

Deep Learning- Based Surveillance System using Face Recognition

Divya Kapil^{1*}, Aishwarya Kamtam^{1*}, Akhil Kedare^{1*}, Smita Bharne^{1*}

¹Ramrao Adik Institute of Technology, Navi Mumbai, India

Abstract. Surveillance systems are used for the monitoring the activities directly or indirectly. Most of the surveillance system uses the face recognition techniques to monitor the activities. This system builds the automated contemporary biometric surveillance system based on deep learning. The application of the system can be used in various ways. The face prints of the persons will be stored inside the database with relevant statistics and does the face recognition. When any unknown face is recognized then alarm will ring so one can alert the security systems and in addition actions will be taken. The system learns changes while detecting faces automatically using deep learning and gain correct accuracy in face recognition. A deep learning method including Convolutional Neural Network (CNN) is having great significance in the area of image processing. This system can be applicable to monitor the activities for the housing society premises.

1. Introduction

Privacy isn't the state of maintaining secrecy, is the state of being alone and not watched or disturbed by alternative people. Individuals are developing to be profoundly troubled and on edge of their surroundings attempting to guarantee that they stay secured consistently. Surveillance videos have a major contribution for the security purpose. CCTV cameras are implemented in all places where security is having much importance. A surveillance system consists of monitoring of different activities through the detection and recognition of faces. As a significant part of biometric distinguishing proof innovation, face acknowledgment innovation has the attributes of helpful procurement and high unwavering quality that has broadly utilized in the fields of data security, national security and traffic monitoring. Face recognition is the category of biometric programming that maps and people's facial highlights scientifically and stores the information as face print [3].

Hence the surveillance system which captures the individual user face, detect and compare with the existing user is the utmost requirement while monitoring. Thus, the face recognition techniques having great importance while monitoring through the surveillance system. The technique like Deep Convolutional Neural Network (CNN) algorithm is used to compare a live capture or digital image to the stored face print in order to verify an individual's identity. The convolution neural system can communicate the qualities of the chain of importance and can communicate the attributes of the

picture well. The present convolution neural system has been applied to protest classification, picture division, and video recovery and different fields.

In the profundity of learning, the convolution neural system is the best application. It can utilize an enormous number of labelled data, extract the attributes of various levelled structure, and can get the profundity of the picture articulation qualities, can be communicated all the more distinctive picture features [4].

2. Literature Survey

In paper [1] Xiujie Qu, et.al, proposed "Fast Face Recognition System" using Field Programmable Gate Array board with the help of Convolutional Neural Network that could recognize face instantly. It helped in increasing the computational speed of the network in real time processing system to give accurate results. Power consumption was more in Field Programmable Gate Array (FPGA) and the programmers did not have control over the optimization of code and the system was appropriately only for low quantity production.

In paper [2] Zhiming Li, et.al proposed "Double Supervision Face Recognition Based on Deep Learning" with the help of double supervised full convolution neural network to identify face. It was implemented using python and resulted in good practical value that could be applied in industrial fields. With the help of this system faces could be classified and then verified for the purpose of identification. It can extract many features of face and is superior compared to traditional method. This system consumed large amount of power which resulted in lowering the processing speed.

* Corresponding author: divyakapil98@gmail.com, aishu.kamtam@gmail.com, akhil.kedare@gmail.com, smita.bharne@rait.ac.in

In paper [3] Aniwat Juhong, et.al proposed “Face Recognition on Facial Landmark Detection” using Supervised learning with human annotated data which was used in detecting the facial landmarks and recognizing face. It was implemented using python to provide robust, fast processing system and less expensive. This system could not provide accuracy since it used very less database.

In paper [4] Sanjay Thakre, et.al, proposed “Secure Reliable Multimodal Biometric Face Recognition” using Principal Component Analysis for multimodal feature extraction to integrate the face image pixels intensity and local entropy of the image. The advantage of the system is for a static database is proven as an efficient application with very lower, negligible percentage of false acceptance and rejection rate.

From literature survey, it has been understood that the various systems are proposed along with the usage of one-of-the-kind algorithms with some advantages and limitations. The recent paper said that the system was made using Field-Programmable Gate Array (FPGA) which helped in determining the face pixels but due to this large amount of power was consumed. In paper [2] used double supervised full convolution neural network to identify face but it could work accurately and also it was good only for low quantity of production. In paper [3] used Supervised Learning with human annotated data to achieve the best classification but the results were always not accurate. In paper [4] used multimodal feature extraction to integrate the face image pixels intensity and local entropy of image but it resulted in inaccurate results. Considering all the above methods this system is generated using the FaceNet Model.

