Face Recognition based on Spatio Angular Using Visual Geometric Group-19 Convolutional Neural Network

[1]M. Tamilselvi , [2]S. Karthikeyan , [3]G.Ramkumar

[1]Research Scholar, Department of Electronics and Communication, Sathyabama Institute of science and technology, Chennai, Tamilnadu, India.
E-Mail-ID: tamilselvivlsi@gmail.com
[2]Associate Professor, Department of Electronics and Communication, Sathyabama Institute of science and technology, Chennai, Tamilnadu, India.
E-Mail-ID:skarthi1879@gmail.com
[3]Associate Professor, Department of Electronics and Communication, Saveetha School of Engineering, Chennai, Tamilnadu, India
E-Mail-ID : pgrvlsi@gmail.com

ABSTRACT

In the booming era of face recognition, various challenges are arising owe to the changes in the characteristics of biometric identifications. In order to handle the uncontrolled situations caused by the variations of light, there are many algorithms are in practice. To capture the image with light intensity and also to detect the travelling path of light ray, plenoptic camera is used. The proposed work uses a framework of double-deep spatio-angular learning to recognize light filed based images that are the sample of spatial and angular information with the sequence of two deep networks. This proposed framework added the inputs from VGG-Face descriptions and computed the VGG-19 based convolutional neural network. The 2D Sub Aperture(SA) images, that are derived from the images captured from plenoptic camera are used to extract the VGG-Face spatial descriptions in various angles. Then Long-Short Term Memory(LSTM) is used to analyze the VGG-face spatial description sequence.The images are then compared with the data base and the result obtained result is efficient and accurate when compared to existing system.

Keywords: Convolutional Neural Network(CNN), Double deep learning, LSTM, VGG-19

INTRODUCTION

These days, naturally perceiving the character of an individual is of primary significance. Face recognition is one of the biometric methodology which is broadly utilized in home gadgets, yet additionally in real time and business applications, by ensuring good worthiness in all aspects. During the past decade, the technology of face recognition becoming an emerging trend because of its various applications towards various fields[1-6]. In addition to this, the upcoming sensor technologies are also taking part in the development of this field to recognize face in an accurate manner. The frame work of face recognition is an innovation fit for distinguishing or checking an individual from an advanced picture or a video outline from a video source. Though, the face recognition is working with numerous strategies, the chosen facial highlights from given picture with faces inside an information base. It is additionally portrayed as a Biometric Artificial Intelligence based application that can exceptionally recognize an individual by dissecting designs dependent on the individual's facial surfaces and shape [7-10]. It is commonly utilized as access control in security framework and can be contrasted with other biometrics, for example, finger impression or recognition of iris.Even though, the recognition of facial framework accuracy is lesser than the other unique identifications such as iris and finger print recognition, this technique is broadly implemented in due course of non-invasive and touchless steps when compared to other algorithms [11-14].

PROPOSED SYSTEM

This proposed system of face recognition shown in fig 1 that is constructed on deep learning based convolutional neural network(CNN) using visual geometric group-19 is better in accuracy and speed

