International Journal of Engineering and Information Systems (IJEAIS)
ISSN: 2643-640X
Vol. 6 Issue 2, February - 2022, Pages:57-65

Multi-Face Recognition System in Surveillance Video

Ibtisam Mohammed Humaid Al Bahri[1], Arockiasamy Soosaimanickam [2], Sallam Osman Fageeri[3]

Department of Information Systems, College of EMIS, University of Nizwa, Sultanate of Oman,
E-mail: 6498134@uofn.edu.om [1], arockiasamy@unizwa.edu.om [2], sallam@unizwa.edu.om [3]

Abstract—*An automatic multi-face recognition system is one of the artificial intelligent systems. It helps to manage enormous number of tasks such as Access control & surveillance systems, monitoring management, and so other detection applications. Multi-face recognition system in surveillance video (MFRS) or automatic multi-face recognition. It is a system that have a capability to detect and read human face from surveillance video automatically and immediately and then sends them to database, afterwards result readable by machines. These kinds of systems have been widely used to recognize human face by using several algorithms and methodologies, including optical dimensions' recognition, convolutional neural network, artificial neural network or deep neural network, morphological operations, and facial features edge detection. This paper aims to discuss the multi-face recognition system using surveillance video. It aims at understanding and analyzing the concept of the human face recognition system, essentially those systems that don't required any human resources support to accomplish their functions and trying to evaluate those human face recognition systems by using real-time generated human faces dataset. For this purpose, this research uses an analytical and theoretical comparison between various previous research works in face recognition field to understand which deep learning algorithm is providing much accurate results. Also, there is a practical evaluation by using real human face images. In this study, there are three different stages of the human face recognition system. Initial operation starts from images aggregation, face detection, human face recognition, and finally identity recognition. In each processing phase, there is a sub preprocessing procedures and deep learning algorithms used. For instance, morphological operations were used to detect and extract face region and using convolutional neural network in face feature recognition level. The preprocessing levels are from the insert human image step until output the recognized data, however deep learning algorithms are distinct from one system framework to another. The performance of recognition operation is evaluated based on different metrics such as: classification accuracy, and precision, and recall.*

Keywords— Face recognition, Convolutional Neural Network, Deep Learning Algorithm, Morphological Operations.

1. INTRODUCTION

Recently authentication technology using biometric became more fame. People deals positively with regard to biometric authentication technologies. Researchers found that 74% of sample respondents are familiar and intimate towards biometric technology. There is a fact that 90% of the public said that they are ready to use biometric identification operation which means World-Readiness [1].
According to IDEMIA company research which included a survey of 2,800 respondents from the whole world, that show little concern between participants due to the non-human technological aspect of biometrics. Iris biometric technology comes in the second order in accuracy and security, whereas Fingerprint gets the highest rate convenient to use, otherwise face recognition technology was a most inconvenience one for participants [1]. There is a direction in information security to replace passwords and smart cards with biometrics systems, which will be used to identify a personal identity and access the data in the database. It's great to mention that, at the beginning of the twentieth century biometrics usage was only restricted to criminal and military persons. Then by the end of the end of twentieth century, private sector and other educational institutions started to realize that this feature of using the technology to detect employer identity in the organizations. At that time, fingerprint was the most popular biometric type [2].
With the fourth industrial revolution, numerous modern technologies were developed. Biometrics has been used to enrich big data performance. Currently, biometric systems established in several domains like security, identification, and judicial science [3].
In Sultanate of Oman, there is rapid advancement using this technology. As part of that, the idea of applying biometrics science in the students' attendance system has been started. A biometric system will provide reliability and efficient student identity management.
Researchers in Gnanamani College of Technology proposed a fingerprint recognition system for student attendance using minutia-based fingerprint algorithms and GSM tracking system. They worked to establish an embedded system to use for security applications. Biometric applications have been increased; it contributes to facilitating educational institutions operations. The prototype idea uses identity sensor to track the location of the missing person's location [4].
According to [5], they build face recognition model using the PCA algorithm. With experiments to measure system effectiveness using eigenface approach. Paul, and Al Sumam mention that PCA algorithm can be upgraded to detect the facial expression and a human gender. Minor facial changes do not affect PCA efficiency, furthermore, simplicity and accelerate performance.

