU represents the anger, disgust, fear, happy, sad, surprise and neutral expressions, respectively. In this paper, hybrid texture features have been extracted by using Gabor and LBP. Random Forest classifier has been used for prediction of expressions.

For experimentation, three different types of experiments have been conducted by using different ratios of training and testing data like 60-40 ratio mean 60 percent for training and 40 percent for testing similarly 50-50 and 40-60 percent so that there should no biasness in the results. These tests have been conducted 10 times and then results have been shown by taking average of all these differently 10 times. Results have been shown in Table 1 on JAFFE database and Table 2 shows results on MUG database.



Fig. 5. Sample Images from datasets.

TABLE. I. CONFUSION MATRIX ON JAFFE DATASET

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	96.0%	1.0%	0.0%	0.0%	1.0%	1.0%	0.0%
DIS	5.89%	87.66%	2.25%	1.20%	2.0%	1.0%	0.0%
FEA	2.12%	0.0%	91.63%	1.0%	1.0%	4.24%	0.0%
HAP	0.0%	1.0%	0.0%	95.38%	1.50%	0.0%	2.22%
SAD	3.22%	0.0%	0.66%	3.22.0%	91.0%	2.0%	0.0%
SUR	0.0%	0.0%	0.0%	2.0%	1.33%	91.0%	5.67%
NEU	0.0%	1.0%	0.67%	1.0%	3.33%	0.0%	94.0%
						Average	92.38%

Table 1 shows results of proposed method on JAFFE dataset. It shows that approximately it achieves good results for all expression. As table shows that on anger accuracy is 96%, disgust 87.66%, fear 91.63%, happy 95.38%, sad 91%, surprise 91% and neutral 94%. Average results of all these has been shown that 92.38% that is good for this problem. These results has been shown on three experiments and then taken average results of all those three experiments. This shows that proposed method works well specially by extracting features of hybrid Gabor LBP features. These GLBP features plays important role in good accuracy and recognition rate.

TABLE. II. CONFUSION MATRIX ON MUG DATASET

	ANG	DIS	FEA	HAP	SAD	SUR	NEU
ANG	95.15%	0.64%	0.0%	0.0%	1.91%	2.30%	0.0%
DIS	2.55%	91.08%	0.0%	3.18%	1.0%	0.91%	1.27%
FEA	0.0%	0.0%	92.07%	1.0%	1.0%	2.75%	3.18%
HAP	2.4%	1.27%	0.0%	96.09%	0.0%	0.0%	0.60%
SAD	3.18%	0.0%	1.47%	1.91%	92.14%	0.0%	0.0%
SUR	0.0%	1.64%	3.46%	3.18%	1.0%	90.72%	0.0%
NEU	2.18%	1.0%	0.0%	1.0%	2.83%	1.99%	91.0%
						Average	92.60%

Table 2 shows results of proposed method on MUG dataset. It shows that approximately it achieves good results for all

expression. As table shows that on anger accuracy is 95.15%, disgust 91.08%, fear 92.07%, happy 96.09%, sad 92.14%, surprise 90.72% and neutral 91%. Average results of all these has been shown that 92.60% that is good for this problem. These results have been shown on three experiments and then taken average results of all those three experiments. This shows that proposed method works well specially by extracting features of hybrid Gabor LBP features. These GLBP features plays an important role in good accuracy and recognition rate.

Table 3 shows comparison of different methods on both JAFFE and MUG datasets. It has been compare with existing methods like F. Wallhoff used DCT features with SVM and hidden Markova model (HMM) and it achieved 61.7% on JAFFE and 63.5 on MUG datasets. Similar other results have been shown in this table. Proposed method also shows 92.38% on JAFFE and 92.60 on MUG datasets that is the highest from all other existing methods.

TABLE. III. COMPARISON WITH EXISTING METHODS

Methodology	Recognition (%) JAFFE dataset	Recognition (%) MUG dataset
Wallhoff, F [21] (DCT+SVM-SFFS+HMM)(5-fold)	61.7	63.5
Xiao, R [22] (Multiple Manifold)	78.5	79.2
Samad, R [23] (Gabor+PCA+SVM)	81.7	82.4
Samad, R [24] (Edge-Gabor+PCA+SVM)	91.7	92.2
Zhang, L [25] (SIFT_FAP+AdaBoost)	63.6	65.4
Khan, R.A [26] (LBP-Pyramid+SVM)(10-fold)	91.36	92.13
Khan, R.A [26] (LBP-Pyramid+2NN)(10-fold)	91.28	92.20
(Proposed System) (GaborLBP) (10-fold)	92.38	92.60

V. CONCLUSION

Facial expressions are the most of the important part of the non-verbal communication and it is the most promising way for people to communicate their feelings and emotions to represent their intentions. Pervasive computing and ambient intelligence required to develop human-centered systems that actively react to complex human communication happening naturally. Therefore, Facial Expression Recognition (FER) system is required that can be used for such type of problem. In this paper, FER system has been proposed by using hybrid texture features to predict the expressions of human. Existing FER systems has a problem that these systems show discrepancies in different cultures and ethnicities. Proposed system also solves this type of problem by using hybrid texture features which are invariant to scale as well as rotation. For texture features, Gabor LBP (GLBP) features have been used to classify expressions by using AdaBoost Classifier. Experimentation has been performed on different facial

databases that demonstrate promising results. In the future, deep learning will be explored in this domain.

