Hindawi
Mathematical Problems in Engineering
Volume 2019, Article ID 9185481, 12 pages
https://doi.org/10.1155/2019/9185481

Research Article

Recognition of Facial Expressions under Varying Conditions Using Dual-Feature Fusion

Awais Mahmood,[1] Shariq Hussain,[2] Khalid Iqbal,[3] and Wail S. Elkilani[1]

[1]*College of Applied Computer Science, King Saud University, Al Muzahimiyah Campus, Riyadh, Saudi Arabia*
[2]*Department of Software Engineering, Foundation University Islamabad, Islamabad, Pakistan*
[3]*Department of Computer Science, COMSATS University Islamabad, Attock Campus, Attock, Pakistan*

Correspondence should be addressed to Shariq Hussain; shariq@fui.edu.pk

Received 23 July 2019; Accepted 4 August 2019; Published 21 August 2019

Guest Editor: Marco Perez-Cisneros

Facial expression recognition plays an important role in communicating the emotions and intentions of human beings. Facial expression recognition in uncontrolled environment is more difficult as compared to that in controlled environment due to change in occlusion, illumination, and noise. In this paper, we present a new framework for effective facial expression recognition from real-time facial images. Unlike other methods which spend much time by dividing the image into blocks or whole face image, our method extracts the discriminative feature from salient face regions and then combine with texture and orientation features for better representation. Furthermore, we reduce the data dimension by selecting the highly discriminative features. The proposed framework is capable of providing high recognition accuracy rate even in the presence of occlusions, illumination, and noise. To show the robustness of the proposed framework, we used three publicly available challenging datasets. The experimental results show that the performance of the proposed framework is better than existing techniques, which indicate the considerable potential of combining geometric features with appearance-based features.

1. Introduction

Facial expression recognition (FER) has emerged as an important research area over the last two decades. Facial expression is one of the immediate, natural, and powerful means for humans to communicate their intentions and emotions. The FER system can be used in many important applications such as driver safety, health care, video conferencing, virtual reality, and cognitive science etc.

Generally, facial expression can be classified into neutral, anger, disgust, fear, surprise, sad, and happy. Recent research shows that the ability of young people to read the feeling and emotion of other people is getting reduced due to the extensive use of digital devices [1]. Therefore, it is important to develop a FER system which accurately recognizes facial expression in real time.

An automatic FER system commonly consists of four steps: Preprocessing, feature extraction, feature selection, and classification of facial expressions. In the preprocessing step, face region is first detected and then extracted from the input image because it is the area that contains expression-related information. The most well-known and common algorithm used for face detection is the Viola–Jones object detection algorithm [2]. Subsequently, in the feature extraction step, distinguishable features are extracted from the face image. The two popular approaches for feature extraction are geometric-based feature extraction and appearance-based feature extraction. In the geometric-based techniques, the facial landmark points are first detected and then combined into a feature vector, which encodes geometric information of face from the position, distance, and angle [3]. The appearance-based techniques characterize the appearance information brought by different facial movements. Next, a subset of relevant features is selected in the feature selection step which contains more discriminatory power to classify different classes. In the last classification

step, classifiers like K-nearest neighbor (KNN) [4] and support vector machine (SVM) [5] are first trained and then used to classify the input data.

Although a lot of work has been done to develop a robust FER system, we find that several common problems still exist in the real-time environment which hinder the development of the FER system: (i) The extracted features are sensitive to the change in illumination, occlusion, and noise. That means a slight change in illumination, occlusion, and noise may influence the recognition accuracy rate. (ii) The large data dimension is another problem which deteriorates the performance of such systems.

The contributions of the proposed work are as follows:

(i) A dual-feature fusion technique is proposed in this work for effective and efficient classification of facial expressions in the unconstrained environment.

(ii) The proposed framework is based on local and global features, which make the proposed framework robust to change in occlusions, illumination, and noise.

(iii) Feature selection process is used to obtain the discriminative features, where the redundant features are discarded. The reduction in feature vector length also reduces the time complexity which makes the proposed framework suitable for real-time applications.

The rest of the paper is organized as follows: Section 2 presents the related work. Section 3 provides the description of the materials and methods. Experimental results are presented in Section 4. Finally, conclusion is provided in Section 5.

2. Related Work

Numerous methods for facial expression recognition have been developed due to its increased importance. These methods are mainly categorized into geometric-based and appearance-based methods based on feature extractions.

In geometric-based methods, information such as shape of the face and its components are used for feature extraction. The first important and challenging step in the geometric-based method is to initialize a set of facial points as the facial expression evolves over time. The study presented in [6] employed the elastic bunch graph matching (EBGM) algorithm for initialization of facial points. The discriminative features are also selected from triangle and line features with the multiclass AdaBoost algorithm. Sun et al. [7] proposed an effective method for the selection of optimized active face regions. They used convolution neural network (CNN) to extract features from optimized active face regions. The method used by Hsieh et al. [8] was based on the active shape model (ASM). They employed ASM to extract different facial expression regions. Similarly, Zangeneh and Moradi [9] first used the active appearance model (AAM) to reveal the important facial points, and then differential geometric features are extracted from those facial points. In the geometric-based features extraction techniques, it is difficult to track and initialize facial feature points in real time. If the error occurs during facial point initialization process, then this error deteriorates the overall feature extraction process.

On the contrary, appearance-based features extraction methods encode the face appearance variations without taking muscle motion into account. Chen et al. [10] introduced the multithreading cascade of Speeded Up Robust Features (McSURF), which improve the recognition accuracy rate. Cruz et al. [11] explore the temporal derivative and adjacent frames by using new framework known as temporal patterns of oriented edge magnitudes. The cases of out-of-plane head rotations are handled using rotation-reversal invariant HOD, presented by Chen et al. [12]. They also developed the cascade learning model to boost the classification process. Alphonse and Dharma [13] employed the maximum response-based directional texture pattern and number pattern for feature extraction. The performance is tested in the constrained and unconstrained environments. Recently, the work proposed in [14] employed spatiotemporal convolution to jointly extract the temporal dynamic and multilevel appearance feature of facial expressions. Another promising method to enhance the performance of random forest is proposed in [15]. They reduce the influence of various distortions like occlusion and illumination by extracting the robust features from salient facial patches. Sajjad et al. [16] presented a model integrating the histogram-oriented gradient with the uniform-local ternary operator for the extraction of facial features. The performance of the proposed method was tested on facial expression images which contains noise and partial occlusions. In another interesting approach, the authors proposed a new framework named local binary image cosine transform for computationally efficient feature extraction/selection [17]. Munir et al. [18] proposed a merged binary pattern code (MBPC) to represent the face texture information. They performed experiments on real-time images. In order to normalize the illumination effects, they preprocessed the images using the fast Fourier transform and contrast limited adaptive histogram equalization. Liu et al. [19] made use of deep network to learn the midlevel representation of face. They tested the effectiveness of their proposed method both on wild environment images and lab-controlled data.