when comparing to the existing system. The proposed method involves in recognizing the face with various processes such as aligning the face and also extracting the features and additionally stresses the significance of the face arrangement, hence the exactness and genuine positive, false negative rate is seen via utilizing the mentioned procedure.
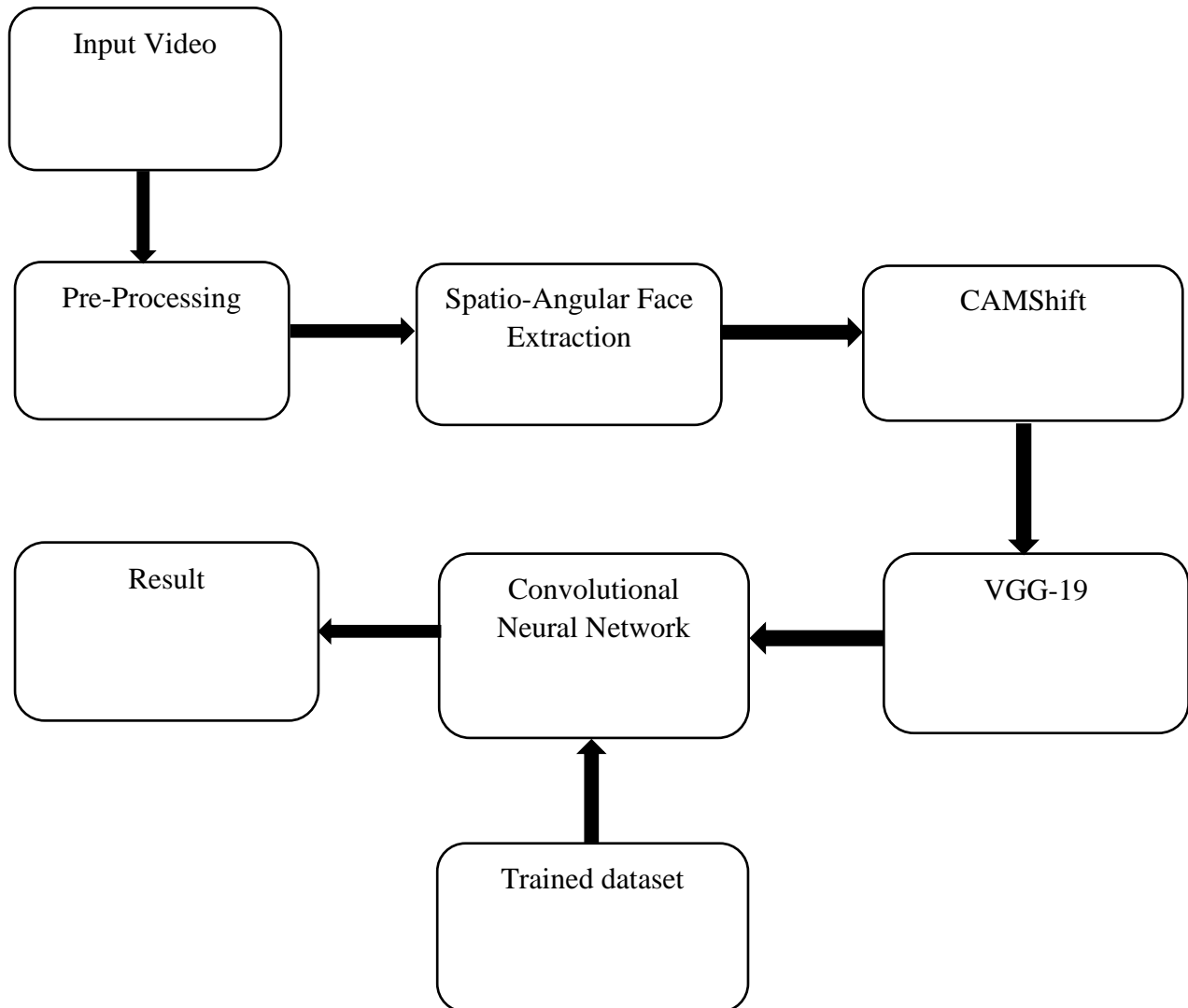

FIG 1 BLOCK DIAGRAM FOR PROPOSED SYSTEM

Pre-processing is a typical name for doing various operations on images at the least degree of reflection when input and output images are intensity images. These notable pictures are same as the reference images that are captured with the sensors where the intensity images are represented in a matrix structure. The main significance of doing pre-processing is to reduce the noises from the images and to increase its perfections in all aspects that are essential for further processing of images.

