Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 218 (2023) 741–749

Procedia Computer Science

www.elsevier.com/locate/procedia

International Conference on Machine Learning and Data Engineering

Face Detection and Recognition Using Face Mesh and Deep Neural Network

Shivalila Hangaragi[a], Tripty Singh[b], Neelima N[a,b,*]

[a] *Department of Electronics and Communication Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India*

[b] *Department of Computer science and Engineering, Amrita School of Engineering, Bengaluru, Amrita Vishwa Vidyapeetham, India*

Abstract

Face detection and recognition are emerging and active research areas in computer vision and deep learning. Face detection and recognition have a wide range of applications such as recognizing the people in particular areas namely stores, and banks. Identifying people in a particular database (police database). To control people's entry into restricted areas or grant access to ATMs or computers. In this paper, the proposed model can detect and recognize the face using Face mesh. Due to Face mesh, the model operates in a variety of conditions such as varying illumination and background. The model can also handle non-frontal images of males and females of all ages and races. The Labeled wild face (LWF) dataset images and images captured in real-time are used to train the deep neural network of the model. During testing, if the face landmarks of the test image match with the face landmarks of any of the training images the model gives the name of the person else model outputs as "unknown".94.23% accuracy are achieved for face recognition by the proposed model.

Keywords: Neural Network; Face mesh landmarks; Face Recognition; Face Detection; Computer Vision

1. Introduction

Human face detection and recognition are very important aspects of security and surveillance. Face recognition technology can compare human faces from an image and the stored database of faces. The model extracts the facial features from the input image to authorize users through Identity verification services. These models are implemented in both the private and public sectors. Some of the applications include:1) Security and law enforcement 2) Healthcare 3) Banking and retail. One of the major applications of face recognition is Facial biometrics. It is the most preferred

* Shivalila Hangaragi.
E-mail address: bl.en.r4ece21008@bl.students.amrita.edu

10.1016/j.procs.2023.01.054

biometric benchmark because of its speed. It is also easy to deploy and implement without Human Interaction. It is the method of authenticating and authorizing the identification of the person from the extracted facial features. It captures, analyses, and matches patterns based on the person's facial features. The steps involved in the process are:1) Detecting and locating the human faces in the image 2) Extracting the facial features 3) Comparing the detected face with the database.

Algorithms such as Eigen Faces[16], Local Binary Pattern Histograms[17], Fisher faces[18], Scale Invariant Feature Transform, and Speed Up Robust Features are used for extracting the feature. Face landmarks such as the distance between the two eyes, the depth between the eye sockets, the contour of the face, and the shape of the lips and cheekbone are used for comparing the query input image with the database images. The major benefit of facial recognition systems is that they can identify in bulk without human interaction intervention. There are many applications of such systems. Some applications include installing facial recognition systems in public places such as malls, airports, and offices for biometrics. Factors such as Expression, illumination, pose, and noise during image capture can alter the performance of facial recognition systems.

The illumination variation and variation in facial pose have a greater impact on the accuracy of these systems. The orientation of illumination leads to shifts in the position & shape of the shadow, changes in highlights, and also the reversal of contrast gradients. The proper orientation and alignment of the face pose lead to the extraction of facial features that are Robust to geometric variations.

The proposed model overcomes the problems encountered in the face recognition system and gives the most appropriate results. The model uses Face mesh for identifying and detecting the face. Due to Face mesh, the model operates in a variety of conditions such as varying illumination and background. The model can also handle non-frontal images of males and females of all ages and races. The Labeled wild face (LWF) dataset images and images captured in real-time are used for cognition of the model. During testing, if the face landmarks of the test image match with the face landmarks of any of the training images the model gives the name of the person else model outputs as “unknown”. The proposed model achieves an accuracy of 94.23% for face recognition.

In the introduction to the need for the proposed model, the drawbacks of the existing models are briefed. The related work outlines the significant work done by prominent researchers in the same field. Followed by data acquisition and system description. The results of the experiment conducted, Novelty, and comparison with other existing algorithms is accomplished in the latter section of the paper.

2. Related Work

Many researchers have worked in the field of face recognition some significant contributions are reported.

