Facial Emotion Recognition using Neighborhood Features

Abdulaziz Salamah Aljaloud[1], Habib Ullah[2], Adwan Alownie Alanazi[3]
College of Computer Science and Engineering
University of Ha'il, Ha'il, Saudi Arabia

***Abstract*—We present a new method for human facial emotions recognition. For this purpose, initially, we detect faces in the images by using the famous cascade classifiers. Subsequently, we then extract a localized regional descriptor (LRD) which represents the features of a face based on regional appearance encoding. The LRD formulates and models various spatial regional patterns based on the relationships between local areas themselves instead of considering only raw and unprocessed intensity features of an image. To classify facial emotions into various classes of facial emotions, we train a multiclass support vector machine (M-SVM) classifier which recognizes these emotions during the testing stage. Our proposed method takes into account robust features and is independent of gender and facial skin color for emotion recognition. Moreover, our method is illumination and orientation invariant. We assessed our method on two benchmark datasets and compared it with four reference methods. Our proposed method outperformed them considering both the datasets.**

Keywords*—Haar features; feature integration; emotion recognition; face detection; localized features; multiclass SVM classifier

I. Introduction

Classification of emotion in different classes is a field of significant attention nowadays. The most important of this field is related to human facial emotion classification which is demonstrated as a chain procedure to recognize various human emotions via facial skin expressions (shown in Fig. 1), verbal expressions, different gesture and body movements, and different physiological signals measurement methods. The importance of people feelings in the research of latest technology gadgets is well-known. In today's world, the analysis and recognition of human emotion recognition has an extensive range of significance in wide majority of applications including machine learning based human-computer interaction, online automated tutoring systems, image and video retrieval, smart environments for health-care, and automated driver warning systems as narrated by Seyedehsamaneh et al. [1]. In addition to what has been mentioned above, facial emotion recognition plays very important role in finding various mental health conditions by doctors, psychiatrists and psychologists. In the past few decades, scientists and researchers from multidisciplinary fields have proposed different approaches and methods to identify emotions from facial features, speech signals, and many other sources. However, it is worth noticing that it is still a difficult issue in the field of machine learning, deep learning, computer vision, psychology, physiology due to the nature of its complexity. Facial recognition started nearly 80's [41]. Scientists and researchers agreed that facial expressions are the most influential part in recognizing human emotion. But, it is difficult to interpret human's emotion by utilizing facial expression characteristics due to the sensitivity to the external noises for example illumination conditions and dynamic head motion Kwang et al. [2]. Moreover, the final results for emotion classification based on facial expressions still need to be improved. For this purpose, different research investigations have been made and it is found out that the clue lies in the baseline or the backbone of most of the methods based on the initial step of face recognition. This fact was further investigated by Jiankang et al. [49]. They discovered that if a robust technique is used to detect faces, then the complexity of next steps can be reduced substantial and the effectiveness of these next steps improve significantly. Ray and Mishra [12] investigated EEG signals and on top of that they considered different techniques to measure the performance of emotion recognition capabilities.

To handle these problems, we introduce robust technique for human facial emotion classification into various states using facial features in the localized regions. Our proposed technique does not rely on the postulation of a specific gender or skin color of different human beings. The proposed technique is illumination and orientation invariant to prevail robustness to these changes. In fact, the proposed technique is characterized by the compact representation of spatial information as illustrated by Manisha et al. [3] that effectively combines human facial emotion features. We fuse the characterizations of both face detection and human facial emotion classification into a unique framework. The proposed technique follows the inspiration of investigating the local structure of facial image with a different technique of unification of localized features. It is important to mention here that the proposed technique is motivated by smaller computational overhead. This characteristic of the method makes this method very feasible to be placed in practice for any handheld device, for example, smart phones and other smart portable devices. The flow and complete process of our proposed technique is outlined in Fig. 2. We identified faces in the images using famous Haar features. Subsequently, we then formulate localized regional descriptor (LRD) and exploit multi-class SVM to classify different human facial emotion. Our contribution lies in the development of localized regional descriptor that motivates us further enhance the proposed method with experimental analysis from different aspects.

Fig. 1. Facial Emotions. different Human Emotion are Depicted from their Faces in the Provided Sample Images.

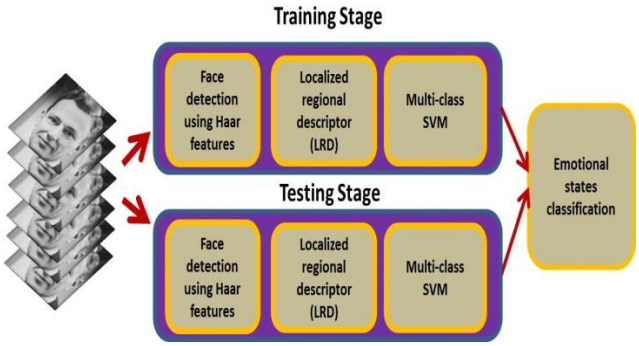

Fig. 2. Emotional States Classification using M-SVM Classifier. The Proposed Method Detect Faces in Video, and then Extract Localized Regional Descriptor (LRD) from the Detected Faces which are used to Train a Multiclass SVM During Training Stage. The Same Features are used to Classify Facial Emotions into different States During the Testing Stage.

In the rest of the paper, we present literature reviews in Section II and our proposed method in Section III for classification of facial emotions into seven different classes. Results are presented in Section IV, discussion is presented in Section V, and conclusion is presented in Section VI.

