2021 12th International Conference on Computing Communication and Networking Technologies (ICCCNT) | 978-1-7281-8595-8/21/$31.00 ©2021 IEEE | DOI: 10.1109/ICCCNT51525.2021.9580098

FACIAL RECOGNITION SYSTEM FOR AUTOMATIC ATTENDANCE TRACKING USING AN ENSEMBLE OF DEEP-LEARNING TECHNIQUES

Venugopal A, Rahul R Krishna,*Rahul Varma U
Department of Computer Science and Applications
Amrita Vishwa Vidyapeetham
Amritapuri, India
Email: venugopalasok2001@gmail.com, rahulrkrishna781@gmail.com, *rahulvarmau@am.amrita.edu

***Abstract*—Face recognition technology has made countless contributions in improving and changing the world. Attendance Systems utilizing Real-Time Face Recognition technology is a solution that can efficiently carry out the procedure of marking student attendance. Face recognition-based attendance system is the process by which we mark the attendance of the students present in the classroom by utilizing the facial data that is acquired from a surveillance camera. The proposed system captures the face of students attending the lecture by first detecting a face from the video input and with the help of an ensemble of deep learning models recognize the student and mark his/her attendance in the database [1]. This system uses an ensemble of facial recognition models such as VGG-FACE, Facenet, Openface, DeepFace so that it may be able to yield a much higher accuracy while identifying the subject.**

***Index Terms*—Facial Recognition, Attendance System, Face Detection, Deep Learning, Ensemble Learning, VGG-Face, OpenFace, Facenet, DeepFace**

I. INTRODUCTION

Face recognition technology is utilized in many different fields of education. Here we use the same technology to efficiently automate and manage the attendance system. Attendance is considered as an important factor for a student as it is considered as a part of his/her academic record during evaluations. It is a difficult task to manage, with the continuous increase in the number of students that enroll into the institution, but with the help of such technologies we can easily overcome this hurdle. Such digital attendance system are really helpful as they are an appropriate system for approving and maintaining the attendance record and these systems are consistent.

Generally the process of attendance Tracking can be categorized into two types:

1.Manual Attendance Tracking:

Manual student attendance can be defined as the type of attendance marking which is usually done by a person that is the teacher. It can be considered as a time-consuming process. Such a method can also be termed as unreliable as the attendance record of some students may be missed or incorrectly marked due to human error, there are also cases of students intentionally calling attendance of others even if they are not present in the classroom.

To solve such type of issues we make use of an automatic attendance system or (AAS).

2.Automated Attendance System:

In automated attendance systems we utilize the help of computers and cutting-edge image processing and facial recognition technology to process the attendance of students that are present in a particular classroom. This is achieved with the help of machine learning, machine learning is the process by which a machine can train itself with the help of datasets and a process known as supervised learning. A similar implementation of an automated attendance system that uses deep learning can be seen in [2], [3] shows another approach to an automated attendance using a combination for facial and fingerprint biometrics for identification. But in this implementation we will be focusing on an ensemble of deep learning techniques with which we can increase the accuracy of facial recognition.

II. LITERATURE SURVEY

A. Ensemble Learning

Ensemble learning is a method by which we increase the accuracy of prediction or classification by collectively using the data from multiple models or classifiers. These models are setup together in a way that they may be able to solve a particular type of computational problem. Ensemble learning is primarily used to improve the classification, prediction of an implementation. The works done in [4] give an in-depth and thorough explanation as to how ensemble models work.

B. LightGBM

LightGBM is short for Light Gradient Boosting Machine, it is used to increase the accuracy of classification or prediction with the help of a gradient boosting framework. As the name suggests the model is light so it is a fast, it uses a tree based learning algorithms known as gradient boosting, which can

used to create a strong prediction models that are an ensemble of multiple weak ones.