Yaswanthram et al. [1] Face recognition is the method in which the system compares the facial patterns, features, and contour with a stored database to authenticate the person. This face recognition technology is used in biometric systems in offices to identify and authenticate people. The dimensionality reduction has a greater impact on the accuracy of machine learning algorithms in face recognition. Experiments were conducted on various algorithms such as Random Forests, SVM, Linear Regression, Logistic Regression, and KNN. It was analyzed that Logistic shows the better performance with 97% accuracy without PCA (Principle Component Analysis) whereas, with PCA, Logistic Regression achieves an accuracy of 93%. Nehru et al. [2] Face recognition is a widely used technology to authorize and authenticate people without human intervention. These systems fail to produce accurate results in some conditions. The factors affecting the accuracy of the system are non-frontal images and variation in illumination. An illumination invariant face detection model using the viola jones algorithm is proposed. The proposed model was trained to detect human faces irrespective of illumination conditions. Many different filters are employed by Viola-Jones Cascade Object Detector to detect facial features.

Sravya et al. [3]Blind people face daily life challenges to do day-to-day activities due to blindness. One of their difficulties is recognizing people for safety and security. An efficient and cost-effective model for blind people is developed. The model comprises a hat embedded with a camera and headphones. The model alerts the user if any person is approaching, through a voice message, and speaks out whether the person is known or unknown. Model alerts the user about the surrounding. [4] Nowadays Road accidents have increased in great numbers due to the recklessness and negligence of the driver. A smart driver alerting system to avoid road accidents can be built by extracting facial features. The driver monitoring system with an alarm and camera was developed. The camera focuses on the user's face. The variation in the facial expression is observed & processed to find the restlessness of the driver. The alarm rings to alert the driver. Aishwarya et al. [5] Human emotion recognition is a very important technique to interact intelligently with machines. An emotion recognition system through facial expression is developed. HOG(Histogram of oriented gradients) and Haar cascade algorithm are used to extract the facial features. The extracted feature is used to categorize the emotions like anger, disgust, fear, happiness, neutral and sad. An accuracy of 65.5% proves that the HOG approach is more robust than other existing approaches.

Kalturi et al. [6] Face recognition technology has a wide range of applications. One of the major advantages of such systems is to identify and authenticate the users. An AI-based intelligent door unlocking model using face recognition was designed for the Blind. The text messages are sent to the enrolled phone number and buzz the alarm for safety and security. The model is trained with images of authorized users to grant access through the door. Access will be denied to unauthorized users. Singh et al. [7] Image segmentation is the key area of research. It is used for processing the images to extract the necessary features. An image segmentation-based technique for spotting the cancerous regions in medical images was developed. The double mask method is used for the identification of the region of interest. Experiments were conducted on medical images with various modalities. The proposed model experimented with different images and information such as Entropy, Energy, mutual Information, & Peak SNR computed. Rahul et al.[8] Blindness comes with its difficulties and challenges. One of the major causes of early blindness is diabetic retinopathy. The technique for the detection of diabetic retinopathy to prevent early blindness is proposed. The appearance and structure of blood vessels in retinal images are the main problems in the diagnosis of eye diseases. The Canny edge detection technique is used for edge detection. The proposed method performs well in extracting the vascular pattern than traditional edge detection techniques. Pramod et al[9] It is a difficult and very important task to reconstruct the parts of the head and face in medical imaging. Tetrahedral mesh is generated for such tasks. These images are stored in binary format. Sampling points and surface boundaries are extracted from these images for internal and external tetrahedralization. Classification & labeling of elements of different tissue is accomplished. Existing face recognition and detection systems fail to recognize faces accurately if tested under various conditions such as varying illumination and background, and non-frontal human faces of different ages and races. The proposed model efficiently recognizes non-frontal human faces of different ages and races. If the test image is faced in different orientations proposed model will reconstruct the complete face from the extracted face landmarks using face mesh. If the face matches with the database it is tagged with the recognized name.

