Cluster Comput
DOI 10.1007/s10586-017-0921-5

Facial expression recognition using histogram of oriented gradients based transformed features

Muhammad Nazir[1,2] · Zahoor Jan[1] · Muhammad Sajjad[1]

Received: 23 March 2017 / Revised: 3 May 2017 / Accepted: 9 May 2017

Abstract Facial expression recognition has been an emerging and long standing research problem in last two decades. Histograms of oriented gradients (HOGs) have proven to be an effective descriptor for preserving the local information using orientation density distribution and gradient of the edge. A robust powerful approach of HOG features has been investigated in this paper. In particular, this paper highlights that the transformation of HOG features to frequency domain can make this descriptor one of the most suitable to characterize illumination and orientation invariant facial expressions. Discrete cosine transform (DCT) is applied to transform the features into frequency domain and obtain the most important discriminant features. Finally, these features are fed to the well-known classifier to determine the underlying emotions from expressive facial images. To validate the proposed framework, we used MMI, Extended Cohn-Kanade dataset (CK+) and cross dataset. The results indicate that the proposed framework is better as compared to other methods in terms of classification accuracy rate with utilization of minimum number of features.

Keywords Facial expression recognition · Histograms of oriented gradients · Discrete cosine transform · Cross dataset

✉ Muhammad Sajjad
muhammad.sajjad@icp.edu.pk

Muhammad Nazir
muhammad.nazir@hitecuni.edu.pk

Zahoor Jan
zahoor.jan@icp.edu.pk

1 Digital Image Processing Laboratory, Department of Computer Science, Islamia College, Peshawar, Pakistan

2 Department of Computer Science and Engineering, HITEC University, Museum Road, Taxila, Pakistan

1 Introduction

Facial expression recognition has been extensively studied in the areas of computer vision and pattern recognition over the past decades due to its potential applications like human robot interaction and intelligent affective computing. It contains the cues of the non-verbal communication to help in understanding the intended meaning of the verbal communication.

Various frameworks and different styles of data are used to recognize facial expressions. Generally, the data uses from static images and image sequences [1,2] for facial expression recognition. We chose the peak image frames for experiments because they contain sufficient information of the specific expressions. Facial expression recognition system can be divided into three steps: (1) Face acquisition (2) Feature extraction and representation (3) Facial expression classification.

Face portion is extracted in face acquisition step. Feature extraction technique is grouped into two categories i.e. (1) appearance based features and (2) geometric based features. Different face components including nose, mouth and eyebrows are used to represent the geometric features of face. HOG [3] and Local binary pattern [4] descriptors are used as appearance features. Finally, in classification step, different classifiers like K-nearest neighbor (KNN), (SMO) [5] and Random forest (RF) [6] are used to recognized facial expressions.

Most of the previous studies focus on recognizing face expressions from still facial images. However, in real world application it is more convenient to use videos as it provide more information as compared to expression image. The key issue of facial expression recognition in video is how to extract illumination and orientation invariant features which more clearly represent face structure. Similarly, low

Published online: 30 May 2017

resolution and large data dimensions makes this task even harder.

We have applied two main algorithms in facial expression recognition work. Histogram of gradient orientation [3] is first used to store the local information of face image and discrete cosine transform is performed to select the prominent face features and reduce the data dimensions. Gradient of the HOG features corresponds to the first derivative of the image and it also reflects the shape information of the image which makes this descriptor robust to illumination and orientation. To the best of our knowledge, HOG has been used for facial expression recognition in few studies.

Kar et al. [7] used HOG descriptor to extract the robust face features. Linear discriminant analysis (LDA) and principal component analysis (PCA) are applied to obtain the important features. Back-propagation neural network (BPNN) classifier is trained and tested on the extracted features. In order to extract more efficient features, Salient areas on the faces is first defined in [8] and then HOG and LBP descriptors are used to extract the face features. Next, fusion of LBP and HOG feature is performed to achieve better results and then PCA is utilized to reduce the data dimensions. Experiments are performed on JAFFE and CK+ datasets. Shape and appearance-based features are combined in [9] to generate hybrid feature vector. LBP is utilized to extract the appearance features and pyramid of histogram of gradients (PHOG) is used for shape features. To capture the information of major changes during different expressions, active facial patches are located and then hybrid features are extracted from these facial patches. Linear discriminant analysis is applied to select the more discriminative face features and then fed these features to support vector machine for classification. Wang et al. [10] proposed a hybrid scheme by combing the weber local descriptor and HOG features. First the image is divided into different regions and then features are extracted from these regions. Nearest neighbor and chi-square distance is used to classify the weighted fused histograms for different expressions. Average classification accuracy rate of 95.86% has been obtained for CK+ database. Facial components like mouth, nose and eyebrows are first located in [11], and then HOG is used to encode these components. To perform different experiments, cell size is set to 8×8 and histogram bins to 9. The effectiveness of the proposed framework is tested on CK+ and JAFFE database. Donia et al. [1,2] performed experiments on both static and video database for facial expression classification. HOG is utilized to perform the micro expression analysis. The face image is first divided into six basic components which are the most representative regions for expression recognition. SVM is used for training and testing and the average accuracy rate of 95 and 80% is reported on static and videos respectively. Fan and Tjahjad [12] formulated the 3-dimensional facial features by extending the histogram of gradients to spatio-temporal domain and then integrate with dense optical flow which contribute to both dynamic and spatial motion of facial expressions. Experiments are performed on MMI and CK+ database with leave-one-out cross validation strategy and multi-class support vector machine is used for classifying different facial expressions. A novel framework has been presented in [13] to handle low resolution images. The local binary pattern (LBP) has been extended to preserve the texture and spatial layout of an image. The extended version of the LBP is known as pyramid of local binary pattern (PLBP). PLBP was used to extract the face features from only salient region and thus computationally efficient. PLBP equally work well on high and low resolution images. Similarly, a novel framework known as Weber local binary image based cosine transform features (WLBI-CT) has been presented in [14] to handle the multi-scaling and multi-orientation problem.

