Int. J. Nonlinear Anal. Appl. 13 (2022) 1, 1067-1079
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.5648

Face and facial expression recognition using local directional feature structure

Vidyarani H.J.[a], Shrishail Math[b]

[a]Deptartment of ISE, Dr. Ambedkar Institute of Technology, Bengaluru, India
[b]Deptartment of CSE, Shri Krishna Institute of Technology, Bengaluru, India

(Communicated by Madjid Eshaghi Gordji)

Abstract

The face expression recognition is used in countless application areas for example security, computer vision and medical science etc. The facial expressions are used to communicate in a non-verbal way (i.e., using eye contact, facial expressions etc.). Emotions play an important role in facial expression recognition which helps to identify what an individual is feeling. Various AI research in the field of facial recognition system is being carried since a decade. Many of the machine learning algorithms are also being used to identify the facial expression which helps them to train and test using the facial expression to get a correct output of the given expression. This paper presents a new facial expression recognition system, local directional feature structure (LDFS). LDFS uses different features of the face (i.e., eyebrows, nose, mouth, eyes). The face is detected and aligned using the edge detection. The task of the edge detection is to detect the face, face alignment and position variations of the face. The edge detection extracts the specific features for the identification of the emotions. Two types of datasets have been used for the qualitative and quantitative experiments on the face expression mainly the CK+ and Jaffe dataset. This approach for our model shows an improvement when compared to the existing system.

Keywords: Face and facial expression, local directional.

1. Introduction

The automatic identification of the emotion plays an essential part in the region of machine vision system since its growth in the market. One of the methods to identify or recognize the emotions automatically is the method of facial expression recognition. The field of FER has fascinated

Email addresses: hjvidyarani@gmail.com (Vidyarani H.J.), shri_math@yahoo.com (Shrishail Math)

Received: May 2021 *Accepted:* June 2021

incredible attention since the last few years and is popular in the widespread applications for example video-communication, e-learning, e-health, behaviour analysis, human machine interaction. The aim of facial expression is to recognize how the person is feeling from a single image. As there are many studies currently being performed in this field, the computer vision and Machine Learning Algorithms are being used to enhance the model. Facial expression can help an individual to know how another individual may feel. There are many examples and fields where the face recognition is being used in real-world. Some of the examples for facial recognition would be the face unlock in the mobile phones. Another example would be of the driver drowsiness. In the driver drowsiness the facial expression recognition plays a vital role. The drowsiness can be recognized by the facial recognition which recognizes the emotions of the driver and alerts the driver to stop the car or it is slowed down by some the sensors.

Human-beings are the only species who can distinguish an expression, but with the advancement in the technology even machines can learn how to detect an emotion as well. Different factors are taken when the facial recognition is currently executed. Some of the factors include motion and positions of the facial muscles situated below the skin and movement of those is taken as an input for some models. Another factor may be taking a set of features such as the movement of the mouth to identify whether the person is happy or sad. Different models have been created for the facial expression to identify certain emotions. This paper uses the seven general emotions: Happy, Sad, Surprise, Neutral, Angry, Fear and Disgust to identify what a person is feeling using a set of images. The paper introduces to a model on local directional feature structure (LDFS) for face expressions where the different structures of face are taken to give an output of one of the seven general emotions. The research significance of proposed model is given below:

- We introduce a new face expression recognition in which the edge detection is used which helps to extract the main features of the face and discards the meaningless ones from the facial image.
- We propose a new method for the coding scheme that improves the classification results.
- Further, LDFS is trained using support vector machine for automatic detection of face and facial expression with good classification accuracies.

2. Literature Survey

Facial Recognition is currently being performed in the RGB photos [6]. A dataset of huge size of recordings having different expressions, facial elements, 3D pose variations and characters trained under the CNN model performs well in identifying the basic emotions. A technique called FERC [8], trained under the CNN first removes the background of the image and the extracts the facial expressions. The removal of the background helps to avoid multiple problems such as distance from the camera, faces of another person coming in the way and other factors. A two-stage technique can be used for identifying the facial expressions [10] in a given set of images. Initial step is to extract the face area and then check the face area using the different kinds of emotions. An improved version of Local Motion Pattern [1] which checks the face deformations and face skin temporal elasticity. Then it uses the method of micro and macro expressions recognition which recognizes more features from the face. After which the small changes in the head pose or any other variations are identified.