2. RELATED WORKS

Researchers [6] certify that the face is a complicated 3-dimensional visual model which required a difficult computational system to perform face recognition operation. Devi, N. S., & Hema chandran, explained in their paper that suggests merging these two-stages.

Applying recognition operation feed-forward back-propagation Artificial Neural Network (ANN), after face feature has been extracted by principal component analysis (PCA) and this method has been experimented in Oracle Research Lab (ORL). Test score gave an adequate recognition rate of (99.50%). The main aim is to eliminate judgment of facial expressions pose, situations like open/closed eyes, with and without a face accessory.

Whereas in [7], proposed with patch-based principal component analysis (PB-PCA). The previous research papers on Eigen face method which extracts features as pixels or columns. Researchers from The University of Electronic Science and Technology of China said that face recognition can be outlined more clearly by face patches. The operation goes through many steps starting by splitting face into patches that transformed into column vectors which form "image-matrix" with 2D PCA structure. Then they calculate the correlation using total scatter and enhance the result by optimizing the total scatter. Lastly, they calculate the nearest neighbor classifier of the projected matrix. This patch preserves the accuracy, compared to the PCA algorithm. In addition to that, patch is an essential unit for classification.

The PCA algorithm performs multi-tasks, i.e. trait extraction, forecast, eliminate duplication, and data compression. As mentioned before, there are issues to be considered, firstly whether if the photo hosted to the system is a not a face. Secondly, hosted face not stored as a known face? The initial Eigen face with quality face filter, give assistance to measure the degree of image correlated with itself. The second issue can solve by discarded images with low correlation. The research was based on a face database in Cambridge AT&T Laboratories, where ten images in different facial expressions situation [8].

In the previous research, they focus on the wide application of face recognition algorithms from 10 to 15 past years. Therefore, Sadeghipour, & Sahragard, proposed to apply improved SIFT algorithm on ORL database, and compare it other pattern recognition algorithms. They performed key-points extraction operators besides feature extraction operators in SIFT, to increase the efficiently and eliminate computational abundant. The outcome shows that SIFT algorithm accuracy and run-time measurement were better from other algorithms [9].

3. RESEARCH METHODOLOGY

Computer vision field has gone through many qualitative developments and an artificial intelligence contributed to developing biometric systems. Face detection systems depend on artificial neural networks which contains many hidden inner layers. Each layer is concerned with a specific task, for example, first layer is concerned with defining the edges and external geometric structures of face image, and then this layer passes the results to the next layer. Second layer responsible of performing colors and shapes specification, after that compare what does database have learned and what had trained previously. This process will continue until image information are completed, at that point, system will determine what is the core features of the image with a high accuracy degree [10]. The proposed work presented is simplified in three objectives as multi-face video surveillance detection, recognition and validate, applying, and benchmark the effectiveness and accuracy of the system. In support of this system Open CV library, prebuild-in dataset and python programming is used for computer vision mechanisms involved.

A. Procedure of proposed model

The proposed work of Multi-Face Recognition in Surveillance Video system is shown in Figure.1. Initially, a camera is used to capture the human face and detect the precise location of face by adding a bounding box direction for the face real-time detection. The process takes part of face detection using Haar cascade algorithm with open CV library. Haar cascade features used to perform human face detection. There is different outer situation for each image, detected Images have different external effects, such as illumination degree, objects, and landscapes etc. At this stage of process, human face is detected, then face features are extracted, at that point data will be stored in the face recognition database.