Previous recognition methods tend to focus only on movement of facial landmarks, not on the analysis of the effect of multi-scaling, illumination and variations in the facial shape. In this paper we utilize HOG to extract the dynamic and orientation features from an image and then discrete cosine transform (DCT) is used to select the high variance features from HOG image.

The main contributions of the proposed framework are summarized as follows:

- Proposed framework is reliable for multi scale images and work adequately on videos.
- The selected features using DCT maximize the between class variations and minimize the within-class variation of expressions.
- Proposed framework is computationally efficient and can be used for real time applications.
- Proposed framework is robust to illumination and orientations.

This article has been organized as follows: In Sect. 2, we provide an overview of the proposed method. Empirical results and related discussion are provided in Sect. 3. We conclude our work with Sect. 4.

2 Overview of the proposed technique

The steps of the proposed contrast enhancement framework are outlined below:

(1) Take input human facial image representing facial expressions.
(2) Face Detection using Viola and Jones Algorithm.
(3) Apply HOG for large gradient features and then normalize the features
(4) Image is formulated from the normalized features.

(5) Transformation of feature image into frequency domain using DCT
(6) High variance features are extracted from the frequency domain in a zigzag manner to form a feature vector.
(7) Feature vector is fed to KNN, SMO and MLP for classification of expressions.

Figure 1 illustrates the proposed framework steps.

Algorithmic steps of the proposed framework are given below;

Gradients thus obtained are then utilized to calculate gradient magnitude and angular orientations using the Eqs. 3 and 4.

$$m(x, y) = \sqrt{g_v^2(x, y) + g_h^2(x, y)} \tag{3}$$

$$\theta(x, y) = \tan^{-1}\left(\frac{g_v(x, y)}{g_h(x, y)}\right) \tag{4}$$

It divides the image into cells. A block is formed consisting of various cells and features are constituted by convolving

ALGORITHM

1. **Input:** Facial images $I(I_1, I_2, I_3, ..., I_N)$ representing facial expressions
2. I_f = Face Detection using Viola and Jones Algorithm from I
3. Compute $HOG_I(I_f)$ using equation(1-4).
4. $[f_1, f_2, f_3, ..., f_M] = HOG_I(I_f)$ where M = number of HOG features calculated using equation (5).
5. Compute feature set $OGDCT_{fs} = DCT[HOG_{img}]$ high variance feature extraction in zig-zag manner from HOG-image equations (6-7).
6. Store $OGDCT_{fs}$ features
7. Repeat steps (2-6) for N input images
8. Train classifiers for all expressions
9. Perform Classification

Symbol	Description
$I(I_1, I_2, I_3, \ldots, I_N)$	N input images.
I_f	Facial Image extracted using Viola and Jones Algorithm.
HOG	Histogram Oriented Gradient function.
$f_1, f_2, f_3, \ldots, f_M$	m-number of Features formed using HOG.
HOG_{img}	Image constructed after applying HOG.
$OGDCT_{fs}$	Resultant feature set extracted using DCT for HOG image.

2.1 Materials and methods

Histogram of oriented gradients (HOG) and discrete cosine transform (DCT) are presented in this section.

2.1.1 Histogram of oriented gradients (HOG)

HOG descriptor was developed by Dalal and Triggs [3] and has been successfully utilized by many researchers working in the domain of computer vision. Mostly it has been used for human detection, object identification and pedestrian identification. HOG is computed using magnitude and orientation. Horizontal and vertical gradients of the input image are computed using the following Eqs. 1 and 2.

$$G_x = I_f * [-1, 0, 1] \tag{1}$$

$$G_y = I_f * [-1, 0, 1]^T \tag{2}$$

the blocks. These blocks are overlapped. Orientations related to the same cell are quantized and are integrated into final histogram bins. These histogram bins are sorted and are combined into final histogram. Let T_{hog_fs} represents the total number of features computed using HOG descriptor. B_{img} denotes blocks per image, block size is represented B_s and $N_b$denotes the number of bins used. Then total number of features can be computed using the following formula given in Eq. 5.

$$T_{hog\_fs} = B_{img} * B_s * N_b \tag{5}$$

2.1.2 Discrete cosine transform (DCT)

DCT is a well-known feature extraction technique being used by computer vision research community for solving various problems. We can take advantage of its discrimination capabilities in the areas of energy compaction, orthogonality, decorrelation and high separability. These properties make