A new feature, Local Shape Pattern (LSP) [5] has been introduced where the structure of the nearest pixel based on the prominent directional information is taken to identify the local distortion and noise of the image. A Generative Adversarial Network [19] deep learning model has been proposed where the facial expression recognition and facial image synthesis factors have been taken for modelling which increase the model's performance of the facial expression recognition. Similar images of the faces are not used in the model which makes their proposed system more flexible for any given dataset. In this model, the facial expressions are encoded in separate ways which assists to combine the captured images of the face having random expressions which in turn helps to identify a certain feature in the facial expression.

A model (IE-DBN) [17], extracts identity, features that are affecting the expressions in the image and then aggregates using the bilinear aggregation which improves the intra-class similarities and inter-class variations. Another similar model of GAN with an improved version [13], (LPG-GAN) has been introduced where the model extracts and creates the structure for the critical face areas. It uses the local system for the extraction of the different features from the critical region of the face and also a global system to study the entire data of the face. It combines the results of the local network and global network to give the final output of the facial expression. Similarly, another model Two-branch Disentangled Generative Adversarial Network (TDGAN) has been proposed [14], which uses the specific face expression attributes to identify an expression from the face given as an input. Facial Expression recognition can be captured using the different head poses [18]. A review on different type of deep facial expression recognition has been done [7], where the different datasets, data selection process and estimation of the dataset has been explained. The review has a background knowledge and suggestions for the execution of the facial expression recognition. A dataset (FENP) [16], which has the face expression of different categories i.e., crying expression, severe pain expression, calmness expression, and mild pain expression which can be used for medical research FER.

The model (DSAN) [4], takes the information such as race, gender, age-related which helps to recognize the face expressions of a given person more easily. It uses multiple convolutional layers to get a particular precision for the model. From survey it can be seen facial expression detection is a challenging task; effective facial expression feature representation model is required to capture eliminate outlier and retain good spatial information. In next section this paper present an effective face and expression detection mechanism considering above challenges.

3. Face and Facial Expression Recognition Using Local Directional Feature Structure

Facial expressions are the most natural, powerful and quick means for an individual to express their feelings, emotions and intentions. The skill to evaluate the facial expression is an important factor in non-verbal communication. If an individual only evaluates what another individual is speaking and does not concentrate on how the other individual face says, then there can only be a part of the story. Human-beings are the only species who can differentiate between expressions and emotions. With the advancement in the technology the computers can learn how to identify these emotions as well. There are mainly seven general human expressions which are used for classifying the emotions: Happy, Sad, Surprise, Neutral, Angry, Fear and Disgust. These seven factors help us to detect the given expression on the face of any individual.

The Proposed model Local Directional Feature Structure (LDFS) contains an eight-bit structure code which is allocated to respective picture element of the captured face image. In face detection method, as the expressions in the face change the facial features also change which is the most significant factor than the whole captured image, as the borders of the face expression have more features and high edge magnitude. Hence, the technique of edge detection is used as an edge operator to check the calculation of the edge responses efficiently. The edge detection method mainly takes two directions at each picture element to show the local edge outlines. The proposed technique, differentiates the required directional structure and removes the not required structure of the image using the edge magnitude The Local Directional Feature Structure (LDFS) has some properties which help us for the edge detection. The properties are discussed below:

- The gradient direction is used for the identification of the face for emotion-related features
- The Edge detection is effective because of its symmetry
- The ternary structure codes the edge sign data, and separates the regions which are smooth and near the edges which solves the problem of the edge structure in the smooth regions

In this method, the LDFS codes are generated using the edge responses from the captured image and the edge detection $\{N_0, ..., N_3\}$ is calculated, then the coding of both the direction i.e., primary and secondary directions and their matching feature structure are coded correspondingly. The Edge detection method generates the same degree of responses with all the different kind of signs in different directions using the four masks i.e., N_0 to N_3 to search the main direction, which helps to reduce the time for the calculation of the main direction. This technique also helps to effectively represent the symmetrical facial features. The representation of the symmetric facial feature is done using the calculated directional code and the information of their sign to create a feature structure. This feature structure then gives the information of the direction using three edge response conditions which can be positive strong edge response or negative strong edge response or either it can be weak edge response. As the symmetrical mask is used and the information of the sign and the degree is in the feature structure, which helps to differentiate the edge regions and smooth regions, the edge direction $\{N_0, ..., N_3\}$ is coded using only half of the edge detection. To code the primary and secondary directional number 2 bits are assigned to both of these directions. Each of the directional number contains 2 bits to the respectively feature structure. The Edge detection method is used on the entire captured image which creates a set of response degree which are linked to the four directions, using the equation:

$$S_i = N_i * I, 0 \leq i \leq 3 \tag{3.1}$$

where I defines the actual captured information from the image, N_i represent i^{th} edge detection, and S_i defines i^{th} response information from the image. After this, this work calculates j^{th} maximal total significance E_j of entire four edge detection responses, which is defined by:

$$E_j(x, y) = \arg \max_i^j \{\mid S_i(x, y) \mid : 0 \leq i \leq 3\} \tag{3.2}$$

where $\arg\max_i^j$ defines a function for returning the identifier (i.e., key) i of the j^{th} maximal significance in the features collected. Further this work need to find out direction for the 1st and 2nd that is $j \in \{1, 2\}$, the direction response is converted to the feature structure correspondingly. The process codes the edge response by means of three stages i.e., equal, positive and negative and also gives the edge response sign information. Therefore, the feature structure shows that is the direction in edge region or in the smooth region. The encoding is done using:

$$U_j(x, y) = \begin{cases} 2 & if \ S_{\hat{\imath}}(x, y) < -\sigma, \\ 1 & if \ S_{\hat{\imath}}(x, y) < \sigma, \\ 0 & if \ -\sigma \leq S_{\hat{\imath}}(x, y) \leq \sigma \end{cases} \tag{3.3}$$

where U_j is the feature structure of the degree of j^{th} direction at the location (x, y), $S_{\hat{\imath}}(x, y)$ defines the response of the edge of i^{th} direction considering location (x, y), $i = E_j(x, y)$ represent j^{th} primary direction at location (x, y), and σ is a parameter used for defining threshold parameter. The data is divided according to the threshold in three sections, lower, upper and in between. Here the lower means strong negative edge response, upper means strong positive edge response and in between means a weak edge response. From these different threshold values, it is very easy to differentiate and separate the edge response data and remove the directional data from the smooth regions. Using the value of U_j, the minimum total value E_j is kept or discarded. Therefore, for all the directions of $E_j(x, y)$ the different rules are used. Suppose the feature structure of the first direction gives the result as 0, then it means the picture element (x, y) is in the smooth region. So, from this step we check the codes from which the codes of the LDFS are produced in the captured image. Similarly, if the feature structure of the second direction gives the value as 0, then the value of the first direction is considered. Therefore, the code is formed by grouping the values of the directions and the two feature structures. The process of the grouping can be showed using the following equation:

$$LDFS(x, y) = 2^6 E_1(x, y) + 2^4 U_1(x, y) + 2^2 E_2(x, y) + U_2(x, y) \tag{3.4}$$

The $LDFS(x, y)$ is the coding value for each picture element (x, y) with respect to captured image, E_1 and E_2 defines respective directional identifier of the major and minor directions (from 0 to 3) from the two maximum mask responses of corresponding neighbourhood of the pixel (x, y), and U_1 represent the 1st feature structure direction of E_1 and U_2 represent the 1st feature structure direction of E_2. As our proposed method uses the edge degree and encodes feature extraction to remove the not required data, this makes our model different than the compared existing models which use only the code of the directions.

4. Face Description for LDFS

A. Issues of Face Description using standard histogram model

Different techniques have been in use for the extraction of the face descriptions using the histograms. Using the histogram method, we are dividing the region of the face image into different areas, $\{S1, ..., SN\}$, and a histogram H^k of each area R^k is composed of number of windows using its unique image intensities. Therefore, face description model are formed using the different areas applying following equation

$$H^k(c) = \sum_{(x,y)\in R^k} \delta(P(x, y), c),\ \forall c, \tag{4.1}$$

$$\delta(a, b) = \begin{cases} 1 & a = b, \\ 0 & a \neq b \end{cases} \tag{4.2}$$

where $P(x, y)$ is a calculated structure code in (x, y), c is the structure code, (x, y) is a picture element location with the segment R^k. The histogram model is simple and robust, but it requires more sufficient codes for the calculation and it sometimes loses the data inside the region even if the regions contain the positional data. When the captured picture is partitioned into many parts then the spatial data of the image increases but the sample codes of the different other regions show a decrease because of the small dimension of that region. This is the problem due to which the histogram has limitation to expand the necessary spatial information.

