JOURNAL OF INFORMATION SCIENCE AND ENGINEERING 34, 1237-1249 (2018)
DOI: 10.6688/JISE.201809_34(5).0008

Robust Facial Expression Recognition using Local Haar Mean Binary Pattern

MAHESH GOYANI[1] AND NARENDRA PATEL[2]
[1,2]*Department of Computer Engineering*
[1]*Charotar University of Science and Technology*
Changa, 388421 India
[2]*Gujarat Technological University*
V. V. Nagar, 388120 India
E-mail: mgoyani@gmail.com; nmpatel@bvmengineerring.ac.in

In this paper, we propose a hybrid statistical feature extractor, Local Haar Mean Binary Pattern (LHMBP). It extracts level-1 haar approximation coefficients and computes Local Mean Binary Pattern (LMBP) of it. LMBP code of pixel is obtained by weighting the thresholded neighbor value of 3×3 patch on its mean. LHMBP produces highly discriminative code compared to other state of the art methods. To localize appearance features, approximation subband is divided into $M \times N$ regions. LHMBP feature descriptor is derived by concatenating LMBP distribution of each region. We also propose a novel template matching strategy called Histogram Normalized Absolute Difference (HNAD) for histogram based feature comparison. Experiments prove the superiority of HNAD over well-known template matching techniques such as L2 norm and Chi-Square. We also investigated LHMBP for expression recognition in low resolution. The performance of the proposed approach is tested on well-known CK, JAFFE, and SFEW facial expression datasets in diverse situations.

Keywords: affective computing, appearance based feature, local binary pattern, Gabor filter, support vector machine

1. INTRODUCTION

Facial Expression Recognition (FER) plays a vital role in social communication and in conveying emotions [1]. In the earlier development stage, the scope of FER was confined to psychological studies only, but nowadays it covers a broad range of applications including human-computer interfaces (HCI), industrial automation, surveillance systems, sentiment identification, *etc.* Precise recognition of facial expressions can become a driving force for the future automation interfaces like car driving, robotics, driver alert systems, *etc.*

According to input, expression recognition systems can be classified as static or dynamic. In static approaches, features are computed from still image. Whereas in dynamic approaches temporal relationships between features over the image sequence is extracted. In last decade, many video-based methods have been studied [2]. Research is also focused on detecting micro-expressions [2-5], recognition of spontaneous expressions [6], analysis of multi-views or profile views [7] and fusion of geometric and appearance features [8-10]. Nowadays, deep learning based approaches also getting significant attention [2, 11-13].

Received February 27, 2017; revised May 9, 2017; accepted July 14, 2017.
Communicated by Chung-Lin Huang.

In this paper, we present a novel hybrid face descriptor called Local Haar Mean Binary Pattern (LHMBP), which efficiently represents the local intensity variation and local structural information around the pixel. It produces highly discriminative code compared to other state of the art methods. We have also proposed a novel template matching technique, Histogram Normalized Absolute Difference (HNAD). Experiments show that HNAD outperforms the traditional template matching techniques like L2 norm and Chi-square measures.

Rest of the paper is organized as follows: A brief review of related work is presented in Section 2. LHMBP operator is introduced in Section 3. HNAD measure is explained in Section 4. Result analysis, feature selection methods and expression recognition in the low resolution are presented in section 5. Finally, the work is summarized in Section 6.

2. RELATED WORK

In order to recognize the facial expressions, it is necessary to detect the key parameters which best describe the expression. The set of such parameters is called *feature descriptor*. The success of any pattern recognition system depends on the amount of discriminative information carried by the feature descriptor.

Gabor filters are widely used to extract changes in facial appearance as a set of multi-scale and multi-orientation coefficients. It mimics the behavior of simple cell receptive fields in cat striate cortex [14] and provides an excellent base for image analysis. Interesting studies on Gabor-based systems can be found in [2, 15-18] *etc.* Local Binary Pattern (LBP) [19] is another powerful texture analysis descriptor, which is extensively studied in texture analysis. It encodes the micro patterns in the local neighborhood of the pixel using the knowledge of neighbor pixel intensity. Many variants and extensions of LBP are proposed, and they are detailed in [20-22]. Jabid *et al.* [23] introduced a local direction pattern (LDP) descriptor, which derives 8-bit binary pattern by encoding edge response value, in 8 directions using Kirsch edge detector. Kirsch detects directional response precisely compared to Prewitt or Sobel edge detectors. The response of eight masks is collected in a 3×3 matrix. 8-neighbor responses are thresholded with respect to m_kth directional response. The value of k is chosen through experiments. LBP code of this gives the LDP value for the center pixel.

