JOURNAL OF BASIC SCIENCE AND ENGINEERING

GRADIENT DIRECTIONAL EDGE CODING (GDEC) FOR EXPRESSION RECOGNITION FROM FACIAL IMAGES

B. Srinivas, Prof.L.Nirmala devi, Dr.A.Nageswar Rao, K.V Subba Reddy

Abstract:*Facial Expression Recognition (FER) requires an effective Expression descriptor that can provide sufficient discrimination between different facial expressions. However, the existing FER methods are susceptible to noise, distortions and not able discriminate flat regions from noisy regions. Hence, this paper proposes a new Face descriptor called as Gradient Directional Edge Coding (GDEC) which encodes the expression components through the directional edges. Initially, GEDC finds the gradients for each pixel and then encodes them with their neighbor pixel's support. The support is assessed based on the deviation in the direction of corresponding neighbor pixel with the mean direction of a local region. Each pixel is encoded a 7-bit code word among which the six bits are belongs to the directions of neighbor pixels and one bit is sign bit. After describing the expression, the classification is accomplished through Support Vector Machine at different kernels. Experimental validation on Standard CK+ dataset shows an accuracy of 94.6300% which is outstanding compared to the state-of-the-art methods.*

Keywords: Face Expression Recognition, Gradient Directions, Compass masks, Edge Detection, Accuracy.

I. INTRODUCTION

From the past few years, the Facial Expression Recognition has been gone through a huge research because it has a widespread applicability in different computer vision [1] related applications. Majorly the FER is employed in Human Computer Interaction (HCI), Security, Automatic Counseling Systems, face expression synthesis, lie detection, music for mood, automatic tutoring systems, driver fatigue detection etc. [2]. The main reason behind the significance of FER is the presence of facial cues which contributes more and more information towards expression synthesis. Among different cues of expression recognition like face expressions, hand gestures, and voices, the face expression contribute approximately 55% for human to human interaction. For instance, the robots can communicate easily with the people if they can analyze their emotional state by analyzing their facial expressions. Similarly, the FER can also assist for medical experts in health care systems to determine the mental statuses of patients such that they can improve their quality. In general, there exists two types of facial expressions, namely negative and positive expressions. The Former category includes anger and sad while the latter category includes happy and pleasure etc. The negative expression represents the unhealthy condition while positive expression reflects the healthy. Thus the FER can enhance the medical treatment for the patients by the analysis of their behavior [3], [4].

Typically, the FER system is established through two major components, they are (1) Face Expression Descriptor and (2) Classifier. The first component describes the features of expressions in facial image while the second component classifies it based on the knowledge available. The provision of knowledge is done through different machine learning algorithms and deep learning algorithms. The major research is done at the first component and researchers develops new and effective descriptors such that the recognition system can become robust for different real time issues like noise, illuminations, distortions, etc. Based on the methodology followed to describe an expression from facial images, the entire methods are categorized as Appearance based methods and geometry based methods [5-7]. The method in the latter category [8-10] describes the expression through facial geometry that includes the key elements such as Mouth, eye, ear and noise. Initially, these key components are extracted and their shapes, position, geometrical relations through Euclidean distances are used to describe the expression. But, they are not reliable and the tracking of expression through geometric features is a complex task. On the other hand, the appearance based methods uses texture as the features of expressions. For texture extraction, local and global filters are applied on the local region and overall facial region respectively. Compared to the geometric features, appearance features express more information and also much stable in nature.

Appearance based methods are again categorized as global methods [11-14] and local methods [15-17]. In the former category, the entire facial region is considered as input and some fitters are applied over it to represent the expression. However, they are not able to explore the variations in local appearance at key component such as mouth, eyes etc. On the other hand, the local method initially segments the image into several segments and then each segment is applied to represent the expression based on the local statistical or texture information. Local appearance features are more effective in describing the expression as they encode all the expression attributes like edges, corners, curves those incur at key components. Thus, they have gained a superior performance in FER. Local Binary Pattern (LBP) [18] is one of the most prominent local appearance face descriptor which encodes the target pixel based on the intensities of neighbor pixels. Even though it is a simple, it experienced a hue information loss. Moreover, it is much sensitive to noises and non-monotonic illuminations. Even though some variants of LBP like Local Mean Binary Pattern (LMBP) [19] are proposed for FER, they also experienced the same problems. Moreover, LBP and LMBP were not concentrated on the specific discrimination between flat and noise pixels. They applied a common texture encoding for all the pixels in the facial image.

Local Directional Number Pattern (LDN) is proposed by A. R. Rivera et al. [20] as an edge specific encoding technique which applies a preprocessing method to determine the edges before encoding the pixel. They applied Kirsch Compass Mask and Gaussian Masks to find out the edge responses of facial image. Further, the edge responses are encoded independently through LDN and fed to SVM for classification. However, the LDN produces similar codes for noisy and flat pixels as it is more prone to noise in the local region. Similar, M. T. B. Iqbal [16] also employed Kirsch Compass Kernel for edge responses extraction and discriminate the flat pixels from noise through the top two directions of edge responses. However, they allowed a constant threshold for encoding which is not robust for varying noise and contrast levels [21]. Broadly speaking, the major problems observed are as follows;

1. 8-Bit code: almost all the methods encoded each pixel in the facial image with an eight bit pattern.

2. Less discrimination between flat, noisy and local distorted pixels: Even though edge coding techniques are effective, they didn't provide a perfect discrimination between flat and noisy pixels.

In order to sort out these problems, we propose a new face descriptor called as Gradient Directional Edge Coding (GDEC) that encodes each pixel in the facial image with a 7-bit code. Initially, the proposed method determines the strongest edge responses by applying different edge filters over facial image. Next, each pixel is encoded with the help of its neighbor pixels support. The directions of two neighbor pixels those have maximum support with center pixel are encoded as the GDEC code. Further, the GDEC coded image is described through a global face descriptor through histogram computation and then fed to classification. Here Support Vector Machine is employed for classification.

The remaining paper is structured as follows; section II explores the particulars of literature survey. Section III explores the particulars of proposed GEDC method. Section IV explores the particulars of experimental validations and section V concludes the paper.

II. LITERATURE SURVEY

2.1 Handcrafted Methods

Recently, N. T. Cao et al. [22] firstly separated the facial image into equal sized non-overlapping blocks. After successful division, each block is texture encoded using LBP method. Then they classified the emotion through Support Vector Machine (SVM) algorithm. They have been used three different types of images to validate the simulation results such as small, medium, & large. In [23], the authors concentrated on prime features of face. They are eyes, mouth, &nose and measured LBP for each feature. Later exclusive coding pattern is used for encoding and finally Multi-Layer Perceptron Neural Network (MLP-NN) classifier is used to classify the emotion.