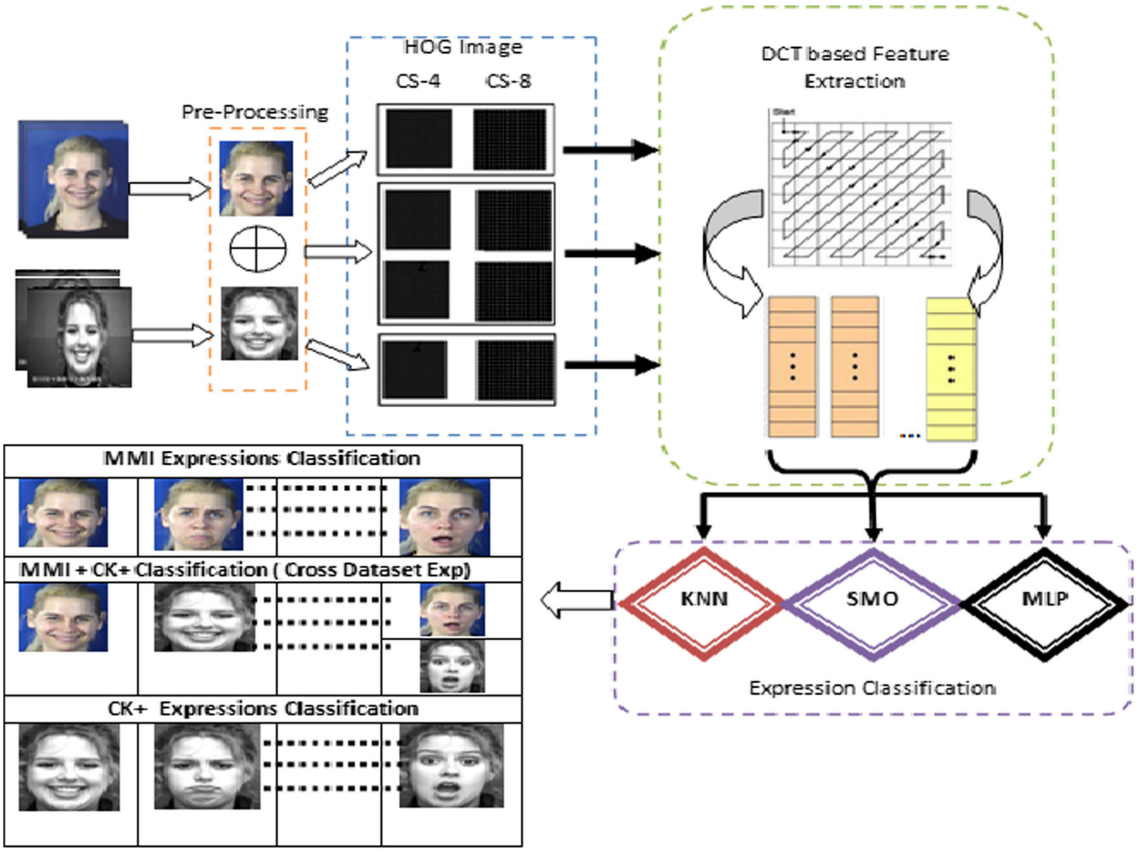

Fig. 1 Proposed framework

it robust in solving complex problems like facial expression recognition. We have applied DCT on the image formed after applying normalized HOG using the following Eq. 6. Features having high variations lie near top left corner and carry most significant information and realizes the effectiveness of the DCT for feature extraction. These features are extracted in zigzag manner starting from top left corner leaving the mean value. DCT of an input HOG image $f(x, y)$ of size NxM can be calculated using the Eqs. 6 and 7.

$$F(u, v) = \alpha(u)\,\alpha(v) \sum_{x=0}^{N-1} \sum_{y=0}^{M-1} \cos\left[\frac{\pi u}{2N}(2x+1)\right] \times \cos\left[\frac{\pi u}{2M}(2y+1)\right] f(x, y) \quad (6)$$

where $f(x, y)$ represents the pixel intensities at point (x, y) and $u = \{0, 1 \ldots N-1\}$ and $v = \{0, 1 \ldots M-1\}$. The function $\alpha(u)$ and $\alpha(v)$ are defined as;

$$\alpha(u),\ \alpha(v) = \begin{cases} \sqrt{\frac{1}{N}}\ for\ u,v{=}0 \\ \sqrt{\frac{2}{N}}\ for\ u,v{\neq}0 \end{cases} \quad (7)$$

2.1.3 Classification

K-nearest neighbor(KNN) classifier is considered to be the simplest but an efficient algorithm and has been applied successfully by research community for solving both supervised and unsupervised learning problems. It classifies an input sample among its k -nearest neighbors in the training data. Distance measure used for classification should be selected carefully in order to get the improved performance. Task of classification using KNN is done by taking a feature vector f_v and calculating the distance between f_v and feature vectors of known training samples. The distance between the training set and the test set can be calculated by using one of the distance metrics i.e. Euclidean distance, the Manhattan distance or cosine distance measure. For classification purpose the dataset is divided into training set and test set using different train to test ratios. Let the training set be represented by $T = \{(s_i, c_1), (s_2, c_2), \ldots, (s_n, c_n)\}$, where n is the number of input feature vectors with their output class labels $c_i (i = 1, 2, \ldots, n)$. The test sample t_i as feature vector f_v is assigned the class label with majority of k-neighbors having least distance. We have used the Euclidean distance (D_e) using the formula given in Eq 8 defined as follows [15].

$$D_e(s,t) = \sqrt{\sum_{i}^{l} (s^2(i) - t^2(i))} \tag{8}$$

Here s_i is the number of feature set used for training and t_i is the feature set used for testing. l is length of the feature vector.

3 Experiments

Three set of experiments are performed on CK+, MMI and cross dataset including both MMI and CK+ images. 10-fold cross validation scheme is employed to all experiments. The original sequences are down sample to obtain the low resolution images. Three different resolutions (128×128, 64×64 and 32×32) are used to test the proposed framework. Table 1 shows the parameter settings for KNN, SMO and Random forest classifier.