B. Active Structures

In the histogram model, the data of the image is extracted using the 2D grid. This strategy is not much efficient as it gives only importance for the facial features mapped data. The mapped data allocates special features to the emotion related facial expression of the captured image. These are called as LDFS codes active. From the dataset (CK+), we have extracted the facial features through which we select the region where the emotion changes the facial expressions. We extracted 42 emotion related facial expressions using the different features of the face such as eyes, nose, mouth to select the active structure. For the extraction of the most often occurring patterns from the histograms, we use the following equation:

$$d_n^r = \arg\max_n^c\{\mid \mathbb{H}^r(c) \mid: c \in \mathbb{LDFS}\} \tag{4.3}$$

Where d_n^r is group of respective nth maximal c behavior (codes) with respect to corresponding r^{th} segment (i.e.,$r \in \{eyebrows, eyes, upper\ nose, mouth\}$), $\arg\max_n^c$ defines a function that retrieves n^{th} maximal significance c behavior within histogram $\mathbb{H}^r$ of corresponding r^{th} region, and $\mathbb{LDFS}$ defines useful LDFS behaviors. Finally, this work combine d_n^r into a distinct structure applying following equation

$$\mathbb{D}_n = \bigcup_{r=0}^{R} d_n^r \tag{4.4}$$

where $\mathbb{D}_n$ represent distinct active structure which comprises of nth best behavior among different segment. Theoretically, these patterns are represented by $\mathbb{D}_n$ without changing the original codes. To secure the stability of the codes, the codes are extracted using

$$\mathbb{S}_n = \arg\max_c^j\{\mid \mathbb{H}^m(c) \mid: c \in \mathbb{LDFS}, c \notin \mathbb{D}_n\} \tag{4.5}$$

Where $mathbbS_n$ defines jth maximal behavior with respect to average histogram $\mathbb{H}^m$, as expressed below

$$\mathbb{H}^m(i) = \frac{1}{R}\sum_{r=0}^{R-1} \mathbb{H}^r(i), \forall i \tag{4.6}$$

where R defines total segments considered, without containing $\mathbb{D}_n$, j defines the total patterns considered for selection, and is expressed as follows

$$j = R_n - \mid \mathbb{D}_n \mid \tag{4.7}$$

where $\mid \mathbb{D}_n \mid$ defines the total size of elements of the set and R is the total size of regions, in this case $R = 4$. The concluding set of active structure $\mathbb{D}_n$ is given using following equation

$$D_n = \mathbb{D}_n \cup \mathbb{S}_n \tag{4.8}$$

From n, we acquire the different active structure.

C. Face Description for LDFS

In the model, the Local Directional Feature Structure (LDFS) of each region is generated. As the spatial data of the active LDFS codes is more important for the recognition of the facial expression, the local-regions is split into different sub-regions so that each region has a different label. By combining the positional label, more mapped data is added to the active LDFS

$$LDFS_{N\ M}^n(x,y) = \begin{cases} 2^8 l_{x,y} + c_{x,y}, & c_{x,y} \in D_n \\ c_{x,y}, & c_{x,y} \notin D_n \end{cases} \tag{4.9}$$

Figure 1: (a) Angry (b) Disgust (c) Happy (d) Neutral (e) Sad

where $LDFS^n_{N\ M}(x,y)$ is the value or pattern of every picture element (x,y) in the segment which is separated into $N \times M$ sub-segments with definite active LDFS codes D_n, $l_{x,y}$ defines the identifier of $N \times M$ sub-segments identifier from 0 to $(NM-1)$, $c_{x,y}$ defines LDFS patterns with respect to (x,y), and D_n is determined. If the sub-region is not divided or D_n contain a null set, then $LDFS^n_{N\ M}$ is similar to the LDFS. Then, the histogram H^k and $LDFS^n_{N\ M}$ is created.
From the sub-regions and the active structure, it can be shown that the mapped data is more efficient when compared to the existing histogram models. Using this method, the mapped data can be allocated to the active structure using the sub-regions as the active structure are more complex to the position data. The active structure are robust for the calculation of the error as the have more accuracy. Therefore, this approach is more efficient and uses spatial information to compare with existing histogram-based model.
Finally, spatial region aware global histogram ($SRAGH$) is calculated by grouping all the histograms