The CNN model uses VGG 16 to match the face feature from the database and recognize with the Identity of the detected face. Faces are recognized from the stored database, addendum to that embedding vectors which used to perform as compared to identify or detect the face. Distribution platforms use Anaconda, with python 3.8 software in face detection processing, recognition, and classification phase. whereas the image features located in the database dlib. At the beginning, face is detected, afterwards recognized with the database stored features, and then matching using convolutional neural network (CNN) model training and testing database, as shown the below figure 9, which indicates the recognizing the human face identity. VGG 16 network architecture is built with CNN model with the purpose of performing huge database recognition. The structured network model has honeycomb 3 x 3 layers, where each two connected layers have 4096 nodes accompanied by Softmax classification. The image pixels are identified by setting a threshold measurement, also the outcomes are represented in form of a binary number. To perform this LBPH uses 4 parameters. [11]

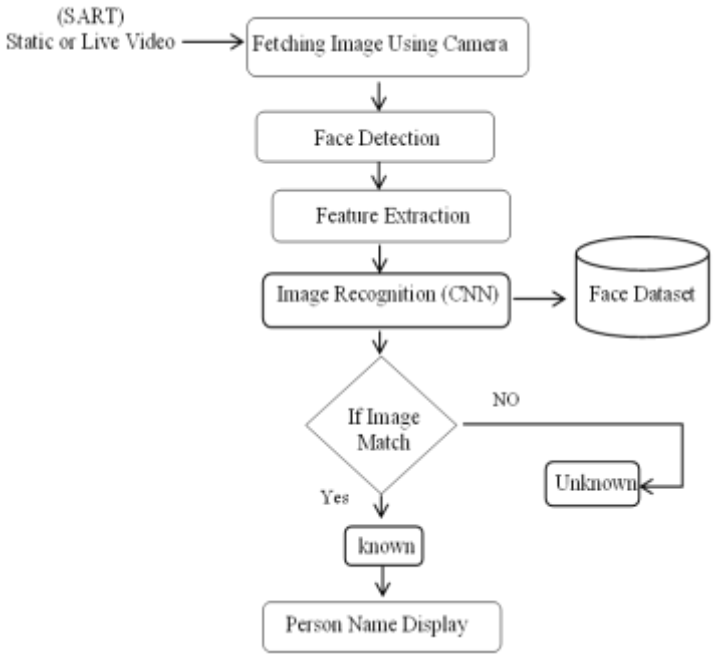

Fig.1. Human face recognition system

B. Generating dataset

First, prepare data sets by collecting images. For this purpose, Open-CV package must be used python pip install OpenCV-python, and haarcascade_frontalface_default.xml file as a classifier, to get the coordinate of faces, consequently we crop it then utilize it to generate the dataset. During that time, we convert images from RGB (Red, Green, Blue) to grayscale. The RGB image has 3 channels whereas the grayscale image has only 1 channel, which reduces a lot of complexity if we convert it. Subsequent, using face classifier and using detect Multiscale acquisition, to get the coordinates. it is required to proceed with scale factor and minimum neighbors.

C. Data exploration

Efficiently extracting knowledge from data. Investigate the datasets is the main procedure in deep learning to depict how to portray dataset attributes like size, amount, quality to comprehend this information. Pick which exactness metric and which model are smarter to utilize. Investigate the number of training and testing tests are in this dataset, the width and tallness aspect sizes.

D. Import python libraries

Import the libraries used in the model:

Keras is a well-known and simple to-utilize library for building deep learning models. Keras library upholds all known sort of layers: input, convolutional, dense, rendered convolution, reform, actuation, dropout, and normalization. Each layer plays out a specific procedure on the information.

Lambda: Lambda layers provide appropriate method for bundling libraries and different dependencies which can be used with Lambda functions. Layers minimize the size of transferred sending files and makes code deployment operation speeder, additionally, its advance code sharing and division of obligations to iterate quicker on composing business logic.

Dense layer: a neural network layer that is linked deeply, where each neuron in the dense layer receives input result from neurons of its prior layer. It`s the most popularly used layer in for models. The dense layer executes a matrix-vector multiplication. The figures used in the matrix are in fact parameters that can be updated and trained with the help of backpropagation algorithm.