Further, Min Huet al. [24] proposed a local feature descriptor called a Center-Symmetric Local Octonary Pattern (CS-LOP) which encodes the pixel by considering the gray value between center pixel and neighbor pixels in all eight directions along with the four pairs of center symmetric pixels. Along with CS-LOP, they extracted diverse features from pre-processed image, gradient feature maps. The classification is done with the SVM algorithm through RBF kernel.

Like LDN, In order to find out the directions of edges A. Vijaya Lakshmi, P. Mohanaiah [25] applied Gaussian & Gabor filters over the facial image.They used Gaussian attributes to remove noise sensitivity and Gabor attributes to remove view variance. They called the above method as Gradient Based Compact Binary Coding (GCBC). Like GCBC method Sammaiah et al. [26] introduced analogous method named as Edge Adaptive Local Direction Binary Pattern (EALDBP). In EALDBP method, maximum top positive & negative responses are used to encode the facial image. Further, Robinson & Gaussian Compass masks are considered to determine the edge responses. GCBC & EALDBP methods have three common disadvantages: (i) the length of the code is 8-bit (lengthy) (ii) effect of noise is more in local region (iii) There is no distinction made between flat and noisy pixels.

In [27], the authors extended the LBP method by including the covariance matrix of the input facial image. Extended LBP (ELBP) method is used for texture encoding & covariance matrix is used for reducing the dimensionality. The K-L Transform (KLT) is used for classification, followed by SVM. Directional wavelet transform (DIWT) is introduced by M. Abdul and R.S. Holambe [28] through which facial image is decomposed into directional wavelet sub-bands. Further, adaptive directional band is selected from each sub-band and denoted as LBP features. To abolish the noise from facial images, I.M. Revina et al. [29] proposed an Enhanced Modified Decision based Unsymmetric Trimmed Median Filter (EMDBUTMF). Further, LDN and Dominant Gradient Local Ternary Pattern (DGLTP) [30] is used for feature extraction and SVM is used for classification. Finally, the simulation results are validated using JAFFE and CK databases.

Sumeet Saurav et al. [31] introduced a different form of LTP named as Dynamic LTP (DLTP) [32]. It is used to extract the feature related to expression from facial image. Initially, facial region is located and registered for recognition system. Feature extraction is done using DTLP method & Principal Component Analysis (PCA) [33] is method is used for reducing the dimensionality. Finally, Multi-Class Kernel Extreme Learning Machine (K-ELM) [34] classifier is employed to recognize the face expression and validated the simulation results with RaF, JAFFE, RAF-DB, & CK databases. In [35], the authors proposed a new encoding method named as Center-Symmetric Local Magnitude Pattern (CS-LSMP) and hybrid feature extraction method. Facial patches, feature maps belong to Orientational Magnitude, Negative and Positive Magnitude, Gabor filter [36] are considered in CS-LSMP method. A super vector is formed by combining all these features and it is forwarded through SVM for classification with polynomial kernel.

2.2 Deep Learning Methods

Recently few authors are concentrated on Deep Learning algorithms to extract the features and classification because of their flexibility & supremacy. In the direction of such, H. Zhang et al. [37] proposed Image Edge Computing with Convolutional Neural Network (CNN) technique to recognize the facial expression. After the image normalization, each edge layer is extracted through CNN and then masked over facial image thereby the edges are preserved. Max-pooling layer is used to reduce the dimensions and softmax layer is used for classification.

In [38], the authors collected the active units (AUs) of a facial image that were strongly associated to expression, and the resultant image is known as a partial image. Further, Expression metric Loss Function is used to get the dissimilarities between intra & inter classes. Finally, the classification and expression metric losses are optimized as part of the joint optimization.

Chang li et al. [39] proposed the combination of a series of "Long-Term Short Memory networks with Attention Mechanism (ALSTMs)" with "Spatial Attention CNNs (SACNN)" to enhance the discriminative power in FER. SACNN extracts the features from static images and ALSTMs explores the significance of landmarks in FER. The landmarks are categorized into seven groups and each group is subjected to the extraction of geometric features.

The deep model based three stage FER system is proposed by Jun Liu et al. [40]. At the beginning, to decrease intra-class differences, a pose-guided facial alignment is performed. Next, LBP, Edge Histogram Descriptor (EHD), & Pyramid Histogram Orientation Gradient (PHOG) are used as part of a hybrid feature extraction method. ResNet [41] and VGG-16 [42] are considered for the purpose of classification.

By considering feature extraction, Deep learning methods are easy when compared to handcrafted methods but they are limited in their ability to describe the expression in various conditions such as distortions, occlusions, & sounds. Further, there is no definite technique to distinguish noisy & flat pixels. Moreover, Deep learning features add more complexity when compared to the handcrafted features.

III. PROPOSED METHODOLOGY

3.1 Overview

As depicted in Figure.1, the proposed methodology consists of two phases; they face expression representation and classification. The former phase represents the facial image through gradients followed by histograms. Under gradient formation, we employed totally three edge detection filters; they are Prewitt Filter, Kirsch Compass Mask (KCM) and Robinson Compass Mask (RCM). At first, the facial image is convolved with these three filters and based on the obtained edge responses, a gradient magnitude image and gradient direction image is constructed. Further, with the help of these two images, each pixel is encoded as a 7-bit code word. For encoding each pixel, we consider the support of its eight neighbor pixels and a priority is give more to the pixels those contribute more towards the edge passing through center pixel. The 7-bit code word consists of directions of two pixels with maximum Weightage. After, the encoded image is subjected to classification through SVM classifier.