$$SRAGH = \prod_{k=1}^{k} H^k \tag{4.10}$$

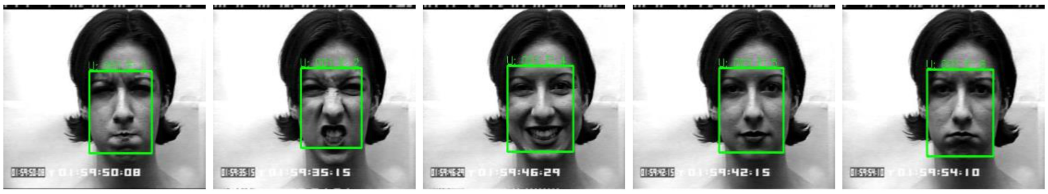
where k represent histogram merging function, K is the number of segments considering which every facial image is segmented, and H^k is the histogram computed using $LDFS^n_{N\ M}$ instead. The $SRAGH$ function is used for the feature vector as it represents the image of the face in the FER. The final feature extracted is trained using Support vector machine similar to model presented in [11]. The LDFF-SVM based classification model can be used for performing both binary and as well as multi-label classification.

5. Results and Discussions

Here experiment is conducted for validating performance and accuracy achieved using Local directional feature structure (LDFS)-based facial and expression detection method over existing face and expression detection methods. Different datasets have been used for the testing purpose.

a. CK+ Dataset

This dataset has 593 video successions having 123 distinct subjects. These subjects range from 18 to 50 years having various individuals. All the video sequences or successions have a change in the facial expressions which have been taped at 30 FPS. The dataset contains seven expression classes: Happy, Sad, Surprise, Neutral, Angry, Fear and Disgust. The dataset has been tested and the results for the given testing are given below in the Fig. 1, Fig. 2 and Fig. 3. The Fig. 1 shows the expressions of the individual i.e., angry, disgust, happy, neutral and sad. The Fig. 2 shows the extraction of the FE features and the FE is given as E in the following figure. Then finally in the Fig. 3 shows the clear extracted data of the facial expression.



(a) (b) (c) (d) (e)

Figure 2: Extraction of data from (a), (b), (c), (d), (e)

(a) (b) (c) (d) (e)

Figure 3: Extracted data of (a), (b), (c), (d), (e)

The Table 1 shows the accuracy of the existing system vs the proposed system. The DCNN, ExpNet, AFER, LMP, sLSP+LB, LDTP have the accuracy 96.25%, 90.4%, 96.37%, 97.52%, 96.67%, 94.19% respectively. The accuracy of the proposed system has an accuracy of 98.55%

Table 2 shows the accuracy of the existing system compared to the proposed system of the CK+ dataset using the number of expression/class.

b. Jaffe Dataset

This dataset contains 213 different pictures having different face expressions of ten Japanese woman individuals. All the individuals have 7 expressions and all the picture were explained using the normal semantic evaluation on every expression by sixty individuals. The Jaffe dataset can be downloaded from the $https : //zenodo.org/record/3451524\#.YRtXy4gzZPY$. The seven facial expressions are Happy, Sad, Surprise, Neutral, Angry, Fear and Disgust. The Fig. 6 shows the images of the dataset. The Fig. 7 shows the extraction of the facial features from the dataset and the facial expresion of the face is given as E .In the Fig. 8 the extracted features of the facial expressions is shown.

The Fig. 9 shows the testing dataset of the JAFFE dataset. The figure has 7 facial expressions. The Fig. 10 shows the extraction process of the FE of the dataset, and the FE is given as E. In the Fig. 11 the results of the facial expression are obtained with a clear view of the facial expression.

Table 1: Accuracy of the existing system and the proposed system using the CK+ Dataset

Model	Accuracy
DCNN [6], 2020	96.25
ExpNet [3], 2018	90.4
AFER [15], 2018	96.37
LMP [1], 2019	97.52
sLSP+LB [5], 2020	96.67
LDTP [18], 2017	94.19
LDFS (Proposed System)	**98.55**