3. Data Acquisition

Face detection and recognition experiments were conducted on the Labeled Faces in a Wild (LFW) dataset and images captured in real-time. Labeled Faces in the Wild (LFW)[15] is a database of facial pictures developed for studying the problem of face recognition in an unconstrained environment. The database consists of 13,233 images of 5,749 people which are identified by the Viola-Jones face detector [11] and gathered from the Internet. Experiments were conducted on 1,680 images of people having two or various photographs in the dataset combined with 20 images captured in real-time, making 1700 images. The 3D face reconstruction experiments are carried out using the BU3DFE [14] database.

Table 1: Sample Labeled images of LFW dataset and real-time images

				
Aaron_Eckhart	Aaron_Peirsol	Abel_Aguilar	Tripty Singh	Shivalila Hangaragi

4. System Description

The block diagram of the proposed model is shown in Fig 1.

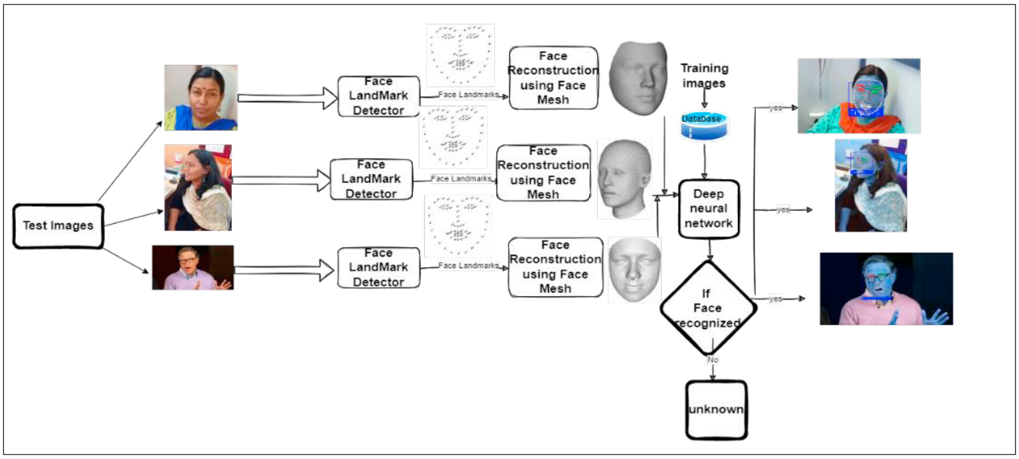


Fig1: Block diagram of the proposed model

The image of Aaron Sorkin, the famous screenwriter, and Hollywood director is taken as an example to illustrate the working of the proposed model as shown in Fig 2. Two different images of Aaron Sorkin are taken as a test image and training image for the experiment. Mediapipe an open-source, cross-platform machine learning framework is used for face detection. The MediaPipe face detector detects faces and extracts key points called landmarks to pass on to a Deep neural network to match with the database. The name of the person is displayed if the face is recognized else “unknown” is displayed.

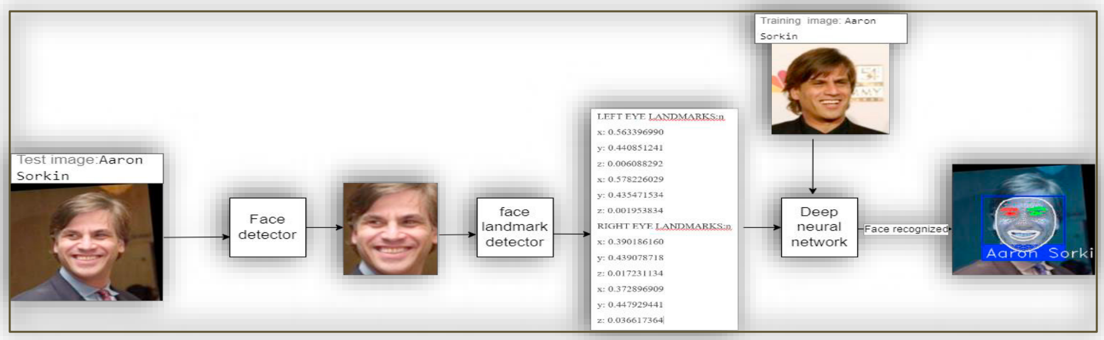


Fig 2: Evaluation of the proposed model using test images

4.1 Face landmarks

Face landmark detection is a computer vision task where the key points in the human face are detected as shown below in Fig3.