C. Convolution Neural Network(CNN)

CNN or convolution neural networks is a deep learning neural network that is used to analyze visual imagery. These algorithms work by converting the images into feature maps. Most of the latest and advanced facial recognition models and systems use CNN for facial recognition. We can see from [5] and of review section of [6]on how CNN based recognition models are much more faster and efficient than other methods shown in it.

D. EXISTING RECOGNITION SYSTEMS

- Fingerprint ID Attendance System: Fingerprint recognition is the process of identifying an individual by comparing two fingerprints. When the individual scans his fingerprint the unique attributes of his fingerprint is extracted and compared with a dataset to identify the person. Fingerprint recognition is the most commonly used example of automated attendance systems. An example of an attendance system based on this method can be seen here [7].The problem with such a system is that it is expensive to implement and maintain such scanners and, it also over-complicates the whole process.
- RFID Attendance System: RFID based attendance system has an RFID reader, RFID Tag, that helps in the establishing the identity of students. In such a system the students will carry around such an RFID tag that they can scan while entering a class or while exiting one [8]. The problem with this type of system is that if the student loses the RIFD tag it can cause problems and there is no check whether the actual owner of the RFID took the tag through the scanner.
- Iris Recognition Attendance system: Iris recognition is the process of identifying a person by using mathematical pattern recognition algorithms on the individuals eyes. These algorithms look for certain unique pattern's around the iris, pigments in the pupil etc to uniquely identify the individual. Theses systems require a lot of expensive equipment to be implemented and it results in over-complicating the whole ordeal of attendance tracking. The design and implementation of such a system can be referred from [9]
- Similar Implementations: Most similar implementations of facial recognition-based attendance systems can only identify students one at a time, this causes a problem during attendance tracking of a huge number of students. This creates problems similar to the one's we can see for the other recognition systems like fingerprint based and RFID based. Hence it becomes a hassle to take the attendance after every lecture. The proposed system is capable of recognising students while they are attending the lecture and it can identify multiple students at the same time that too with a high accuracy. The system also utilises an ensemble of multiple facial recognition models rather than the single model approach so as to achieve high accuracy in recognition.

III. PROPOSED SYSTEM

The proposed system is used to capture the face of students attending the lecture and store it in a database. The face of the subject must be captured in such a way that all the unique feature of the face such as the size of their eye's, the distance between the two eye's, the distance between the eye's and nose of the person etc. And by attaining the unique feature the system is able to identify the subject and mark his/her attendance onto a digital database. Similar systems and implementations can be seen in [10] [11]and [12].

A. Video Capture Module

In this module a camera is used to capture image and video data of a student to be processed and analyzed for recognition. Image Processing is the procedure of manipulating the image such as changing its tones converting to gray-scale etc, in order to get an enhanced image so as to extract useful information from it. The captured video is run through an algorithm that converts the video into frames per second and each frame is then converted into feature maps for feature extraction.

B. Face detection and Facial Recognition Module

In this module the system takes in the processed image and runs it through a face detection model to separate the face portion from different parts of the body and so as to use facial recognition algorithms on the image. We can see multiple methods and techniques by which facial detection can be achieved from [13] [14]and [15]. The face is uniquely identified from the processed data using a technology known as deep learning, in this implementation we use an ensemble of multiple deep learning methods to uniquely identify the face of a person in an image. There are multiple ways or algorithms that have already been developed for facial recognition and there are multiple ways in approaching this problem as referred in [16] and [17]. We use an ensemble of facial recognition models in this implementation to increase accuracy of recognition.

C. Attendance Module

This is the module that is tasked with the function of attendance tracking. The system must store this data in a way that it can be easily accessible to the authorized personnel and it can be easily understood or interpreted by them. Storing this data can be as simple as a csv file or complex as cloud based databases.