Apart from the appearance-based or geometric-based feature extraction, fusion of this two-feature extraction method is also a promising trend. Zhang et al. [20] combined both texture and geometric-based features to maintain reasonable amount of tolerance against noise and occlusion. They used an active shape model and SIFT for geometric and appearance-based feature, respectively. To inherit the advantages of geometric and appearance information, Yang et al. [21] fused deep geometric features and LBP-based appearance features. They also proposed an improved random forest classifier for effective and efficient recognition of facial expressions. In the method of Tsai and Chang [22], features are extracted via Gabor filter, discrete cosine transform, and angular radial transform. In the work of Ghimire et al. [23], first, the face local specific regions were

selected, and then central moments were normalized. A local binary pattern descriptor is used for the extraction of geometric and appearance-based features, respectively.

In this paper, different from other methods, we select the facial informative local regions instead of dividing the face image into nonoverlapping blocks. Such representations can improve the classification performance compared with the block-based image representation. The appearance-based feature is computed from local face regions and also from the whole face area. These features are then fused which provide more robust features.

3. Materials and Methods

The working of the proposed framework based on dual-feature fusion is illustrated in Figure 1. Initially, the face portion is detected and extracted from input images using the Viola–Jones algorithm [2]. For dual-feature fusion, we first detect the facial landmark point on the face image and then the important local regions are located. The Weber local descriptor (WLD) excitation and orientation image is also generated from the input images. In next step, DCT is used to select the high variance features from local regions along with excitation and orientation image of WLD. In order to improve the performance, both types of features are then fussed using the score-level fusion.

3.1. Face Detection and Landmark Position Estimation. In order to extract the region of interest (i.e. face portion), we utilized the Viola–Jones algorithm [2] in our study which is mostly cited in literature and also considered as a fast and accurate object detection algorithm [24].

The spatial misalignment usually occurs due to the expression and pose variations in the face image. Division of the face image into nonoverlapped blocks or exploiting holistic features cannot resolve this issue [25]. Admittedly, the intraclass difference is increased due to variation in face appearance because of expressions and facial poses. In that case, the local features are more robust to these changes as compared to holistic features. There are some reliable and stable regions which preserve more useful information to deal with these changes. That is why, in this study, we extract the features from inner facial landmarks rather than extracting the features from whole face image.

For this purpose, we used the method presented by Kazemi and Sullivan [26] in which the face landmark position is estimated from subset of pixel intensities using ensemble of regression trees. This method is highly effective to locate the landmark position not only in the face with neutral expression but also in the face with variation in different expressions.

After landmark position estimation, we use the facial point location to divide the face image into 29 local regions. The local feature is extracted from all these local regions. In order to reduce the data dimensions, we do not require exhaustive search technique as performed in [23] to search for a subset of local regions among 29 local regions because our feature selection method is more efficient and effective.

3.2. Construction of WLD Excitation and Orientation Image. The Weber local descriptor is proposed by Chen et al. [27] which is inspired from Weber's law. WLD consist of two main components, namely, differential excitation and gradient orientation. The differential excitation component represents the intensity differences of the neighbor pixel and the center pixel where the gradient orientation of the center pixel is described by the gradient orientation component. Both the components provide the local texture description of an image.

Formally, the differential excitation component can be defined as

$$\xi_{\mathrm{m}}(x_{\mathrm{c}}) = \arctan\left(\alpha \sum_{i=0}^{p-1} \frac{x_i - x_{\mathrm{c}}}{x_{\mathrm{c}}}\right), \tag{1}$$

where the arctangent is used to suppress the noise side effect and also to avoid the output of being too large. The neighbor pixels are denoted as $x_i\,(i = 0, 1, 2, 3, \ldots, p-1)$, while x_{c} represents the center pixel. Similarly, the differential orientation component of an image can be defined as follows:

$$\xi_{\mathrm{o}}(x_{\mathrm{c}}) = \arctan\left(\frac{x_1 - x_5}{x_3 - x_7}\right), \tag{2}$$

where the intensity difference is indicated by $x_3 - x_7$ and $x_1 - x_5$ in the x and y directions.

Figures 2 and 3 illustrate the WLD excitation and orientation component images.

3.3. DCT-Based Feature Selection and Fusion. We can compute the DCT of an input scanned image $d_{x,y}$ of size $M \times N$, by using the expression as defined in equation (3) [28]. For all values of $u = 0, 1, 2, \ldots, M-1$ and $v = 0, 1, 2, \ldots, N-1$, the expression of equation (1) must be evaluated. Also, given $D_{u,v}$, for $x = 0, 1, 2, \ldots, M-1$ and $y = 0, 1, 2, \ldots, N-1$, $d_{x,y}$ can be obtained by using the inverse DCT transform which is mentioned in equation (4). Note that both equations (3) and (4) consist of a two-dimensional pair of DCT, where x and y are spatial coordinates and u and v refers to frequency variables:

$$D_{u,v} = \rho(u)\rho(v) \sum_{x=0}^{M-1} \sum_{y=0}^{N-1} d_{x,y} \cos\left[\frac{(2x+1)u\pi}{2M}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right], \tag{3}$$

$$d_{u,v} = \sum_{u=0}^{M-1} \sum_{v=0}^{N-1} \rho(u)\rho(v)\, D_{u,v} \cos\left[\frac{(2x+1)u\pi}{2M}\right] \cos\left[\frac{(2y+1)v\pi}{2N}\right], \tag{4}$$

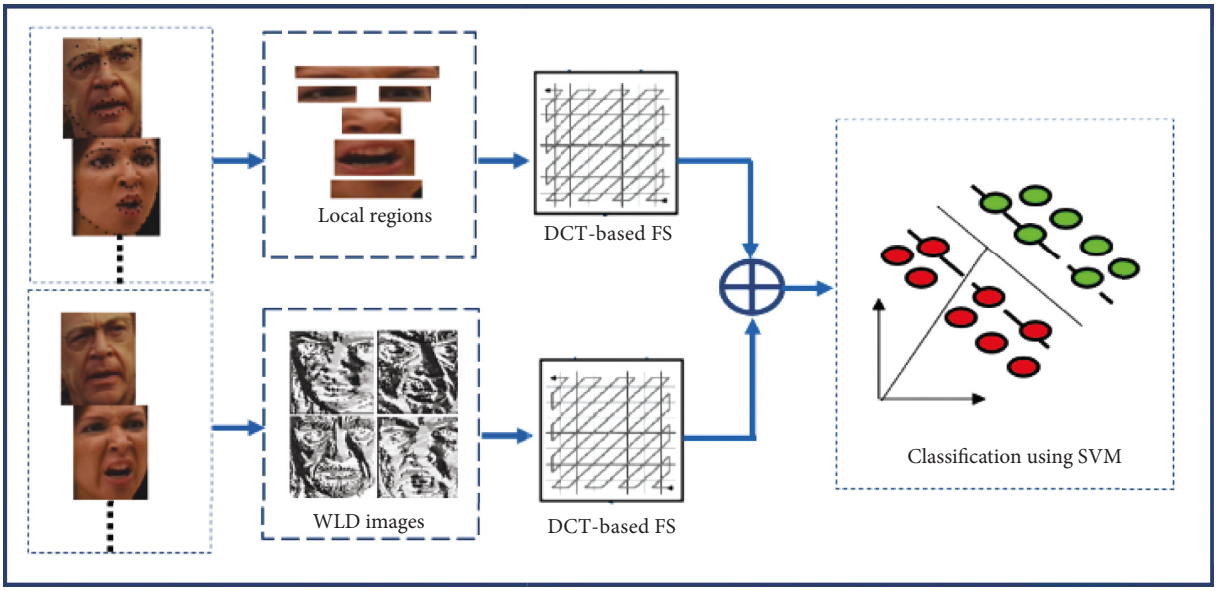

Figure 1: Proposed framework flow diagram.