3. LOCAL HAAR MEAN BINARY PATTERN

LMBP assigns an eight-bit binary code to each pixel, which represents the local structure. Multi-scale modeling of LMBP is very similar to LBP. In presented LMBP, the mean value of 3×3 neighborhood is compared with eight neighbor values. LMBP code of central pixel (x_c, y_c) is computed as follow:

$$LMBP(x_c, y_c) = \sum_{n=0}^{7} 2^n * S(i_n - i_m), \qquad S(x) = \begin{cases} 1, & x \geq 0 \\ 0, & x < 0 \end{cases}. \tag{1}$$

Where, i_n is the intensity of neighbor of center pixel (x_c, y_c), and i_m is the mean intensity of 3×3 image patch. Contrast to LBP, LMBP computes the weighted code by

thresholding the patch with respect to its mean, and hence it suppresses the effect of noise and non-monotonic illumination by diluting the intensity variation. It provides stability to generated LMBP code. Working of basic LMBP operator is described in Fig. 1.

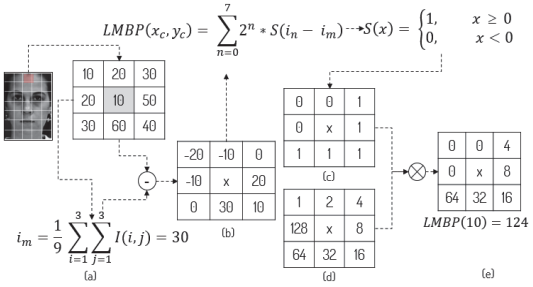

Fig. 1. Derivation of LMBP code.

The mean value of the patch is subtracted from eight neighbors (Fig. 1 (b)). Positive neighbors are assigned value 1, while negative neighbors are assigned value 0 (Fig. 1 (c)). The weight matrix shown in Fig. 1 (d) is applied to thresholded patch. Summation of masked output is the LMBP code of center pixel. LDP generates the feature descriptor by considering the edge response using Kirsch masks, which makes it precise in comparison to LBP. But due to fixed mask, LDP cannot be extended for larger neighborhood.

3.1 Robustness of LMBP

LMBP binarizes 3×3 neighbor on its mean, and hence it is less susceptible to the noise and non-monotonic gray level changes. Thus, robustness is inherent in LMBP operator. Patch in Figs. 2 (b)-(d) shows the intensity after adding Gaussian noise of zero mean and 0.01 variance, Gaussian white noise of local variance 0.5 and zero mean, and multiplicative speckle noise of zero mean and variance 0.1, respectively. Results show that LBP and LDP codes are sensitive to these noises but LMBP suppresses those effects, and LMBP code remains unaltered in all the three cases. LBP binarizes the neighbor values with respect to center pixel only, so the introduction of noise in any of the neighbor pixel can affect the local binary pattern. LDP mask is weighing three pixels by value 5, and five pixels by value −3, which makes it vulnerable to small change in intensity.

55	31	50
29	30	30
20	50	30

LBP = 63
LDP = 253
LMBP = 37
(a)

56	32	50
30	31	30
21	51	31

LBP = 55
LDP = 125
LMBP = 37
(b)

56	32	50
30	31	30
21	51	30

LBP = 39
LDP = 125
LMBP = 37
(c)

55	31	50
30	30	30
20	51	31

LBP = 191
LDP = 127
LMBP = 37
(d)

Fig. 2. Robustness of LMBP vs. LBP and LDP: (a) LMBP, LBP, and LDP value of original patch; Respective code after adding; (b) Gaussian; (c) Local variance; and (d) Speckle noise.

3.2 LHMBP Feature Descriptor

LHMBP is hybrid statistical feature extractor, which first finds the Haar decomposition of the query image. LMBP codes are computed from approximation sub-band in order to extract LHMBP features. Feature descriptor is obtained by computing 256 bin histogram of LHMBP features. While encoding the whole face, the global position of local texture is encoded, but its precise location is unknown. The relation between feature and its spatial location plays a vital role in texture analysis.