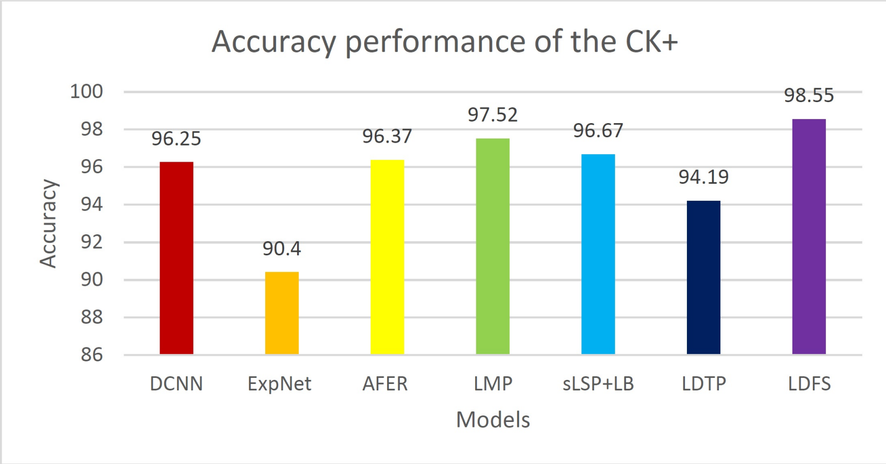


Figure 4: Accuracy performance of the existing system vs the proposed system

Table 2: Accuracy of the exisiting system vs proposed system of CK+ using the number of expressions

No of expression/class	DCNN [6], 2020	ExpNet [3], 2018	AFER [15], 2018	LMP [1], 2019	sLSP+LB [5], 2020	LDTP [12], 2017	**LDFS**
6	97.31	91.54	91.25	97.98	96.77	95.6	**99.8**
7	96.45	90.4	89.3	97.25	95.13	94.2	**98.94**
8	94.86	87.66	88.52	95.64	93.81	93.4	**98.55**

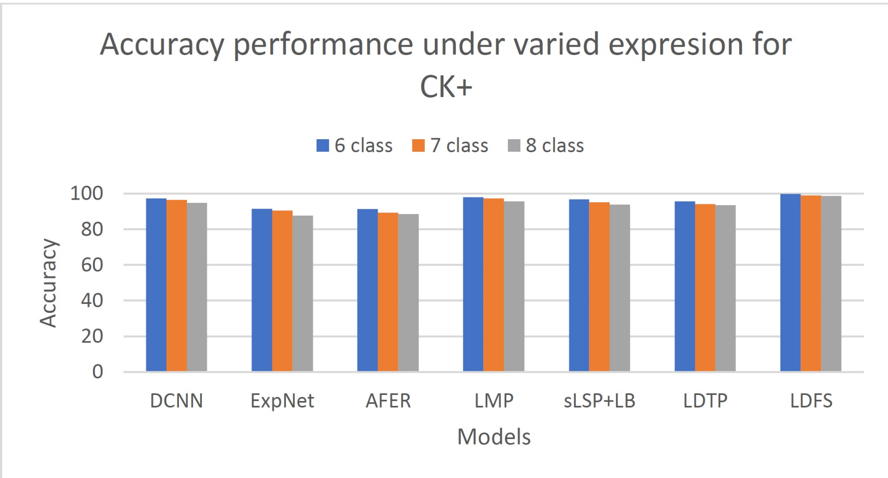
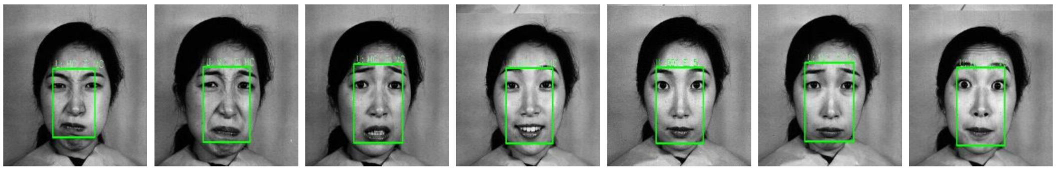


Figure 5: Accuracy performance of the exisiting system vs the proposed system in terms of number of expressions/class for the CK+

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

Figure 6: (a) Angry (b) Disgust (c) Fear (d) Happy (e) Neutral (f) Sad (g) Surprise