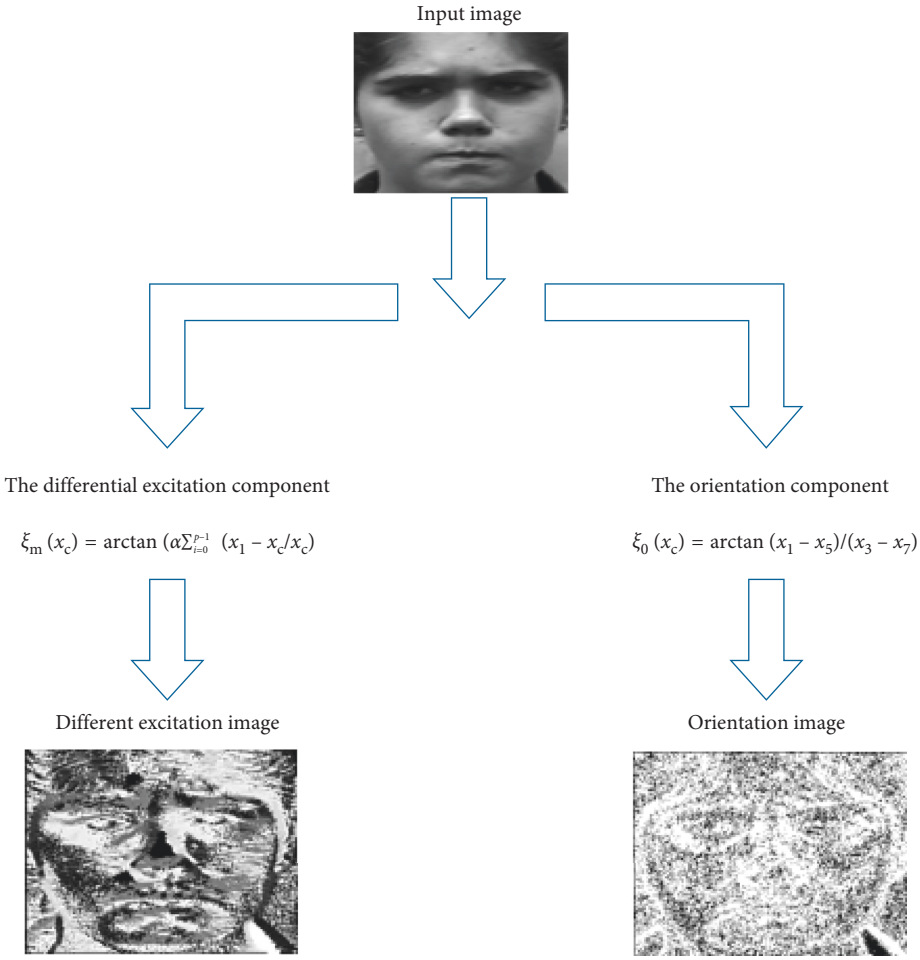

Figure 2: WLD excitation and orientation component.

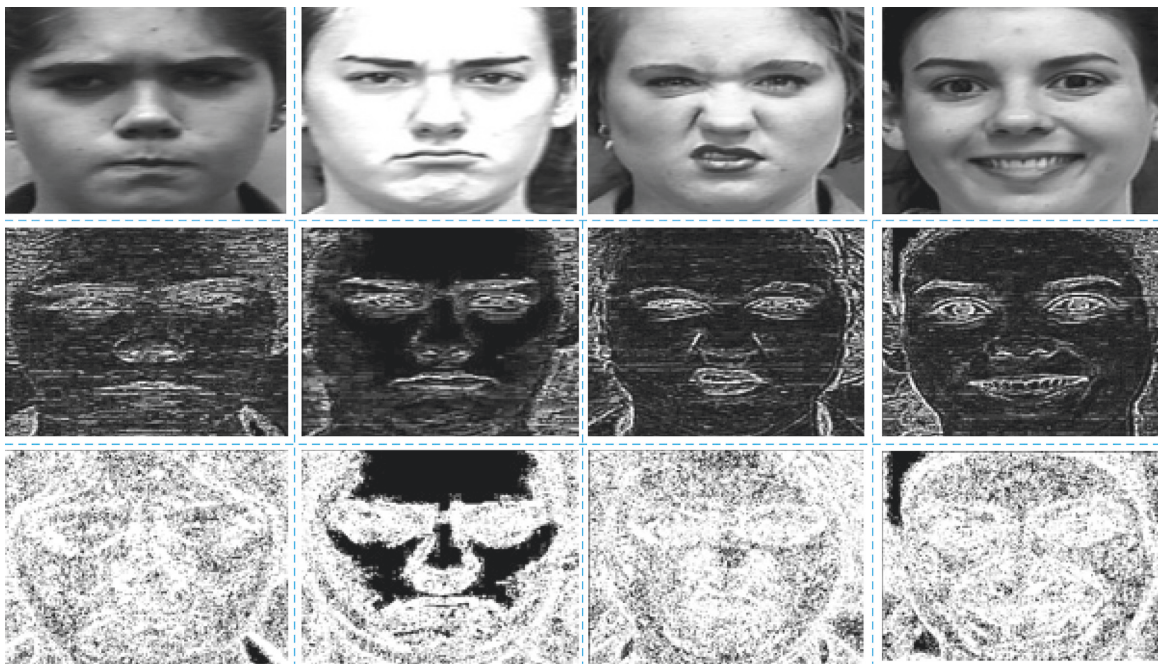

Figure 3: First row represents the original image, second row is the sample images of excitation component, and third row depicts orientations of component images.

$$\rho(u) = \begin{cases} \sqrt{\frac{1}{M}}, & u = 0, \\ \sqrt{\frac{2}{M}}, & u = 1, 2, 3, \ldots, M-1, \end{cases}$$
$$\rho(v) = \begin{cases} \sqrt{\frac{1}{N}}, & v = 0, \\ \sqrt{\frac{2}{N}}, & v = 1, 2, 3, \ldots, N-1, \end{cases} \quad (5)$$
$$\begin{cases} u = 0, 1, 2, \ldots, M-1, \\ v = 0, 1, 2, \ldots, N-1, \\ x = 0, 1, 2, \ldots, M-1, \\ y = 0, 1, 2, \ldots, N-1, \end{cases}$$