3. Proposed Methodology

3.1 System Design

Figure 1 shows the flow of proposed system. The user face is scanned and stored in the database. We first train the data set for our surveillance system using one shot learning using FaceNet [7][11]. The matches the faces and based on FaceNet is a pre-trained CNN which embeds the input image into a 128-dimensional vector encoding. It is trained on several images of the face of different people. Although this model is pre-trained. But it still struggles to output usable encoding for unseen data. We want this model to generate encoding such that there is less distance between encoding of the images of same person, and more distance between encoding of the different persons.

To achieve this goal images, we will train FaceNet model on Triplet Loss function [7][11]. This function takes three images as anchor points as positive and negative. and Positive means images of same person whereas negative means the image of a different person.

Now when a user tries to pass through the surveillance system his face image is captured. Then the image is fed

into the Deep learning CNN architecture. The output is 128D vector encoding. This output is then matched with the data set if it matches the user is allowed through the surveillance system else his entry is rejected.

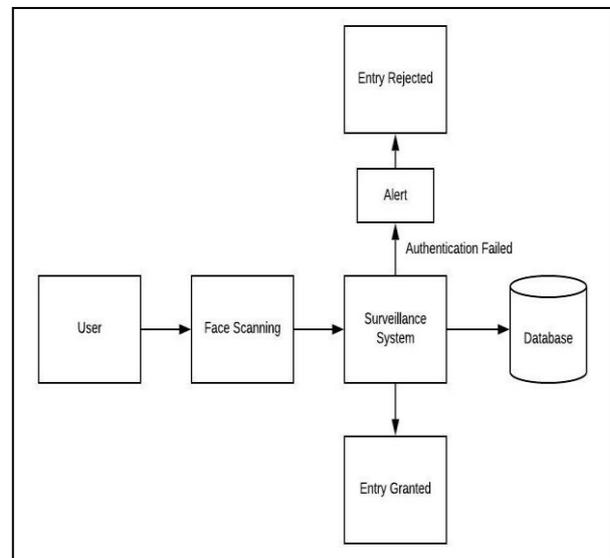


Fig. 1. Proposed System Design

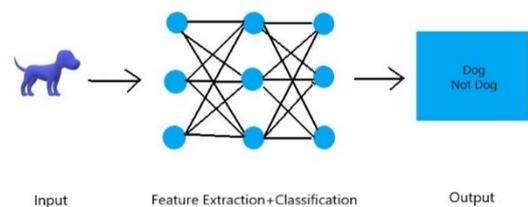


Fig. 2 : Deep Learning

Figure 2 shows the concept of deep Learning. A deep CNN algorithm is used to compare images stored in a face prints to verify individuals. The extracted features from these images can be used to train a classifier which can be verify the individual person’s identity. While registration it takes 50 images per person. This model takes 50 images for the purpose of registration. While capturing image of the user it gets train with FaceNet model. The FaceNet model extracts the features of the registered image and then gives to the CNN module for face recognition. CNN module will match the images with the registered images in the database and then identify whether the person is authenticated user or not.

3.2 FaceNet and Feature Extraction

The model utilized in our assessment is the FaceNet [7], which changes the face into 128D Euclidean space like word implanting. FaceNet is the pre-prepared CNN model. The FaceNet model having been prepared with triplet misfortune for various classes of countenances to catch the similitudes and contrasts between them, the 128-

dimensional installing returned by the FaceNet model can be utilized to groups faces successfully. It can be used to clusters faces effectively. Vectors of the space are created for recognition, verification and clustering may be easily implemented with the use of standard strategies with FaceNet. Smaller distance indicates closer for similar faces and similarly away for non-similar faces.[11] FaceNet utilizes a particular misfortune strategy called Triplet Loss to compute loss. Triplet Loss limits the separation between a grapple and a positive, pictures that contain the same personality, and expands the separation between the stay and a pessimistic, pictures that contain various characters.

Figure 3 shows the triplet loss mechanism. The triplet-based loss function used to learn the mapping of a classifier with same or different persons. It uses the network trained using aggregate of classification and verification loss. The verification loss they use is much like the triplet loss used to examine the mapping utilized by FaceNet in that it minimizes squared L2 distances between snapshots of faces from the same person and enforces a margin separating photographs of faces from a unique individual, however its unique in that only pairs of photographs are compared, whereas the triplet loss encourages a relative distance constraint by searching at three at a time[7][11]. We want to ensure that an image $x(a)$ of a specific person is closer to all images $x(p)$ of that same person than is closer to all other images $x(p)$ of that same person than it is to any image $x(n)$ of any other person by a margin where $x(a)$ is the anchor point of the input image, $x(p)$ is the positive point of the image which is similar or close to the input image and $x(n)$ is the negative point which is the dissimilar image.