The proposed face recognition that is based on light field is formulated by combining a deep VGG-Face descriptor using convolutional neural network, hence in the proposed solution the double deep term is utilized. The VGG and CNN pair can be used to study the models of spatio-temporal that has been used for various processes such as facial classification, task description that includes recognition of action, classifying different facial expressions or capturing video or image. But this pair fails to predict the features in detailed manner in the images captured that are light field based and also taken over a instance of temporal at once. As for as the case above mentioned state, the proposed double-deep system had the unique sections to select SA image and to scan and target of creating a sequence that has been formulated using VGG-Face descriptions with the SA images that are taken at various

positions. From the chosen SA images, the face descriptors of VGG were removed and then to the CNN architecture, description sequence is given as input. This is the reason for claiming that the proposed algorithm of CNN-VGG will be an efficient algorithm to deal with the light field images that are having angular spatial evidences.

The CAMshift algorithm is employed for face detection where this algorithm is working over a scanning window to find the various face alignments in every frame. In this step itself, the preceding frame size and the scanning window size have been calculated by this CAMshift algorithm.

In the proposed system, the face is detected using CAMshift algorithm where the various distributions such as Hue, saturation and value are utilized. The main significance of hue is to get rid of the variations over the skin colour of human and also the contextual utilizing while the frames are taken. This CAMshift can be expanded as Continuously Adaptive Mean Shift, that is derived from the mean shift algorithm which has been carried over repeatedly in order to get aligned with the continuously varying probability distribution of colours in the frame while converting from series of videos.

The CAMShift calculation, Continuously Adaptive Mean Shift, typifies that of the Mean-Shift in a circle fluctuating the size of the window until intermingling. In every instance of repetition, this algorithm is allowed to ride over a given window size. After combination of the mean move, the method is re-iterated with another window, fixated on the position found by the mean move, yet of size contingent upon the zero request snapshot of the spatial dissemination of the skin tone likelihood recently determined by the mean move. In every instance of repetition, this algorithm is allowed to ride over a given window size. The same procedure is followed for an another window, that has got the place discovered with shift of mean after the conjunction of shift of mean. Then the probability of skin color which was calculated earlier based on the order of moment by shift of mean. The still images are segmented when applied with CAMshift algorithm, later the conjunction of the shift of mean, and the window height can be chosen randomly that is appropriate to segment a face in frontal position. Based on the object type, the aspect ratio can be varied or modified.

A VGG-19 is a Convolutional Neural Network shown in fig 2, which utilizes 19 layers having been trained on millions of Image samples and utilizes Architectural style for Zero-Centre normalization on Images Convolution, ReLU, Max Pooling, Convolution etc. It is a convolutional neural network (CNN) architecture to classify an image using deep learning. Zero-Centre normalization is the Centralization and Normalization. It normalizes and reduces dimensions - to keep scale centralized - in terms of when we will perform Convolutions. The reason this is important is to bring everything down in line in a normalized, streamlined and orderly fashion - so we have some sense of normality condition. As in we want the general structure of what we are parsing to be normalized and centralized so that we have a pre-defined boundary that we are being relative towards.