IV. SYSTEM ARCHITECTURE

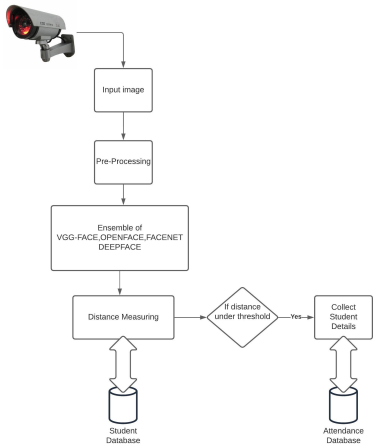

Fig. 1. Overall Working of Proposed System

V. SYSTEM DESCRIPTION

A. Video Capturing and Processing

The following model uses a simple camera for taking a video input. We use OpenCV library to handle the conversion of the video input to frames, we can see an illustration of this process in Fig:2. We also use OpenCV to display the data directly onto the video frame. The experiments conducted in [18] show how well OpenCV library works for video and Image processing. Once the video is converted to frames the model checks whether a face is present in the obtained frame. If a face is detected then it extracts the embedding's from the face and passes it on to the Neural network for further processing and identification.

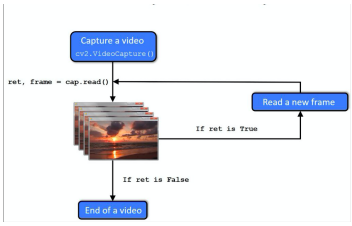

Fig. 2. Video to frame Conversion Using OpenCV

B. Feature Extraction from Image

There are multiple ways in which we can obtain the face encoding's/unique attributes of the face from the image. We have used a simple python library to test the accuracy of the recognition using the facial recognition library. We have also used multiple other neural network models such as VGG-FACE, OpenFace, Facenet etc. Different models have multiple and unique methods in identifying or recognizing a face, some process images in its data sets and extract encodings from it which are later used to compare with the test image to find that match. Some use even more efficient methods such as converting the unique features of the face into vectors as we can see in Fig:3 and assessing the cosine similarity between them or euclidean distance between these vectors, and the pair that shows the least amount of distance or is under a particular threshold of distance is identified as the matching pair. [19] shows a system that uses euclidean distance classification for emotion classification of image and how it classifys images based on distance between input image and the image in the dataset.

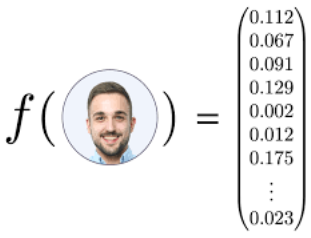

Fig. 3. Models Convert Facial Features to Vectors

C. Attendance Logging

Once a particular pair of images have been found, the image in dataset is saved with name of the student, the number of image (since there are multiple images of the same person in the dataset), for reference student name.jpg we can use the OS python package and its functions to extract the names of the students and pass it onto the database for recording the attendance. For reference purposes we have given an example of a sample attendance logging that is done on a simple .csv file in Fig:4.

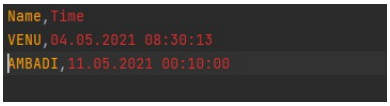

Fig. 4. Sample Attendance Record On Excel File

VI. ALGORITHMS

A. VGG-FACE

VGG-Face is a CNN-based facial recognition model. The model consists of 22 layers and 37 deep units as we can see illustrated in the Fig:5 below. VGG-Face is a collection of multiple models that were developed by the Visual Geometry Group at Oxford University. The model uses keras to construct a face model from the input image. When we process and image using the VGG-FACE model we get a vector representation of the image as a 2622 dimensional vector. These vectors are later used to assess similarity using distance measuring techniques.