Flatten layer: The classification model required a processed data that should be convenient input to the model. Flatten layer converting the data into a 1-dimensional array for loading it to the following layer.

NumPy: a python library used complicated level of mathematical equations for large multiple dimensional arrays and matrix processing for instance: linear algebra.

Matplotlib: it is a dataset representation and two-dimensional plotting library used to generate 2D diagrams, for example, outlines, histograms, and bar figures.

Glob Module: The glob module is a helpful task of the Python standard library. Glob is utilized to return all document paths that match a particular pattern. Glob use to look for a particular document design, or maybe more helpfully, look for records where the filename matches a specific manner by utilizing wildcard characters.

E. Load the data

To evaluate the accuracy of face recognition it requires to import dataset into the system. Human face dataset has 3100-face images for 31 participants divided into 31 classes, the age of participants between 3 and 50 years. The dataset is divided into two sets: a training dataset of 1300 images and 100 images in each class. The testing dataset contains 620 images, and, in each class, we have 20 images which is shown in Figure 2.

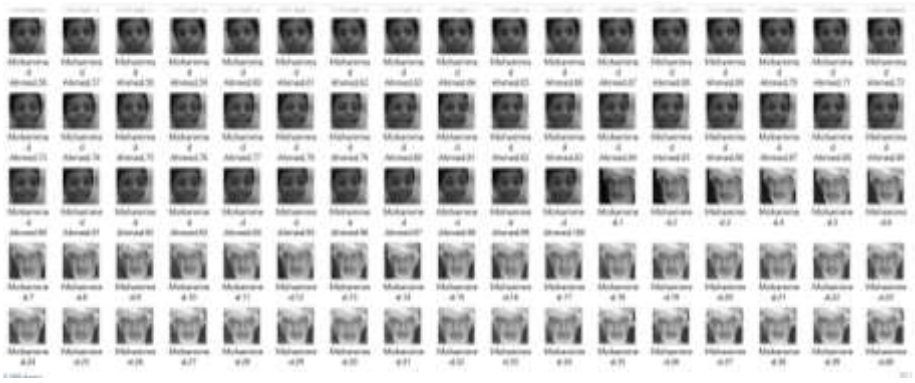

Fig. 2. Human face dataset

F. Data Preprocessing

Adding preprocessing layer to the front of VGG. Employ Include_top=False to remove the classification layer that was trained on the (ImageNet) dataset and establish the model as not trainable. Additionally, the transfer learning preprocess input function from VGG16 to normalize the input data has been used. VGG16 model is shown in Figure 3.

Model: "functional_1"

Layer (type)	Output Shape	Param #
input_1 (InputLayer)	[(None, 224, 224, 3)]	0
block1_conv1 (Conv2D)	(None, 224, 224, 64)	1792
block1_conv2 (Conv2D)	(None, 224, 224, 64)	36928
block1_pool (MaxPooling2D)	(None, 112, 112, 64)	0
block2_conv1 (Conv2D)	(None, 112, 112, 128)	73856
block2_conv2 (Conv2D)	(None, 112, 112, 128)	147584
block2_pool (MaxPooling2D)	(None, 56, 56, 128)	0
block3_conv1 (Conv2D)	(None, 56, 56, 256)	295168
block3_conv2 (Conv2D)	(None, 56, 56, 256)	590080
block3_conv3 (Conv2D)	(None, 56, 56, 256)	590080
block3_pool (MaxPooling2D)	(None, 28, 28, 256)	0
block4_conv1 (Conv2D)	(None, 28, 28, 512)	1180160
block4_conv2 (Conv2D)	(None, 28, 28, 512)	2359808
block4_conv3 (Conv2D)	(None, 28, 28, 512)	2359808
block4_pool (MaxPooling2D)	(None, 14, 14, 512)	0
block5_conv1 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv2 (Conv2D)	(None, 14, 14, 512)	2359808
block5_conv3 (Conv2D)	(None, 14, 14, 512)	2359808
block6_pool (MaxPooling2D)	(None, 7, 7, 512)	0
flatten (Flatten)	(None, 25088)	0
dense (Dense)	(None, 31)	777759