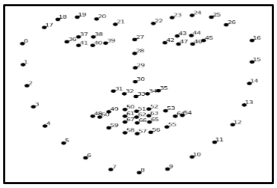


Fig 3: The human face landmarks

The human faces are represented as a list of 468 face landmarks and each face has X, Y, and Z components. The media pipe is based on the Blaze face model for detecting the face key points. It detects the face in the images and marks 468 face landmarks around the detected face.

4.2 3D Face reconstruction

The reconstruction accuracy of a 3D face is calculated by using Root Mean Square Error (RMSE) [12], defined as

$$\mathrm{RMSE} = \frac{1}{N_T}\Sigma_{i=1}^{N_T}(\|s_i^* - \hat{s}_i\| / n) \qquad (1)$$

Where, N_T is the total number of samples tested, s_i^* and $\hat{s}_i$ are the ground truth and 3D face shape reconstructed of i^{th} testing sample. The face reconstruction is depicted in Fig 4 below

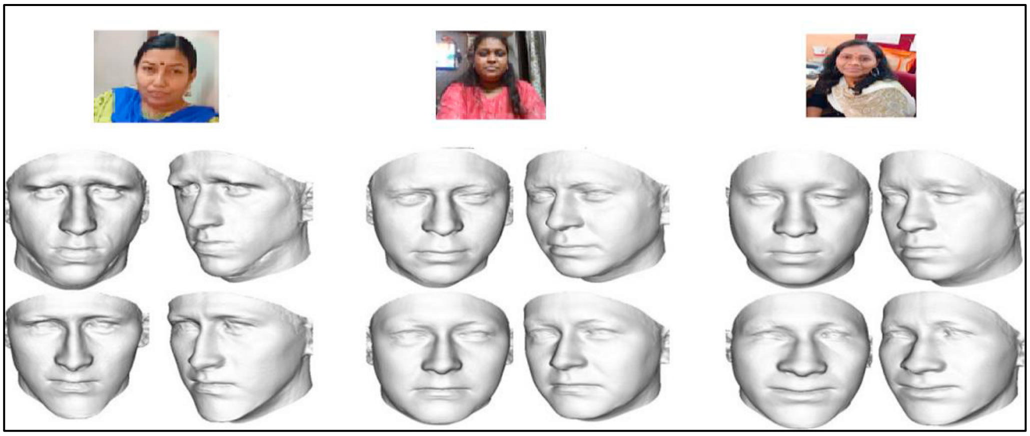


Fig 4:The 3D face reconstruction from the sample images

4.3 Face Mesh

The representation of 3D mesh can be done by connected undirected network G=(V,E) where $V \in R^{nx3}$ denotes set of n vertices, $\mathcal{E}$ is the set of edges and $W \in \mathbb{R}^{nxn}$ is an adjacency matrix encoding the connection status between vertices. The Laplacian L is analyzed using Fourier bases $U=[u_0,\ldots,u_{n-1}] \in \mathbb{R}^{nxn}$ such that $L=U\Lambda U^T$ where $\Lambda=\mathrm{diag}([\lambda_0,\ldots\lambda_{n-1}]) \in \mathbb{R}^{nxn}$. In Chebyshev spectral graph CNN, the graph convolution operation is defined in eq(2) as

$$g_\theta(\Lambda) = \sum_{k=0}^{k-1} \theta k T_k(\Lambda^{\sim}) \qquad (2)$$

where $\theta \in \mathbb{R}^k$ is a vector of Chebyshev coefficients and $T_k(\Lambda^{\sim}) \in \mathbb{R}^{nxn}$ is the Chebyshev polynomial of order k evaluated at a scaled Laplacian $\Lambda^{\sim} = 2\,\Lambda / \lambda_{max}\, I_n$. The spectral convolution can be defined in eq(3) as

$$yj=\sum_{i=1}^{Fin} g\theta i,j(L)xi \qquad (3)$$

Where xi & yj are the input and output feature maps respectively.