AdaBoost face detection algorithm proposed by Voila and Jones [16] is firstly used to detect all the face images. After detection of face image the original image of size 256×256 is reduced to 128×128 sizes. All the extracted face images are passed to HOG descriptor for feature extraction. The HOG divides the image into 16×16 block size and overlap of 50% is used. It results into total $15 \times 15 = 225$ block. Each block consists of 2×2 cells each of size 8×8. Gradient orientations range is set between [0–180] and is quantized using histograms of 9 bins. In this way, the total number of HOG features for input image of size 128×128 becomes 8100. HOG features $= 15 \times 15 \times 2 \times 2 \times 9$

Normally two main parameters are used to describe characteristics of the HOG descriptor i.e. orientation bines and cell size. The dimension of the patch involved in the single histogram computation is represented by cell size. Appropriate selection of parameters plays a vital role and should be selected with care for better performance of classification algorithm. Features extracted using HOG are affected by varying cell size. In case of using larger cell size, Spatial information of certain region from facial image is squeezed into the unit cell histogram, and the contribution becomes less significant. High resolution analysis on the other hand can be done by using smaller cell size, but it results in more detailed information and in order get most significant information from larger features becomes computationally expensive and also can be contributed towards reduced accuracy. Similarly, while calculating HOG features, selection of number of orientation bins can play important role. It is related to the quantization levels of the gradient information. Significant information can be lost by using less number of orientation bins. Whereas, using more quantization levels, classification performance get affected because the information can be scattered along the bins. We have analyzed the classification performance on the dataset both with respect to cell size and orientation bins. Figures 2, 3 and 4 shows the happy face image with varying cell size and different image resolutions. After detailed experiments, we have found maximum performance on cell size 8×8 and bin size is set to be 9 with block size 2×2.

HOG used gradient histogram which is not scale invariant [17]. Choosing the scale invariant coefficient for reliable facial expression recognition is a challenging task. The proposed framework integrating the DCT provides scale invariance for input image having different resolutions. DCT is applied to HOG image to select the high variance features of different length in zigzag manner. DCT method is computationally efficient as it did not require complex arithmetic.

3.1 Experiments on CK+ database

In order to evaluate the performance of the proposed algorithm extended CK+ dataset (CK+) [18] has been applied. In most of the work done in this area CK+ is being used extensively and is publicly available. The dataset consists of 593 image sequences taken from 120 subjects. Ages of the subjects used in the dataset ranges from 18 to 30 years. One of the major characteristic of this dataset is that the subjects in the dataset belong to different gender and races. Which makes it versatile and algorithm trained become more generalized. Ratio of different subjects consists of approximately 65% female, 81% Euro-American, 13% Afro-American, and 6% of other racial groups. Seven basic facial expressions, namely Anger, Contempt, Disgust, Fear, Happiness, Sadness and Surprise are given in the dataset. Images with resolution 640×480 or 640×490 pixels having 8-bit gray scale are used. We used 540 image sequences of six basic expressions. Neutral expression has been excluded. Sample of the images used is shown in the Fig. 5 and the number of images used per expression is illustrated by Table 2.

Table 1 Classifiers parameter settings

Parameter settings for KNN		Parameter settings for SMO		Parameter settings for random forest	
Set of nearest neighbors	2	Complexity parameter ‘C’	1.0	Number of iterations	100
nearest neighbor search algorithm	Euclidean distance	Kernal	Poly kernel	Bag size	100
batch size	100	Number of folds	-1	Seed	1