II. Literature Review

We provide details of literature review in this section. We have partitioned state-of-the-art techniques into 3 parts to explain human facial emotion classification considering speech signals, physiological signals measurements, and human facial expressions based recognition methods.

One dimensional signal, namely, speech which is a complicated signal providing a lot details, for instance, about the data to be communicated, speaker, language, region, and emotions. Therefore, we want to mention that speech processing is a significant field in digital signal processing and it presents a number of different applications including human computer interfaces using machine learning techniques, telecommunication between peer, assistive technologies for health-care, and security and safety associated with different places of people gatherings. The sound and speech/acoustic properties of the speech signal represent feature and the procedure through which some data is extracted from the speech signal and this is called feature extraction as introduced by Likitha et al. [4] and they utilized Mel Frequency Cepstral Coefficient (MFCC) method for human facial emotion recognition through speech signals representing different properties. Lotfidereshgi et al. [5] introduced an algorithm that uses the speech signal directly from the provided data through various speech collection devices. Therefore their technique fuses the robustness of the traditional source filter model of human speech generation with those of the currently presented liquid state machine (LSM) which is also called as biologically-inspired spiking neural network (SNN). Tzirakis et al. [6] presented a technique consisting of a Convolutional Neural Network (CNN). This model formulates features from the unprocessed signal, and concatenates them together to present them to a 2-layer Long Short- Term Memory (LSTM) network. Taking into account speech emotion from multiple sources i.e., in multiple speech emotion, the rate of identifying emotion will be decreased due to the expansion of emotional confusion. To fix this issue, Sun et al. [7] presented a speech emotion recognition technique considering the decision tree support vector machine (SVM) algorithm with Fisher feature selection bottom-up approach. Liu et al. [8] introduced a speech emotion recognition technique considering an enhanced version of brain emotional learning (BEL) algorithm, which is motivated by the emotional processing procedure of the limbic system in the brain of human beings. The outcome results of BEL algorithm is affected and improperly adapted by the reinforcement learning rule. Moreover, human emotions classification considering speech signals suffer from the unavailability of information and features because they don't provide improved interaction between human and machine in the form of a computer. To enhance the robustness of speech signals information itself, still a very large amount of technical space should be completed and addressed by the researchers in the same field.

Now we consider different category where emotions are classified using signal measurement procedure. For instance, physiological signals measurements are engendered by the physiological process of human beings, e.g., heart-beat rate (electrocardiogram or ECG/EKG signal of brain in the human), respiratory rate of human and content (capnogram), skin conductance (electro thermal activity or EDA signal on the body), muscle current (electromyography or EMG signal taken via different hardware sources available in the market), brain electrical activity (electroencephalography or EEG signal that can be measured using different electrodes on human skull). The aforementioned ways of signals collection help in finding emotion of human beings due to various mental and physical activities. For instance, Ferdinando et al. [9] used LDA technique (Linear Discriminant Analysis feature method), NCA (Neighbourhood Components Analysis feature method), and MCML (Maximally Collapsing Metric Learning for feature assessment) for the supervised monitoring and decreasing of different features in human emotion recognition based on ECG signals collected via electrodes. Kanjo et al. [10] presented a technique that removes the requirements for manual feature extraction by using multiple learning methods, for example, a hybrid method considering a deep model namely Convolutional Neural Network and another deep model, namely, Long Short-term Memory Recurrent Neural Network (CNN-LSTM) on the unprocessed sensor information based on phones and wearable devices easily available in the marked. Nakisa et al. [11] fixed the problem related with the high-dimensionality of EEG signals by presenting an algorithm to effectively search for the optimal subset of EEG features in

EEG signals. For this purpose, they used evolutionary computation (EC) methods. Moreover, taking into account signal pre-processing and emotion classification, their technique divides a huge set of emotions and combines extra features. Ray et al. (2019) introduced a method by using computational intelligence algorithm e.g., discrete wavelet transform and Bionic Wavelet Transform (BWT) for the evaluation of EEG signals Ullah et al. [13]. Jirayucharoensak et al. [14] investigated the usage of a deep learning network (DLN) to find out undiscovered feature correlation between input signals from various sources. The DLN is used with a stacked auto encoder using hierarchical feature learning technique. It is worth mentioning that the physiological signals measurement based techniques for human emotion classification face several issues as illustrated by Egon et al. [15]. These issues are obtrusiveness of physiological sensors, unreliability of physiological sensors, for example, due to movement artifacts of multiple reasons, not fixed bodily position, changing air temperature, and varying humidity. In addition to that, these signals have many-to-many relationship issues; that is, multiple physiological signals can partially serve as indicators for multiple conventional biometric features of human emotions. These signals also present varying time windows where measurements could differ.

