IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

A Survey on Video-Based Face Recognition System

[1]Mr. Yogesh M. Rohit, [2] Prof. Mahasweta J. Joshi, [3]Dr. Hemant D. Vasava

[1]M.Tech. Scholar, [2]Assistant Professor, [3]Assistant Professor
[1]Computer Department,
[1]BVM Engineering College, V.V. Nagar, India

Abstract: Face recognition has been important research in the field of computer vision technology for a very long time. It is an efficient and one of the most preferred biometric technique, The proposed algorithms in this field are endless, and the accuracy that can be achieved higher than ever due to vast advancement in deep learning technology, video-based facial recognition it plays a major role in real-life applications but there are some challenges also, Deep learning techniques archives state-of-the-art results in the field of computer vision, At present deep learning techniques are one of the dominant technique compare to other technique, This paper presents a comprehensive survey on the different issues related to Video-based face recognition systems also to solving these issues are analyzed by presenting existing techniques that have been proposed in the literature and various Real-world applications of video-based face recognition.

Index Terms - Face recognition, Deep learning, Video-based.

I. INTRODUCTION

The 21st century is the modern digital era in which lots of processes have been automated, a lot of progress has been achieved to expedite humans for accomplishing their tasks. Nowadays technology is part of life, computers are used for tons of applications, from simpler to complex tasks, among such face recognition technology. The wide use of deep learning technology and rapid growth of Artificial Neural Network (ANN) had powerful features extraction capability on image data that increased Advancement in face recognition, Video-based face recognition is the technique of identifying one or multiple persons present in a video, based on their facial features. Comparatively, videos produce more information than still images. These things can be helpful for the video-based facial recognition system, it has been one of the most researched areas in the field of pattern recognition and computer vision because research is no longer limited to still images.

video-based face recognition system (VFR) receives attention more than ever and attracts the researcher's to work on this kind of technology through the face recognition widely studied and video-based face recognition is still in its infancy phase [1]. However, many factors affect the system efficiency and accuracy. There are many applications like Face tracking, Emotion recognition, pose estimation, access control, security system [14], surveillance system and now it being used at daily life gadget. The organization of the paper is as follows: Section II. Provides literature Survey, Section III. Provide various deep learning based approaches, Section IV. Describes Popular Dataset that often used in deep learning based approaches, Section V. Provide the Conclusions.

II. LITERATURE SURVEY

Chao Guo et al. (2019) at [1] have proposed a system in which they effectively dealing with the problems that are lacking in the target sample data and the low video quality, which are common in the practical application of face recognition system divided into submodule which are pre-processing module, learning module, an identification module, A data pre-processing module is to extract the face images in the video and provide data for the next face recognition module, FFmpeg is used to extracts the keyframes of the video to be detected and the unrelated video, and the face images in keyframes are detect by MTCNN, The ultimate goal of the learning module is to create a face recognition network model that can be used to identify the target characters , appearing in the video for that they have used facenet the identification module uses SVM classifiers to identify faces, the performance of the system is with own dataset with accuracy of 94.9% and authors have also tested on CASIA-Web face dataset with 90% of data and getting 84.7% accuracy.

Samadhi Wickrama et al. have proposed (2019) at [2] A Framework for Real-Time Face-Recognition in which they have discussed how deep neural network effectively recognize faces under unconstrained conditions, system have divided into submodules are face detector it detects the face along with face alignment, embedding generation unit maps face, feature points to features vector of fixed dimensions and the Euclidean distance measures the face features point with features vector point, IOU based tracking mechanism is used for face tracking, Face detection – alignment unit. implemented on MTCNN architecture EGU is DCNN model of inception ResNet VI trained with VGGFace2 data set here authors have performed transfer learning with 99.75% accuracy.

In 2017 Musab Coşkun et al. at [3] proposed facial recognition systems based on convolutional neural network. It starts with pre-processing stage: colour space conversion and resize the images, continues with the extraction of facial features, and is classified afterward. Authors have proposed a modified Convolutional Neural Network (CNN) architecture by adding two normalization operations to the layers. The normalization operation which is batch normalization providing an} acceleration of the network. CNN architecture extracts distinctive facial features and uses Softmax classifier to classify faces in the fully connected layer of Convolutional Neural Network. In the experiment part, Georgia Tech Database showed that the proposed approach is an improved face recognition algorithm with improving performance with better accuracy 94.8 % and with better recognition results.