Total params: 15,492,447
Trainable params: 777,759
Non-trainable params: 14,714,688

Fig.3. VGG16 model summary

- **Set optimizer**

In the last decade, different optimization frameworks in assorted programming languages were improved. nevertheless, multi-objective optimization covered only partially. In general, the decision of a suitable framework for an optimization task is consider a multi-objective problem. Optimization is a critical component in deep learning to reduce errors, weights, and calculate

components of the loss function result. Classical optimization process is frequent until it reaches the minimum range of the loss function and maximum range of the accuracy function. Using Adam optimizer. [12]

- **Training the model**

Using human face dataset in different ages and features in CNN module training and testing to observe the variance of accuracy and loss and presenting the curves of accuracy and loss percentage results of training and validation.
Epochs is a hyperparameter that outlining the number times that the learning equation will work through the entire training dataset [13].

G. Performance evaluation

Using deep learning (CNN) algorithm for face recognition

Convolutional Neural Network is the most approved Deep learning algorithm for human face image recognition, feature extraction operations for instance image classification, pattern recognition, and other operations. There are many varieties of CNN algorithm. But essentially, two types are presented human face recognition to explain the CNN algorithm. One is feature extractor and the second is the classifier [14].

The name of (CNN) comes from a mathematical linear equation between two matrices defines as convolution. In CNN, one matrix represents the image, and the other is the kernel (operator). The image is an uncomplicated single channel (gray scale) or three channels (colour image) matrix, each notation in this grid is one pixel of the image. The dimension of the image framework is (AxBxC). Here, A=height, B=width and C=RGB channel for RGB colour image. The image grayscale channel contains just one, and the colour image channel contains three RGB colour channels. The kernel (operator) is a matrix that has dimension of (TxRxD). Here T and R are mandatory, however the most common kernels for example, Edge detectors or that operators which use 3x3 size kernel. Here, D means the depth distance or dimension of the kernel. The dimension of the kernel is analogous to the image colour channel. Figure 4. describes the basic architecture of CNN. In Human Face Recognition systems, CNN depicts an excellent performance and many models have been built from CNN algorithm architecture [15].

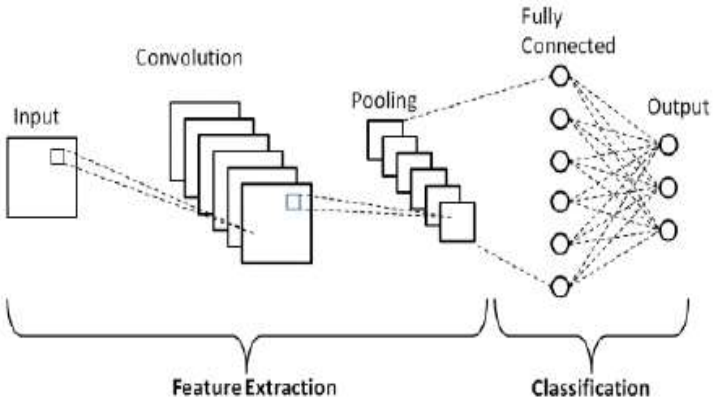

Figure.4. Basic CNN Architecture

Results of multi-face detection and recognition

This section of research paper discusses the results of recognition system starting from input the human face image until output the results of recognition. Here are some examples with images. The below table shows the preprocessing levels of human face detection and recognition. Table1 describes some results as examples of preprocessing steps of human face detection and recognition. The results are collected from different face angles, different image resolution and different distances.

Table 1: Examples of human face images

	Preprocessing steps	Results "examples of human face images"	
1	Input the face image, from different angles.		
2	Human face edge detection		
3	Detected human face		
4	RGB to gray conversion		

4. RESULTS AND DISCUSSION

Training the model

Assigning the length of training for the neural network by specifying the number of epochs to train.

Batch size: number of samples defines by hyperparameter used for each epoch [16].