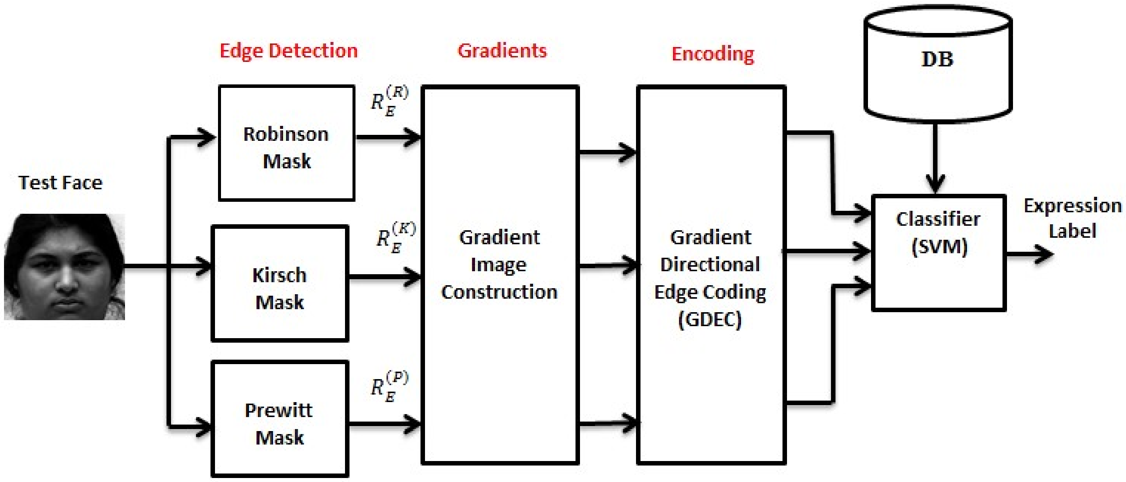

Figure.1 Block diagram of proposed method

3.2 Edge Detection

Edges signify the local pixel intensity changes in the facial image. An edge is defined as set of connected pixels those formulate a boundary between two disjoint regions. Generally, there exists three types of edges; they are Vertical Edges, Horizontal Edges and Diagonal Edges. Edges play a significant role in the representation of expressions from a facial image. During the occurrence of expression, the facial components deform into different shapes and such kind of deformations can be regarded as edges. These edges can be extracted through edge detection filers and broadly there exists two types of filters; they are Gradient and Gaussian Filters. The gradient based edge filters regards the first order deviations of the facial image as edges and applies different filters including Robert, Prewitt and Sobel operators for their extraction. On the other hand, the Gaussian based edge filters regards the edges as second order deviations of the facial image and applies different filters like Canny, Laplacian of Gaussian (LoG) for their extraction. As the facial expression related components are more directional in nature, the extraction of such kind of attributes is possible with Gradient based filters only. Hence we employed totally four types of edge detection filters namely Prewitt, Sobel, kirsch and Robinsons operators. The details of these operators are explored din the following subsections;

3.2.1 Prewitt

Prewitt operator [43] is generally employed for the detection of edges in images. It detects the edges in two directions; they are vertical direction and horizontal direction. Prewitt operator applies derivative masks to derive the edges from images. Hence, it can also be called as derivative masks or derivative operators. The derivatives should be derived in such a way the sum of all elements is equal to zero and opposite sign must be there in the mask. Since we need to extract the edges in differ directions, the basic Prewitt operator is rotated with an angel of 45^0. Thus, we get totally eight derivative masks for Prewitt operator, as shown in Figur.2. At each rotation, one mask is derived based on the coefficients in the earlier derived mask. The edge responses at each mask are shown in Figure.3.

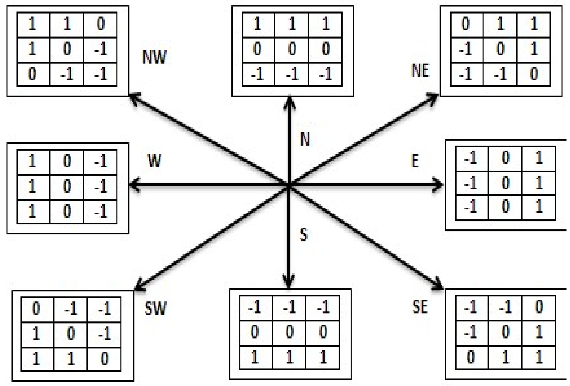

Figure.2 Prewitt mask in eight directions

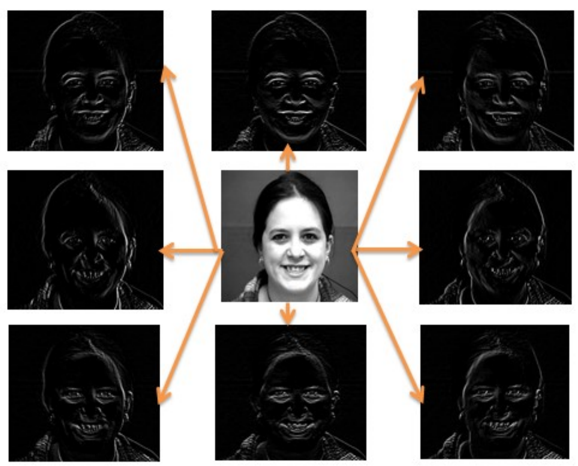

Figure.3 Edge responses of Prewitt operator in eight directions

3.2.2 Kirsch Compass Mask (KCM)

KCM [44] is an edge directional mask and regarded as a non-linear edge detection filter. KCM determines the maximum edge response after the accomplishment of its eight kernels in eight directions. KCM takes one mask as a base mask and rotates it to get the masks in different directions such as (1) North East (NE), (2) East (E), (3) South East (SE), (4) South (S), (5) South West (SW), (6) West (W), (7) North West (NW) and (8) North (N). The edge magnitude is calculated as the maximum magnitude among the obtained edge responses in eight directions. The kernel size is uniform and it is kept as 3×3. Figure.4 shows the different masks of KCM and their edge responses are shown in Figure.5.