Saibal Manna et al. at [4] has proposed their work on video based face recognition in 2020 here authors have used pre trained model which is facenet here goal of author is to achieve high accuracy when face image maps from Euclidean space, in which the distances directly correspond to an approximation of the face's similarity. When space is generated, different tasks can be easily accomplished. Face verification, identification and clustering use regular FaceNet embedding methods as the function vectors. Triplets are used for training the framework. Here authors have not tried to create work from scratch instead they are focusing on accelerating the process with their own created dataset with accuracy of 90%.

In 2016 S.V.Tathe et al. at [5] introduced to reduce the human intervention and increased the overall systems accuracy system is segregate into three different category’s motion detection face detection and face recognition, in motion detection model work on background subtraction but it has many challenges to encounter like poor law camera quality, camera jitter, noisy environment, illumination changes, natural movements of stationary objects like trees shaken by wind etc. so these problem could be solved by Gaussian Mixture Model (GMM) for face Detection simple method for face detection is based on skin color information viola jones method is used for corrected detection LPA and LBP method used to extract textual features with LDA dimensionality that can be classified using svm classifier next is face recognition module Face recognition systems can be classified as sub-class of pattern recognition systems. face dimension can be reduced by principal component analysis before feed to Neural Network to enhance recognition process and make improve robustness of system.

Mohamed Heshmat et al. at [6] proposed system consists of three stages: first step is skin-like regions detection in CIE-Luv color space, the second step is face detection based on skin-like regions, contour detection and geometrical properties such as face shape. The third step is face verification, in which, each face is compared with known faces and the location of the best matched one is returned Face verification step uses variance formula and skin to non-skin percentage in each facial feature to compare the test face and the faces images in the known database. The proposed system was tested on many different videos with different numbers of persons in the video. The detected faces are compared with a preset database of known images, So the first it read input video it goes to skin detection step convert to CIE-Luv color spec and detect skin like region after that detect and select contours contain skin like region by rectangle in stage one after that in stage two checks the face criteria and extract faces and in third stage calculate and compare variance value for test face and all faces in code book it goes to selected faces from codebook whose variance close to test face variance and in last process stage compare facial regions of test faces and close codebook faces by Euclidean distance and display matched faces on codebook. The proposed method has been tested using single face and multi-faces video effectiveness of the proposed system and its ability to recognize a variety of different faces in spite of different pose, expression, zooming and illumination conditions. And accuracy with different frames on video it may vary accuracy with 98.4%.

Aswathy et al have proposed face recognition at [7] and tagging the proposed system which is implemented on one of application which is attendance, system using face as a biometrics, System first registered persons detail such as name, email, mobile number and gender of the person after that through webcam capture persons face around 200 images capture and stored in database, model will be trained on that dataset, from web cam it record the video and detect the faces and match those faces with stored images in database if match is found mark the attendance of that person and if doesn’t then update database , CNN used as a model and achieve accuracy of 95%.

Maheen Zulfiqar et al [8] have proposed system at on Biometric authentication using facial recognition, system which uses viola jones algorithm for identification and for classification CNN model is used authors have used here pre-trained CNN which is squeeze net authors have compare on their dataset with some state of the art techniques like VGG and ResNet50 and they find squeeze net have lower computational cost and higher accuracy , dataset made of 30 different object with different lighting condition and varying different angle around 9000 images dataset contain and achieve 98.76% of accuracy.

Bruce poon et al at [9] have proposed a system which encounter their previous research problem which is illumination problem for that applying technique for gradient face at preprocessing stage that greatly improves recognition rate specially at those images under various lighting condition , but not only that it also worked well on noisy and blurry images, therefore author have used same database which is Asian face database with greatly increased rate of 6.25% to 60.75% on previous work which is performance evaluation and methods for distorted images.

Reshma and Kannan at [10] proposed their work on partial face recognition, partial face recognition is having application in a broad spectrum of different fields. The different approaches used for partial face recognition are the keypoint-based approach, region-based approach, and CNN-based approach. In keypoint-based, the popular method was MKD-SRC. In the region-based partial face recognition approach, the prominent model is MR-CNN. And the CNN-based approaches are Dynamic Feature matching (DFM). A comparatively CNN-based approach is more precise compare to others as of now. According to the authors, Dynamic Feature matching (DFM) is having a promising application in video recognition approaches in the future.