$P(u, v)$ is the power spectrum of image $d_{x,y}$ and can be defined as

$$P(u, v) = \left|D_{u,v}\right|^2. \quad (6)$$

After selection of appearance-based and geometric-based features, we employed a score-level fusion strategy to combine these features. Feature-level fusion and score-level fusion are the two fusion strategies which are used widely in the literature. In the feature-level fusion, different feature vectors are simply concatenated after normalization process. In contrast to the feature-level fusion, a distance-based classifier is used in the score-level fusion to compute the distance between the feature vector of training and testing samples. The feature-level fusion mainly produces large data dimension [29] that is why we prefer score-level fusion in this study. In the score-level fusion, the extracted appearance- and geometric-based DCT features are stored in $\mathrm{FS}^{\mathrm{ap}}$ and $\mathrm{FS}^{\mathrm{geo}}$, respectively. These features are computed for all training $\mathrm{FS}_{\mathrm{tr}}$ and testing $\mathrm{FS}_{\mathrm{te}}$ samples. Afterward, score vectors, namely, S^{ap} and S^{geo}, are produced by computing the distance between training samples and all the testing samples of appearance and geometric feature sets. In order to perform normalization, the min-max method of normalization [30] is used which is described as

$$S_i' = \frac{S_i - \mathrm{Min}(S)}{\mathrm{Max}(S) - \mathrm{Min}(S)}, \quad (7)$$

where the original score ith entry is represented by S_i. The minimum and the maximum values of the score is denoted by $\mathrm{Min}(S)$ and $\mathrm{Max}(S)$. Finally, the product rule or the sum rule method is used to normalize the score vectors [30].

The procedure of feature extraction and fusion is presented in Algorithm 1.

3.4. Support Vector Machine (SVM) for Expression Classification. For multi and binary classification problem, the SVM [31] acts as a more powerful tool. The SVM draws a hyperplane between the two classes by maximizing the margin between the closest points of the class and hyperplane. The decision function for class labels $y_i = \mp 1$ and training data $x_i\,(i = 1, 2, 3, \ldots, N)$ can be formulated as [23]

$$f(x) = \mathrm{sign}\left(w^T x + b\right), \quad (8)$$

where the hyperplane separation is denoted by $w^T x + b = 0$. In order to handle the multiclass problem, we have used SVM with radial basic function kernel, implemented as libsvm [32] and is publicly available for use.

4. Experimental Results and Discussions

To evaluate the performance of the proposed framework, we used 3 publicly available benchmarking databases, namely, MMI database, extended Cohn-Kanade (CK+) and static face in the wild (SFEW).

(i) MMI database: this image database [33] contains both video sequences and static images which include head movements and posed expressions. It consists of images of high resolutions of 88 subjects and over 2900 videos of male and female. For our experiment, we have selected different video sequences and extracted a total of 273 images from these sequences.

(ii) Extended Cohn-Kanade (CK+): this database contains 593 video sequence of 123 subjects [34]. The subjects are origins from Latino, Asian, and African-American and aged from 18 to 30 years. We have selected different video sequences and obtained 540 static images of six basic expressions.

(iii) Static face in the wild (SFEW): the SFEW [35] contains real-time movie images which are captured in unconstrained settings. The images are having different variations like noise, pose variation, and high illumination changes. We have taken 291 images from the available 1394 images in the database.

Sample images of each database is shown in Figure 4, and Table 1 illustrates the number of images taken from MMI, CK+, and SFEW database.

To make maximum use of the available data, we employed 5-fold and 10-fold cross validation for all the experiments. To get the better picture of the facial expression recognition accuracy, average accuracy rate and confusion matrices are given across all the three datasets.

4.1. Experiment on MMI, CK+, and SFEW Database. This section shows the results obtained using MMI, CK+, and SFEW datasets. MMI dataset contained most of the spontaneous expressions. The proposed framework achieved an average recognition accuracy of 96% and 98.62%, respectively, for MMI and CK+ database. The confusion matrix of classifying 7 facial expressions for MMI dataset and 6 basic expressions for CK+ is shown in Tables 2 and 3, respectively.

In Table 2, among the seven facial expressions, neutral and sad expressions are the easiest with an average recognition accuracy rate of 100%, which is followed by happy and surprised. In contrast, angry and fear are the most difficult expressions for classification. As shown in the table, the fear expression is mostly confused with neutral and surprised, which is expected because of the structural similarities [36]. Furthermore, the anger facial expression is mostly misclassified with disgust and neutral expressions. This is probably because of the wrinkles of the forehead in anger expression, which is also the characteristics of disgust expression.

The confusion matrix in Table 3 depicts that disgust, sad, and happy expressions are classified with 100% recognition accuracy rate which is followed by surprised and anger expressions. The recognition accuracy for fear expression is slightly deviated at 95%. The results indicate that the fear expression misclassified either as anger or disgust emotion. The reason is that the fear, disgust, and anger expressions demonstrated similar muscle activities [37]. Moreover, it is also observed that the average recognition accuracy rate of the CK+ dataset is slightly higher than the MMI dataset. This is because the CK+ dataset contains more expressive emotions.

The confusion matrix for SFEW results is shown in Table 4. The performance on the SFEW database is low as compared to MMI and CK+ databases. This is because the images of the SFEW database are captured in the uncontrolled environment (real-world images) and are more challenging to classify as compared to other datasets. The average recognition accuracy rate of 50.2% is obtained using the SFEW database. By inspecting the recognition accuracy rate of each expression, we observed that sad, fear, and happy expressions are more accurately recognized. However, the disgust expression obtained the smallest recognition accuracy of 31.7%.

Table 5 illustrates the comparative assessment of the proposed method with the existing state-of-the-art [6, 10–14] methods. In literature, the FER system presented in [11] has achieved the highest recognition accuracy rate of 93.66% which works on the nonoverlapping patches. But in their method, the length of their code is controlled by a new coding scheme which makes their process more complex for real-time FER systems. The results show that the performance of our proposed method is superior as compared to existing techniques in terms of average recognition accuracy. Furthermore, it is also notable that recognition accuracy rate per expression of our proposed method is also high as compared to other methods.

In Table 6, the results for CK+ database are compared with the state-of-the-art methods. The average recognition accuracy rate of our method is highly competitive with other methods. Although the performance of the method presented in [14] is 1.11% higher than our method, the use of 3D convolution neural network makes their method computationally more expensive.

Figure 5 illustrates the comparative assessment of the proposed method with other methods on the SFEW database. It is evident from the results that our proposed method achieved better results as compared to existing methods in the literature. The average recognition accuracy rate of our proposed method is 50.2%. For the same dataset present in the studies [13, 19, 20, 38–40], the average accuracy rates were 26.1%, 30.14%, 33.8%, 44.0%, 49.31%, and 48.3%, respectively. The results depict that our strategy of the dual-feature fusion is more appropriate for FER in the uncontrolled environment. The recognition accuracy rate is significantly degraded on SFEW as compared to the results on MMI and CK+ due to its challenging condition, e.g., change in illumination and large pose variations.