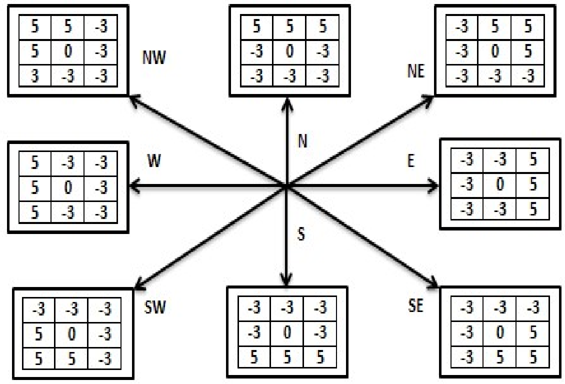

Figure.4 Different Kernels of KCM

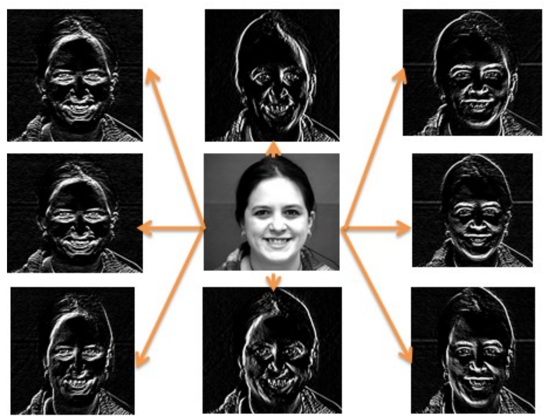

Figure.5 Edge responses of KCM in eight directions

3.2.3 Robinson Compass Mask (RCM)

Similar to KCM, RCM [45] also one of the most popular edge directional mask and employed totally in eight directions such as N, NW, W, WS, S, SE, E, and NE. The major difference between KCM and RCM is filter coefficients. As the KCM consist of multiple types of values, the RCM consists of only 0, 1 and 2. RCM is symmetric in nature hence eth edge responses obtained in one direction are symmetric with the edge responses obtained in the opposite direction. For instance, if the edge response of a pixel in the second direction is 123, then the edge response of a pixel in the sixth direction is -123. Thus out of eight edge responses, we consider only four responses and the size of each mask is kept as 3×3. Figure.6 shows different masks of RCM at different directions and the corresponding edge responses are shown in Figure.7.

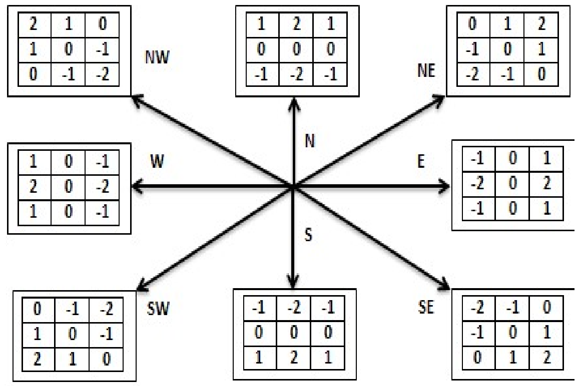

Figure.6 Different masks of RCM

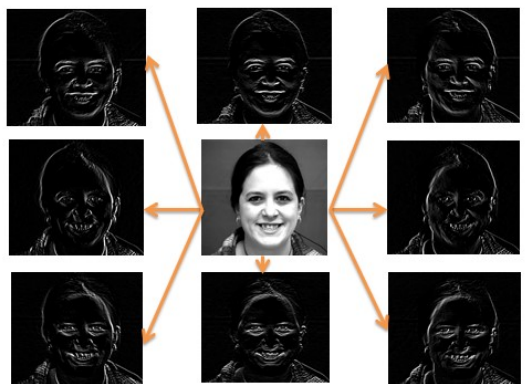

Figure.7 Edge responses of RCM in eight directions

3.3 Gradient Attributes

The proposed encoder looks for the Edge properties passing through the pixel. So our method initially calculates the gradient magnitude and gradient directions of every pixel of the input facial image.Consider a facial image P of size $M \times N$. In the above edge detection process through different filters, we get tally four responses for each pixel. Consider $R_E^i(m,n)$ is the response obtained by the convolution of i^{th} kernel and $R_E^j(m,n)$ be the response obtained by the convolution of j^{th} kernel, the gradient magnitude ($G_M^{ij}(m,n)$)and Gradient direction ($G_\theta^{ij}(m,n)$) of the pixel at $(m,n)$is calculated as

$$G_M^{ij}(m,n) = \sqrt{\left(R_E^i(m,n)\right)^2 + \left(R_E^j(m,n)\right)^2} \quad (1)$$

$$G_\theta^{ij}(m,n) = \tan^{-1}\left(\frac{R_E^i(m,n)}{R_E^j(m,n)}\right) \quad (2)$$

Further $R_E^i(m,n)$ and $R_E^j(m,n)$are calculated as

$$R_E^i(m,n) = P(m,n) * K_i \quad (3)$$

And

$R_E^j(m,n) = P(m,n) * K_j$ (4)

Where K_i and K_j are the two masks orthogonal to each other. Since there exists eight directions, each mask have an orthogonal mask and thus each pixel can be represented by four gradient magnitudes and four gradient directions.From these four gradient magnitudes and directions, the final gradient magnitude and gradient direction are calculated as

$G_M(m,n) = \max_i G_M^i(m,n) \forall i \in [1,2,3,4]$(5)

And

$G_\theta(m,n) = \max_i G_\theta^i(m,n) \forall i \in [1,2,3,4]$ (6)

Where $G_M(m,n)$ is the final gradient magnitude and $G_\theta(m,n)$ is the final gradient direction of the pixel at position (m,n). The size Gradient magnitude image and Gradient direction image is equal to the size of input image P.

3.4 Encoding

The proposed encoding scheme looks for the presence of properties of edge passing through center pixel. For this purpose, we consider the gradient magnitude and orientation from above calculations. To encode each pixel, we compute the Weight of Gradient Direction (WGD) at each of its eight neighbor pixels. The proposed encoder searches for the edge segments those were passing through center pixel and neighbor pixels. These kinds of pixels are termed as neighbor edge pixels. The proposed method concentrates at the determination of support of neighbor edge pixels and it is determined with the help of their gradient orientations. To determine such kind of edges, the proposed encoder defined eight template orientations based on the eight compass masks [46]. If the gradient orientation at any neighbor edge pixel is analogous to the template orientation, the encoder considers it as a supporting pixel. At this study, we use the Gaussian distribution to ensure larger weight for the neighbor pixel if it is closer to the template orientation. The Gaussian weight for a gradient orientation at i^{th} neighbor pixel is measured as

$W_\theta^i = \frac{1}{\sigma\sqrt{2\pi}} e^{-(\theta_i-\mu)^2/2\sigma^2}$(7)

Where θ_i is the gradient orientation at i^{th} neighbor pixel and μ is the mean andσ is standard deviation of gradient orientations of the given template of size 3×3. According to the earlier discussion, the gradient magnitude has significant importance in signifying the edgeness of a pixel because the higher value of gradient magnitude indicates the presence of prominent edge. Hence, we combined the Gradient magnitude with gradient direction and derived a new measure called as WGD Score (WS), as

$$W\_S_i = \begin{cases} G_m^i \times W_\theta^i, & if\ G_m^i \geq \varphi \\ 0, & otherwise \end{cases} \quad (8)$$

Where φ is referred as threshold based on which the neighbor edge pixels are determined. φ is used to filter the edge responses from flat featureless regions which have very less contribution towards face expression recognition. Eq.(8) obtains a larger value for a neighbor pixel with gradient magnitude larger than the threshold and gradient orientation matching with template orientation.