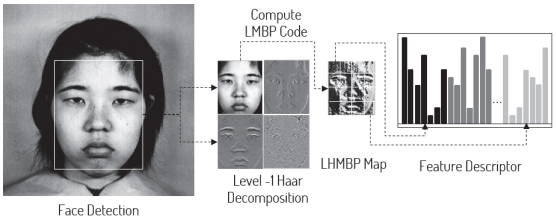

Fig. 3. LHMBP descriptor is derived by concatenating local and global LMBP histograms of approximation coefficients of level-1 Haar decomposition of the query image.

As shown in Fig. 3, to localize the features, we divide the level-1 haar approximation sub-band into $M \times N$ small regions, R^1, R^2, …., R^{MN} and 256-bin LHMBP^i histogram of each region R^i is computed and concatenated to derive local texture descriptor. Histogram of the whole face approximates the global shape of the face. To incorporate texture and shape both, we append the global histogram of face LHMBP^G to local texture descriptor. This feature vector gives accurate facial representation. Concatenated histogram is computed as,

$$H_c = \prod \left\{ \left[\prod_{i=1}^{MN} LHMBP^i \right] LHMBP^G \right\} \tag{2}$$

where Π is the concatenation operator.

4. HISTOGRAM NORMALIZED ABSOLUTE DIFFERENCE

Over the time, various classifiers have been presented for pattern discrimination. Chi-square, nearest neighbor, neural network, support vector machines are few of them. Comparative analysis of different classifiers for facial expression recognition is discussed in [21]. Proposed HNAD measure normalizes absolute difference of histograms by the histogram union. HNAD distance between two *k*-bin histograms X and Y is defined as,

$$HNAD(X,Y) = \sum_k \frac{abs(X(k) - Y(k))}{\max\{X(k), Y(k)\}}. \tag{3}$$

Unlike Chi-square and L2 norm, it does not involve squaring or square root opera-

tions. HNAD is computationally straightforward and efficient. Results of all three measures on different datasets are compared in Section 5. Rigorous experiments on various histogram patterns show that HNAD outperforms both the measures.

5. RESULTS AND DISCUSSION

5.1 Experimental Setup

Experiments are performed on widely used datasets: CK [24], JAFFE [25] and SFEW [26]. We considered neutral (NE) and basic six prototypic expressions, anger (AN), disgust (DI), fear (FE), happy (HA), sad (SA) and surprise (SU) for analysis.

Cohn-Kanade (CK) [24] dataset contains image sequence of 97 university students having a 7:13 Male: Female ratio. The dataset contains people of different ethnicity, with age group of 18-30 years. Each subject was instructed to perform series of 23 facial displays, out of which six were based on prototypical expression description. Each sequence contained 12-16 frames. Each image sequence starts with a neutral expression and ends at apex level.

Japanese Female Facial Expression (JAFFE) database was planned and assembled by Lyons *et al.* [25] in 1998. This database is primarily used to evaluate the facial expression recognition systems. It may also be used for face recognition. JAFFE dataset is not large enough compared to CK. JAFFE contains 213 images of 10 Japanese female, with two, three or four samples of each of the seven basic expressions.

The Static Facial Expression in the Wild (SFEW) [26] is a subset of Acted Facial Expressions in the Wild (AFEW) [27] and created by selecting static frames from AFEW. SFEW contains spontaneous expressions with the different head pose, age range, and occlusions.

For CK and JAFFE, we used K-fold cross validation technique with $K = 10$. For SFEW training and test sets are already provided with dataset release. Reported results are averaged over 10 executions. The region of interest from the input image is extracted using Viola-Jones face detector and is normalized to 150×110 pixels. Samples and details of a number of images used in the experiment are listed in Table 1.

Table 1. Snapshots and number of images used in experiments.

Dataset	Expression	AN	DI	FE	HA	SA	SU	NE
CK	Snapshots							
	#Image	110	120	100	280	130	220	320
JAFFE	Snapshots							
	#Image	30	29	32	31	31	30	30
SFEW	Snapshots							
	#Image	104	81	90	112	92	86	98

5.2 Optimal Parameter Selection

In the prototypic facial expression, textures such as wrinkles, bulges, furs play a crucial role. To extract the local texture features, we divided face image into $M \times N$ regions. Rigorous experiments are conducted to find the optimal number of blocks. Larger regions size fails to capture the small texture and results in lower recognition rate. Small regions effectively capture local and spatial relationship. However, after a certain point, the smaller size of regions introduces unnecessary computation and feature vector becomes too big to train the classifier efficiently. Performance behavior on CK dataset for a different number of blocks is stated in Table 2.