Figure 4: Sample images taken from MMI, CK+, and SFEW database.

Input: Training sample images I_{train} with size $M \times N$
Testing sample images I_{test}
Output: $\text{Fused}_{\text{feat}}$
Procedure
(1) **For** each I_{train} **do**
(2) Compute WLD images $I_{\text{tr}}^{\text{ap}}$ and local region images $I_{\text{tr}}^{\text{geo}}$
(3) **For** each $I_{\text{tr}}^{\text{ap}}$ and $I_{\text{tr}}^{\text{geo}}$ **do**
(4) Compute $\text{FS}_{\text{tr}}^{\text{ap}}$ and $\text{FS}_{\text{tr}}^{\text{geo}}$ using equations (3) and (4)
(5) $\text{FS}_{\text{tr}}^{\text{ap}} = \langle \text{FS}_1^{\text{ap}}, \text{FS}_2^{\text{ap}}, \ldots, \text{FS}_{\text{sap}}^{\text{ap}} \rangle$, $\text{sap} : \text{size}(\text{FS}^{\text{ap}})$
(6) $\text{FS}_{\text{tr}}^{\text{geo}} = \langle \text{FS}_1^{\text{geo}}, \text{FS}_2^{\text{geo}}, \ldots, \text{FS}_{\text{sap}}^{\text{geo}} \rangle$, $\text{sap} : \text{size}(\text{FS}^{\text{geo}})$
(7) **End For**
(8) **End For**
(9) **For** each I_{test} **do**
(10) Compute WLD images $I_{\text{te}}^{\text{ap}}$ and local region images $I_{\text{te}}^{\text{geo}}$
(11) **For** each $I_{\text{te}}^{\text{ap}}$ and $I_{\text{te}}^{\text{geo}}$ **do**
(12) Compute $\text{FS}_{\text{te}}^{\text{ap}}$ and $\text{FS}_{\text{te}}^{\text{geo}}$ using equations (3) and (4)
(13) $\text{FS}_{\text{te}}^{\text{ap}} = \langle \text{FS}_1^{\text{ap}}, \text{FS}_2^{\text{ap}}, \ldots, \text{FS}_{\text{sap}}^{\text{ap}} \rangle$, $\text{sap} : \text{size}(\text{FS}^{\text{ap}})$
(14) $\text{FS}_{\text{te}}^{\text{geo}} = \langle \text{FS}_1^{\text{geo}}, \text{FS}_2^{\text{geo}}, \ldots, \text{FS}_{\text{sap}}^{\text{geo}} \rangle$, $\text{sap} : \text{size}(\text{FS}^{\text{geo}})$
(15) **End For**
(16) **End For**
(17) **For** each I_{train} **do**
(18) $S^{\text{ap}} = \text{Compute\_Distance}(\text{FS}_{\text{tr}}^{\text{ap}}, \text{FS}_{\text{te}}^{\text{ap}})$
(19) $S^{\text{geo}} = \text{Compute\_Distance}(\text{FS}_{\text{tr}}^{\text{geo}}, \text{FS}_{\text{te}}^{\text{geo}})$
(20) **End For**
(21) **For** each I_{train} **do**
(22) Normalize S^{ap} and S^{geo} using equation (7)
(23) **End For**
(24) $\text{Fused}_{\text{feat}} = \text{Score\_Level\_Fusion}(S^{\text{ap}}, S^{\text{geo}})$

Algorithm 1: The procedure of feature extraction and fusion.

4.2. Robustness against Noise and Occlusions. In the uncontrolled environment, noise, and occlusions are the main factors to degrade the image quality and reduce the facial expression recognition accuracy rate. It is required for any FER system to perform well in the presence of noise and partial occlusions. In this section, we examine the robustness of our proposed method in the presence of noise and partial occlusions.

To check the robustness against noise, we randomly added salt and pepper noise of different levels to the images of MMI and CK+ database. This type of noise is composed of two components.

The first component is the salt noise which occurs as a bright spot in the image, and the second component is the pepper noise which appears as a dark spot. As shown in Figure 6, the noise density was increased up to 0.05 level

Table 1: Number of selected images per expression from MMI, CK+, and SFEW database.

Dataset	Expression							
	Neutral	Fear	Disgust	Angry	Surprised	Sad	Happy	Total
MMI	36	41	39	45	39	34	39	273
CK+	N/A	90	90	90	90	90	90	540
SFEW	N/A	50	41	50	50	50	50	291

Table 2: Confusion matrix of recognition accuracy for MMI database.

	Neutral (%)	Fear (%)	Disgust (%)	Angry (%)	Surprised (%)	Sad (%)	Happy (%)
Neutral	**100**	0	0	0	0	0	0
Fear	4.88	**92.7**	0	0	2.44	0	0
Disgust	2.56	0	**94.9**	2.56	0	0	0
Angry	4.44	0	4.44	**91.1**	0	0	0
Surprised	0	2.56	0	0	**97.4**	0	0
Sad	0	0	0	0	0	**100**	0
Happy	0	2.56	0	0	0	0	**97.4**

Table 3: Confusion matrix of recognition accuracy for CK+ database.

	Fear (%)	Disgust (%)	Angry (%)	Surprised (%)	Sad (%)	Happy (%)
Fear	**95.0**	2.8	2.2	0	0	0
Disgust	0	**100.0**	0	0	0	0
Angry	0	0	**97.8**	0	2.22	0
Surprised	0	0	0	**98.9**	1.11	0
Sad	0	0	0	0	**100.0**	0
Happy	0	0	0	0		**100.0**

Table 4: Confusion matrix of the recognition accuracy for the SFEW database.

	Fear (%)	Disgust (%)	Angry (%)	Surprised (%)	Sad (%)	Happy (%)
Fear	**64.0**	0.0	6.0	14.0	8.0	8.0
Disgust	7.3	**31.7**	17.1	12.2	19.5	12.2
Angry	6.0	10.0	**42.0**	10.0	14.0	18.0
Surprised	22.0	0.0	16.0	**42.0**	12.0	8.0
Sad	8.0	8.0	8.0	2.0	**64.0**	10.0
Happy	10.0	4.0	14.0	8.0	10.0	**54.0**

Table 5: Confusion matrix of recognition accuracy for MMI.

Method	Fear (%)	Disgust (%)	Angry (%)	Surprised (%)	Sad (%)	Happy (%)	Mean (%)
Chen et al. [10]	68.40	65.30	69.50	82.60	68.20	83.90	73.00
Cruz et al. [11]	91.36	92.27	88.44	97.63	93.53	98.75	93.66
Ghimire et al. [6]	70.00	80.00	70.00	90.00	73.33	92.50	79.305
Chen et al. [12]	76.50	60.40	70.20	84.20	62.10	81.20	72.40
Alphonse and Dharma [13]	81.30	81.30	82.00	90.00	76.70	83.33	82.44
Yu et al. [14]	81.24	88.21	83.24	85.29	85.77	93.22	86.16
Proposed method	92.70	94.90	91.10	97.40	100.00	97.40	95.58

because in the real-time system, the average noise of this level is normally observed [16].