5. Result Analysis

Face detection and face recognition experiments were conducted on pictures captured in real-time and images picked from the LWF dataset. The proposed model results are shown in table 1 by picking four images as an example from the 1700 images experimented with. The experiment is conducted by taking training images and test images captured under different constraints and are explained below.

As depicted in fig(5a) of Fig 4 in the training image the person is wearing glasses and in the test image the person is not wearing the glasses but still, the proposed model can recognize the face with the name.
In fig(5b) the face in the test image is turned to the left side and the training image is faced to the camera. The proposed model reconstructs the complete face using face mesh to acquire the face landmarks of the complete face. The acquired landmarks are matched with the face landmarks of the faces in the database and the face is recognized by name.
In fig (5c) the training image and test image are captured in different backgrounds and with different illumination. Irrespective of the environment under which the image is captured the model can recognize the face with the name.
In fig(5d) the model is trained with the image of the Palestinian militant Abdel_Nasser_Assidi picked from the LFW dataset. The training image has only one face of the Abdel_Nasser_Assidi but the test image is given with three faces.

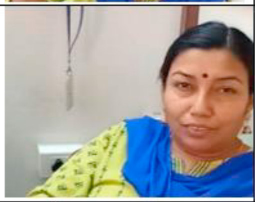
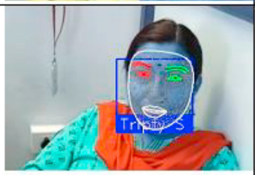
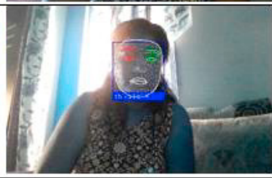
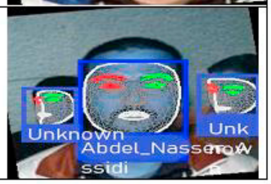
Name of the person	(a) Tripty S	(b) Neelima	(c) Shivalila	(d) Abdel_Nasser_As sidi
Training image				
Test image				
Output				

Fig 5:Face detection and recognition results on pictures captured in real-time and images from the LWF dataset

The model recognizes the face of the militant by matching the marked face landmarks with the face landmarks of the faces in the training database. The other two faces are identified, but the model could not recognize the face due to the database's non-availability of matching face landmarks. Hence the two faces are tagged as "unknown".

Table 2: Table depicting the similarity function between labels and names tagged by the proposed model for the examples considered in Table 1

Sl.No	Labels of the image to be recognized	Number of images	Name detected by the proposed model	Similarity Function
1	Tripty S	3	Tripty S	1
2	Neelima	3	Neelima	1
3	Shivalila H	2	Shivalila H	1
4	Abdel_Nasser_Assidi	2	Abdel_Nasser_Assidi	1

$$\mathrm{Sim}(\mathrm{T1},\mathrm{T2}) = 2 * \frac{\mathrm{M}}{\mathrm{T}} \qquad (4)$$

Where M = Number of characters matching with the position in both strings and T =total number of elements in both strings.

$$Accuracy\ in\ \% = \frac{Number\ of\ images\ correctly\ recognized}{Total\ Number\ of\ images} \mathrm{x}100 \qquad (5)$$

Table 3 Accuracy of the proposed model on the LWF dataset

Total number of images	1700
Number of images correctly recognized	1602
Accuracy in %	94.23%

Experiments were conducted by feeding multiple sets of training and test images to the model. The images are captured under different orientations, illumination, background, and environment. The model matches the test and training image if the match is found the face is tagged with the label of the image. The Similarity Index[17] is calculated for the label of the image and the names tagged by the model after the face recognition as shown in Table 2 using eq(4). The accuracy of the model is computed using eq(5). Experiments were conducted on 1,680 images of people having two or more distinct photos in the dataset combined with 20 images captured in real-time, making 1700 images. The model attains an accuracy of 94.23% as shown in Table 3.

5.1 Face reconstruction Accuracy

The BU3DFE database contains 3D faces displaying various emotions such as Anger, Grief, Joy, Excitement, etc. Table 5 shows the reconstruction error on various poses with different methods. It is observed that face reconstruction accuracy is less for the proposed model and there is no substantial increase in the error value for various poses. The proposed model is efficient and robust to a variation in the pose as listed in Table 4.