Now we will provide details of methods based on facial emotion recognition aspects. Facial expression based emotions classification moves the next level the fluency of the environment, accuracy and genuineness of interaction taking place in the surroundings, especially to demonstrate human-computer interaction complications as illustrated by Rota et al. [16] in his method related to particle groupings. To take into account these considerations, both scientists and researchers from the community are contributing important efforts to facial expression based emotion classification techniques and the literature is increasing with the passage of time. Jain et al. [17] introduced an algorithm based on advance and latest Deep Convolutional Neural Networks (DNNs) that is made of various layers performing different functions and deep residual blocks to achieve different tasks of interest. Wang et al. [18] proposed a technique considering stationary wavelet entropy to discover robust features, and used a single hidden layer feed forward neural network as the classifier for facial expression classification. Jaya method is presented to block the training of the classifier fall into local optimum regions that would ultimately compromise the overall performance. Yan et al. [19] introduced a novel and robust discriminative multi-metric learning approach for facial expression classification in multiple video. Orientation feature descriptors from many directions for each face video are discovered to illustrate facial appearance and motion data from dynamic aspects. These metrics driven by multiple features are subsequently learned with these extracted multiple features in a unified fashion to use complementary and discriminative data for emotion classification. Sun et al. [20] introduced a multi-channel deep neural network that learns and puts together the spatial-temporal descriptors for facial expressions identification in static frames. The important concept of the algorithm is to discover and collect optical flow from the difference among the peak expression face frame and the neutral face frame as the temporal data of a specific facial expression, and consider the grey-level frame of peak expression face as the spatial data. A Deep Spatial-Temporal feature Fusion neural Network is investigated to collect the performance of the deep feature extraction and combination from the frames and images. Lopes et al. [21] introduced a robust algorithm for facial expression identification that uses a unification of Convolutional Neural Network and some novel pre-processing factors for the same purpose. Chen et al. [22] proposed a robust method to handle the key challenge of face motions by considering a robust set of features namely Histogram of Oriented Gradients from three perpendicular planes to collect features associated with textures from video data. For the consideration and utilization of facial appearance variations, a robust geometric feature Ullah et al. [23] is introduced from a novel transformation of facial landmarks. Discovering the strengths of facial features based emotions classification techniques, people in the field paid attention to facial expression based emotion classification techniques for handheld smart devices including mobiles. To this end, smart mobiles and smart wrist watches are fully equipped with different types of sensors, for instance, accelerometer, gyroscope, fingerprint Sensor, heart rate sensor, and microphone. Alshamsi et al. [24] investigated a method driven by sensor technology and cloud computing for identification of emotion in both speech and facial expression. Hossain et al. [25] introduced a framework that puts together the strengths of emotion-aware big data and cloud technology towards 5G. In fact, they fused together facial and verbal descriptors to introduce a bimodal technique for big data emotion classification. Grünerbl et al. [26] presented a method considering smartphone sensors for the identification of depressive and panic mental states and recognize state variations of people targeted by bipolar disorder disease. Sneha et al. [27] introduced the textual content of the message and user typing behaviour to make a model that easily divides the future instances. Hossain et al. [28] introduced a method in which Bandlet transform is used on the face areas, and the resultant subband is partitioned into non-overlapping sections. Additionally, a local binary pattern is investigated for each section. The Kruskal-Wallis feature selection is used to choose the most discriminative bins of the fused histograms, which are provided to Gaussian mixture model-based classifier to find different human emotion. Sokolov et al. [29] presented a cross-platform system for human emotion identification. Their system is based on convolutional neural network. Their system can effectively identify human emotions on arousal-valence level of measurement. Lee et al. [42] proposed deep networks for context-aware emotion recognition that consider both human facial expression and context data in a combined fashion. Mao et al. [43] introduced three HMM based frameworks and compared throughout the current paper. Han et al. [44] investigated and summarized the ideas and categories, techniques and applications of transfer learning briefly, and studies the combination of transfer learning and deep learning, and the application of speech emotion recognition. Borra et al. [45] presented an attendance system using partial facial recognition. Nhuong et al. [46] propose an algorithm for feature extraction for the purpose of face recognition. Imen et al. [47] introduce sequence kernels for emotion recognition. Erfana et al. [48] present a survey about the emotion intelligence of different algorithms in the field.

The literature is very limited due to the associated challenges of developing a reliable technique with low computational requirements. The aforementioned methods require huge computational powers since most of them are based on deep models. These methods are modelled for very narrow and specific emotions and they are not extendable easily to consider other emotional states. Therefore, we propose an efficient method for emotions classification into a set of different states using facial features. Our method is independent of gender class, skin colour, illumination changes, and face orientations. Our proposed method presents compact representation of spatial information Verma et al. [3] that effectively encodes emotion information. We integrate the strengths of both face detection and emotion classification into a unified model. Additionally, our method is driven by low computational complexity. Therefore, it can be implemented easily on any handheld device including smart phones.

III. Proposed Method

Feature modeling for facial emotion classification has been an active area in the fields of image processing and computer vision. The motivation for fast face modeling for realistic facial recognition and classification has led scientists to discover different model-based methods. The techniques in the literature for facial expression modeling and recognition differ in various aspects depending on the application under observation, computing efficiency, type of sensors, cost, and required accuracy. Some researchers proposed 3D generic face deformation for smartphone applications, where they use a single image to adapt the generic model to the face in the video frame captured via smartphones. Different methods can be adapted facial features extraction from video frames. Some researchers use stereo to model face features using differential geometry. However, this kind of technique requires prior knowledge about the shape of the surfaces of the face and its differential geometry for accurate performance. Parallel stereo images can also be used that rely on manually selected corresponding feature points to compute the rotation and translation matrices that are used to fit the model to the computed feature points. For facial emotion recognition we model features which are illumination and orientation invariants. For classification we use multiclass SVM which is a powerful and accurate classifier. Multiclass SVM presents good performance on many problems including non-linear problems. Due to the classification strengths of multiclass SVM, our method avoids both overfitting and underfitting. Multiclass SVM renders good performance by training it even with small samples. Considering our proposed features, this makes the classifier ideal for different personality traits, and high segmented facial expression. The multiclass SVM presents generalization capability; therefore, our proposed method can handle unseen data. The generalization capability of our method is determined by complexity and training of the multiclass SVM.