In general, any facial expression is expressed with the help of several edge components like line-end, Curve, Corners and Straight edge. For example, the eyes and mouth looks like corners while the lips, eyelids and eyebrows looks like curved edges. So, to characterize an edge component, we need two pixels and hence the proposed encoder determines two neighbor edge pixels those are maximum close direction with the direction of center pixel. Based on the obtained W_S_i, we choose two most prominent edge directions through

$$\theta_d = argmax2(W\_S_i; i = 0\ to\ 7)(9)$$

The above Eq.(9) returns indices of two maximum weights from the given set of eight weights. These two indices indicate the directions of neighbor pixels in the given template. Consider d_p as θ_d^1 and θ_d^2, as primary direction and secondary directions. As discussed earlier, the darker pixels lies at one side of edge and brighter pixels lie on the other side, an indicator need to be assigned to differentiate the direction. For this purpose, a sign bit is added to the directionθ_d. The sign information is added before the Most Significant Bit (MSB) of θ_d as

$$\theta_d = \theta_d + sign(10)$$

Here, the direction θ_d is represented with seven bits (three bits for positive direction and three more bits for negative direction a done bit for sign). The first bit signifies the sign, second three bits signifies the primary direction and last three bits signifies the secondary direction. In most of the cases, the sign of secondary direction is same as the sign of primary direction, we didn't consider it. Hence, we add sign for primary direction such that the code redundancy can be achieved. Finally, the code of center pixel is obtained as

$$C = 2^3 \times \theta_d \ (11)$$

Where C is a 7-bit code which discards the effect of insignificant spurious codes in the classification stage. For the pixels in flat regions, we assign a default value of 128 = 2^7. Since the remaining edge pixels are encoded with 7-bit code pattern, to ensure the uniformity, the flat regions are also encoded with 7 bits and the equivalent decimal value of 2^7 is 128.

IV. EXPERIMENTAL ANALYSIS

Herein the following section, we explore the performance of proposed FER system by validating it over several standard benchmark and publicly available databases. For the experimental validation, we used MATLAB tool. Herein this section, initially, we demonstrate the details of databases used and then the details of performance metrics followed by comparative analysis.

4.1. Databases

For experimental validation, we used standard and publicly available and benchmark facial expression database named as Cohn-Kanade (CK+) [47]. The details of the database, i.e., number of samples for each expression, resolution, number of expression types etc. are discussed here. CK+ dataset is one more popular and facial expression dataset which consists of totally 486 sequences. All these images under this dataset are acquired with the help of 97 subjects. It composed of both posed and non-posed expressions and the resolution of each image is observed as 490×640. All the images are of .PNG format and some samples images are shown in Figure.8. Here, every sequence has a neutral expression at starting and it rises to the corresponding expression after some time. The emotions present in this dataset are 8, they are namely Neutral, Contempt, Disgust, Fear, Sad, Happy, Angry and Surprise.

(a) (b) (c) (d) (e) (f) (g)

Figure.8 Samples from CK+ dataset (a) Surprise (b) Angry (c) Happy (d) Sad (e) Fear (f) Disgust and (g) Neutral

4.2 Results

Under the simulation with CK+ dataset, we conduct two kinds of simulations; they are 7-class expression simulation and 6-class expression simulation. Under first simulation, the total number of expressions considered is seven; they are namely Contempt, Disgust, Fear, Sad, Happy, Angry and Surprise. For the second simulation study, we excluded Contempt expression and the reaming expressions are considered as it is.Moreover, the simulation is done for three different combinations; such as Prewitt with GDEC, RCM with GDEC and KCM with GDEC. For all these simulations, the total number of facial images used for testing is 1488 (Angry – 231, Contempt – 165, Disgust – 193, Fear – 187, Happy – 267, Sadness – 130 and Surprise - 315). The confusion matrices obtained at these simulation studies are shown here;

Table.1Confusion matirx of 7-class expression simulation with KCM plus GDEC over CK+ dataset

	Angry	Contempt	Disgust	Fear	Happy	Sadness	Surprise	Total
Angry	**208**	2	4	5	6	4	2	**231**
Contempt	3	**148**	3	3	1	7	0	**165**
Disgust	6	0	**184**	3	0	0	0	**193**
Fear	5	2	0	**173**	0	7	0	**187**
Happy	6	0	1	0	**256**	0	4	**267**
Sadness	4	2	1	6	1	**116**	0	**130**
Surprise	2	0	2	0	0	0	**311**	**315**
Total	**234**	**154**	**195**	**190**	**264**	**134**	**317**	**1488**

Table.2 Confusion matirx of 7-class expression simulation with RCM plus GDEC over CK+ dataset

	Angry	Contempt	Disgust	Fear	Happy	Sadness	Surprise	Total
Angry	**195**	2	9	10	8	5	2	**231**
Contempt	5	**140**	5	5	1	9	0	**165**
Disgust	5	2	**180**	4	1	0	1	**193**
Fear	10	0	4	**163**	0	10	0	**187**
Happy	5	0	0	0	**260**	0	2	**267**
Sadness	3	2	2	7	0	**116**	0	**130**

Surprise	8	2	6	2	2	0	**295**	**315**
Total	**231**	**148**	**206**	**191**	**272**	**140**	**300**	**1488**

Table.3 Confusion matirx of 7-class expression simulation with Prewitt plus GDEC over CK+ dataset

	Angry	Contempt	Disgust	Fear	Happy	Sadness	Surprise	Total
Angry	**187**	4	9	10	10	8	3	**231**
Contempt	5	**136**	8	4	2	9	1	**165**
Disgust	10	4	**165**	2	4	2	6	**193**
Fear	11	1	2	**160**	1	11	1	**187**
Happy	9	1	1	1	**249**	1	5	**267**
Sadness	4	2	2	9	1	**111**	1	**130**
Surprise	9	3	8	3	4	1	**287**	**315**
Total	**235**	**151**	**195**	**189**	**271**	**143**	**304**	**1488**

According to the confusion matrix shown in Table.1, the diagonal values denotes the True Positives for every expression, means the correctly recognized samples. For instance, consider the firs row, the total number of samples used for testing are 231 and the total number of sample recognized correctly are 208. Similarly, the total number of correctly recognized surprise samples is 311 out of 315 input samples. Here the 311 is called as True positive (TP) and the summation of remaining sample sin each row is called as False Negative (FN). For a given input facial image with one expression, if the system recognized it as it is then it is counted under TP otherwise it is counted under FN. From the above tables, we can observe that the surprise has gained more number of TPs while the least number of TPs are gained for Sadness. Since the sad expression resembles its facial features with Fear and neutral, the recognition system misclassifies it and results in more FNs.Next, from the three tables, we can also observe that for each expression, the TP count is high for the combination of KCM with GDEC while it is low for the combination of Prewitt with GDEC. For example consider the expression, contempt, the total TPs for KCM with GDEC are 148 while for RCM with GED and Prewitt with GEDC it is observed as 140 and 136 respectively. From all these values, we compute the F-score which resembles the expression level detection performance. A higher value of F-score denotes the better detection capability while the lower value of F-score denotes the poor detection capability. For the 7-class expression, the obtained F-score at three different simulations is shown in Figure.9 and for 6-class expression simulation, it is shown in Figure.10. From these two figures, we can observe that the maximum F-score is obtained for 6-class expression simulation. As there exists less number of expression in 6-classexpression simulation, the recognition system would have a clear knowledge about the expression features and it recognizes every expression accurately.