Convolutions is the functional operation of performing concatenation of Functional Curves so that they add up to encapsulate how they affect each other in terms of multiplicative relationships. ReLU is a rectifying linear unit. A unit that is made to rectify and compensate for signal parsing errors such as when we are forced to encapsulate epsilon (infinitely small) and we have to work around that. So ReLU Steers back the signal to where it's supposed to be headed so that the signal does not explode or vanish. Max Pooling is when you pool together the largest sample you can find - in an average area of taking strides. Strides is the average functional kernel mapping that you have in a square diagram plot - that averages out the value samplings of a certain space. A Fully Connected Layer is a one where the layers are connected fully that is utilized for Classification. Since this is used so sparingly it's at the end when the rough feature extraction and averaging of Functional relationships have had its sequence. Softmax is an applicative algorithm that normalizes a distribution dynamic from K amount of composite functional concatenations. This means that instead of having a spread where the values can be anything - 0, -1, 1, 2, etc. They are all normalized to be in line to a distribution (as the Softmax act as a regularize over a distribution)
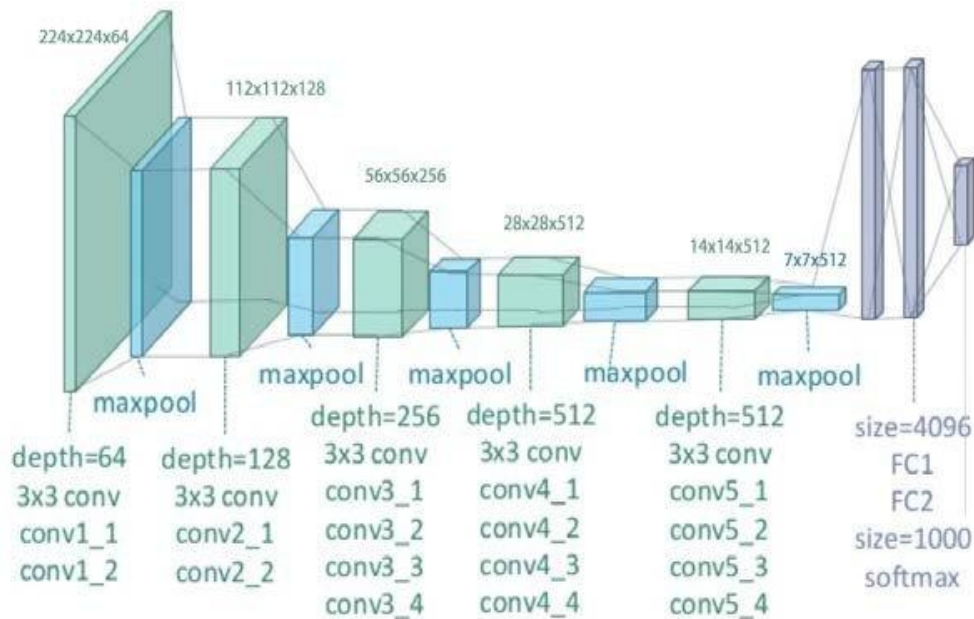
FIG 2 Architectural structure of VGG-19

A convolutional neural network is a package of various layers such as an input, output layer and many layers that are hidden. The sequence of convolutional layers that are convolved using a dot product or multiplying term are the major part of the hidden layers. The Rectified Linear Unit is responsible for activation function which has been carried out over the added convolutions with the layers like pooling layer and fully connected layer. They are termed as hidden layers since the input and output is covered with activation function.

In spite of the fact that the layers are conversationally mentioned by convolutions, this is simply done with convention. Numerically, it is in fact a sliding dab item or cross connection. This has importance for the lists in the grid, in that it influences how weight is resolved at a particular list point.

The tensor input along with figure (count of pictures) x width of the image) x (height of image) x (depth of the image) is the input to a CNN while programming. At that point subsequent to going through a convolutional layer, the picture gets disconnected to an element map, with shape (number of pictures) x (highlight map width) x (include map tallness) x (highlight map channels). A convolutional layer inside a neural network ought to have the accompanying credits

- A breadth and height can be described by kernels of Convolutional (hyper-parameters).
- The count of I/O channels.(hyper-parameter).
- In a feature map of input, the channel count and convolution filter deepness are equal.

Convolutional layers convolve the info and pass its outcome to the following layer. This is like the reaction of a neuron in the visual cortex to a particular boost. Each convolutional neuron measures information just for its responsive field. Even though, these feed forward networks are used for feature learning and for data classification, practically they cannot be applied to the architectural pictures. A high number of neurons would be vital, even in a narrow (inverse of profound) design, because of the extremely huge info sizes related with pictures, where every pixel is an important variable.For example, a completely associated layer for a (little) picture of size 100 x 100 has 10,000 loads for every neuron in the subsequent layer.The convolution activity carries an answer for this issue as it diminishes the quantity of free boundaries, permitting the system to be more profound with less boundaries.For example, paying little heed to picture size, tiling locales of size 5 x 5, each with similar shared loads, requires just 25 learnable boundaries. As the case mentioned above, the back propagation rectifies the diminishing problem of gradients while preparing the multiple layers neural networks.