Table 2. Behavior of various classifiers on different block sizes on CK dataset.

#Blocks	1×1	2×2	3×2	3×3	4×3	4×4	5×4	5×5	6×5	6×6	7×6	7×7	8×7	8×8	9×8	9×9
L2-Norm	81.4	86.5	87.3	88.2	89.5	91.1	91.4	92	92.1	92.3	**92.5**	91.7	91.5	90.4	90.3	87.6
Chi-Square	82.2	85.8	87.5	89.1	89.5	90.9	91.2	92.7	92.8	93	**93.1**	92.8	91.9	91.1	91	88.4
HNAD	84.1	87.6	88.2	92.7	93.2	94.7	95.5	96.8	96.9	97.4	**98.2**	96.2	93.4	92.8	92.5	91.1
SVM(L)	83.4	85.6	87.7	92.4	93.8	94.9	95.7	95.6	96.6	96.9	**97.4**	92.1	92.6	91.3	90.4	89.2
SVM(P)	84.4	85.9	87.8	93.4	94.8	95.1	95.8	96.2	96.6	97.0	**98.1**	97.2	93.5	90.8	90.4	89.7
SVM(R)	86.1	89.3	89.4	94.6	95.5	95.8	96.1	97.0	97.2	98.3	**99.3**	97.3	95.7	94.9	93.3	81.7

Three template matching schemes are adopted: L2 norm, Chi-square, and HNAD. In recent time, LS-SVM [28] has emerged as first choice classifiers for researchers. We also conducted experiments using LS-SVM with three different kernels: linear, polynomial and RBF. Facial expression classification is a multi-class problem. We used One-versus-One SVM to extend the functionality of basic binary SVM to the multiclass problem. From results, we chose a number of regions to be 7×6, as it gives best recognition rate among all test combinations.

5.2 Subject Independent Tests

The accuracy of the system is evaluated in two different test conditions: subject independent and cross database evaluation. In the subject independent experiment, the dataset is split into k folds, and experiment is repeated k times. We chose $k = 10$ in our experiment. In each iteration, $k - 1$ folds are used for training and remaining one is used for testing. Results are averaged over k executions and are compared in Table 3 for all discussed datasets. From results of Table 2, we choose HNAD template matching and SVM with RBF kernel for classification with 7×6 blocks.

Table 3. Comparison of subject-independent test with state-of-the-art methods.

Dataset	CK	JAFFE	SFEW
HNAD	98.2 ± 0.9	88.0 ± 3.8	45.0 ± 8.8
SVM (RBF)	99.3 ± 0.5	90.3 ± 2.5	52.0 ± 6.5

5.2 Cross Database Evaluation

In this section, we test the generalization ability of LHMBP to unseen data. In cross

database evaluation, one dataset is used for testing, and remaining all are used for training. Due to the unique signature of each dataset, the cross-dataset task is much more challenging compared to subject-independent test. Table 4 shows the accuracy of cross database experiments. Results are reported using SVM with RBF kernel. The level of difficulty of the cross-dataset experiment can be seen from the results. In cross-dataset experiments, faces are not aligned, which introduces feature alignment and ultimately results in poor accuracy. Thus, the pose registration is essential to develop a robust system.

Table 4. Average recognition rate (%) on cross-dataset.

Test Dataset	AN	DI	FE	HA	SA	SU	NE	Average
CK	52.2	57.5	56.5	65.3	51.2	69.3	63.4	59.3
JAFFE	50.3	49.3	47.3	53.5	50.7	61.5	49.6	51.7
SFEW	38.2	37.5	33.3	43.3	35.5	47.2	36.6	38.8

5.4 Expression Specific Recognition Rate

Results discussed so far were averaged over all seven expressions. To get a better idea about recognition rate of individual expression, confusion matrices for the individual dataset are shown in Table 5. Results in confusion matrices are derived using HNAD template matching.

Table 5. Confusion matrices for CK, JAFFE, and SFEW.