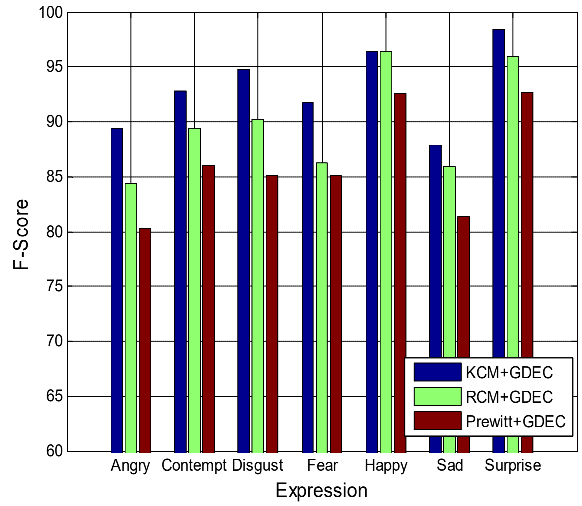

Figure.9 F-score for 7-class expression simulation

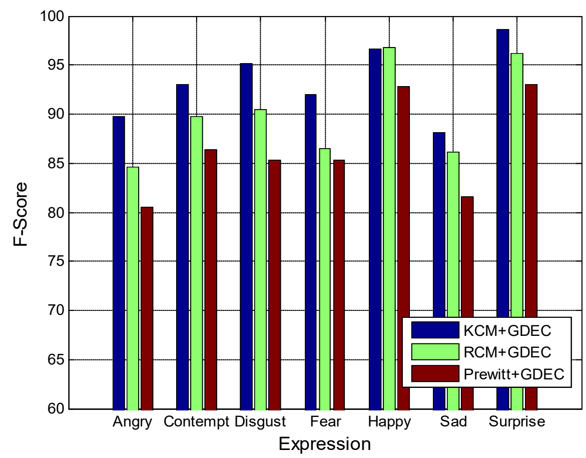

Figure10 F-score for 6-class expression simulation

As the number of expressions increases, the similar features also increases which makes the system confused and results in the higher false positives. For 7-class expression, the average F-score for the combination KCM with GDEC is observed as 93.0848% while for remaining combinations, it is observed as 89.8110% and 86.1579% for RCM with GEDC and Prewitt with GEDC respectively. Similarly,for 6-class expression, the average F-score for the combination KCM with GDEC is observed as 95.8963% while for remaining combinations, it is observed as 91.6223% and 89.3525% for RCM with GEDC and Prewitt with GEDC respectively.

Figure.11 shows the overall recognition accuracy at two simulation studies with varying combinations. From the results, the maximum accuracy is observed for KCM with GDEC and minimum accuracy is observed for Prewitt with GDEC. The average accuracy of KCM with GDEC for 7-class expression is observed as 93.4520% while for 6-class expression simulation, it is observed as 95.8214%, hence the average accuracy is 94.6367%. For the remaining combination such as RCM with GDEC and Prewitt with GDEC, the average accuracy is observed as 90.4575% and 89.2235%. From the results, we can see that the accuracy of RCM and Prewitt is nearly same as the edge responses are almost same for these two filters. The edge filter coefficients of KCM are larger than the RCM, they boosted the edge features much effectively such that the key components of facial expression are highlighted.

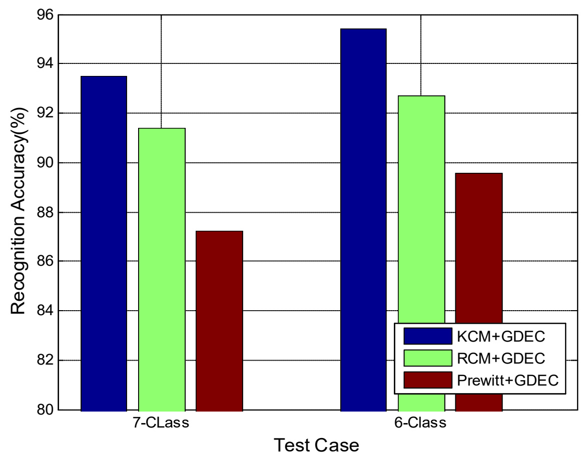

Figure.11 Recognition Accuracy at different simulations

C. Comparison and Discussion

The comparison is carried out here to explore the significance of proposed contribution in the improvisation of recognition accuracy in FER system. LBP [22] is the simple and effective method which can represent the facial features based on their textures. However, due to the comparison with center pixel's intensity, it lasts much information. Moreover, they don't explore even a small hint about the movements of facial expressions. LDP [15] and LDN [20, 29] concentrated on the inclusion of directional movements of facial expressions, and encoded each pixel with the corresponding gradient magnitude and direction. However, they didn't highlight the difference between noisy and flat pixels. As they are similar in pixel intensities, they can be encoded with same code words and creates an ambiguity for recognition system. Some authors employed transformations before subjecting the image for encoding. M. Guoet al. [27] applied KL Transform which breaks the correlation property of image thereby results in improper discrimination between expressions. Similarly, Sumeet Saurav et al. [31] employed PCA which is a dimensionality reduction technique that discards most of the information form facial image. Les information can't make the recognition system robust, means it become subject specific or environment specific or gender specific. Such kind of system can't be called as robust. Unlike all these methods, [25] and [26] encoded the each pixel of facial image through their gradients after processing them through different compass masks. However, they encoded each pixel with a 8-bit code word which has slightly more computational complexity. Further, they didn't discriminate between edges, corners, and curves etc. which are main components in the FER. Unlike all the above methods, the proposed GEDC encode the directions of edges passing through center pixel thereby it can encode almost all texture primitives such as corners, edges and lines. The GEDC considered each facial expression as different set of edges and proposed an optimal encoding technique for every primitive. Moreover, the GEDC is a 7-bit code and reduces the storage burden as well as computational complexity and both training and testing. Hence, it had gained superior recognition accuracy than all the state-of-the-art methods.