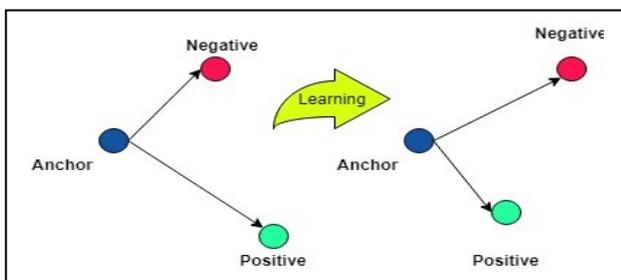


Fig. 3 : Triplet Loss

4. Results Analysis and Discussion

The dataset is used by the system where in one shot 50 frames per user is captured and then feed to the system. The sample images of some users are shown in fig 4(a). Total 2000 images were taken of 50 users and then trained into the model i.e. 4 classes were generated. The images of user2 were trained with spectacles

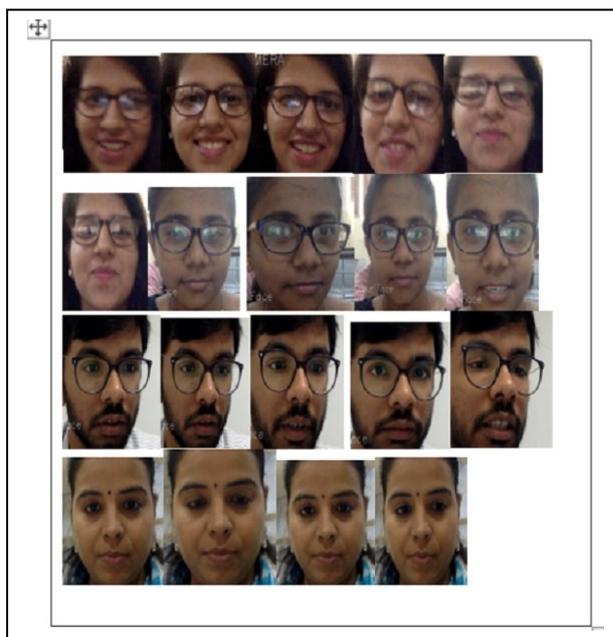


Fig. 4(a) : Input Data (Training data -Sample Images)

The images of user2 were then tested with different specifications such as with spectacles and without spectacles. The system is able to identify the faces with the learned specifications. Also using deep learning system is able to identify the faces with the variations of feature points.

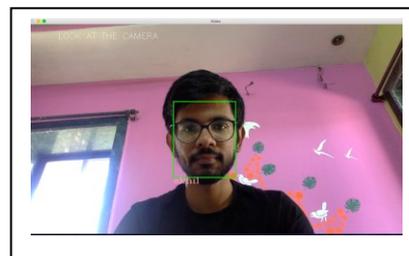


Fig. 5(a) : Test case1 with Glasses

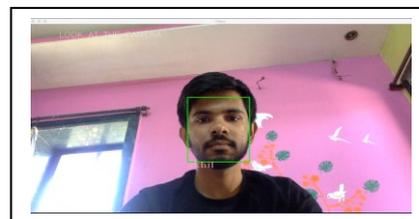


Fig. 5(b) : Test case1 without Glasses

The images of user-2 have been tested with different specifications such as with spectacles and without spectacles. Figure 5(a) and 5(b) shows the test case with glasses and without glasses respectively. Table 1 shows the accuracy results with glasses.

Table 1. Results for Facial recognition of User-2 with Glasses.

Test Cases	User1	User2	User3	User4	Accuracy
1.	0.03367	0.87418	0.06516	0.02698	87%
2.	0.03937	0.85495	0.06732	0.03833	85%
3.	0.04253	0.84192	0.07869	0.03685	84%
4.	0.04634	0.82040	0.10254	0.03071	82%
5.	0.03270	0.85782	0.07884	0.03062	85%

The results of the table 1 shows an average of 84.6% accuracy where the user-2 was with spectacles.

Table 2 shows an average of 68% accuracy where the user-2 was without spectacles.

Figure 7 shows the accuracy graph of the system. Results generated results by the proposed system is able to detect the faces with the different complexities. This feature makes our system more robust comparing with existing system. Literature survey shows the accuracy of existing system from 80 to 85 % [7][5]. In our system we are achieving accuracy up to 87% and also with the complex-additive features for same user up to 68%. For same user feature mention in table 1 and table 2 makes our system differentiable from the existing system.