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

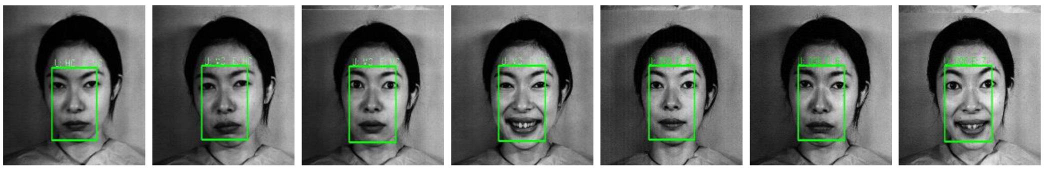
Figure 7: Extraction of data from (a), (b), (c), (d), (e), (f), (g)

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

Figure 8: Extraction data of (a), (b), (c), (d), (e), (f), (g)

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

Figure 9: (a) Angry (b) Disgust (c) Fear (d) Happy (e) Neutral (f) Sad (g) Surprise



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

Figure 10: Extraction of data from (a), (b), (c), (d), (e), (f), (g)

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

Figure 11: Extraction data of (a), (b), (c), (d), (e), (f), (g)

Table 3: Accuracy of the exisiting system vs proposed system of CK+ using the number of expressions

Model	Accuracy
AFER [15], 2018	95.43
HOG [9], 2017	92.75
LDTP [12], 2017	94.8
LDFS (Proposed System)	**98.98**

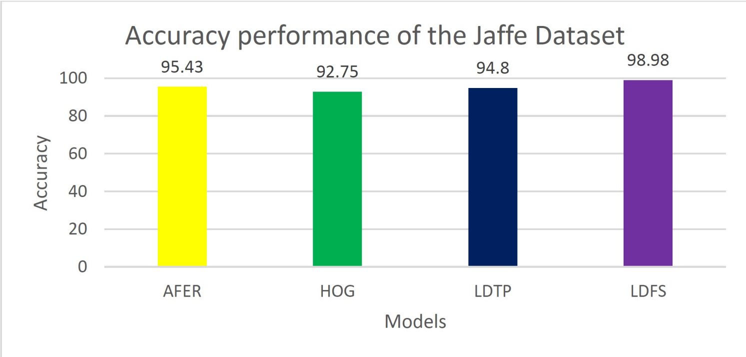


Figure 12: Accuracy performance of the Jaffe Dataset

The Table 3 shows the comparison results of the existing system vs the proposed system. Different models of the facial recognition system have been compared with the LDFS model to compare the accuracy. The accuracy of the LDFS model achieved the accuracy of 98.98%.

The Table 4 shows the comparison on the basis of the number of expression/class of the existing system vs the proposed model of the JAFFE Dataset.

6. Conclusion

Our paper proposes a new method through which the trough which the recognition or identification of the facial expression can be done. The Local Directional Feature Selection (LDFS) model uses the edge detection to detect the face and give an appropriate result. The edge detection helps to detect the edge, smooth regions and the noise sensitive areas. The spatial region aware global histogram helps to calculate the given result of the following expression and to attain a higher accuracy. In this paper we have compared the results with the existing systems and have reached to a conclusion that our model has a higher accuracy and better performance than the previous models. For the future work, different other datasets can be used to check the performance and accuracy of the system.

Table 4: Accuracy of the existing system vs proposed system of Jaffe Dataset using the number of expressions

No of expression/class	AFER [15], 2018	HOG [9], 2017	LDTP [12], 2017	**LDFS**
6	95.3	91.57	94.8	**99.24**
7	94.6	92.75	93.2	**98.98**

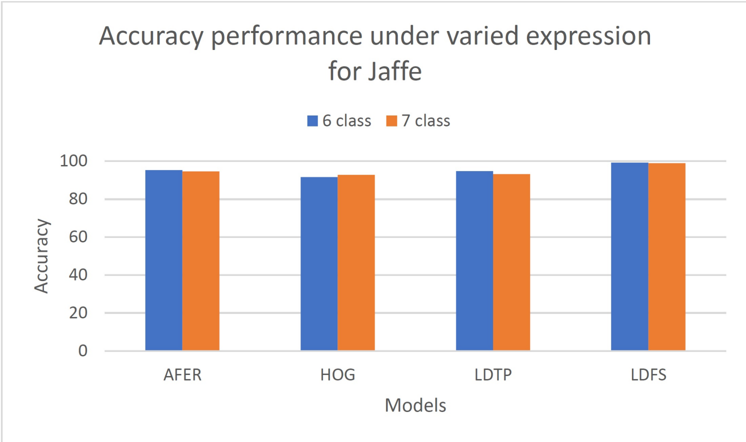


Figure 13: Accuracy performance of the existing system vs the proposed system in terms of number of expressions/class for the Jaffe Dataset

References

[1] B. Allaert, I.M. Bilasco and C. Djeraba, *Micro and macro facial expression recognition using advanced local motion patterns*, IEEE (2019) 1–12.