Epochs: number of iterations times performed for training neural network algorithm through the entire dataset training process [16].

Human face dataset

Train the model using batch size=32 on 5 epochs and 10 epochs to check the accuracy of Human face dataset. Figure 5 and Figure 6 explains the training model of 5 epochs and 10 epochs of Human face dataset. Figures below depicts four-line plots, the top two plots showing the sparse cross entropy loss over epochs for the training (orange) and validation loss(blue), in addition to the bottom two plots showing classification train accuracy over epochs.

In this case, the plot curve shows a stable coverage of the model over training according to loss and classification accuracy, whereas the accuracy rate is dramatically increased and improved when the training epochs number are increased as figure below show that. The accuracy test is increased from 97.5% on 5 epochs to 99.1% on 10th epochs in human face dataset.

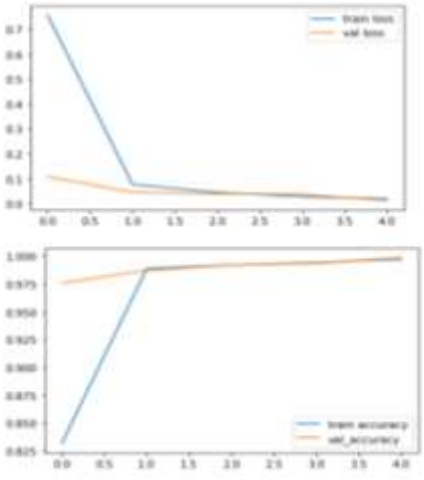

Fig5. *Train model on 5 epochs*

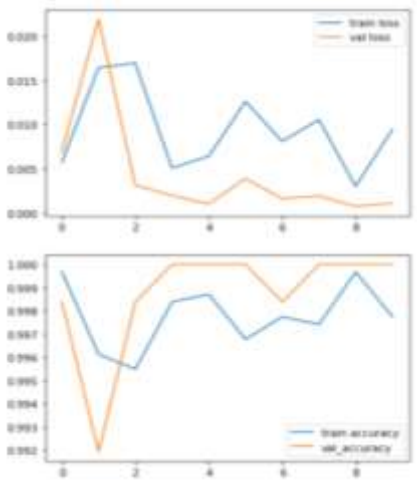

Fig. 6. *Train model on 10 epochs*

Train the model using batch size=32 on 15 epochs and 20 epochs to check the accuracy of Human face dataset. Figure 7 and Figure 8 respectively explains the training model of 15 epochs and 20 epochs of the dataset. It depicts four-line plots, the top two plots showing the sparse cross entropy loss over epochs for training result (orange) and validation(blue) and the other bottom two plots showing classification accuracy over epochs. In this condition, the plot diagram shows an adequate coverage of the model over training data according to train loss and classification accuracy. The accuracy ratio is increased and improved whenever the training epochs are raised. The test accuracy is increased from 99.01% on 15 epochs to 99.8% on 20 epochs in human face dataset model result.