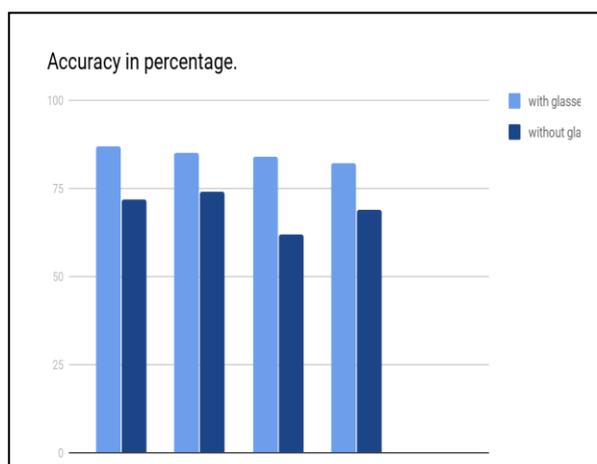


Fig 7: Accuracy Graph

Table 2. Results for Facial recognition of User2 without glasses

Test Cases	User-1	User-2	User-3	User-4	Accuracy
1.	0.0574	0.72486	0.0509	0.1667	72%
2.	0.0670	0.74153	0.0591	0.1323	74%
3.	0.0940	0.64932	0.0509	0.2056	64%
4.	0.0586	0.69176	0.0533	0.1963	69%
5.	0.0710	0.61811	0.0565	0.2602	61%



Fig 8: Surveillance system detecting multiple faces

Figure 8 shows the frame which is captured from a live surveillance system detecting and identifying authorised users.

The proposed system uses images of face from different angles to train the model giving it an upper edge over existing systems. The proposed system can work with normal CPU but can give better performance with GPU.

5. Conclusion

This work has been carried for recognizing facial images which can be used in different real time applications like forensics analysis, residential, system security etc. The photo can be highly contrasting or conceal for the above-built up usage. Considering housing society most of the information stored manually can be altered which leads to misleading actions. Thus, the system uses the deep learning which identifies unknown faces entering into the residential and maintain proper data regarding relevant information of the visitors and people residing in it. The proposed method tries to provide better solutions and helpful towards the betterment of the people

References

- [1] Xiujie Qu, Tianbo Wei, Cheng Peng, Peng Du, "A Fast Face Recognition System Based On Deep Learning", 2018 11th International Symposium on Computational Intelligence and Design.
- [2] Zhiming Li, Yongzhong Tang, "Double Supervision Face Recognition Based on Deep Learning", 2017 International Conference on Computer Systems, Electronics and Control (ICC- SEC)
- [3] Aniwat Juhong, C. Pintavirooj, "Face Recognition Based on Facial Landmark Detection", The 2017 Biomedical Engineering International Conference (BMEiCON-2017)
- [4] Sanjay Thakre, Ambikesh Kumar Gupta, Shilpi Sharma, "Secure Reliable Multimodel Biometric Fingerprint and Face Recognition", 2017 International Conference on Computer Communication and Informatics (ICCCI -2017).
- [5] Umme Aiman, Virendra P. Vishwakarma, "Face Recognition using modified deep learning neural network", IEEE – 40222, July 2017.
- [6] Virendra P. Vishwakarma, "Deterministic Learning Machine for Face Recognition with Multi-model Feature Extraction", 2016 Ninth International Conference on Contemporary Computing (IC3), Aug 2016.
- [7] "FaceNet Based Face Sketch Recognition",2017 International Conference on Computational Science and Computational Intelligence (CSCI) by Hyo Jong Lee, Weiguo Wan, Dec 2017.
- [8] Phat Nguyen Huu, Loc Hoang Bao, Hoang Lai The. "Proposing an Image Enhancemen Algorithm Using CNN for Applications of Face Recognition System", Journal of Advances in Mathematics and Computer Science, 2020
- [9] Thida Nyein, Aung Nway Oo. "University Classroom Attendance System Using FaceNet and Support Vector Machine", 201 International Conference on Advanced Information Technologies (ICAIT), 2019
- [10] P. Jothi Thilaga, B. Arshath Khan, A.A. Jones N. Krishna Kumar. "Modern Face Recognition with Deep Learning", 2018 Second International Conference on Inventive Communication and Computational Technologies (ICICCT), 2018
- [11] Guodong Guo, Na Zhang, "A Survey on Deep Learning Based Face Recognition"2019 Computer Vision and Image Understanding 189(2019) 102805.
- [12] by Qiangchang Wang, Guodong Guo. Benchmarking deep learning techniques for face recognition", J.Vis. Commun. Image R.65 (2019)102663.