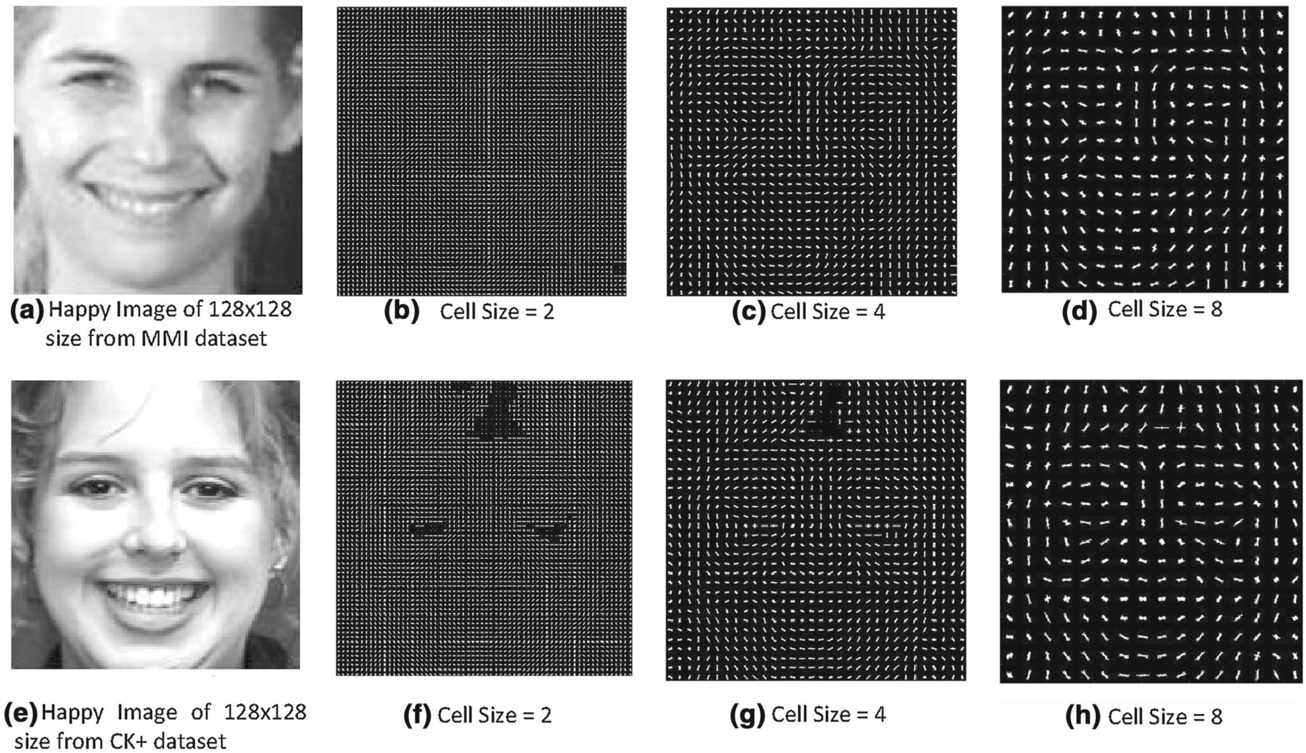

Fig. 2 HOG processed CK+ and MMI images using different cell size and 128 × 128 image size

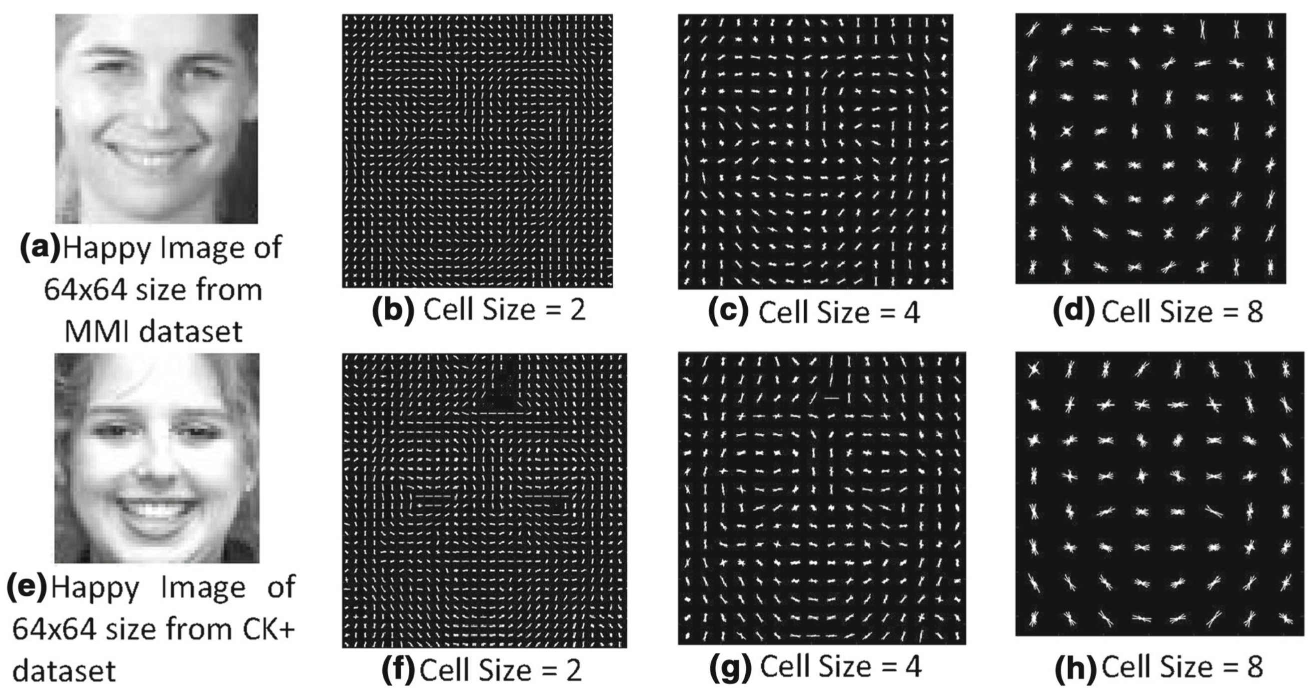

Fig. 3 HOG processed CK+ and MMI images using different cell size and 64 × 64 image size

Figure 6 illustrates the average recognition accuracy of the proposed framework with different feature vector size (FV-32, FV-48, FV-64, FV-80, FV-96, FV-112).

It has been observed that the proposed framework work amicably on low resolution images by utilizing less number of features as shown in Fig. 6. We stop increasing the