B. OPENFACE

OpenFace is a very lightweight model when compared to other similar face recognition models. OpenFace model uses RGB images as its input, after processing the image the model

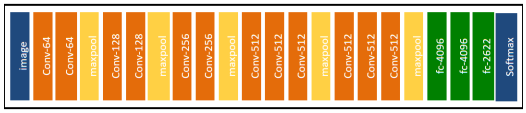

Fig. 5. VGG-FACE Model Multiple Layers

represents the image in a 128 bit dimensional hypersphere. This makes further classification or similarity detection tasks much more easier. Eventhough the model has a hard time measuring up to heavyweight models like VGG-Face, Facenet it is still very effcicient and accurate.The Openface model is very minimalist to the point where its training weights or test weights are usually much smaller in size when compared to VGG-FACE. Hence it is much faster.

C. FACENET

Facenet is another CNN based facial recognition model similar to VGG-FACE and OpenFace. Facenet was developed by google and is a much more advanced model than VGG-FACE. Eventhough the facenet also employs CNN for feature extraction. But it is diffrent from other such models in a way that the model trains to create face embeddings, rather than extracting them from an intermediate layer. These embeddings are then passed onto the triplet loss function to calculate the loss as we can see in the Fig:6. The goal behind this process is to decrease the distance between two images that have the same identity and increase the distance between two images that have different identities.The model has a 128bit output similar to OpenFace.

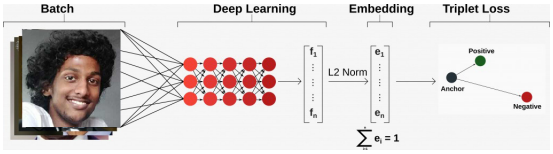

Fig. 6. Facenet Model

D. Facebook DeepFace

DeepFace is a facial recognition model that was designed and created by Facebook, it uses deep learning to process images. The model uses over 120 million connection weights for processing the image. It is one of the most advanced and most powerful facial recognition models in existence. It boasts accuracy rates of 97.35 percentage on the LFW dataset. The model works by creating a 3D model of the face present in the image so that it becomes a forward looking face even if it is in a diffrent orientation in the input image. Then it uses an 8 layered CNN to extract the face embeddings from forward-looking face 3D model. Fig:7 shows an illustration of the models working. Eventhough the number of layers in this model is less, the parameters size is much larger than other models.

Fig. 7. Recognition Model Extraction Facial Features From Input Image

E. Deepface-Ensemble

This is a custom facial recognition model that uses an ensemble of multiple models and distance metrics and feeds these distance metrics as features to a boosting method to radically increase the overall accuracy of the overall prediction of the subject. It utilizes the help of models which are VGG-FACE, OPENFACE, FACENET and Facebook Deepface paired with distance metrics such as Euclidean Distance, Cosine Similarity, Euclidean L2, with an overall output of 12 model-metric pairs. The multiple distance measures that we get form these pairs are then passed onto a LightGBM model as features. An implementation of a similar system can be seen in [20] where ensemble methods can increase the overall efficiency of a model. We can see the working of this model illustrated in Fig:8.

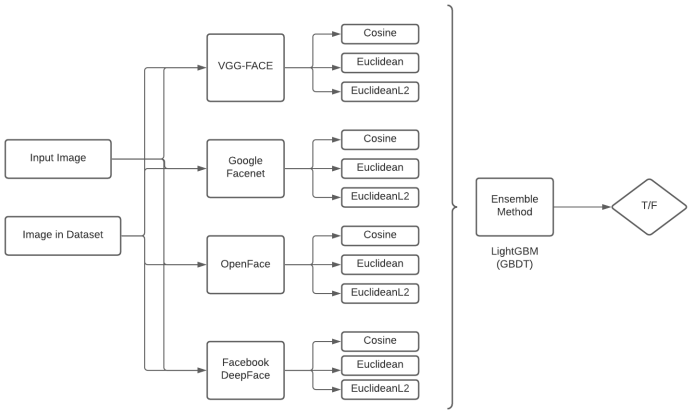

Fig. 8. Deep Face Ensemble Model Working

VII. Experimental Results

We tried implementations that used the basic Facial recognition library that is available in python packages that uses HOG(Histogram of Gradients method to detect the face form the image and process it to attain the face encoding's from both the images in the dataset and the image in real-time video input which is then used to compare with each other. Finally using the SVM classifier we attain the name of the person from the database that has distance measures closest to the image and it classify's that image in the video frame as that particular person as seen in Fig:9. Finally we can see the the multiple distance measures between the two images in Fig:10