Face detection considering Haar feature-based cascade classifiers is a famous face detection model Aguilar et al. [30] and Viola et al. [31] due to its simplicity and robustness. Inspired by the mode, where we train a cascade function considering ground truth faces with their labels. In fact, the model entails a lot of positive labels for faces and negative labels for non-faces to train the classifier. Subsequently, we extract Haar features which resemble convolutional kernel. Each feature is a single value calculated by subtracting sum of pixels under a rectangle from sum of pixels under a different rectangle considering a video frame under observation. Due to different rectangles, we exploit different sizes and locations of each kernel to obtain a lot of features. For this purpose, the concept of integral image is exploited.

$$\Phi(x,y) = \sum_{x' \le x, y' \le y} \Gamma(x', y') \xi(x,y) = \xi(x, y-1) + \xi(x-1, y) + \Gamma(x,y) \quad (1)$$

$$\Phi(x,y) = \Phi(y-1) + \Phi(x-1) + \xi(x,y)$$

Where Φ is the integral image and Γ(x', y') is the original image. ξ(x, y) is the cumulative row sum. The integral image can be obtained in one pass over the original image. Additionally, we explore Adaboost model to filter out irrelevant features. To remove irrelevant features, we consider each and every feature on all the training images. For each feature, we investigate the optimal threshold which will classify faces and non-faces. We choose the features with smallest error rate since these features classify the faces and non-faces in optimized way. In the beginning, each image is rendered an equal weight. After each classification, we increase the weights of misclassified images and repeat the same procedure. We then calculate new error rates and new weights. We found that in each video frame and image, significant section consists of irrelevant areas. Therefore, if part of a window does not contain face, we remove it. To consider the concept of Cascade of Classifiers is modeled, which fuse the features into different stages of classifiers and use them one-by-one instead of applying all the features on a window. We remove the widow if it does not qualify the first stage. Therefore, we do not explore the remaining features. If the window qualifies the first stage, we apply the second stage of features and continue the procedure. A window qualifying all stages is a face region.

We than extract localized regional descriptor (LRD) which represents the features of a face based on localized appearance encoding. The LRD formulates different pattern based on the relationships between local areas themselves instead of considering only intensity information. For appearance information, we use localized regions in numerous directions and scales to compute regional patterns. We find the correspondence between localized areas by using the extrema on appearance magnitudes. We want to efficiently summarize the local structures of face by using each pixel as center pixel in a region under observation. Considering a detected face, for a center pixel Δc and neighboring pixels Δn (n=1,2,..,8), we compute the pattern number (ω) as,

$$(\omega)m,n = \sum_{n=0}^{M-1} 2^n xy \Xi 1(\Delta n - \Delta c) \quad (2)$$

$$\Xi 1 = \begin{cases} 1, & if \ \ (\Delta n - \Delta c) < 0 \\ 0, & Otherwise \end{cases} \quad (3)$$

where M and N are the radius of neighbors and number of neighbors for the pattern number. After calculating the ω of face, histogram is computed as formulated in the equation,

$$\Upsilon 1(1) = \sum_{x=1}^{M} \sum_{y=1}^{N} \Xi 2(\omega x, y, 1) : 1 \in [0, 2^M - 1] \quad (4)$$

Research deanship at the University of Ha'il, Saudi Arabia.

$$\Xi 2(a,b) = \begin{cases} 1, & if \quad a = b \\ 0, & Otherwise \end{cases} \tag{5}$$

Relationship between regions in terms of these pixels has been used, and a pattern number is assigned. We model histogram to represent the face in the form of LRD. For regional pixels Δ n and a center pixel Δ c, LRD can be formulated as,

$$\eta_1^n = \begin{cases} \eta_1^n = \Delta_8 - \Delta_n . \eta_2^n = \Delta_{n+1} - \Delta_n, & for \ n = 1 \\ \Delta_{n-1} - \Delta_n, \eta_2^n = \Delta_{n+1} - \Delta_{n,}, \forall, n = 1,2, \ldots .8 \\ \eta_1^n = \Delta_{n-1} - \Delta_n, \eta_2^n = \Delta_1 - \Delta_n, for, n = 8 \end{cases} \tag{6}$$

We find the difference of each region with two other regions in ηn1 and ηn2. Considering these two differences, we assign a pattern number to each region,

$$\Xi_3(\eta_1^n - \eta_2^n) = \begin{cases} 1, if \ \eta_1^n \geq 0, and, \eta_2^n \geq 0 \\ 1, \ if \ \eta_1^n < 0, and, \eta_2^n < 0 \\ 0, \ if \ \eta_1^n \geq, and, \ \eta_2^n < 0 \\ 0, \ if \ \eta_1^n < 0, \ and, \eta_2^n \geq 0 \end{cases} \tag{7}$$

For the central pixel Δc, LRD can be found using the above numbers and the histogram for LRD map can be calculated in the equations,

$$LRD(\Delta c) = \sum_{n=1}^{8} 2^{n-1} x \Xi_3(\eta_1^n - \eta_2^n) \tag{8}$$

$$\Upsilon_2(LRD) = \sum_{x=1}^{M} \sum_{y=1}^{N} \Xi_2 \left(LRD_{x,y,}1\right): 1 \in [0, 2^8 -]$$

The LRD represents robust features which are calculated by extracting the relationship among local regions by considering them mutually. The LRD finds the relationship of local regions with central region. In the proposed method, face detection and LRD are fused as they complete each other on the basis of characteristics they represent individually.