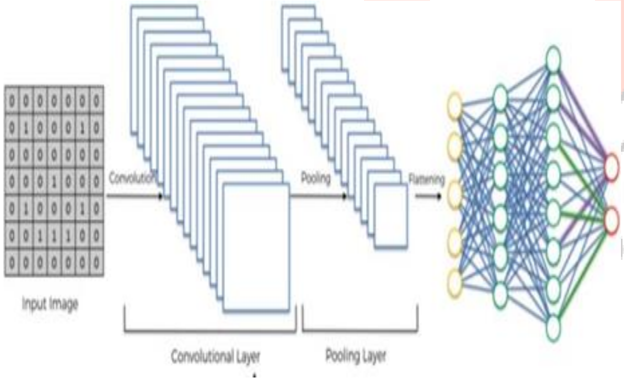
In the work by Katarina Knezevic et al at [13] An influence of blur and motion blur on face recognition performance. For the face detection they have used pre trained Haar classifier and for the face recognition they have used Local Binary Patterns Histograms algorithm (LBPH), here they have done two sets of experiments Gaussian blur and motion blur on original images but in that they have found Gaussian blur is much harder problem compare to motion blur , In case of motion blur edges are more visible another thing they conclude here that enhanced images are more likely to be recognized by system compare to blurred images.

Yingcheng Su et al. at [15] proposed their work on face recognition with Occlusion. The author has considered all those scenarios, in a real-life occlusion are quite common especially in the uncooperative scenario, through regression analysis we can recover clean images from the degraded or occluded image by using a clean training sample, to recover or reconstruct occluded image introduced noise with it so authors have present a new occlusion, detection method by combining both raw and residual image, Using the non – occluded part for face recognition has a better result using the reconstructed image. Here the authors approach NMR method is first used to obtain a reconstructed image and residual images, by repeating the same operation to get the raw image and residual image i.e. diving face into an upper part and lower part, for feature, extraction using Gabor wavelet and PCA, for feature extraction and dimensionality reduction respectively and classifying raw and residual image first two levels of SVM classifiers are used. After experimenting with the non-occluded facial region of the occluded face can still provide more precise information than reconstructing face images.

Jeffrey R.et al at [12] proposed differentiating identical twins by face recognition, identification of twins it's a very complex task then identification of two different identities. Identification of twins more complex due to variation in illumination, expression, age, or a year later, the author has done a comparison between seven different algorithms. On the same dataset which is Twins Day Dataset, it consists of 17486 images of 126 pairs of identical twins. The experimental results measured the performance when a face was collected on the same and one year apart. Experiment results can be affected by images are taken in a controlled environment like a studio setup or an uncontrolled environment like an outside home environment. The baseline error rate ranged from <0.1% to 2.4%, The highest performing algorithm had an error rate from 4.1% to 17.4%. the experiment shows that it is possible to distinguish identical twins under ideal condition (same-day acquisition, studio-like illumination, consistent expression).

III. Deep Learning Based Approaches

Deep learning has won so many contests in machine learning, pattern recognition competition over the past few years. Deep learning is a subset of machine learning technique, its uses hidden layer for information processing and based on that it classify/predict output.



0	0	0	0	0	0	0
0	1	0	0	0	1	0
0	0	0	0	0	0	0
0	0	0	1	0	0	0
0	1	0	0	0	1	0
0	0	1	1	1	0	0
0	0	0	0	0	0	0

Figure 1. Generic representation of CNN [10].

The performance of a deep learning method can be increased by increasing variations in the dataset and increasing the size of the dataset [8] many researchers have proven successful results in a diverse use case like speech recognition, voice search, handwriting recognition, image feature recognition and many more. Deep learning can be categorized into three main categories that can be depending on how technique and architecture are used. Unsupervised: Recurrent neural network(RNN) and sum-product network (SPN), Supervised: Convolutional Neural Network(CNN), Hybrid: (Deep Convolutional Neural Network DCNN).

Today's world is data-driven, every second millions of data generated, that data can help us to for prediction or taking the right decision, Convolutional neural network is the best example of supervised learning architecture, face recognition is a popular research area and it's a widely studies, CNN has archive great success in face recognition [11].

The structure of CNN includes a layer of convolution, pooling, rectified linear unit, fully contacted.

Convolutional layer: This is CNN's core building block that aims at extracting features from the input data. Each layer uses a convolution operation to obtain a feature map. After that, feature maps are fed to the next layer as input data [9].

Pooling layer: Reduces the dimensionality of the feature map but still has the important information. The input images are divided into a set of a non-overlapping rectangle, each region is down sampled by a nonlinear operation, such as average or maximum, it better generalized, faster convergence and robust to translate, placement of this layer between convolutional layer [3].