In order to rationalize the fundamental computation, it is necessary to add either local or global pooling layers to the convolutional networks. The neuron outputs from each clusters are combined and made as an odd neuron from subsiding layers to decrease the size of the data. 2x2 size clusters are combined by locally available pooling layers. The global pooling layer acting over the available convolutional layer neurons. A maximum value or an average value can be calculated by the pooling layer which is the added feature of CNN. Max pooling utilizes the highest value from every one of a group of neurons at the earlier layer. Avg pooling utilizes the average value from every one of a group of neurons at the earlier layer.

The neurons present in one layer are connected to the other layer by the use of Fully connected layers as like mentioned in the multiple layers perceptron neural network(MLP). The images are classified by allowing the flattened matrix to go through

Each and every neuron getting its input from nearly various locations from the previous layer. From the previous layer, the input from each and every part is given to the fully connected layer but as for the case for convolutional layer, some specific inputs only given from the prior layers with the size of 5x5. Since the receptive field is the area covered by input, the whole previous layer is considered as receptive field for convolutional layer. Figure 3 showing the sample of the input image given to the CNN network, figure 4 indicates the detection of face using CAMshift algorithm and the figure 5 shows the hardware kit output mentioning the name of the recognized person from the data base.
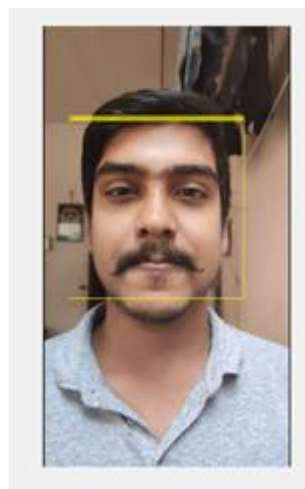

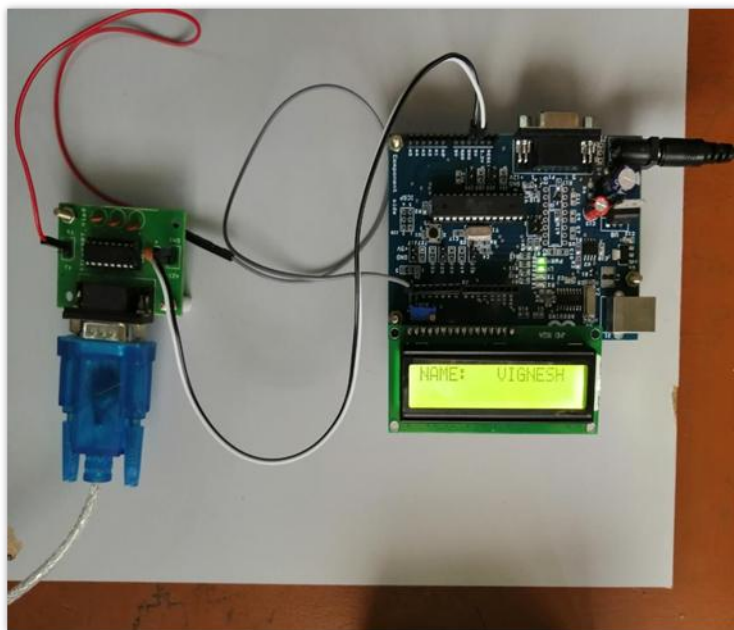**Fig 3: Input Image**

Fig 4: Face Detection

Fig 5 – Hardware output for face recognition

Results and Discussion

The proposed framework is analysed with the real time data set and the Table 1 indicates the comparison of measured output parameters such as accuracy, sensitivity, specificity and precision between VGG-16 and VGG-19.

- Accuracy is the one among the evaluation parameters of convolutional neural networks. It can be achieved by getting the ratio between the count of correct predictions and to the total count of predictions.
- Sensitivity describes about the true positive rates that is how many predictions of positive are made correctly.
- Specificity tells us about the true negative rates that is how many predictions of negative are made correctly.
- Precision indicates the ratio of instances that are relevant and to the instances that are irrelevant.