The results illustrated in Figure 7 shows that the recognition accuracy rate of our proposed method does not significantly reduce with increase in variance of salt and pepper noise. We have also observed that the recognition accuracy rate of the CK+ database is more stable in the presence of noise as compare to the MMI database. This is because the expression of CK+ is more representative.

In order to assess the proposed method performance in the presence of occlusions, we have added a block of random size to the test images. The range of block size starting from $[15 \times 15]$ to $[55 \times 55]$ randomly placed to the face images are shown in Figure 8.

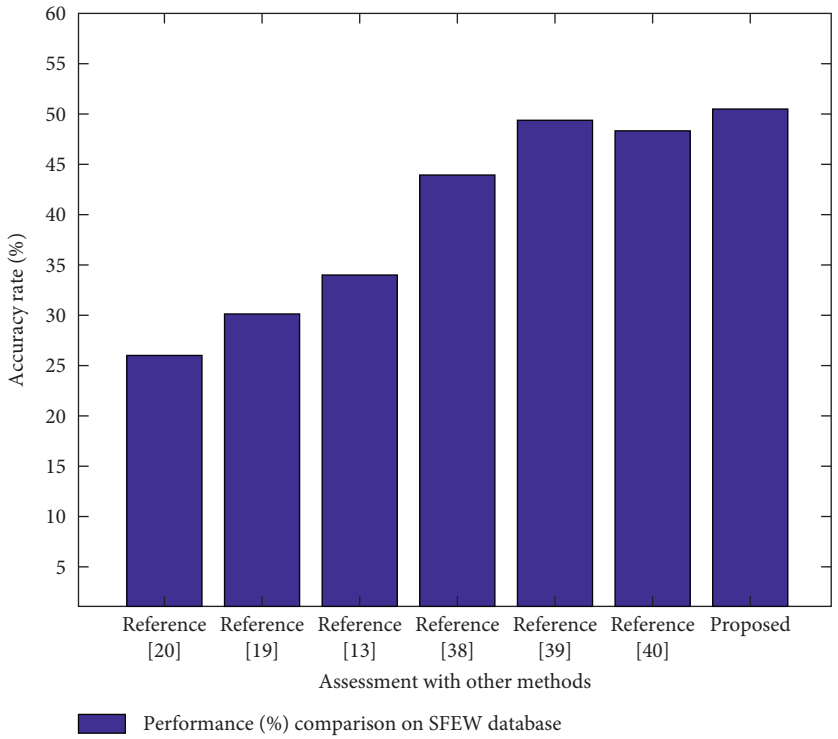

Figure 5: Comparison between existing method and proposed approach based on recognition accuracy.

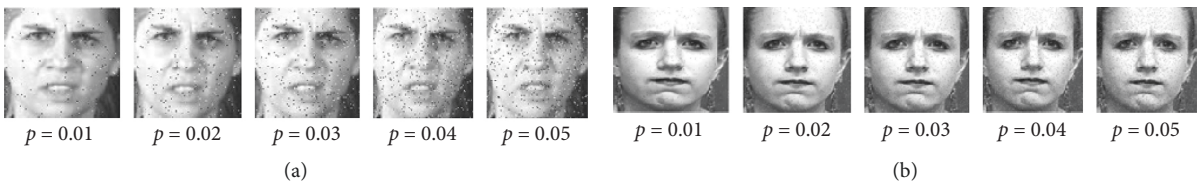

Figure 6: Sample images of salt and pepper noise from (a) MMI and (b) CK+ where p represents the noise density.

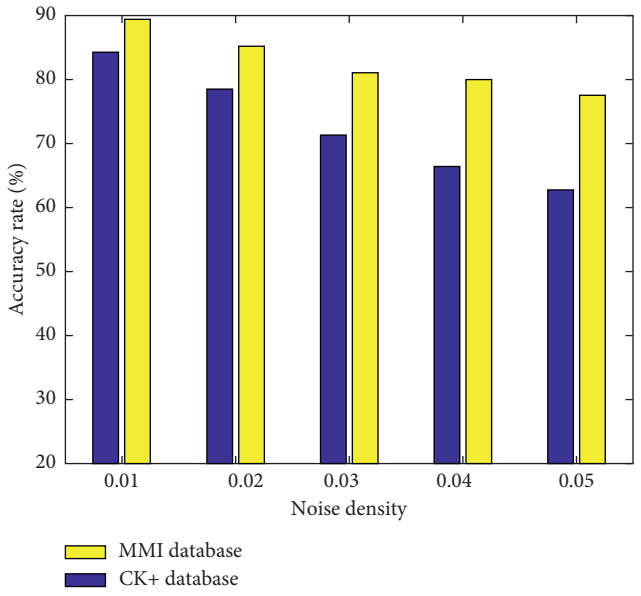

Figure 7: Recognition accuracy of MMI and CK+ databases in the presence of noise.

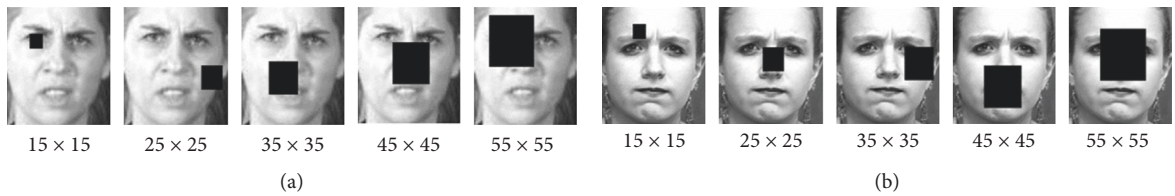

Figure 8: Sample images of occlusion from (a) MMI and (b) CK+ databases with varying block size.

Table 6: Confusion matrix of recognition accuracy for CK+.

Method	Fear (%)	Disgust (%)	Angry (%)	Surprised (%)	Sad (%)	Happy (%)	Mean (%)
Chen et al. [10]	92.50	86.20	96.10	96.40	94.10	98.20	91.20
Cruz et al. [11]	89.33	91.58	93.52	94.75	87.00	100.00	92.69
Ghimire et al. [6]	96.00	96.67	97.50	100.00	93.33	100.00	97.80
Chen et al. [12]	91.70	94.30	95.60	97.50	89.40	95.90	93.80
Alphonse and Dharma [13]	99.23	97.36	92.77	99.55	98.69	98.69	97.715
Yu et al. [14]	99.71	99.68	100.00	100.00	99.14	99.89	99.73
Proposed method	95.00	100.00	97.80	98.90	100.00	100.00	98.62

Table 7: Assessment of MMI and CK+ results in the presence of occlusions.