	CK							JAFFE							SFEW						
	AN	DI	FE	HA	SA	SU	NE	AN	DI	FE	HA	SA	SU	NE	AN	DI	FE	HA	SA	SU	NE
AN	**96.3**	0.3	1.2	0.4	1.2	0.2	0.4	**85.4**	4.5	3.6	1.5	2.5	1	1.5	**40.1**	12.3	12.5	8.7	9.3	8.6	8.5
DI	0.9	**97.2**	0.8	0.2	0.6	0.0	0.3	2.9	**84.7**	5.1	1.5	3.5	1.2	1.1	12.4	**43.2**	10.4	7.7	8.4	8.7	9.2
FE	0.8	0.2	**98.8**	0.0	0.2	0.0	0.0	3.5	2.3	**87.4**	1.2	3.2	1.1	1.3	10.2	9.6	**44.1**	7.8	9.5	10.2	8.6
HA	0.2	0.1	0.4	**99.1**	0.2	0.0	0.0	2.5	1.4	2.7	**90.6**	0.8	1.2	0.8	7.2	8.3	8.8	**51.1**	7.6	9.5	7.5
SA	0.6	0.3	0.3	0.1	**98.1**	0.1	0.5	4.2	2.6	2.5	0.9	**86.3**	0.3	3.2	11.2	9.4	11.4	6.8	**43.1**	7.9	10.2
SU	0.2	0.1	0.2	0.1	0.0	**99.4**	0.0	1.6	1.4	2.1	1.1	0.7	**92.3**	0.8	8.9	8.6	10.3	6.7	7.9	**52.3**	5.3
NE	0.4	0.4	0.1	0.1	0.7	0.1	**98.2**	1.7	1.8	2.7	0.9	3.2	0.5	**89.2**	12.7	8.5	7.9	7.7	13.4	8.7	**41.1**
Avg	**98.2**							**88.0**							**45.0**						

Without any doubt, surprise and happy expressions are visually distinguishable compared to other expressions. Maximum facial muscle deformation takes place during surprise expression. Experimental results also show that surprise and happy expressions have better recognition rate compared to other expressions. The expression pairs *angry-fear* and *sad-disgust* are often confusing, and they have a higher rate of false recognition for counter expression in the pair.

5.5 Expression Recognition in Low Resolution

In certain applications such as home monitoring, surveillance, smart meeting, only low-resolution videos are available [29]. FER in low resolution is almost unaddressed

area. Even for a human, recognizing facial expressions in low resolution is challenging. In our experiment, we studied the performance of LHMBP operator in four different resolutions: 150×110, 75×55, 48×36 and 37×27 and results are portrayed in Table 6. Low-resolution images are derived by down-sampling the original images.

Table 6. Expression recognition in low resolution using LS-SVM with RBF kernel.

#Resolution	150×110	75×55	48×36	37×27
CK	98.2	96.7	96.2	95.3
JAFFE	88.0	87.9	87.2	84.3
SFEW	45.0	42.5	41.7	39.8

Shan *et al.* [21] conducted the experiment on a low resolution with LBP and Gabor features derived by convolving the image with 40 Gabor filter banks. Tian [29] selected 375 image sequences from CK dataset. He used geometric and appearance features. Geometric features were computed using feature tracking and feature selection. Appearance features were derived using 40 Gabor filter banks. Jabid *et al.* [23] derived LDP features using 8 Kirsch edge detectors. They achieved remarkable accuracy using SVM. Due to down sampling and smaller size of the image, geometric features do not make sense, as detecting and tracking geometric feature is also challenging and vulnerable in such environment. This fact motivated us to choose only appearance features in our framework. The performance of LHMBP operator at different resolutions is compared with various existing approaches in Table 7.

Table 7. Performance comparison of expression recognition in low resolution on CK.

Resolution	150×110	75×55	48×36	37×27
Feature Descriptor				
Gabor [29]	92.2	91.6	–	77.6
Gabor [30]	89.1 ± 3.1	89.2 ± 3.0	86.4 ± 3.3	83.0 ± 4.3
LBP [21]	92.6 ± 2.9	89.9 ± 3.1	87.3 ± 3.4	84.3 ± 4.1
Gabor [21]	89.8 ± 3.1	89.2 ± 3.0	86.4 ± 3.3	83.0 ± 4.3
LDP [23]	96.4 ± 0.9	95.5 ± 1.6	93.1 ± 2.2	90.6 ± 2.7
LHMBP + HNAD	**98.2 ± 0.9**	**96.5 ± 1.1**	**96.2 ± 1.1**	**95.3 ± 1.4**
LHMBP + SVM (RBF)	**99.3 ± 0.5**	**98.7 ± 0.7**	**97.5 ± 1.1**	**96.4 ± 1.3**