Rectifier linear unit (ReLU) Layer: it is an elementwise operation, Transform function only activates a node if the input is above a certain quantity, it is applied per pixel and reconstitutes all negative values in the feature map by zero.

Fully connected layer (FC): it refers to that every filter in the previous layer is connected to every filter in the next layer. [3] The high-level reasoning in the neural network is done via fully connected layers' After applying various convolutional layers and max-pooling layers and ReLU [11].

3.1 Popular CNN Architecture

3.1.1 LeNet

LeNet is a kind of earliest convolutional model promoted for development in deep learning around 1988-1989, after some leading work, and some of the changes it named LeNet-5, proposed in 1998 by Lecun et al. t is a capable CNN for digit recognition. LeNet is consider as the backbone of modern CNN. [14] The LeNet-5 architecture consists of 2 sets of convolutional, and average pooling layers, followed by a flattening convolutional layer, then two fully-connected layers, and finally a softmax classifier.

3.1.2 AlexNet

AlexNet was designed by Alex Krizhevsky in collaboration with Ilya Sutskever and Geoffrey Hinton.ImageNet Large Scale Visual Recognition Challenge (ILSVRC)-2012 using AlexNet achieves record-breaking results and compare to traditional machine learning approaches and [11] AlexNet were outperformed by Microsoft Research Asia's very deep CNN with over 100 Alexnet layers, which won the ImageNet contest in 2015. It contains a total of eight-layer, in that first five of that convolution layer followed by max-pooling layer and the last three were fully connected.

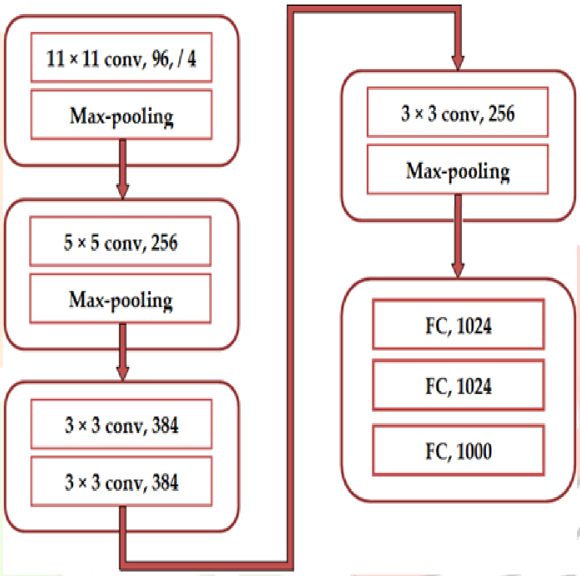


Figure 2. Basic Block diagram of AlexNet [11].

3.1.3 VGG Net

VGG Net are a type of CNN Architecture proposed by Karen Simonyan & Andrew Zisserman of Visual Geometry Group (VGG), Oxford University, it won the 2014 image net competition, Network depth was increased to 16-19 weight layer, in 16 depths there are 13 convolutional layers and 3 are FC, in 19 there are 16 convolutional layers and 3 FC and five pooling layer in both [11].

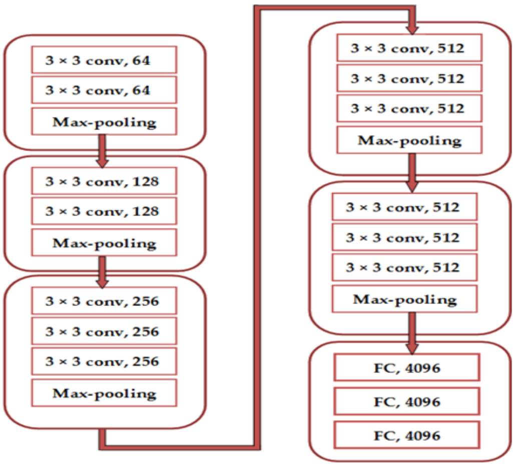


Figure 3. Basic Block diagram of VGG Net [11].

3.1.4 Google Net

The winner of ILSVRC-2014 was the 22-layer GoogleNet, a model proposed by Szegedy et al. (2014), {To minimize computational complexity compared with the standard CNN model. It introduced an "inception module," containing variable receptive fields generated by different kernel sizes. [11] It has Several convolutions (1 × 1, 3 × 3, and 5 × 5) and (3 × 3) max-pooling layers.