To classify LRD features into various classes, we use the M-SVM classifier Liu et al. [32] and Du et al. [33]. The M-SVM consists of different parameters which are a combination of different predictors. The M-SVM classifier takes the input features, classifies them with every set of parameters in the classifier, and provides the class label that obtained the majority of votes. The classifier is trained with the same parameters considering the training sets which are produced from the original training set using the bootstrap process. For each training set, the classifier identifies the same number of features as in the original set. The features are chosen with replacement. It means that some features will be taken more than once and some will be ignored. At each iteration of the algorithm, the classifier does not use all the variables to compute the best split, but a unpredictable subset of them. With each set of parameters a new subset is generated. The M-SVM classifier does not require any performance estimation process, such as cross-validation or bootstrap, or a separate test set to get an approximation of the training error. In fact, the error is calculated internally during the training. In fact, in machine learning, M-SVMs are supervised learning models with associated learning algorithms that analyze data used for classification and regression analysis. Given a set of training samples, each labeled as associated to one or the other of two categories, an SVM classifier sets up a model that assigns new samples to one class or the other, making it a non-probabilistic binary linear classifier. An M-SVM classifier is a representation of the samples as points in space, mapped so that the samples of the separate classes are isolated by a clear gap that is as wide as possible. New samples are then mapped into that same space and predicted to associate to a class based on the side of the gap on which they fall. M-SVM can efficiently perform a non-linear classification using what is called the kernel trick, implicitly mapping their inputs into high-dimensional feature spaces.

IV. Results

For experimental evaluation, we consider static facial expressions in the wild (SFEW) 2.0 Dhall et al [34] and Dhall et al. [35] dataset and real world affective faces (RAF) dataset Li et al. [36]. The dataset namely static facial expressions in the wild (SFEW) has been collected by choosing frames from AFEW part of the collection which is popular among the community of facial emotion recognition. The database presets a lot of images and frames representing unconstrained facial expressions, varied head poses, various age ranges, different occlusions, various focus, different resolution of face and close to real world illumination in both the background and foreground. These chosen facial frames are takes from AFEW sequences and labeled based on the label of the sequence. In summary, SFEW consists of many images and that have been marked for six different facial expressions including angry, disgust, fear, happy, sad, surprise and the neutral class and was labeled by two independent participants. Similarly, real-world affective faces database is very big facial expression dataset consisting of very diverse facial frames downloaded from the Internet. Based on different annotation technique, each individual frame has been independently marked by a huge number of participants. Frames in this dataset are of great varieties that include changes in age, gender and ethnicity, head poses, lighting conditions, occlusions, and post-processing operations. This dataset has large aforementioned diversities considering different factors, large quantities, and rich annotations. Additionally, we perform comparison with many state-of-the-art methods and reported the results in term of both confusion matrices and total accuracies. We consider seven facial emotion classes including sad, happy, angry, disgust, fear, neutral, and surprise.

We compare our proposed method with four reference methods over two datasets. These reference methods include implicit fusion model Han et al. [37], biorthogonal model Dong et al. [38], higher order model Ali et al. [39], and bioinspired model Vivek et al. [40]. The comparison results are listed in Table I in term of total accuracies. Our proposed method achieved promising results and performed better than four reference methods. Our method still has some limitations. For example, we did not exploit geometric features. Our method is applicable to treat and diagnosis patients with emotion issues. It is worth noticing that the anger facial expression is tense emotional outcome when the human considers that his/her personal limits are violated. Persons in this kind of emotion generally take the gestures including intense stare with eyes wide open, output uncomfortable sounds, bare the teeth, and attempt to physically seem larger. The staring with eyes wide

open is a significant hint for computers to recognize anger considering other facial emotions. There are also other face related elements including V-shape eyebrows, wrinkled nose, narrowed eyes, and forwarded jaws. All these important elements help to recognize anger emotion.

In the facial expression, happiness indicates an emotional state of joy. In this emotional state, the reader can find that the forehead muscle relaxes and the eyebrows are pulled up slowly. Apart from that, both the wrinkled outer corners of eyes and pulled up lip corners represent unique representation. In fact, the neutral facial emotion relaxes the muscles of the face and other facial emotions all need to use extensive muscles of face. The other six facial emotions in the datasets are more extreme.

We have also provided the confusion matrix for SFEW dataset in Table II. As can be seen, our proposed method presents encouraging results regarding the facial emotions.

We have also provided the confusion matrix for RAF dataset in Table III. As can be seen, our proposed method presents encouraging results regarding the facial emotions.

A great diversity of approaches has been proposed to solve the problem of facial emotion recognition. However, most of them are designed to work for specific emotion, where different representations of structures and appearance are analyzed with different models. In this paper we consider spatial properties of faces considering different emotions. The facial emotions are complex spatial representation with unexpected appearance or spatial patterns. For facial emotion recognition, we propose a novel method where we compute localized regional descriptor from the face images. Considering these facial emotions, we design a set of robust features combined into a unified LRD descriptor. For compact encoding of spatial patterns in these faces, we explore regional pixels which represent distinguish spatial patterns of faces. In fact, localized regional features are mid-level characteristics to fuse the distance between low-level and high-level features for capturing facial emotions. For classification, we exploited M-SVM which is a set of supervised learning methods used for classification and regression. Provided a set of training samples, the SVM classifier builds a model that finds the class of new unseen samples. This classifier is very significant in both machine-learning and data-mining curriculums and is frequently used by researchers. Besides, its utilization spans to a wide variety of applied research fields including but not limited to neuroscience, text categorization, and finance. The effectiveness of M-SVMs classification tasks in a wide variety of fields, such as text or image processing and medical informatics, has inspired researchers to do research on the execution performance and scalability of the training phase of serial versions of the algorithm. Since we describe a facial emotion from the view of a set of features, our method can be widely exploited in different applications. What's more, our modeling does not limit the type of features or the type of scenes, which helps us to extend the proposed technique to broader research fields. Experimental results demonstrated that our proposed approach is effective for the detection of various facial emotions.