From the values mentioned in the table, the proposed framework with VGG-19 showing the better percentage of values in all the output parameters. Figure 6 represents the graphical analysis of the comparison between VGG-16 and VGG-19. Table 2 indicates the comparison of rate parameters that includes the true and false detection rates for both the models VGG-16 and VGG-19. The percentage of values showing the better increase in the proposed framework and the same can be visually recognized using the graphical analysis of rate parameters that has been shown in figure 7.

TABLE 1 MEASURED OUTPUT PARAMETER

MEASURED PARAMETERS	VGG-16	VGG-19
ACCURACY	75%	90%
SENSITIVITY	83%	93%
SPECIFICITY	42%	48%

PRECISION	86%	96%
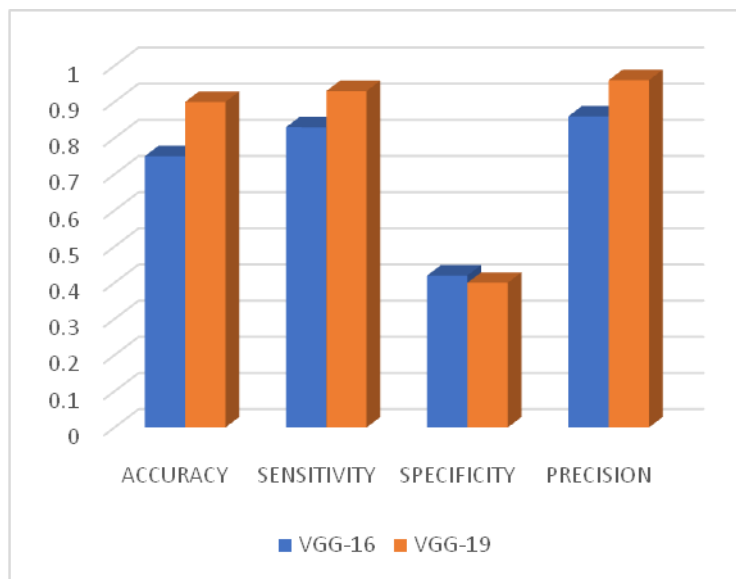
Fig 6: Graphical representation of comparative analysis

TABLE 2 RATE PARAMETERS

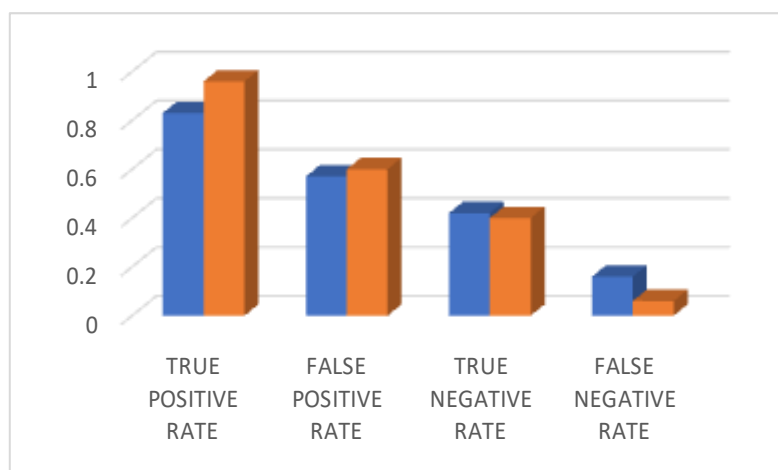| RATE PARAMETERS	VGG-16	VGG-19
TRUE POSITIVE RATE	83%	96%
FALSE POSITIVE RATE	57%	60%
TRUE NEGATIVE RATE	42%	40%
FALSE NEGATIVE RATE	16%	6%

Fig 7: Graph representing the rate parameters

CONCLUSION

This paper proposes a novel technique named Visual Geometry Group -19(VGG-19) which is equally able to cope with illumination variations and heterogeneous face recognition. The key idea of our

method is to recognize the face in uncontrolled environment condition. The earlier used models are not robust and they are not completely infallible. The existing methods are high in false detections. In order to overcome the issues due to environment condition mentioned earlier, proposed a novel method based on spatio-angular VGG19 convolutional neural network that provides a 19-layer deep introspection of the input image whereas VGG16 provides only a 16-layer deep information. Results obtained from the proposed method are outperform the existing framework in terms of high recognition rate. Another important aspect of our method is the processing time which is far much better than the state-of-the-art systems.