5.6 Comparison with State of the Art Methods

In this section, we compared the performance LHMBP feature descriptor with the recent state of the art methods. SFEW is a comparatively new dataset, and very few researchers have utilized it in their experiment. CK and JAFFE are designed under a controlled environment, and hence it is not difficult to localize the features after face registration. However, due to considerable variability in pose, illumination, size, and

occlusion, face registration is challenging in SFEW. The ethnic difference also plays a vital role in the success of expression recognition systems. Subjects in JAFFE belong to the same ethnicity. In CK, 15% of subjects belong to African-American background, and 3% subjects belong to Asian or the Latino-American background. Subjects in SFEW belong to largely separated geographical areas. The performance of LHMBP with various state of the art methods is compared in Table 8 for all three used datasets.

Table 8. Performance comparison with state of the art methods.

Dataset	[21]	[23]	[31]	[10]	[32]	[18]	[12]	[33]	[34]	[35]	**LHMBP + SVM**	**LHMBP + HNAD**
CK	88.9	93.4	96.6	94.1	45.7	92.3	97.8	–	96.6	95.4	**99.3**	**98.2**
JAFFE	80.7	84.9	–	91.8	95.2	–	–	–	98.8	–	**90.3**	**88.0**
SFEW	–	–	–	–	–	–	47.7	24.7	–	54.3	**52.0**	**45.0**

5.2 Feature Extraction Time

To localize the features, we divide the image in $M \times N$ blocks and LHMBP features are extracted from each block. We analyze the effect of a number of blocks on feature extraction time. Features are extracted from an individual block, so it is evident that feature extraction time increases with increase in a number of blocks. Extraction of haar is straightforward, it only finds the running average and running difference of original signal. The procedure involves only arithmetic addition, subtraction, and multiplication. Haar-based feature extraction does not involve any neighborhood processing, and hence it takes considerably less time.

LMBP computes the mean of 3×3 patch for each pixel, neighbor pixels are thresholded with respect to mean, and finally, the decimal code is calculated from a thresholded binary string. Histogram of this decimal code is used as a feature vector. This will make it computationally expensive compared to haar.

LHMBP is a combination of both, the feature extraction time of LHMBP is somewhere in the midst of both. In LHMBP, LMBP is applied on haar approximation subband of the query image, which is four times smaller than the original image. Hence the feature extraction time of LHMBP is less compared to LMBP. Feature extraction time for different numbers of blocks is compared in Table 9.

Table 9. Feature extraction time (second) of each step for different block size.

#Blocks	1 × 1	3 × 3	5 × 5	7 × 6	9 × 8
Haar	0.0041	0.0166	0.0382	0.0620	0.1043
LMBP	0.5053	0.5058	0.5080	0.6048	0.6098
LHMBP	0.1280	0.1300	0.1310	0.1330	0.1363

6. CONCLUSIONS

Deriving discriminative feature descriptor is an important task in pattern recognition problems. Haar wavelets are straightforward and accurate. They can effectively express the signal with comparatively fewer dimensions. The ability of haar to preserve the sig-

nal energy with very few coefficients makes it a choice for image representation. Local and global histogram of LHMBP is concatenated to encode texture and shape of the face. The strength of LHMBP lies in its robust response to noise and non-monotonic gray level change. Gabor involves extensive computation due to large filter bank. LBP is sensitive to both, noise and non-monotonic gray level change. LDP is computationally expensive as it requires eight masks to compute a single LDP code. As LHMBP performs the thresholding with respect to mean of the patch, it can sustain against noise and illumination changes stronger than other descriptors.

The average accuracy of 99.3% for the 7-class problem using LS-SVM on CK dataset is really encouraging, compared to the 91.4% accuracy reported by Shan *et al.* [21]. Bartlett *et al.* [30] achieved the highest recognition rate of 93.3% using a subset of Gabor filters with Adaboost feature selection and SVM classifier. Using three layer neural network, Tian [29] achieved 94% recognition rate with Gabor and geometric feature. Results show that LHMBP is consistent, it achieves good recognition rate with minimum standard deviation from the average value. With template matching using HNAD measure, presented method achieves a recognition rate of 98.2%, 88.0% and 45.0% for CK, JAFFE and SFEW dataset. And with SVM, highest achieved results are 99.3%, 90.3% and 52.0% for respective datasets. In the low-resolution environment, LHMBP achieves a recognition rate of 96.4% for 37×27 pixel image, which is remarkable.