Table.4 Accuracy Comparison between proposed and existing methods

Reference	**Method**	**Accuracy (%)**	**Database**
N. T. Cao et al. [22]	LBP and SVM	81.3000	CK+
T. Jabid et al. [15]	LDP	92.7552	CK+
A. R. Rivera et al. [20]	LDN with SVM	86.5500	CK+
M. Guo et al. [27]	Extended LBP with KL transform	92.7552	CK+
I. M. Revina, and W.R. Sam Emmanuel [29]	LDN with DGLTP	88.6300	CK+
S. Sammaiah, and K. V. Rao [26]	EALDBP with SVM	92.5922	CK+
A. Vijaya Lakshmi, P.	GCBC with	92.6971	CK+

Mohanaiah [25]	SVM		
Sumeet Saurav et al. [31]	PCA with DTLP and KELM	66.1000	CK+
Proposed	Prewitt with GEDC and SVM	89.2235	CK+
	RCM with GEDC and SVM	90.4575	
	KCM with GEDC and SVM	**94.6367**	

V. CONCLUSION

FER system is mainly dependent on the face expression descriptor which can provide sufficient knowledge about the expression as well as discriminates between different expressions. Towards such intention, this paper proposed a new face texture descriptor called as GEDC which encodes the directions of facial movements. The GEDC is a 7-bit code word constructed based on the direction of two neighbor pixels those have maximum support with the target pixel. To identify direction, a sign bit is added and made it as 7-bit code word. As a preprocessing, the proposed method applied different edge detection filters like KCM, RCM and Prewitt. Each edge responses is encoded with GEDC and processed through SVM for classification. The obtained results on CK+ dataset shows that the KCM can encode the expression much effectively as it gained better accuracy compared to other filters. Moreover, the comparison of proposed method with other state-of-the art methods had proven its outstanding performance.

References

[1] R. Cowie, E, D. Cowie, N. Tsapatsoulis, G. Votsis, S. Kollias, W. Fellenz, J. G. Taylor, “Emotional expression recognition in human–computer interaction”, *IEEE Signal Process. Mag.,* Vol.18, pp.32-80, 2001.

[2] J. Kumaria, R. Rajesha, K. M. Poojaa, “Facial expression recognition: A survey”, In: *Proc. ofSecond International Symposium on Computer Vision and the Internet*, pp.486 – 491, 2015.

[3] E.B. McClure, K. Pope, A.J. Hoberman, D.S. Pine, E. Leibenluft, “Facial expression recognition in adolescents with mood and anxiety disorders”, *Am. J. Psychiatry*, Vol. 160, pp. 1172–1174, 2003.

[4] S. Wallace, M. Coleman, A. Bailey, “An investigation of basic facial expression recognition in autism spectrum disorders”, *Cogn. Emot.*, Vol. 22, pp. 1353–1380, 2008.

[5] Z. Zeng, M. Pantic, G.I. Roisman, T. S. Huang, “A srvey of affect recognition methods: Audio, visual, and spontaneous expressions”. *IEEE Trans. Pattern Anal. Mach. Intell.*, Vol. 31, pp. 39–58, 2009.

[6] N. Wang, X. Gao, D. Tao, H. Yang, X. Li, “Facial feature point detection: A comprehensive survey”, *Neurocomputing*, Vol. 275, pp. 50–65, 2018.

[7] B. Fasel, and J. Luettin, “Automatic facial expression analysis: a survey”, *Pattern Recognition*, Vol.36, No.1, pp.259-275, 2003.

[8] N. G. Bourbakis and P. Kakumanu, “Skin-based face detection-extraction and recognition of facial expressions,” In *Applied Pattern Recognition*, New York: Springer-Verlag, pp. 3–27, 2008.

[9] A. Cheddad, D. Mohamad, and A. A. Manaf, “Exploiting Voronoi diagram properties in face segmentation and feature extraction,” *Pattern Recognit.*, Vol. 41, No. 12, pp. 3842–3859, 2008.

[10] X. Xie and K. M. Lam, “Facial expression recognition based on shape and texture,” *Pattern Recognition*, Vol. 42, No. 5, pp. 1003–1011, 2009.

[11] M. Turk and A. Pentland, “Eigenfaces for recognition,” *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[12] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, “Eigenfacesvs. fisherfaces: Recognition using class specific linear projection,”*IEEE Transactions on pattern analysis and machine intelligence*, vol. 19,no. 7, pp. 711–720, 1997.

[13] J. Yang, D. Zhang, A. F. Frangi, and J. y. Yang, “Two-dimensionalPCA: a new approach to appearance-based face representationand recognition,” *IEEE transactions on pattern analysis and machine intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[14] K. Etemad and R. Chellappa, “Discriminant analysis for recognitionof human face images,” *JOSA A*, vol. 14, no. 8, pp. 1724–1733,1997.

[15] T. Jabid, M. H. Kabir, and O. Chae, “Robust facial expression recognitionbased on local directional pattern,” *ETRI journal*, vol. 32,no. 5, pp. 784–794, 2010.

[16] M. T. B. Iqbal, B. Ryu, G. Song, J. Kim, F. Makhmudkhujaev, andO. Chae, “Exploring positional ternary pattern (PTP) for conventionalfacial expression recognition from static images,” pp. 853–855, 2016.