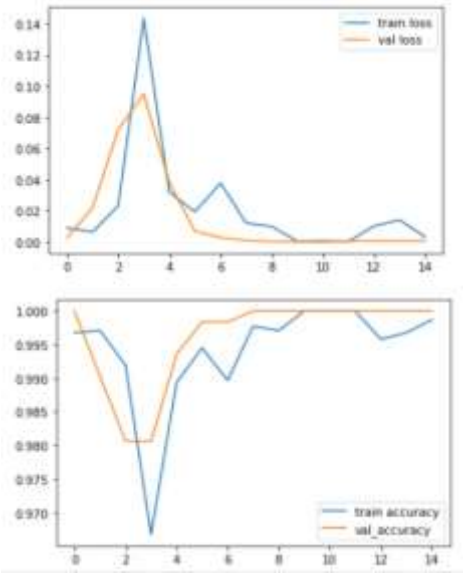

Figure. *7.* *Train model on 15 epochs*

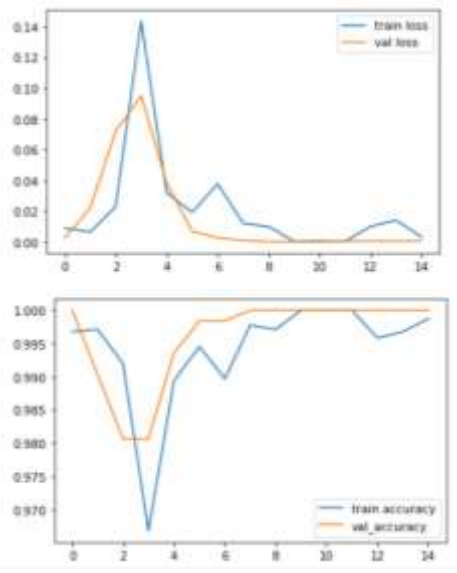

Fig. 8. *Train model on 20 epochs*

Prediction results

Prediction results of human face dataset

Figure 9 explains all metrics (precision, recall, f-score and support) with each 31 class of human face images in dataset and. The average metrics of 31 human face classes dataset is 95% in precision, 93% in recall and f1-score metrics.

	precision	recall	f1-score	support
AbdulAlmalik	1.00	0.33	0.50	3
AbdulAziz	1.00	1.00	1.00	3
Ahmed	0.50	0.67	0.57	3
Alin	1.00	1.00	1.00	3
Almotasim	1.00	0.67	0.80	3
Alreem	1.00	1.00	1.00	3
Assad	1.00	0.67	0.80	3
Azzan	1.00	0.67	0.80	3
Fadilah	1.00	1.00	1.00	3
Fahad	1.00	1.00	1.00	3
Faisal	1.00	1.00	1.00	3
Fatima	1.00	1.00	1.00	3
Halima	1.00	0.67	0.80	3
Ibrahim	1.00	1.00	1.00	3
IbrahimHamed	1.00	1.00	1.00	3
Ibtisam	0.75	1.00	0.86	3
Khalifa	1.00	1.00	1.00	3
Lila	1.00	1.00	1.00	3
Maha	1.00	1.00	1.00	3
Mohammed	1.00	1.00	1.00	3
Mohammed Ahmed	0.60	1.00	0.75	3
MohammedH	1.00	0.67	0.80	3
MohammedI	1.00	1.00	1.00	3
Moza	1.00	1.00	1.00	3
Muhra	1.00	1.00	1.00	3
Noor	1.00	0.67	0.80	3
Noorm	1.00	1.00	1.00	3
Rahma	1.00	1.00	1.00	3
Raja	1.00	1.00	1.00	3
Shaikh	1.00	1.00	1.00	3
Sultan	1.00	1.00	1.00	3
accuracy			0.93	120
macro avg	0.95	0.93	0.93	120
weighted avg	0.95	0.93	0.93	120

Fig. 9: *Accuracy metrics of human faces*

5. CONCLUSION

The aim of this research has been achieved through and investigation about the significance use of Multi-Face Recognition System in Surveillance Video. The fundamental objective was to explore the importance of machine and deep learning equations in the fields of artificial intelligence, algorithms pattern recognition, and image processing. The developed system has the ability to extract and recognize multi-face in Surveillance Video System. The system operates by inserting a captured fresh face image into the system, and if this person had his face captured and trained process before, the recognizer will make a prediction, and finally recognizes the person, and returning its name, shows how confident percentage the recognizer is with this match. The input image of human faces which captured from suitable and adequate distances are in .jpg format. Applying connected contours to extract the region of human face by using classifier function (facehaarcascades), passing some very important measurement parameters, as scale factor, number of neighbors, along with minimum size of the detected face. There were different preprocessing operations to re-size all the images, adding preprocessing layer to the front of VGG16, and convert colored captured images to grayscale images white and black for-instance brightness and contrast adjustments. The cost and optimization method to use (adam optimizer). Utilizing thresholding algorithm to extract the images in BGR color and gray mode. Using CNN algorithm in image recognition phase operation.