Table 4 Accuracy of face reconstruction (RMSE) in various angles on BU3DFE dataset

Method	±90°	±80°	±70°	±60°	±50°	±40°	±30°	±20°	±10°	±0°	Avg
VRN	6.96	6.20	6.14	6.01	5.91	5.50	4.93	3.86	3.70	3.66	5.29
3DDFA	2.90	2.88	2.81	2.82	2.77	2.79	2.76	2.73	2.55	2.48	2.75
3DMM-CNN	-	-	-	-	2.30	2.26	2.23	2.22	2.19	2.17	2.23
3DSR	2.11	2.11	2.12	2.13	2.16	2.14	2.12	2.10	2.10	2.09	2.12
PROPOSED	2.08	2.04	2.04	2.03	2.01	1.97	2.00	2.00	1.96	1.95	2.01

Table 5 The face recognition comparison of the proposed method with existing methods

METHOD	Shape	Texture	ACCURACY
3DMM(LFW)	Y N Y	N Y Y	63.34%-68.92% 73.79%-76.07% 73.13%-77.37%
3DDFA(LFW)	Y	N	64.42-69.54%
3DMM-CNN(LFW)	Y N Y	N Y Y	89.19%-91.87% 89.53%-91.67% 91.06%-93.64%
PROPOSED (LFW+IMAGES CAPTURED IN REAL-TIME)	Y	Y	94.23%

5.2 Face Recognition Accuracy

The Face Recognition accuracy comparison of the proposed technique (LFW dataset+ Images captured in real-time) with different methods on (LWF dataset) is shown in Table5. It is observed that the face recognition accuracy varies between 63.34%-68.92% on the LFW dataset with the 3DMM method using only shape. Similarly, the accuracy variation is observed when both parameters are considered or either of the parameters is considered on various methodologies. The proposed method is robust to variation in illumination, pose, and works better for non-frontal images with an accuracy of 94.23% when both the shape and texture are considered as the parameter. The face reconstruction accuracy is calculated using eq(1) on the BU3DFE dataset in different angles and is shown in Table 4. Face recognition is an emerging task in computer vision. There are different methods already existing to accomplish the task[13]. The accuracy comparison between different existing techniques and the proposed method is shown in table 5.

6. Novelty and comparison with other face detection algorithms.

The Viola-jones face detection technique popularly known as the Haar cascade uses edges and lines to detect faces in images or real-time videos. The feature extraction is done by computing the Haar value from the image. The darker areas are marked as one and the lighter areas with a value of zero. The detection of an edge or line is estimated by a change in the intensities of the pixel. The objective is to find the difference between the summation of image pixels in a darker area and the summation of pixels in a lighter area. If the value is close to one edge is detected else there is no edge. In the proposed method face is detected by face mesh. The Blaze model detects the face landmarks in the image and a face mesh is formed. If only a part of the face is visible in the image, the model is efficient in reconstructing the complete face using face key points and detecting the face. The model detects and recognizes the faces in various illumination, and non-frontal images efficiently which other existing algorithms fail.

7. Conclusion

Face detection and recognition experiments were conducted on 1700 images picked from the combination of the LWF dataset and some of the real-time captured images. Face mesh is used to reconstruct the complete face with face landmarks. The face reconstruction accuracy is calculated on the BU3DFE dataset from different angles. Face detection and recognition of non-frontal human faces of different ages and races are accomplished. The comparison of face recognition accuracy with the existing methods is accomplished. The model attains an accuracy of 94.23% for the experiments conducted on various constraints.