[17] A. R. Rivera, J. A. R. Castillo, and O. Chae, “Recognition offace expressions using local principal texture pattern,” *19th IEEE International Conference in Image Processing (ICIP*), 2012, pp. 2609–2612.
[18] H. Zhou, R. Wang, and C. Wang, “A novel extended local-binarypatternoperator for texture analysis,” *Information Sciences*, vol. 178,no. 22, pp. 4314–4325, 2008.
[19] M. M. Goyani, and N. Patel, “Recognition of facial expressions using local mean binary pattern”, *Electronic Letters on Computer Vision and Image Analysis,* Vol.16, No. 1, pp. 54–67, 2017.
[20] A. R. Rivera, J. R. Castillo, and O. O. Chae, “Local directional number pattern for face analysis: Face and expression recognition,” *IEEE transactions on image processing*, vol. 22, no. 5, pp. 1740–1752, 2013.
[21] Md Tauhid Bin Iqbal, M. Abdullah-Al-Wadud, Byungyong Ryu, Farkhod Makhmudkhujaev, and Oksam Chae, “Facial Expression Recognition withNeighborhood-aware Edge Directional Pattern(NEDP*)”, Journal Of Latex Class Files*, Vol. 14, No. 8, August 2015.
[22] N. T. Cao, A. H. Ton-hat and H. I Choi, “Facial Expression Recognition Based on Local Binary Pattern Features and Support Vector Machine”, *International Journal of Pattern Recognition and Artificial Intelligence*, Vol. 28, No. 06, 1456012, 2014.
[23] N. Kauser and J. Sharma, “Facial expression recognition using LBP template of facial parts and multilayer neural network”, In: *Proc. ofInternational Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC*), Palladam, India, pp.445-449, 2017.
[24] Min Hu, Yaqin Zheng, Chunjian Yang, Xiaohua Wang,Lei He, and Fuji Ren, “Facial Expression Recognition Using FusionFeatures Based on Center-SymmetricLocal Octonary Pattern”, *IEEE Access*, Volume 7, 2019, pp. 29882-29890.
[25] A. Vijaya Lakshmi, P. Mohanaiah, “Gradient-Based Compact Binary Coding for Facial Expression Recognition”, *International Journal of Intelligent Engineering and Systems*, Vol.14, No.6, pp.377-390, 2021.
[26] S. Sammaiah, and K. V. Rao, “A Salient Binary Coding Scheme for Face and Expression Recognition from Facial Images, *International Journal of Intelligent Engineering and Systems*, Vol.14, No.2, pp.67-80, 2021.
[27] M. Guo, X. Hou, Y. Ma, “Facial expression recognition using ELBP based on covariance matrix transform in KLT”, *Multimedia Tools and Applications,* Vol. 76, pp. 2995–3010, 2017.
[28] M. Abdul, R.S. Holambe, “Local binary patterns based on directional wavelet transform for expression and pose invariant face recognition”, *Applied Computations And Informatics*, Vol. 15, No.2, pp. 163-171, 2017.
[29] I. M. Revina, and W.R. Sam Emmanuel, “Face expression recognition using LDN and Dominant Gradient Local Ternary Pattern descriptors”, *Journal of King Saud University – Computer and Information Sciences*, pp.1-7, 2018.
[30] X. Tan and B. Triggs, “Enhanced local texture feature sets for face recognition under difficult lighting conditions,” *IEEE Transactions on Image Processing*, Vol. 19, No. 6, pp. 1635–1650, 2010.

[31] Sumeet Saurav, Ravi Saini, and Sanjay Singh, “Facial Expression Recognition Using DynamicLocal Ternary Patterns With Kernel ExtremeLearning Machine Classifier”, *IEEE Access,* Volume 9, 2021, pp. 120844-120868.
[32] S. Parveen, S. Ahmad, N. Abbas, W. Adnan, M. Hana, and N. Naeem,``Face liveness detection using dynamic local ternary pattern (DLTP),"*Computers*, vol. 5, no. 2, p. 10, May 2016.
[33] B. Schölkopf, A. Smola, and K.-R. Müller, ``Kernel principal component analysis," in *Proc. Int. Conf. Artif. Neural Netw.* Berlin, Germany: Springer, 1997, pp. 583_588
[34] A. Iosifidis, A. Tefas, and I. Pitas, ``On the kernel extreme learning machine classifier," *Pattern Recognit. Lett.*, vol. 54, pp. 11-17, Mar. 2015
[35] Min Hu, Chunjian Yang, Yaqin Zheng, Xiaohua Wang, Lei He, and Fuji Ren, “Facial Expression Recognition Based onFusion Features of Center-SymmetricLocal Signal Magnitude Pattern”, *IEEE Access*, Volume 7, 2019, pp. 118435- 118445.
[36] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, ``Coding facialexpressions with Gabor wavelets," in *Proc. 3rd IEEE Int. Conf. Autom.Face Gesture Recognit.*, Apr. 1998, pp. 200-205.
[37] H.Zhang, Alireza Jolfaei, and Mamoun Alazab, “A Face Emotion Recognition Method UsingConvolutional Neural Networkand Image Edge computing”, *IEEE Access*, Volume 7, 2019, pp.159081-159089.
[38] Naigong Yu and Deguo Bai, “Facial Expression Recognition by Jointly PartialImage and Deep Metric Learning”, *IEEE Access*, Volume 8, 2020, pp.4700-4707.
[39] chang li, kaoru hirota, junjie ma,zhiyang jia and yaping dai, “Facial Expression Recognition Using Hybrid
Features of Pixel and Geometry”, *IEEE Access*, Volume 9, 2021, pp. 18876- 18889.
[40] Jun Liu, Yanjun Feng, And Hongxia Wang, “Facial Expression Recognition Using Pose-GuidedFace Alignment and Discriminative FeaturesBased on Deep Learning”, *IEEE Access*, Volume 9, 2021, pp. 69267- 69277.
[41] He, Kaiming; Zhang, Xiangyu; Ren, Shaoqing; Sun, Jian (2016). *Deep Residual Learning for Image Recognition*. *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. Las Vegas, NV, USA: IEEE. pp. 770–778.
[42] Karen Simonyan, Andrew Zisserman, “Very Deep Convolutional Networks for Large-Scale Image Recognition”, arXiv:1409.1556.
[43] Balochin, S., Baloochian, H., “Edge Detection on noisy image using Prewitt operator and Fractional Order Differentiation”, Multimdia tools and Applications, 81, 9759-9770, 2022.
[44]A.R. Venmathi, E.N. Ganesh and N. Kumaratharan, “Kirsch Compass Kernel Edge Detection Algorithm for Micro Calcification Clusters in Mammograms*”, Middle-East Journal of Scientific R*esearch 24 (4): 1530-1535, 2016.
[45] Babu C.R., Sreenivasa Reddy E., Prabhakara Rao B. (2015) Age Group Classification of Facial Images Using Rank Based Edge Texture Unit (RETU). In: Mandal J., Satapathy S., Kumar Sanyal M., Sarkar P., Mukhopadhyay A. (eds) Information Systems Design and

Intelligent Applications. Advances in Intelligent Systems and Computing, Vol. 340. Springer, New Delhi.

[46] L. Shapiro, Computer vision and image processing. Academic Press,1992.

[47] P. Lucey, J. 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: *Proc. of IEEE Computer Society Conference on Computer Vision and Pattern Recognition. Workshops*, San Francisco, CA, USA, pp. 94–101, 2010.