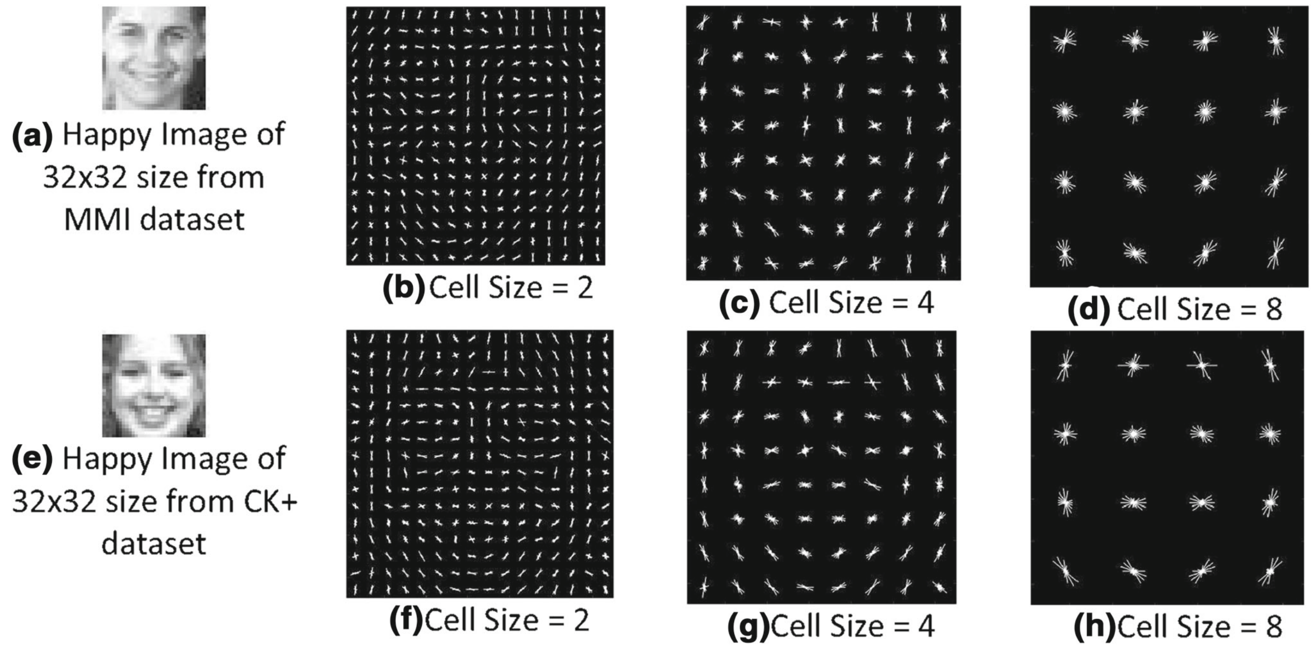

Fig. 4 HOG processed CK+ and MMI images using different cell size and 32 × 32 image size

Fig. 5 The Sample images of CK+ database

Table 2 Number of CK+ image per expression

Expression	No. of images
Happy	90
Sad	90
Surprised	90
Angry	90
Disgust	90
Fear	90
Total	540

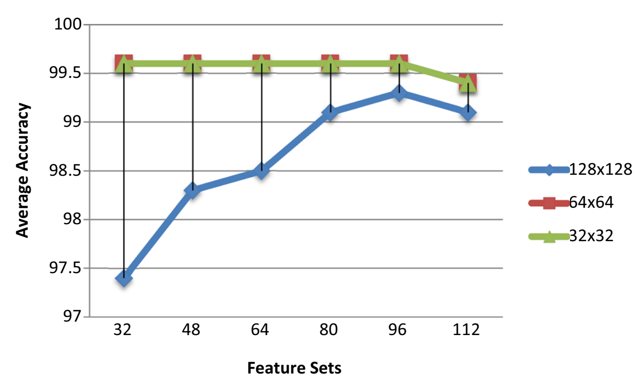

Fig. 6 Average recognition rate with increasing number of features and with different resolution

feature vector size after getting the feature vector of size FV-112 because adding more features will increase the computational time and there is minor improvement in accuracy. The maximum average accuracy of 99.6% has been obtained for images of size 64 × 64 and 32 × 32 using feature vector of size 32. Due to the lower feature vector dimensions the proposed framework can be utilized for real time applications.

In general, the proposed framework also achieved high accuracy using other classifiers like Sequential minimal optimization (SMO) and Random forest (RF) which shows the strength of the extracted feature irrespective to the classifiers. Figure 7 shows the average recognition accuracy rate of fixed feature vector size (FV-32) using KNN, SMO and RF classifiers.

To compare with Khan et al. [14] work, we tested our proposed framework on three different image resolution (i.e. 128 × 128, 64 × 64 and 32 × 32) on CK+ database. Facial image resolutions used in [14] are 144 × 192, 72 × 96 and 36 × 48 which is comparable to the resolutions used in our experiments. Referenced Fig. 8 shows the supremacy of the

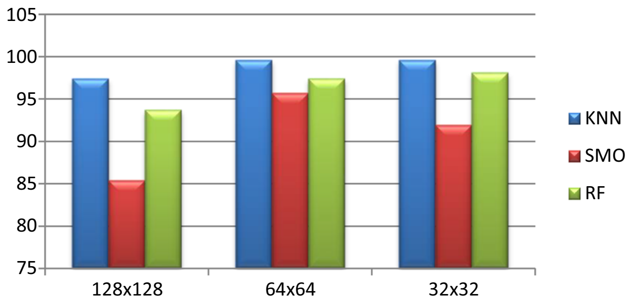

Fig. 7 Classification accuracy using different classifiers

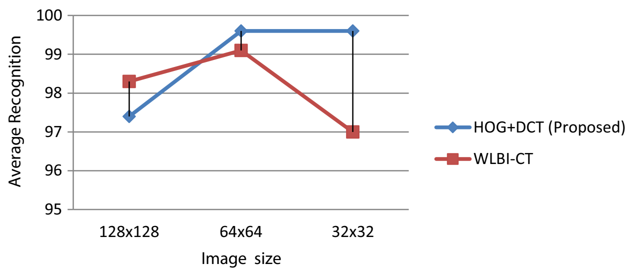

Fig. 8 Proposed framework accuracy rate comparison with other technique

proposed framework on different image resolutions as compared to other methods.

Khan et al. [14] reported the accuracy using feature vector of size 64 which indicate that our proposed framework data dimension cost is low as compared to [14].

3.2 Experiments on MMI database

The MMI dataset [19] was specially designed in 2002 to evaluate the performance of research techniques in the area of facial expression classification. The specialty of this dataset is that it contains full temporal pattern of complete facial expressions along with all existing action units. Subjects used in this dataset also belong to different race and ethnicity and age ranging from 19 to 62. The subjects performed six basic facial expressions. For experimental purpose, we have transformed the original frames into 8-bit grayscale images and extracted 273 image sequences labelled as one of the six basic facial expressions are selected from the MMI dataset.Sample of the images used is shown in the Fig. 9 and Table 3 illustrates the number of expression per image.