Block size	MMI (%)	CK+ (%)
[15 × 15]	91.9	98.1
[25 × 25]	90.8	98.3
[35 × 35]	90.5	90.6
[45 × 45]	88.3	88.5
[55 × 55]	75.1	90.6

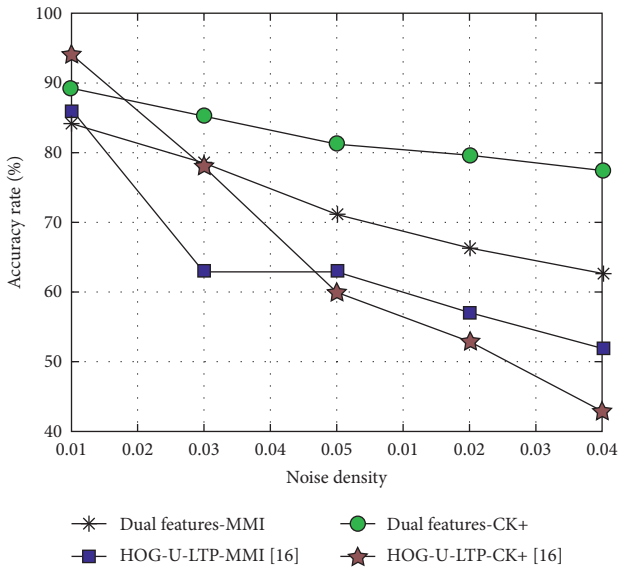

Figure 9: Comparison graph of the proposed method accuracy rate assessment with other methods in the presence of noise.

The average recognition accuracy rates for both MMI and CK+ are illustrated in Table 7. The results of MMI show that the accuracy rate decreased up to 3.6% when the block size increased from [15 × 15] to [45 × 45]. However, the recognition drops down by 17% when the block size [55 × 55] is used. This is because most of the important facial

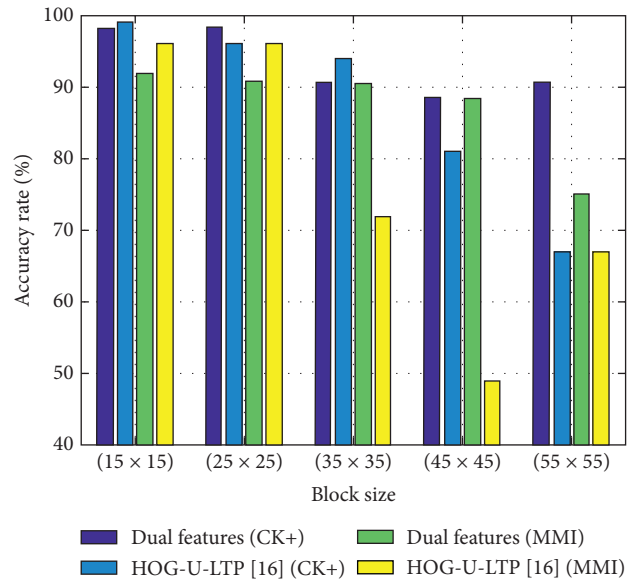

Figure 10: Competitive assessment with the existing method in the presence of occlusions.

points are hidden due to the large block size. In contrast, the recognition accuracy on the CK+ database only decreases by 7.5% when [55 × 55] block size was used in the experiments. It is foreseeable that the recognition accuracy reaches to zero in the presence of total occlusion.

To prove the robustness of our proposed method against noise and occlusions, we also compared the performance with the existing method [16] as shown in Figures 9 and 10. The methods presented in [16] are selected due to their state-of-the-art performance on MMI and CK+ database, and they also used a similar ratio of noise density and block size. From the results, we can easily conclude that our dual-feature fusion method is more robust to noise and occlusions as compared to the methods presented in [16] due to the less decline in recognition accuracy.

5. Conclusion and Future Work

Facial expression recognition in the real-world case is a long-standing problem. The low image quality, partial occlusions, and illumination variation in the real-word environment make the feature extraction process more challenging. In this paper, we exploit both texture and geometric features for effective facial expression recognition. The effective geometric features are introduced in this paper from facial landmark detection, which can capture the facial configure changes. Considering that the geometric feature extraction may fail under various conditions, the addition of texture feature with geometric features is useful for capturing the minor changes in expressions. WLD is utilized for the extraction of texture feature which is more effective to capture the facial subtle changes. Furthermore, we have employed score-level fusion for fusion of geometric and texture features which results in decreasing the number of features. The performance of the proposed approach is evaluated on standard databases like MMI, CK+, and SFEW, and the results are compared with the state-of-the-art approaches. The effectiveness of our proposed dual-feature fusion strategy is verified by different experimental results.

Although WLD works well on the face images for the extraction of salient features, the variation of local intensity cannot effectively be represented by using the standard WLD because it neglects different orientations of the neighborhood pixel. In future work, we are planning to address this issue along with the experimentation with ethnographic datasets.

Data Availability

The authors confirm that the data generated or analyzed and the information supporting the findings of this study are available within the article.

Conflicts of Interest

The authors declare no conflicts of interest.

Authors' Contributions

All the co-authors have made significant contribution in conceptualization, data analysis, experimentations, scientific discussions, preparation of original draft, and revision and organization of the paper.

Acknowledgments

This study was supported by the Deanship of Scientific Research, King Saud University, Riyadh, Saudi Arabia, through the Research Group under Project RG-1439-039.

References

[1] Y. T. Uhls, M. Michikyan, J. Morris et al., "Five days at outdoor education camp without screens improves preteen skills with nonverbal emotion cues," *Computers in Human Behavior*, vol. 39, pp. 387–392, 2014.

[2] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the 2001 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2001 (CVPR 2001)*, pp. 511–518, Kauai, HI, USA, December 2001.

[3] S. Jain, C. Hu, and J. K. Aggarwal, "Facial expression recognition with temporal modeling of shapes," in *Proceedings of the 2011 IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 1642–1649, Barcelona, Spain, November 2011.

[4] N. S. Altman, "An introduction to kernel and nearest-neighbor nonparametric regression," *The American Statistician*, vol. 46, no. 3, pp. 175–185, 1992.

[5] I. Kotsia and I. Pitas, "Facial expression recognition in image sequences using geometric deformation features and support vector machines," *IEEE Transactions on Image Processing*, vol. 16, no. 1, pp. 172–187, 2007.

[6] D. Ghimire, J. Lee, Z.-N. Li, and S. Jeong, "Recognition of facial expressions based on salient geometric features and support vector machines," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 7921–7946, 2017.

[7] A. Sun, Y. Li, Y.-M. Huang, Q. Li, and G. Lu, "Facial expression recognition using optimized active regions," *Human-Centric Computing and Information Sciences*, vol. 8, p. 33, 2018.

[8] C.-C. Hsieh, M.-H. Hsih, M.-K. Jiang, Y.-M. Cheng, and E.-H. Liang, "Effective semantic features for facial expressions recognition using SVM," *Multimedia Tools and Applications*, vol. 75, no. 11, pp. 6663–6682, 2016.

[9] E. Zangeneh and A. Moradi, "Facial expression recognition by using differential geometric features," *The Imaging Science Journal*, vol. 66, no. 8, pp. 463–470, 2018.