TABLE. I. TOTAL ACCURACIES ARE PRESENTED FOR THE REFERENCE METHODS AND OUR PROPOSED METHOD CONSIDERING BOTH THE DATASETS NAMELY SFEW AND RAF

Methods	SFEW Total Accuracy	RAF Total Accuracy
Han et al. [37]	56.4	55.7
Zhang et al. [38]	54.2	56.6
Ali et al. [39]	49.8	52.7
Vivek et al. [40]	53.5	54.9
Prop. method	58.3	59.2

TABLE. II. CONFUSION MATRIX FOR SFEW DATASET IS PROVIDED WHERE ENCOURAGING PERFORMANCE OF OUR PROPOSED METHOD IS SHOWN

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	.54	.05	.06	.05	.04	.10	.02
Disgust	.04	.65	.20	.16	.01	.01	.03
Fear	.31	.10	.52	.01	.01	.04	.03
Happy	.04	.11	.01	.50	.01	.3	.02
Neutral	.09	.03	.10	.02	.63	.19	.05
Sad	.02	.12	.01	.07	.05	.51	.11
Surprise	.01	.02	.10	.01	.06	.13	.66

TABLE. III. CONFUSION MATRIX FOR RAF DATASET IS PROVIDED WHERE ENCOURAGING PERFORMANCE OF OUR PROPOSED METHOD IS SHOWN

	Angry	Disgust	Fear	Happy	Neutral	Sad	Surprise
Angry	.56	.04	.15	.03	.17	.06	.01
Disgust	.13	.58	.01	.01	.02	.15	.11
Fear	.00	.01	.62	.03	.03	.20	.12
Happy	.07	.04	.05	.50	.01	.04	.30
Neutral	.03	.01	.02	.13	.66	.06	.10
Sad	.07	.00	.20	.02	.01	.70	.01
Surprise	.10	.08	.15	.01	.03	.05	.57

V. DISCUSSION

We have presented a new method for facial emotion recognition based image processing and computer vision techniques. It is worth mentioning here that many methods have proposed previously for the same problem as we discussed in the literature review. However, those methods suffer from various problems ranging from limited datasets to limited metrics for the purpose of evolutions. Moreover, our proposed method is invariant to different key challenges as we mentioned in the introduction section. We carried out detail experimental analysis on two benchmark datasets which are considered very challenging for the same problem in the community. Thanks for localized feature descriptor that proved that our method is enriched with robustness to deal with the difficult problem of facial emotion recognition. In the experimental assessment, we used two performance metrics i.e., total accuracy and confusion matrix. Our method showed very promising results considering both aforementioned datasets and performance metrics. In fact, our work can be further extended with many machine learning and deep

learning approaches. However, these advance learning approaches required huge amount of data to process during the training stage. Therefore, we keep it our next step in the future.

VI. Conclusion

We explore a new method for facial emotion classification into seven different states. For this purpose, we detect faces and extract localized regional descriptor (LRD) based on the relationships between neighboring regions. To classify facial emotions into seven different classes, we train a multi-class SVM classifier which recognizes these emotions during the testing stage. We evaluated our method on two benchmark datasets and compared it with four reference methods that show that we outperformed them.

In our future work, we would like to consider publicly available datasets as well as we will collect our own datasets in order to have huge amount of data. Then we will explore a deep learning model for the same problem. A deep learning method will address the weaknesses associated with our method including the usage of limited datasets and the consideration of limited number human emotions for the purpose of classification.

Acknowledgment

This work was supported by the deanship of scientific research, University of Ha'il, Saudi Arabia [BA-1912].

References

[1] S. Shojaeilangari, W.Y. Yau, K. Nandakumar, J. Li, and E. Khwang Teoh (2015), "Robust representation and recognition of facial emotions using extreme sparse learning," IEEE Transactions on Image Processing, vol. 24, no. 7, pp. 2140–2152, 2015.

[2] K. E. Ko and K. B. Sim (2010), "Development of a facial emotion recognition method based on combining aam with dbn," in Cyberworlds (CW), 2010 International Conference on. IEEE, 2010, pp. 87–91.

[3] M. Verma and B. Raman (2018), "Local neighborhood difference pattern: A new feature descriptor for natural and texture image retrieval," Multimedia Tools and Applications, vol. 77, no. 10, pp. 11843–11866, 2018.

[4] M.S. Likitha, S. R. R Gupta, K. Hasitha, and A. U. Raju (2017), "Speech based human emotion recognition using mfcc," in Wireless Communications, Signal Processing and Networking (WiSPNET), 2017 International Conference on. IEEE, 2017, pp. 2257–2260.