Detailed experiments are performed on MMI dataset as shown in Table 4. In order to demonstrate the effectiveness of the proposed technique on different resolutions we have used dataset with different resolutions of size 128×128, 64×64 and 32×32.The pattern shown in the Table 4 clearly shows the performance of the methodology.

The accuracy of the technique increased with increase in number of features initially. It keeps on increasing till 112 features and then it becomes almost stable with further increase in number of features. It is evident from the results shown in Table 4 that proposed techniques perform equally better on images of all sizes. It proves the robustness of the methodology against varying resolutions thus exhibiting multi resolution capability. We have achieved 96% accuracy on dataset size 64×64 by using only 48 features.

In the Table 5 results generated using proposed methodology on MMI dataset are shown by applying different classifiers. Consistency of the technique is obvious with all the three classifiers. We have presented these results by

Table 3 Number of images from each expression of MMI

Expression	No. of images
Happy	39
Sad	34
Surprised	39
Angry	45
Disgust	39
Fear	41
Neutral	36
Total	273

Table 4 MMI images accuracy rate with varying feature vectors and image size

Feature vector (FV)	Images resolutions		
	128×128	64×64	32×32
16	97.4	98.7	99.3
32	97.4	99.6	99.6
48	98.3	99.6	99.6
64	98.5	99.6	99.6
80	99.1	99.6	99.6
96	99.3	99.6	99.6
112	99.1	99.4	99.4

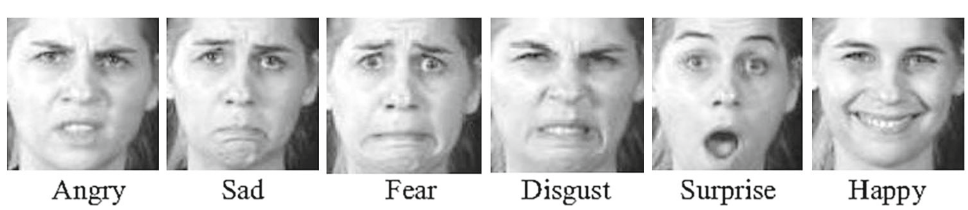

Fig. 9 The Sample images of MMI database

Table 5 MMI images accuracy rate with varying feature vectors and image size

Classifier	Images resolutions		
	128×128	64×64	32×32
KNN	93.8	96	95.2
SMO	79.9	93.8	93
RF	88.6	93	94.1

using only 48 features which are considerable as far as the number of features is concerned. It results in reduced computational cost. KNN performs better in all sizes. The results at low resolutions are comparatively better. Maximum accuracy achieved is by using KNN classifier and is 96% at size 64×64.

3.3 Experiments on MMI and CK+ database (cross database)

Another milestone set in this manuscript was to check the effectiveness of the methodology by using cross dataset. For this purpose, we have combined both datasets and have performed experiments using the proposed framework. Results produced in this way are promising and have boosted the confidence with enhanced accuracy relative to the recently developed techniques in this domain. In this combined dataset, we have used six basic expressions and total number of images used was 777. Results of the dataset used are shown in the Table 6.

The dataset MMI and CK+ have been developed using different environments and with subjects belonging to different regions, gender, race and ethnicity. Using combined dataset results, presented in the Table 6 are quite impressive. It depicts the generalizability of the proposed technique. With the increase in number of features accuracy increases and reaches to its climax 98.2% by using only 208 features which is a significant contribution. The main point in these results is that 96.1% accuracy has been achieved by using only 32 features.

Table 6 Cross dataset images accuracy rate with varying feature vector size

Feature vector (FV)	Accuracy rate (%)
32	96.1
48	96.9
64	97.4
80	97.8
96	97.7
112	97.9
128	97.7
144	97.9
160	98.1
176	98.2
192	98.1
208	98.2

Table 7 Comparison with other methods on CK+ database

References	Technique	Accuracy (%)
2017 [8]	LBP+HOG	96.6
2016 [7]	HOG+PCA+LDA	99.2
2015 [9]	PHOG+LBP	94.6
2013 [10]	HOG+WLD	95.86
2012 [11]	HOG+SVM	88.7
2014 [1,2]	HOG+SVM	80
2015 [12]	PHOG_TOP	83.7
Proposed	HOG+DCT	99.6

3.4 Comparisons with other state-of-the-art techniques

We chose to compare the average recognition accuracy rate with other state-of-the-art techniques using the same database (i.e. Cohn-Kanade database). It can be observed that the recognition accuracy of the proposed framework is better than other techniques presented in the literature as shown in Table 7.

4 Conclusion

Facial expression classification is a vigorous research topic with amplified number of applications. Highly reliable and accurate solution is required to develop such a system. In this paper a hybrid feature extraction technique is presented which is accurate, reliable and equally adept in handling multi-scale and illumination variation problem. Variety of results has been presented to prove the effectiveness of the methodology. System employs only face components extracted using Viola and Jones method. It utilizes the power of Histogram of Oriented Gradients descriptor and is further refined by using high variance features extracted using Discrete cosine transform. System is tested using MMI and CK+ datasets and achieved high accuracy of 99.6% by using only 32 features with KNN as a classifier. Results using other classifiers are also presented. Experiments have also been performed using combined dataset and results are even better on the combined dataset. Comparison with the state-of-the art techniques has also been presented and robustness of the technique is evident from the results and discussion.