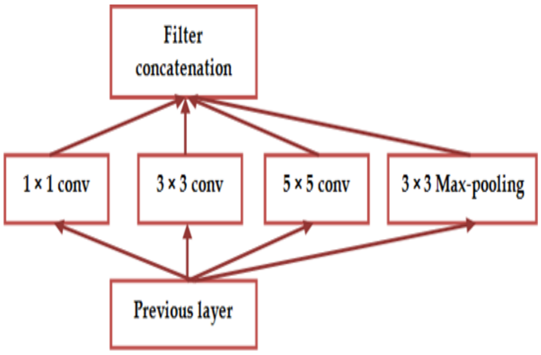


Figure 4. Basic Block diagram of GoogleNet [11]

3.1.5 ResNet

An Introduced a novel architecture named residual neural network (ResNet) to facilitate the training of ultra-deep networks compared to networks already in use. ResNet was the winner of ILSVRC 2015; it was developed with "shortcut connections" and features batch normalization, it was able to train a neural network with various numbers of layers: 34, 50, 101, 152, and even 1202 [11].

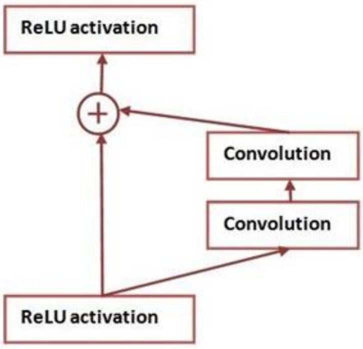


Figure 5. Basic Block diagram of ResNet [11].

3.1.6 SENet

It won first place at ILSVRC-2017 since they proposed the block squeeze-and-excitation (SE), a novel architecture unit, which recalibrates channel-wise feature responses by clearly modeling the inter-dependencies between channels [11]. The SE network (SENet) was developed by stacking a set of SE blocks and can be integrated with standard architecture such as ResNet, improving their effectiveness in numerous datasets and tasks.

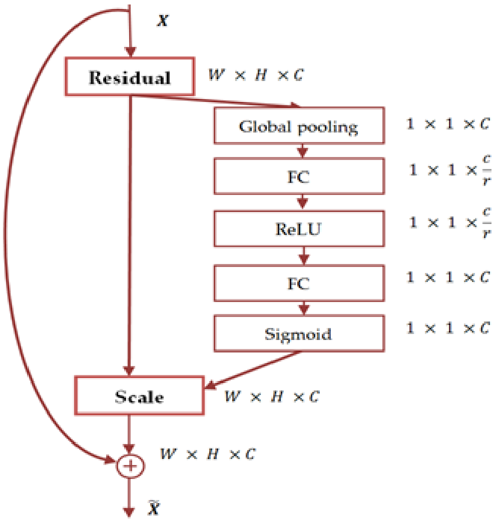


Figure 6. Basic Block diagram of SENet [11].

In the following, we discuss several deep face recognition methods based on CNN, which are typically trained in a supervised manner. Also, there is a significant amount of research done and increasing day by day, researchers are exploring and expanding greater possibility in this direction where some of the methods are focused on extracting appearance variation features from non-frontal images. Many works adopted ideas from metric learning and combined different loss functions and other's use and finding better activation functions.

IV. Popular Datasets

In most well-known/recent 2D face recognition datasets mostly used for training deep face recognition systems.

YouTube faces (YTF) is contains 3425 video of 1595 identity collected from YouTube. With an average of 2 video per identity. And is a standard benchmark for face verification in video [4].

A large-scale dataset for face recognition task, called CASIA-WebFace, was selected from the IMDb website with 10,575 persons and 494,414 facial images. It was built in 2014 by Dong Yi and the team at [11] the Institute of Automation, Chinese Academy of Sciences (CASIA).

Microsoft released the Ms-Celeb-M1 large scale training benchmark in 2016, which contains around 10 million face images from 100k celebrities collected from [11] the web to improve facial recognition technologies.

In 2016 introduced the MegaFace database, which includes a total of 1,027,060 images of 690,572 different subjects. The MegaFace challenge uses a gallery to test the performance face identification / verification algorithms with faces that are not in the test set ('Distractors') [11], by training them from different probe set such as FG-NET and a FaceScrub

VGGFACE (visual geometry group) is a large-scale training database assembled from the Internet by combining automation and humans in the loop. It contains 2.6 M images, over 2.6 K identities. It was created at the University of Oxford in 2016 by Parkhi and the team [11].