References

1. P. Yaswanthram and B. A. Sabarish. (2022) "Face Recognition Using Machine Learning Models - Comparative Analysis and impact of dimensionality reduction," *IEEE Fourth International Conference on Advances in Electronics, Computers, and Communications (ICAC)*: pp. 1-4,
2. M. Nehru and S. Padmavathi. (2017)"Illumination invariant face detection using viola jones algorithm," *2017 4th International Conference on Advanced Computing and Communication Systems (ICACCS)*: pp 1-4.
3. A. S. L. Sravya, K. S. Chinmayi, L. N. Yoshita, S. V and P. V. Manitha,(2019) "A Smart Aid for Visually Challenged in Face Detection and Recognition," *2019 International Conference on Smart Systems and Inventive Technology (ICSSIT)*: pp. 704-708
4. Raju, J.V.V.S.N., Rakesh, P., Neelima, N. (2020). "Driver Drowsiness Monitoring System." In: Reddy, A., Marla, D., Simic, M., Favorskaya, M., Satapathy, S. (eds) Intelligent Manufacturing and Energy Sustainability. Smart Innovation, Systems and Technologies, vol 169. Springer, Singapore. pp 675-683
5. Aishwarya, M., Neelima, N. (2021). "The Analogy of Haar Cascade and HOG Approaches for Facial Emotion Recognition." In: Kaiser, M.S., Xie, J., Rathore, V.S. (eds) Information and Communication Technology for Competitive Strategies (ICTCS 2020). Lecture Notes in Networks and Systems, vol 190. Springer, Singapore. pp 699-707
6. Kalturi, A., Agarwal, A., Neelima, N. (2021). "Smart and Autonomous Door Unlocking System Using Face Detection and Recognition". In: Hemanth, J., Bestakand detects, R., Chen, J.IZ. (eds) Intelligent Data Communication Technologies and Internet of Things. Lecture Notes on Data Engineering and Communications Technologies, vol 57. Springer, Singapore. pp 99-108
7. T. Singh and S. Karanchery,(2019) "Universal Image Segmentation Technique for Cancer Detection in Medical Images," *10th International Conference on Computing, Communication and Networking Technologies (ICCCNT)*:pp 1-7,
8. R. M.S.P., N. A. Mahakalkar and T. Singh,(2019) "Novel Approach For Detection Of Early Diabetic Retinopathy," *Third International Conference on Inventive Systems and Control (ICISC)*: pp 450-455
9. S. P. G. and T. Singh.(2019) "Tetrahedral 3D Mesh Generation for Medical Images," *Third International Conference on Inventive Systems and Control (ICISC)*: pp 443-449.
10. Y. Q. Wang. (2014) "An analysis of the Viola-Jones face detection algorithm", *Image Process. Line*, 4:pp 128-148.
11. A. Bas, W. A. Smith, T. Bolkart, and S. Wuhrer. (2016)" Fitting a 3D morphable model to edges: A comparison between hard and soft correspondences". *In ACCV*: pp 377–391.
12. A. T. Tran, T. Hassner, I. Masi, and G. Medioni. (2017)" Regressing robust and discriminative 3D morphable models with a very deep neural network". *In CVPR*. pp 5163-5172
13. Liu, F., Zhu, R., Zeng, D., Zhao, Q., & Liu, X. (2018). "Disentangling features in 3D face shapes for joint face reconstruction and recognition". In *Proceedings of the IEEE conference on computer vision and pattern recognition*: pp 5216-5225.
14. G. B. Huang, M. Ramesh, T. Berg, and E. LearnedMiller.(2007)" Labeled faces in the wild: A database for studying face recognition in unconstrained environments". *Technical report, Technical Report*: pp 07-49.
15. Agus Makmum and Husni Thamrin.(2018)" Performance of similarity algorithms for statement mapping in a SWAT analysis application". *AIP Conference Proceedings*, pp 20063-20075
16. Turk, M. and Pentland, A. (1991)." Face recognition using eigenfaces". *In Proceedings of the Computer Society Conference on Computer Vision and Pattern Recognition, Lahaina, Maui, Hawaii,* pp. 586– 591.
17. Zhang, G., Huang, Z., Li, S.Z.(2004)" Boosting local binary pattern (LBP)-based face recognition", In Li, S.Z., Lai, J.-H., Tan, T., Feng, G.-C., Wang, Y. (eds.) SINOBIOMETRICS 2004. LNCS, vol. 3338, pp. 179–186.
18. Liu CJ, Wechsler H (1998)" Enhanced fisher linear discriminant models for face recognition". In: Proceedings of the international conference on pattern recognition, pp 1368–1372.