References

1. Donia, M.M.F., Youssif, A.A.A., Hashad, A.: Spontaneous facial expression recognition based on histogram of oriented gradients descriptor. Comput. Inf. Sci. **7**, 31–37 (2014)
2. Hamburger, C.: Quasimonotonicity, regularity and duality for nonlinear systems of partial differential equations. Ann. Mat. Pura Appl. **169**, 321–354 (1995)
3. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (2005)
4. Ojala, T., Pietikainen, M., Harwood, D.: A comparative study of texture measures with classification based on featured distributions. J. Pattern Recognit. **29**, 51–59 (1996)
5. Platt, J.C. : Fast training of support vector machines using sequential minimal optimization. In: Schölkopf, B., Burges, C.J.C., Smola, A.J. (eds.) Advances in Kernel Methods. MIT Press, Cambridge, pp. 185–208 (1999)
6. Quinlan, J.R.: C4.5: Programs for Machine Learning. Morgan Kaufmann (1993)
7. Kar, N.B., Babu, K.S., Jena, S.K. : Face expression recognition using histograms of oriented gradients with reduced features. In: Proceedings of International Conference on Computer Vision and Image Processing, pp. 209–219 (2016)
8. Liu, Y., Li, Y., Ma, X., Song, R.: Facial Expression Recognition with Fusion Features Extracted from Salient Facial Areas. Preprints (2017), 2017010102. doi:10.20944/preprints201701.0102.v1
9. Happy, S.L., Routray, A.: Robust facial expression classification using shape and appearance features. In: Proceedings of 8th International Conference of Advances in Pattern Recognition (2015)
10. Wang, X., Jin, C., Liu, W., Hu, M., Xu, L., Ren, F.: Feature fusion of HOG and WLD for facial expression recognition. In: Proceedings of the IEEE/SICE International Symposium on System Integration, Kobe International Conference Center, Kobe, Japan, pp. 227–232 (2013)
11. Chen, J., Chen, Z., Chi, Z., Fu, H.: Facial expression recognition based on facial components detection and HOG features. In: Scientific Cooperations International Workshops on Electrical and Computer Engineering Subfields, pp. 64–69 (2014)
12. Fan, X., Tjahjadi. T.: A spatial-temporal framework based on histogram of gradients and optical flow for facial expression recognition in video sequences. J. Pattern Recognit. **48**(11), 3407–3416 (2015)
13. Khan, R.A., Meyer, A., Konik, H., Bouakaz, S.: Framework for reliable, real-time facial expression recognition for low resolution images. J. Pattern Recognition Letters **34**(10), 1159–1168 (2013)
14. Khan, S.A., Hussain, A., Usman, M.: Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features. Multimed. Tools Appl. (2017). doi:10.1007/s11042-016-4324-z
15. Zhong, L., Jinsha, Y., Hong, Y., Ke, Z. : Wireless communications, networking and mobile computing. In: WiCOM 2008 4th International Conference, pp. 1–4 (2008)
16. Viola, P. Jones, M. : Rapid Object Detection using a Boosted Cascade of Simple Features. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'01). 1, 511,(2001)
17. Dollar, P., Belongie, S., Perona, P.: The fastest pedestrian detector in the west. In: Proceedings of the British Machine Vision Conference, pp. 68.1–68.11 (2010)
18. Lucey, P., Cohn, J.F., Kanade, T., Saragih, T., Ambadar, Z., Matthews, I.: The Extended Cohn-Kanade Dataset (CK+) A complete dataset for action unit and emotion-specified expression. In: IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 94–101 (2010)
19. Pantic, M., Valstar, M., Radermaker, R., Maat, L.: Web-based database for facial expression analysis. In: Proceedings of the 13th ACM International Conference on Multimedia, pp. 317–321 (2005)

Muhammad Nazir received his M.Sc. degree in computer science from PMAS-University of Arid Agriculture Rawalpindi, Pakistan and his M.S. degree in computer science from National University of Computer and Emerging Sciences, NU-FAST, Islamabad, Pakistan. Currently he is pursuing his PhD from the Islamia College University, Peshawar, Pakistan. He is serving as Assistant Professor at HITEC University, Taxila, Pakistan. His research interests include Digital Image Processing and Computational Intelligence.

Zahoor Jan is currently holding the rank of an associate professor in computer science at Islamia College Peshawar, Pakistan. He received his MS and PhD degree from FAST University Islamabad in 2007 and 2011 respectively. He is also chairman of Department of Computer Science at Islamia College Peshawar, Pakistan. His areas of interests include image processing, machine learning, computer vision, artificial intelligence and medical image processing, biologically inspired ideas like genetic algorithms and artificial neural networks, and their softcomputing applications, biometrics, solving image/video restoration problems using combination of classifiers using genetic programming, optimization of shaping functions in digital watermarking and image fusion.

Muhammad Sajjad received his Master degree from Department of Computer Science, College of Signals, National University of Sciences and Technology, Rawalpindi, Pakistan. He received his PhD degree in Digital Contents from Sejong University, Seoul, Republic of Korea. He is now working as an assistant professor at Department of Computer Science, Islamia College Peshawar, Pakistan. He is also head of "Digital Image Processing Laboratory (DIP Lab)" at Islamia College Peshawar, Pakistan., where students are working on research projects such social data analysis, image super-resolution and reconstruction, medical image analysis, multimodal data mining and summarization, prioritization, image/video quality assessment, and image/video retrieval. His primary research interests include computer vision, image understanding, pattern recognition, and robot vision and multimedia applications, with current emphasis on raspberry-pi and deep learning-based bioinformatics, video scene understanding, activity analysis, and real-time tracking.