REFERENCES

- [1] Lapakko, D., Three cheers for language: A closer examination of a widely cited study of nonverbal communication. *Communication Education*, 1997. 46(1): p. 63-67.
- [2] Mehrabian, A., *Nonverbal communication*. 1977: Transaction Publishers.
- [3] Bull, P., *Nonverbal communication*. *The Psychologist*, 2001. 14: p. 644-647.
- [4] Jaimes, A., N. Sebe, and D. Gatica-Perez. Human-centered computing: a multimedia perspective. in *Proceedings of the 14th annual ACM international conference on Multimedia*. 2006. ACM.
- [5] Zeng, Z., et al., A survey of affect recognition methods: Audio, visual, and spontaneous expressions. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 2009. 31(1): p. 39-58.
- [6] Scherer, K.R., T. Bänziger, and E. Roesch, *A Blueprint for Affective Computing: A sourcebook and manual*. 2010: Oxford University Press.
- [7] Bettadapura, V., *Face expression recognition and analysis: the state of the art*. arXiv preprint arXiv:1203.6722, 2012.
- [8] Ekman, P., *Telling Lies: Clues to Deceit in the Marketplace, Politics, and Marriage (Revised Edition)*. 2009: WW Norton & Company.
- [9] Poria, S., A. Mondal, and P. Mukhopadhyay, Evaluation of the Intricacies of Emotional Facial Expression of Psychiatric Patients Using Computational Models, in *Understanding Facial Expressions in Communication*. 2015, Springer. p. 199-226.
- [10] Valstar, M.F., et al. The first facial expression recognition and analysis challenge. in *Automatic Face & Gesture Recognition and Workshops (FG 2011)*, 2011 IEEE International Conference on. 2011.
- [11] Valstar, M., et al., Fera 2015-second facial expression recognition and analysis challenge. *Proc. IEEE ICFG*, 2015.
- [12] Chin, S. and C.-Y. Lee, Exaggeration of facial expressions from facial motion capture data. *Chinese Optics Letters*, 2010. 8(1): p. 29-32.
- [13] Zhang, L., *Towards spontaneous facial expression recognition in real-world video*. 2012.
- [14] Gangardiwala, A. and R. Polikar. Dynamically weighted majority voting for incremental learning and comparison of three boosting based approaches. in *Neural Networks, 2005. IJCNN'05. Proceedings. 2005 IEEE International Joint Conference on*. 2005.
- [15] Muhlbaier, M.D., A. Topalis, and R. Polikar, Learn. NC: Combining Ensemble of Classifiers With Dynamically Weighted Consult-and-Vote for Efficient Incremental Learning of New Classes. *Neural Networks, IEEE Transactions on*, 2009. 20(1): p. 152-168.
- [16] Kotsia, I. and I. Pitas, Facial expression recognition in image sequences using geometric deformation features and support vector machines. *Image Processing, IEEE Transactions on*, 2007. 16(1): p. 172-187.
- [17] Pantic, M. and L.J. Rothkrantz, Expert system for automatic analysis of facial expressions. *Image and Vision Computing*, 2000. 18(11): p. 881-905.
- [18] Braathen, B., et al. An approach to automatic recognition of spontaneous facial actions. in *Automatic Face and Gesture Recognition, 2002. Proceedings. Fifth IEEE International Conference on*. 2002.
- [19] Berretti, S., et al., 3D facial expression recognition using SIFT descriptors of automatically detected keypoints. *The Visual Computer*, 2011. 27(11): p. 1021-1036.
- [20] Fang, T., et al., 3D/4D facial expression analysis: an advanced annotated face model approach. *Image and Vision Computing*, 2012. 30(10): p. 738-749.
- [21] Wallhoff, F., et al. Efficient recognition of authentic dynamic facial expressions on the feedtum database. in *Multimedia and Expo, 2006 IEEE International Conference on*. 2006. IEEE.
- [22] Xiao, R., et al., Facial expression recognition on multiple manifolds. *Pattern Recognition*, 2011. 44(1): p. 107-116.
- [23] Samad, R. and H. Sawada, Extraction of the minimum number of Gabor wavelet parameters for the recognition of natural facial expressions. *Artificial Life and Robotics*, 2011. 16(1): p. 21-31.
- [24] Samad, R. and H. Sawada. Edge-based Facial Feature Extraction Using Gabor Wavelet and Convolution Filters. in *MVA*. 2011.
- [25] Zhang, L., D. Tjondronegoro, and V. Chandran. Discovering the best feature extraction and selection algorithms for spontaneous facial expression recognition. in *Multimedia and Expo (ICME), 2012 IEEE International Conference on*. 2012.
- [26] Khan, R.A., et al., Framework for reliable, real-time facial expression recognition for low resolution images. *Pattern Recognition Letters*, 2013. 34(10): p. 1159-1168.
- [27] K.J. Archer and R.V. Kimes. Empirical characterization of random forest variable importance measures. *Computational Statistics & Data Analysis*, 52(4):2249-2260, 2008.