[5] R. Lotfidereshgi and P. Gournay (2017), "Biologically inspired speech emotion recognition," in Acoustics, Speech and Signal Processing (ICASSP), 2017 IEEE International Conference on. IEEE, 2017, pp. 5135–5139.

[6] P. Tzirakis, J. Zhang, and B. W. Schuller (2018), "End-to-end speech emotion recognition using deep neural networks," in 2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2018, pp. 5089–5093.

[7] S., S. Fu, and F. Wang (2019), "Decision tree svm model with fisher feature selection for speech emotion recognition," EURASIP Journal on Audio, Speech, and Music Processing, vol. 2019, no. 1, pp. 2, 2019.

[8] Z.T. Liu, Q. Xie, M. Wu, W. H. Cao, Y. Mei, and J.W. Mao (2018), "Speech emotion recognition based on an improved brain emotion learning model," Neurocomputing, 2018.

[9] H. Ferdinando, T. Seppänen, and E. Alasaarela (2017), "Enhancing emotion recognition from ecg signals using supervised dimensionality reduction.," in ICPRAM, 2017, pp. 112–118.

[10] E. Kanjo, E. M. G Younis, and C. Siang Ang (2019), "Deep learning analysis of mobile physiological, environmental and location sensor data for emotion detection," Information Fusion, vol. 49, pp. 46–56, 2019.

[11] B. Nakisa, M. N. Rastgoo, D. Tjondronegoro, and V. Chandran (2018), "Evolutionary computation algorithms for feature selection of eeg-based emotion recognition using mobile sensors," Expert Systems with Applications, vol. 93, pp. 143–155, 2018.

[12] P. Ray and D. P. Mishra (2019), "Analysis of eeg signals for emotion recognition using different computational intelligence techniques," in Applications of Artificial Intelligence Techniques in Engineering, pp. 527– 536. Springer, 2019.

[13] H. Ullah, M. Uzair, A. Mahmood, M. Ullah, S. D. Khan, and F. A. Cheikh (2019), "Internal emotion classification using eeg signal with sparse discriminative ensemble," IEEE Access, vol. 7, pp. 40144–40153, 2019.

[14] S. Jirayucharoensak, S. P. Ngum, and P. Israsena (2014), "Eegbased emotion recognition using deep learning network with principal component based covariate shift adaptation," The ScientificWorld Journal, vol. 2014, 2014.

[15] E. L. V. D. Broek and M. Spitters (2013), "Physiological signals: The next generation authentication and identification methods!?," in Intelligence and Security Informatics Conference (EISIC), 2013 European. IEEE, 2013, pp. 159–162.

[16] P. Rota, H. Ullah, N. Conci, N. Sebe, and F. G. B. De Natale (2013), "Particles cross-influence for entity grouping," in 21st European Signal Processing Conference (EUSIPCO 2013). IEEE, 2013, pp. 1–5.

[17] D. K. Jain, P. Shamsolmoali, and P. Sehdev (2019), "Extended deep neural network for facial emotion recognition," Pattern Recognition Letters, 2019.

[18] S. H. Wang, P. Phillips, Z. C. Dong, and Y. D. Zhang (2018), "Intelligent facial emotion recognition based on stationary wavelet entropy and jaya algorithm," Neurocomputing, vol. 272, pp. 668–676, 2018.

[19] H. Yan (2018), "Collaborative discriminative multi-metric learning for facial expression recognition in video," Pattern Recognition, vol. 75, pp. 33–40, 2018.

[20] N. Sun, Q. Li, R. Huan, J. Liu, and G. Han (2017), "Deep spatialtemporal feature fusion for facial expression recognition in static images," Pattern Recognition Letters, 2017.

[21] A. T. Lopes, E. de Aguiar, A. F De Souza, and T. O. Santos (2017), "Facial expression recognition with convolutional neural networks: coping with few data and the training sample order," Pattern Recognition, vol. 61, pp. 610–628, 2017.

[22] J. Chen, Z. Chen, Z. Chi, and H. Fu, "Facial expression recognition in video with multiple feature fusion," IEEE Transactions on Affective Computing, vol. 9, no. 1, pp. 38–50, 2018.

[23] H. Ullah, A. B. Altamimi, M. Uzair, and M. Ullah (2018), "Anomalous entities detection and localization in pedestrian flows," Neurocomputing, vol. 290, pp. 74–86, 2018.

[24] H. Alshamsi, V. Kepuska, H. Alshamsi, and H. Meng (2018), "Automated facial expression and speech emotion recognition app development on smart phones using cloud computing," in 2018 IEEE 9th Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON). IEEE, 2018, pp. 730–738.

[25] M. S. Hossain, G. Muhammad, M. F. Alhamid, B. Song, and K. Al-Mutib (2016), "Audio-visual emotion recognition using big data towards 5g," Mobile Networks and Applications, vol. 21, no. 5, pp. 753–763, 2016.

[26] A. Grünerbl, A. Muaremi, V. Osmani, G. Bahle, S. Oehler, G. Tröster, O. Mayora, C. Haring, and P. Lukowicz (2015), "Smartphone-based recognition of states and state changes in bipolar disorder patients," IEEE Journal of Biomedical and Health Informatics, vol. 19, no. 1, pp. 140–148, 2015.

[27] H.R Sneha, M. Rafi, M.V. M. Kumar, L. Thomas, and B. Annappa (2017), "Smartphone based emotion recognition and classification," in Electrical, Computer and Communication Technologies (ICECCT), 2017 Second International Conference on. IEEE, 2017, pp. 1–7.