In future, we would like to extend the functionality of LHMBP to larger neighborhood. Rotation invariance property also needs to be tested.

REFERENCES

1. A. Mehrabian, "Communication without words," *Psychology Today*, Vol. 2, 1968, pp. 53-55.
2. B. Hasani and M. H. Mahoor, "Spatio-temporal facial expression recognition using convolutional neural networks and conditional random fields," arXiv preprint arXiv: 1703.06995, 2017.
3. X. Li, X. Hong, A. Moilanen, X. Huang, T. Pfister, G. Zhao, and M. Pietikäinen, "Towards reading hidden emotions: A comparative study of spontaneous micro-expression spotting and recognition methods," *IEEE Transactions on Affective Computing*, 2017, pp. 1-14.
4. J. Parkkinen, M. H. Jaward, R. Parthiban, and S. K. A. Kamarol, "Spatiotemporal feature extraction for facial expression recognition," *IET Image Processing*, Vol. 10, 2016, pp. 534-541.
5. D. H. Kim, W. Baddar, J. Jang, and Y. M. Ro, "Multi-objective based spatio-temporal feature representation learning robust to expression intensity variations for facial expression recognition," *IEEE Transactions on Affective Computing*, 2017, pp. 1-15.
6. D. A. Chanti and A. Caplier, "Spontaneous facial expression recognition using sparse representation," in *Proceedings of the 12th International Joint Conference on Computer Vision*, *Imaging and Computer Graphics Theory and Applications*, Vol. 5, 2017, pp. 64-74.

7. J. Wu, Z. Lin, W. Zheng, and H. Zha, "Locality-constrained linear coding based bi-layer model for multi-view facial expression recognition," *Neurocomputing*, Vol. 239, 2017, pp. 143-152.
8. S. Datta, D. Sen, and R. Balasubramanian, "Integrating geometric and textural features for facial emotion classification using SVM frameworks," *Advances in Intelligent Systems and Computing*, Vol. 459, 2017, pp. 1-10.
9. X. Zhao, J. Yuan, H. Liu, and J. Zhou, "Improved adaboost algorithm for robust real-time multi-face detection," *IEEE Transactions on Image Processing*, Vol. 12, 2016, pp. 53-61.
10. S. L. Happy and A. Routray, "Automatic facial expression recognition using features of salient facial patches," *IEEE Transactions on Affective Computing*, Vol. 6, 2015, pp. 1-12.
11. A. T. Lopes, E. de Aguiar, A. F. De Souza, and T. Oliveira-Santos, "Facial expression recognition with convolutional neural networks: Coping with few data and the training sample order," *Pattern Recognition*, Vol. 61, 2017, pp. 610-628.
12. A. Mollahosseini, D. Chan, and M. H. Mahoor, "Going deeper in facial expression recognition using deep neural networks," in *Proceedings of IEEE Winter Conference on Applications of Computer Vision*, 2015, pp. 1-10.
13. Y. Kim, B. Yoo, Y. Kwak, C. Choi, and J. Kim, "Deep generative-contrastive networks for facial expression recognition," arXiv preprint arXiv:1703.07140, 2017.
14. A. Palmer, "An evaluation of the two-dimensional gabor filter model of simple receptive fields in cat striate cortex," *Journal of Neurophysiology*, Vol. 58, 1987, pp. 1233-1258.
15. Y. Tian, T. Kanade, and J. F. Cohn, "Evaluation of gabor-wavelet-based facial action unit recognition in image sequences of increasing complexity," in *Proceedings of IEEE International Conference on Automatic Face and Gesture Recognition*, 2002, pp. 2-7.
16. Z. Zhang, M. Lyons, M. Schuster, and S. Akamatsu, "Comparison between geometry-based and gabor-wavelets-based facial expression recognition using multi-layer perceptron," in *Proceedings of the 3rd IEEE International Conference on Automatic Face and Gesture Recognition*, 1998, pp. 454-459.
17. T. R. Almaev and M. F. Valstar, "Local gabor binary patterns from three orthogonal planes for automatic facial expression recognition," in *Proceedings of Humaine Association Conference on Affective Computing and Intelligent Interaction*, 2013, pp. 356-361.
18. Y. Sun and J. Yu, "Facial expression recognition by fusing gabor and local binary pattern features," in *Proceedings of International Conference on Multimedia Modeling*, 2017, pp. 209-220.
19. T. Ojala, M. Pietikainen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, Vol. 29, 1996, pp. 51-59.
20. T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 24, 2002, pp. 971-987.