[10] J. Chen, Z. Luo, T. Takiguchi, and Y. Ariki, "Multithreading cascade of SURF for facial expression recognition," *EURASIP Journal on Image and Video Processing*, vol. 2016, no. 1, p. 37, 2016.

[11] E. A. S. Cruz, C. R. Jung, and C. H. E. Franco, "Facial expression recognition using temporal POEM features," *Pattern Recognition Letters*, vol. 114, pp. 13–21, 2018.

[12] J. Chen, T. Takiguchi, and Y. Ariki, "Rotation-reversal invariant HOG cascade for facial expression recognition," *Signal, Image and Video Processing*, vol. 11, no. 8, pp. 1485–1492, 2017.

[13] A. S. Alphonse and D. Dharma, "Novel directional patterns and a generalized supervised dimension reduction system (GSDRS) for facial emotion recognition," *Multimedia Tools and Applications*, vol. 77, no. 8, pp. 9455–9488, 2018.

[14] Z. Yu, G. Liu, Q. Liu, and J. Deng, "Spatio-temporal convolutional features with nested LSTM for facial expression recognition," *Neurocomputing*, vol. 317, pp. 50–57, 2018.

[15] Y. Liu, X. Yuan, X. Gong, Z. Xie, F. Fang, and Z. Luo, "Conditional convolution neural network enhanced random forest for facial expression recognition," *Pattern Recognition*, vol. 84, pp. 251–261, 2018.

[16] M. Sajjad, A. Shah, Z. Jan, S. I. Shah, S. W. Baik, and I. Mehmood, "Facial appearance and texture feature-based robust facial expression recognition framework for sentiment knowledge discovery," *Cluster Computing*, vol. 21, no. 1, pp. 549–567, 2018.

[17] S. A. Khan, A. Hussain, and M. Usman, "Reliable facial expression recognition for multi-scale images using weber local binary image based cosine transform features," *Multimedia Tools and Applications*, vol. 77, no. 1, pp. 1133–1165, 2018.

[18] A. Munir, A. Hussain, S. A. Khan, M. Nadeem, and S. Arshid, "Illumination invariant facial expression recognition using

selected merged binary patterns for real world images," *Optik*, vol. 158, pp. 1016–1025, 2018.

[19] M. Liu, S. Li, S. Shan, and X. Chen, "AU-inspired deep networks for facial expression feature learning," *Neurocomputing*, vol. 159, pp. 126–136, 2015.

[20] L. Zhang, D. Tjondronegoro, and V. Chandran, "Facial expression recognition experiments with data from television broadcasts and the World Wide Web," *Image and Vision Computing*, vol. 32, no. 2, pp. 107–119, 2014.

[21] B. Yang, J.-M. Cao, D.-P. Jiang, and J.-D. Lv, "Facial expression recognition based on dual-feature fusion and improved random forest classifier," *Multimedia Tools and Applications*, vol. 77, no. 16, pp. 20477–20499, 2018.

[22] H.-H. Tsai and Y.-C. Chang, "Facial expression recognition using a combination of multiple facial features and support vector machine," *Soft Computing*, vol. 22, no. 13, pp. 4389–4405, 2018.

[23] D. Ghimire, S. Jeong, J. Lee, and S. H. Park, "Facial expression recognition based on local region specific features and support vector machines," *Multimedia Tools and Applications*, vol. 76, no. 6, pp. 7803–7821, 2017.

[24] M. Kolsch and M. Turk, "Analysis of rotational robustness of hand detection with a viola-jones detector," in *Proceedings of the 17th International Conference on Pattern Recognition, 2004 (ICPR 2004)*, pp. 107–110, Cambridge, UK, August 2004.

[25] Z. Zhang, L. Wang, Q. Zhu, S.-K. Chen, and Y. Chen, "Pose-invariant face recognition using facial landmarks and weber local descriptor," *Knowledge-Based Systems*, vol. 84, pp. 78–88, 2015.

[26] V. Kazemi and J. Sullivan, "One millisecond face alignment with an ensemble of regression trees," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1867–1874, Columbus, OH, USA, June 2014.

[27] J. Chen, S. Shan, C. He et al., "WLD: a robust local image descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1705–1720, 2010.

[28] N. Ahmed, T. Natarajan, and K. R. Rao, "Discrete cosine transform," *IEEE Transactions on Computers*, vol. C-23, no. 1, pp. 90–93, 1974.

[29] Z. Golrizkhatami and A. Acan, "ECG classification using three-level fusion of different feature descriptors," *Expert Systems with Applications*, vol. 114, pp. 54–64, 2018.

[30] M. He, S.-J. Horng, P. Fan et al., "Performance evaluation of score level fusion in multimodal biometric systems," *Pattern Recognition*, vol. 43, no. 5, pp. 1789–1800, 2010.

[31] C. Cortes and V. Vapnik, "Support-vector networks," *Machine Learning*, vol. 20, no. 3, pp. 273–297, 1995.

[32] C.-C. Chang and C.-J. Lin, "Libsvm," *ACM Transactions on Intelligent Systems and Technology*, vol. 2, no. 3, pp. 1–27, 2011.

[33] M. Pantic, M. Valstar, R. Rademaker, and L. Maat, "Web-based database for facial expression analysis," in *Proceedings of the 2005 IEEE International Conference on Multimedia and Expo*, p. 5, Amsterdam, Netherlands, July 2005.

[34] P. Lucey, J. F. Cohn, T. Kanade, J. Saragih, Z. Ambadar, and I. Matthews, "The extended Cohn-Kanade dataset (CK+): a complete dataset for action unit and emotion-specified expression," in *Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 94–101, San Francisco, CA, USA, June 2010.

[35] A. Dhall, R. Goecke, S. Lucey, and T. Gedeon, "Static facial expression analysis in tough conditions: data, evaluation protocol and benchmark," in *Proceedings of the IEEE International Conference on Computer Vision Workshops (ICCV Workshops)*, pp. 2106–2112, Barcelona, Spain, November 2011.

[36] M. Yeasin, B. Bullot, and R. Sharma, "Recognition of facial expressions and measurement of levels of interest from video," *IEEE Transactions on Multimedia*, vol. 8, no. 3, pp. 500–508, 2006.

[37] U. Mlakar and B. Potočnik, "Automated facial expression recognition based on histograms of oriented gradient feature vector differences," *Signal, Image and Video Processing*, vol. 9, no. S1, pp. 245–253, 2015.

[38] W. Sun, H. Zhao, and Z. Jin, "An efficient unconstrained facial expression recognition algorithm based on stack binarized auto-encoders and binarized neural networks," *Neurocomputing*, vol. 267, pp. 385–395, 2017.

[39] I. Gogić, M. Manhart, I. S. Pandžić, and J. Ahlberg, "Fast facial expression recognition using local binary features and shallow neural networks," *The Visual Computer*, pp. 1–16, 2018.

[40] W. Sun, H. Zhao, and Z. Jin, "A visual attention based ROI detection method for facial expression recognition," *Neurocomputing*, vol. 296, pp. 12–22, 2018.