6. FUTURE WORK

Obviously, the result in this study suggests the dimensions for future research. First, there are lot of challenges and limitations that required to overcome. First challenge is diversity in nation's culture from one country to another, and one that needs to customize the software as per each country's policies. The second challenge is to allocate a cyber-security protection of the aggregated data, keeping with a novel and updated technology to manage and supervise it. Furthermore, to enhance the system, the essential future work will focus on the system's accuracy on obtained results and overcoming methodological challenges and limitations mentioned above. It`s great to mention that, this study is related to artificial intelligence applications. The prospective work could be to implement this system to use in real life applications by using novel artificial intelligence technologies with internet of things devices. Finally, comparing the accuracy measure performance between different captured image situations by using different image features collected from other locations such as workplaces, schools, and malls under different conditions and images, classification could be applied to evaluate and to investigate which situation will give accurate performance measurement rate.

7. REFERENCES

[1] INSTITUTIONS, F. (2019). Ten facts you didn't know about biometrics in 2019 | IDEMIA. [online] IDEMIA. Available at: https://www.idemia.com/news/ten-facts-you-didnt-know-about-biometrics-2019-2019-06-25 [Accessed 22 Oct. 2019].

[2] Das, R. (2014). Biometric Technology: Authentication, Biocryptography, and Cloud-Based Architecture. Boca Raton, FL: CRC Press.

[3] Maltoni, D., Maio, D., Jain, A. K., & Prabhakar, S. (2009). Handbook of fingerprint recognition. Springer Science & Business Media.

[4] Chandramohan, J., Nagarajan, R., Dineshkumar, T., Kannan, G., & Prakash, R. (2017). Attendance monitoring system of students based on biometric and gps tracking system. International Journal of Advanced engineering, Management and Science, 3(3).

[5] Paul, L. C., & Al Sumam, A. (2012). Face recognition using principal component analysis method. International Journal of Advanced Research in Computer Engineering & Technology (IJARCET), 1(9), 135-139.

[6] Devi, N. S., & Hemachandran, K. (2014). Face Recognition Using Principal Component Analysis. International Journal of Computer Science and Information Technologies, 5(5), 6491-6496.

[7] Jiang, T. X., Huang, T. Z., Zhao, X. L., & Ma, T. H. (2017). Patch-based principal component analysis for face recognition. Computational intelligence and neuroscience, 2017.

[8] Yang, M. H., Kriegman, D. J., & Ahuja, N. (2002). Detecting faces in images: A survey. *IEEE Transactions on pattern analysis and machine intelligence*, *24*(1), 34-58.

[9] Sadeghipour, E., & Sahragard, N. (2016). Face recognition based on improved SIFT algorithm. Int. J. Adv. Comput. Sci. Appl, 7(1), 547-551.

[10] Ken Nozaki, HisaoIshibuchi 1996. Adaptive fuzzy rule based Classification systems, IEEE transactions on Fuzzy systems, Vol.**4**, No.3.

[11] CUDA Toolkit 10.1 Update 2 available from <https://developer.nvidia.com/cuda-downloads> [2 September 2019]

[12] Sun, R. Y. (2020). Optimization for deep learning: An overview. *Journal of the Operations Research Society of China*, *8*(2), 249-294.

[13] Brownlee, J. (2018). What is the Difference Between a Batch and an Epoch in a Neural Network?. Machine Learning Mastery, 20.

[14] M. Coskun, A. Ucar, O. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in 2017 International Conference on Modern Electrical and Energy Systems (MEES), 2017, pp. 376–379.

[15] Fuad, M., Hasan, T., Fime, A. A., Sikder, D., Iftee, M., Raihan, A., ... & Islam, M. (2021). Recent Advances in Deep Learning Techniques for Face Recognition. *arXiv preprint arXiv:2103.10492*.

[16] Brownlee, J. (2018). What is the Difference Between a Batch and an Epoch in a Neural Network?. Machine Learning Mastery, 20.