21. C. Shan, S. Gong, and P. W. Mcowan, "Facial expression recognition based on local binary patterns: A comprehensive study," *Image and Vision Computing*, Vol. 27, 2009, pp. 803-816.
22. R. Hablani, N. Chaudhari, and S. Tanwani, "Recognition of facial expressions using local binary patterns of important facial parts," *International Journal of Image Processing*, Vol. 7, 2013, pp. 16-170.
23. T. Jabid, M. H. Kabir, and O. Chae, "Robust facial expression recognition based on local directional pattern," *ETRI Journal*, Vol. 32, 2010, pp. 784-794.
24. T. Kanade, J. F. Cohn, and Y. Tian, "Comprehensive database for facial expression analysis," in *Proceedings of the 4th IEEE International Conference on Automatic Face and Gesture Recognition*, 2000, pp. 46-53.
25. M. J. Lyons, "Automatic classification of single facial images," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 21, 1999, pp. 1357-1362.
26. A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: Data, evaluation protocol and benchmark," in *Proceedings of IEEE International Conference on Computer Vision Workshops*, 2011, pp. 2106-2112.
27. A. Dhall, R. Goecke, J. Joshi, M. Wagner, and T. Gedeon, "Emotion recognition in the wild challenge 2013," in *Proceedings of the 15th ACM on International Conference on Multimodal Interaction*, 2013, pp. 509-516.
28. J. A. K. Suykens, T. van Gestel, J. de Brabanter, B. deMoor, and J. Vandewalle, *Least Squares Support Vector Machines*, World Scientific, Singapore, 2002.
29. Y. T. Y. Tian, "Evaluation of face resolution for expression analysis," in *Proceedings of Conference on Computer Vision and Pattern Recognition Workshop*, 2004, pp. 1-7.
30. M. S. Bartlett, G. Littlewort, M. Frank, C. Lainscsek, I. Fasel, and J. Movellan, "Recognizing facial expression: Machine learning and application to spontaneous behavior," in *Proceedings of IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Vol. 2, 2005, pp. 568-573.
31. J. Zhao, Y. Zhao, Y. Yang, Y. Huang, and I. Park, "Facial expression recognition method based on stacked denoising autoencoders and feature reduction," in *Proceedings of International Conference on Engineering Technology and Application*, 2016, pp. 190-194.
32. R. Jiang, A. T. S. Ho, I. Cheheb, N. Al-Maadeed, S. Al-Maadeed, and A. Bouridane, "Emotion recognition from scrambled facial images via many graph embedding," *Pattern Recognition*, Vol. 67, 2017, pp. 245-251.
33. S. Eleftheriadis, O. Rudovic, and M. Pantic, "Discriminative shared gaussian processes for multiview and view-invariant facial expression recognition," *IEEE Transactions on Image Processing*, Vol. 24, 2015, pp. 189-204.
34. M. Z. Uddin, M. M. Hassan, A. Almogren, A. Alamri, M. AlRubaian, and G. Fortino, "Facial expression recognition utilizing local direction-based robust features and deep belief network," *IEEE Access*, Vol. 5, 2017, pp. 1-9.
35. Z. Meng, P. Liu, J. Cai, S. Han, and Y. Tong, "Identity aware convolutional neural network for facial expression recognition," in *Proceedings of the 12th IEEE International Conference on Automatic Face and Gesture Recognition*, 2017, pp. 558-565.

Mahesh Goyani received the master degree in Computer Engineering from Sardar Patel University, India. He is currently pursuing Ph.D. in Department of Computer Engineering, Charotar University of Science and Technology, India. His research interests include pattern recognition, machine learning and analysis of algorithms.

Narendra Patel received his B.E. degree in Electronics Engineering from Maharaja Sayajirao University, Baroda in 1993 and M.E. degree from M.S. University, Baroda in 1997. He received Ph.D. degree from SVNIT, Surat in 2012. His research interests include digital image processing, real time operating systems, distributed systems and computer graphics.