In 2017, a large-scale face database was called VGGFace2. it was created by Cao and team. From the University of Oxford. The database was collected from Google Images search with a wide range of age, pose, and ethnicity. It has 3.31 million images of 9131 identities, with 362.6 images for each identity. The VGGface2 is divided into two subclasses: test and evaluation which is test, training-set, including 8631 classes, and a test set with 500 classes.

V. Conclusion

A significant amount of work is done, On the face recognition system but, there no perfect algorithm that full-filled all the maximum requirements to encounter challenges that appear in real life. Real-world scenarios are quite complicated to solve, but by looking at the growth made by AI deep learning technology, there is no such task to remain unsolved for a long time. Soon, there are more sophisticated algorithms that can meet all the maximum.

References

[1] C. Guo and Y. Yang, "Implementation of a specified face recognition system based on video," 2019 IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chengdu, China, 2019, pp. 79-84, doi: 10.1109/IAEAC47372.2019.8997627.
[2] S. W. Arachchilage and E. Izquierdo, "A Framework for Real-Time Face-Recognition," 2019 IEEE Visual Communications and Image Processing (VCIP), Sydney, Australia, 2019, pp. 1-4, doi: 10.1109/VCIP47243.2019.8965805.
[3] M. Coşkun, A. Uçar, Ö. Yildirim and Y. Demir, "Face recognition based on convolutional neural network," 2017 International Conference on Modern Electrical and Energy Systems (MEES), Kremenchuk, 2017, pp. 376-379, doi: 10.1109/MEES.2017.8248937.
[4] S. Manna, S. Ghildiyal and K. Bhimani, "Face Recognition from Video using Deep Learning," 2020 5th International Conference on Communication and Electronics Systems (ICCES), COIMBATORE, India, 2020, pp. 1101-1106, doi: 10.1109/ICCES48766.2020.9137927.
[5] S. V. Tathe, A. S. Narote and S. P. Narote, "Human face detection and recognition in videos," 2016 International Conference on Advances in Computing, Communications and Informatics (ICACCI), Jaipur, 2016, pp. 2200-2205, doi: 10.1109/ICACCI.2016.7732378.
[6] M. Heshmat, W. M. Abd-Elhafiez, M. Girgis and S. Elaw, "Face identification system in video," 2016 11th International Conference on Computer Engineering & Systems (ICCES), Cairo, 2016, pp. 147-154, doi: 10.1109/ICCES.2016.7821991.
[7] Aswathy V, Lijin Das S, 2019, Video based Face Recognition and Tagging using Deep Learning, INTERNATIONAL JOURNAL OF ENGINEERING RESEARCH & TECHNOLOGY (IJERT) Volume 08, Issue 07 (July 2019),
[8] M. Zulfiqar, F. Syed, M. J. Khan and K. Khurshid, "Deep Face Recognition for Biometric Authentication," 2019 International Conference on Electrical, Communication, and Computer Engineering (ICECCE), Swat, Pakistan, 2019, pp. 1-6, doi: 10.1109/ICECCE47252.2019.8940725.
[9] Chicago Poon, B., M. A. Amin and H. Yan. "Improved Methods on PCA Based Human Face Recognition for Distorted Images." (2016).

[10] M. R. Reshma and B. Kannan, "Approaches on Partial Face Recognition: A Literature Review," 2019 3rd International Conference on Trends in Electronics and Informatics (ICOEI), Tirunelveli, India, 2019, pp. 538-544, doi: 10.1109/ICOEI.2019.8862783.

[11] Adjabi, Insaf; Ouahabi, Abdeldjalil; Benzaoui, Amir; Taleb-Ahmed, Abdelmalik. 2020. "Past, Present, and Future of Face Recognition: A Review" Electronics 9, no. 8: 1188. https://doi.org/10.3390/electronics9081188

[12] J. R. Paone et al., "Double Trouble: Differentiating Identical Twins by Face Recognition," in IEEE Transactions on Information Forensics and Security, vol. 9, no. 2, pp. 285-295, Feb. 2014, doi: 10.1109/TIFS.2013.2296373.

[13] K. Knežević, E. Mandić, R. Petrović and B. Stojanović, "Blur and Motion Blur Influence on Face Recognition Performance," 2018 14th Symposium on Neural Networks and Applications (NEUREL), Belgrade, 2018, pp. 1-5, doi: 10.1109/NEUREL.2018.8587028.

[14] Su, Yingcheng & Yang, Yujiu & Guo, Zhenhua & Yang, Weiguo. (2015). Face recognition with occlusion. 670-674. 10.1109/ACPR.2015.7486587.