[28] M. S. Hossain and G. Muhammad (2017), "An emotion recognition system for mobile applications," IEEE Access, vol. 5, pp. 2281–2287, 2017.

[29] D. Sokolov and M. Patkin (2018), “Real-time emotion recognition on mobile devices,” in Automatic Face & Gesture Recognition (FG 2018), 2018 13th IEEE International Conference on. IEEE, 2018, pp. 787–787.

[30] W. G. Aguilar, M. A. Luna, J. F. Moya, V. Abad, H. Parra, and H. Ruiz (2017), “Pedestrian detection for uavs using cascade classifiers with meanshift,” in Semantic Computing (ICSC), 2017 IEEE 11th International Conference on. IEEE, 2017, pp. 509–514.

[31] P. Viola and M. Jones (2001), “Rapid object detection using a boosted cascade of simple features,” in Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on. IEEE, 2001, vol. 1, pp. I–I.

[32] L. Shuangyin, L. Xu, Q. Li, X. Zhao, and D. Li. "Fault Diagnosis of Water Quality Monitoring Devices Based on Multiclass Support Vector Machines and Rule-Based Decision Trees." IEEE Access 6 (2018): 22184-22195.

[33] D. Shichang, C. Liu, and L. Xi. "A selective multiclass support vector machine ensemble classifier for engineering surface classification using high definition metrology." Journal of Manufacturing Science and Engineering 137, no. 1 (2015): 011003.

[34] D. Abhinav, R. Goecke, J. Joshi, K. Sikka, and T. Gedeon. "Emotion recognition in the wild challenge 2014: Baseline, data and protocol." In Proceedings of the 16th international conference on multimodal interaction, pp. 461-466. ACM, 2014.

[35] D. Abhinav, R. Goecke, S. Lucey, and T. Gedeon. "Collecting large, richly annotated facial-expression databases from movies." IEEE multimedia 19, no. 3 (2012): 34-41.

[36] S. Li, W. Deng, and J. P. Du (2017), “Reliable crowdsourcing and deep locality-preserving learning for expression recognition in the wild,” in Proceedings of the IEEE conference on computer vision and pattern recognition, 2017, pp. 2852–2861.

[37] J. Han, Z. Zhang, Z. Ren, and B. Schuller (2019), “Implicit fusion by joint audiovisual training for emotion recognition in mono modality,” in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 5861–5865.

[38] Y. D. Zhang, Z. J. Yang, H. M. Lu, X. X. Zhou, P. Phillips, Q. M. Liu, and S. H. Wang (2016), “Facial emotion recognition based on biorthogonal wavelet entropy, fuzzy support vector machine, and stratified cross validation,” IEEE Access, vol. 4, pp. 8375–8385, 2016.

[39] H. Ali, M. Hariharan, S. Yaacob, and A. H. Adom (2015), “Facial emotion recognition based on higher-order spectra using support vector machines,” Journal of Medical Imaging and Health Informatics, vol. 5, no. 6, pp. 1272–1277, 2015.

[40] T. V. Vivek and G. R. M. Reddy (2015), “A hybrid bioinspired algorithm for facial emotion recognition using cso-ga-pso-svm,” in 2015 Fifth International Conference on Communication Systems and Network Technologies. IEEE, 2015, pp. 472–477.

[41] B. Arthur L. "The neuropsychology of facial recognition." American Psychologist 35, no. 2 (1980): 176.

[42] L. Jiyoung, S. Kim, S. Kim, J. Park, and K. Sohn. "Context-aware emotion recognition networks." In Proceedings of the IEEE International Conference on Computer Vision, pp. 10143-10152. 2019.

[43] M. Shuiyang, D. Tao, G. Zhang, P. C. Ching, and T. Lee. "Revisiting Hidden Markov Models for Speech Emotion Recognition." In ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 6715-6719. IEEE, 2019.

[44] H. Zhijie, H. Zhao, and R. Wang. "Transfer Learning for Speech Emotion Recognition." In 2019 IEEE 5th Intl Conference on Big Data Security on Cloud (BigDataSecurity), IEEE Intl Conference on High Performance and Smart Computing,(HPSC) and IEEE Intl Conference on Intelligent Data and Security (IDS), pp. 96-99. IEEE, 2019.

[45] S. Borra, K. J. Nazare, S. V. Raju, and N. Dey. "Attendance recording system using partial face recognition algorithm." In Intelligent techniques in signal processing for multimedia security, pp. 293-319. Springer, Cham, 2017.

[46] L. D. Nhuong, G. N. Nguyen, L. V. Chung, and N. Dey. "MMAS Algorithm for Features Selection Using 1D-DWT for Video-Based Face Recognition in the Online Video Contextual Advertisement User-Oriented System." Journal of Global Information Management (JGIM) 25, no. 4 (2017): 103-124.

[47] T. Imen, M. S. Bouhlel, and N. Dey. "Discrete and continuous emotion recognition using sequence kernels." International Journal of Intelligent Engineering Informatics 5, no. 3 (2017): 194-205.

[48] Z. S. Erfana, A. M. Khan, A. K. Srivastava, N. G. Nguyen, and N. Dey. "A study of the state of the art in synthetic emotional intelligence in affective computing." International Journal of Synthetic Emotions (IJSE) 7, no. 1 (2016): 1-12.

[49] D. Jiankang, J. Guo, N. Xue, and S. Zafeiriou. "Arcface: Additive angular margin loss for deep face recognition." In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4690-4699. 2019.