[2] X. Ben, Y. Ren, J. Zhang, S.-J. Wang, K. Kpalma, W. Meng and Y.-J. Liu, *Video-based facial micro-expression analysis: A survey of datasets, features and algorithms*, IEEE Trans. Pattern Anal. Mach. Intell. (2021).

[3] F.-J. Chang, A.T. Tran, T. Hassner, I. Masi, R. Nevatia and G. Medioni, *ExpNet: landmark-free, deep, 3D facial expressions*, Arxiv, (2018) 1–8.

[4] Y. Fan, V. Li and J.C.K. Lam, *Facial expression recognition with deeply-supervised attention network*, IEEE Trans. Affect. Comput. 2020 (2020).

[5] M.T.B. Iqbal, B. Ryu, A.R. Rivera, F. Makhmudkhujaev, O. Chae and S.-H. Bae, *Facial expression recognition with active local shape pattern and learned-size bloc representations*, IEEE Trans. Affect. Comput. (2020) 1–15.

[6] M.R. Koujan, L. Alharbawee, G. Giannakakis, N. Pegeault and A. Roussos, *Real-time facial expression recognition in the wild by disentangling 3D expression from identity*, Arxiv (2020) 1–8.

[7] S. Li and W. Deng, *Deep facial expression recognition: A survey*, IEEE Trans. Affect. Comput. (2020).

[8] N. Mehandale, *Facial emotion recognition using convolution neural networks*, Springer, SN Appl.Sci. (2020) 1–8.

[9] U. Mlakar, I. Fister, J. Brest and B. Potocnik, *Multi-objective differential evolution for feature selection in facial expression recognition systems*, Expert Syst. 89 (2017) 129–137.

[10] H.-D. Nguyen, S.-H. Kim, G.-S. Lee, H.-J. Yang, I.-S. Na and S.-H. Kim, *Facial expression recognition using a temporal ensemble of multi-level convolutional neural network*, IEEE Trans. Affect. Comput. (2019) 1–12.

[11] A.R. Rivera, J.R. Castillo, O. Chae, *Local directional number pattern for face analysis: face and expression recognition*, IEEE Trans. Image Proces. 22(5) (2013) 1740–1752.

[12] B. Ryu, A.R. Rivera, J. Kim, O. Chae, *Local directional ternary pattern for facial expresion recognition*, IEEE 26(12) (2017) 6006–6018.

[13] Y. Xia, W. Zheng, Y. Wang, H. Yu, J. Dong and F.-Y. Wang, *Local and global perception generative adversarial network for facial expression synthesis*, IEEE Trans. Circ. Syst. Video Tech. (2021).

[14] S. Xie, H. Hu and Y. Chen, *Facial expression recognition with two-branch disentangled generative adversarial network*, IEEE Trans. Circ. Syst. Video Tech. (2020) 2359–2371.

[15] Y. Yaddadena, M. Adda, A. Bouzouanea, S. Gabourya and B. Bouchard, *User action and facial expression recognition for error detection system in an ambient assisted environment*, Expert Syst. Appl. 112 (2018) 173–189.

[16] J. Yan, G. Lu, X. Li, W. Zheng, C. Huang, Z. Cui, Y. Zong, M. Chen, Qi. Hao, Y. Liu, J. Zhu and H. Li, *FENP: A database of neonatal facial expression for pain analysis*, IEEE Trans. Affect. Comput. 2020 (2020).

[17] H. Zhang, W. Su, J. Yu and Z. Wang, *Identity-expression dual branch network for facial expression recognition*, IEEE Trans. Cogn. Develop. Syst. 2020 (2020).

[18] F. Zhang, M. Xu and C. Xu, *Weakly-supervised facial expression recognition in the wild with noisy data*, IEEE Trans. Multimedia 2021 (2021).

[19] X. Zhang, F. Zhanga and C. Xu, *Joint expression synthesis and representation learning for facial expression recognition*, IEEE Trans. Circ. Syst. Video Tech. 2021 (2021).