REFERENCE

[1] A. L. Machidon, O. M. Machidon and P. L. Ogrutan, "Face Recognition Using Eigen faces, Geometrical PCA Approximation and Neural Networks," *2019 42nd International Conference on Telecommunications and Signal Processing (TSP)*, Budapest, Hungary, 2019, pp. 80-83.

[2] M. Tamilselvi and S. Karthikeyan, "A Face Recognition System using Directional Binary Code Algorithm and Multi-SVM", International Journal of Engineering and Advanced Technology, 2019.

[3] J Govindaraj, Dr & Logashanmugam, E.. (2019). Multimodal verge for scale and pose variant real time face tracking and recognition. Indonesian Journal of Electrical Engineering and Computer Science. 13. 665. 10.11591/ijeecs.v13.i2.pp665-670.

[4] K. Chang and C. Chen, "A Learning Framework for Age Rank Estimation Based on Face Images with Scattering Transform," in IEEE Transactions on Image Processing, vol. 24, no. 3, pp. 785-798, March 2015.

[5] L. Liu, "Human Face Expression Recognition Based on Deep Learning-Deep Convolutional Neural Network," *2019* International Conference on Smart Grid and Electrical Automation (ICSGEA), Xiangtan, China, 2019, pp. 221-224.

[6] G. Ramkumar & Logashanmugam, E. (2018). Hybrid framework for detection of human face based on haar-like feature. International Journal of Engineering and Technology(UAE). 7. 1786-1790. 10.14419/ijet.v7i3.16227.

[7] M. Tamilselvi and S. Karthikeyan, "Feature Extraction and Facial Expression Recognition using Support Vector Machine," 2019 International Conference on Smart Systems and Inventive Technology (ICSSIT), Tirunelveli, India, 2019, pp. 889-893.

[8] M. Romero, J. Paduano and V. Muñoz, "Point-Triplet Spin-Images for Landmark Localisation in 3D Face Data," *2014* IEEE Workshop on Biometric Measurements and Systems for Security and Medical Applications (BIOMS) Proceedings, Rome,2014.

[9] Sepas-Moghaddam, F. Pereira and P. L. Correia, "Ear Presentation Attack Detection: Benchmarking Study with First Lenslet Light Field Database," 2018 26th European Signal Processing Conference (EUSIPCO), Rome, 2018, pp. 2355-2359.

[10] Sepas-Moghaddam, V. Chiesa, P. L. Correia, F. Pereira and J. Dugelay, "The IST-EURECOM Light Field Face Database," 2017 5th International Workshop on Biometrics and Forensics (IWBF), Coventry, 2017, pp. 1-6.

[11] G. Ramkumar and E. Logashanmugam, "An effectual face tracking based on transformed algorithm using composite mask," 2016 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), Chennai, 2016.

[12] x. shu, J. Tang, Z. Li, H. Lai, L. Zhang and S. Yan, "Personalized Age Progression with Bi-Level Aging Dictionary Learning," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 40, no. 4, pp. 905-917, 1 April 2018.

[13] Govindaraj, Dr & E, Logashanmugam. (2018). Study on impulsive assessment of chronic pain correlated expressions in facial images. Biomedical Research. 29. 10.4066/biomedicalresearch.29-18-886.

[14] Y. Wen, Z. Li and Y. Qiao, "Latent Factor Guided Convolutional Neural Networks for Age-Invariant Face Recognition," *2016* IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, 2016, pp. 4893-4901.