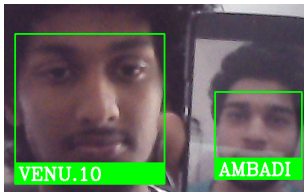

Fig. 9. Realtime Facial Recognition

```
[0.60903882 0.5700143  0.45006412 0.45713803 0.47129129 0.49819425]
[0.57482982 0.54323846 0.42660668 0.42455179 0.46586762 0.44260174]
[0.58664041 0.55717918 0.44280392 0.46512533 0.48768569 0.47757757]
[0.59156472 0.55692511 0.43500809 0.42538607 0.45535456 0.4360217 ]
```

Fig. 10. Distance Measures of Images In The Frame

In the second implementation we used a multiple number of models or an ensemble of models that converts images features to vectors and these vectors are compared with respect to the distance between them. For this purpose 3 distance metrics are used that is Cosine Similarity, Euclidean Distance and EuclideanL2. With the help of these we get distance measures from 12 model-metric pairs which is passed as features onto a LighGBM model that is pre-trained for binary classification of the image to True or False, True in the case that the prediction is above a particular threshold and False when the prediction is below that threshold. After testing the accuracy of this model with two diffrent images of the person we got the result as true and the score for the accuracy of the image calculated by the model in Fig:12. The test images are that were used to yield this result can be seen in Fig:11.

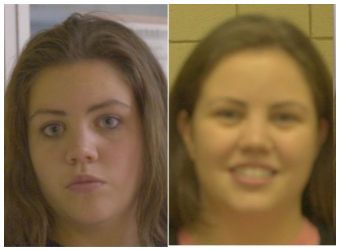

Fig. 11. Images Used For Testing Model Accuracy

```
{'verified': True, 'score': 0.995903915166032,
```

Fig. 12. Ensemble Model Output Of The Image

VIII. CONCLUSION

A. Remarks

The first implementation is a simple method by which we can achieve this particular task the methods used and models used are too simple and it is too slow for real world application the overall accuracy of the predictions differs greatly with respect the overall lighting in the image the quality of the footage etc.

The second implementation works well and is accurate for multiple lighting's, angle's, of the same image but due to processing required by multiple models to achieve this accuracy the compute time is bigger and is harder to run real-time recognition when compared to other single model recognition systems.

B. Future Enhancements

To implement a much more effective system that has the a better balance between accuracy and processing time we must re-evalute the ensemble of models and decrease the number of model metric pair.

An Ensemble model that would have a much better accuracy-processing time balance would be an ensemble of VGG-FACE and Facebook DeepFace paired with cosine similarity and euclidean distance model-metric pair. Therefore decreasing the number of model metric pairs from 12 to 4.

REFERENCES

[1] Tata Sutabri, Ade Kurniawan Pamungkur, and Raymond Erz Saragih. Automatic attendance system for university student using face recognition based on deep learning. *International Journal of Machine Learning and Computing*, 9(5):668–674, 2019.

[2] Sikandar Khan, Adeel Akram, and Nighat Usman. Real time automatic attendance system for face recognition using face api and opencv. *Wireless Personal Communications*, 113(1):469–480, 2020.

[3] S. Anand, Kamal Bijlani, Sheeja Suresh, and P. Praphul. Attendance monitoring in classroom using smartphone wi-fi fingerprinting. In *2016 IEEE Eighth International Conference on Technology for Education (T4E)*, pages 62–67, 2016.

[4] Omer Sagi and Lior Rokach. Ensemble learning: A survey. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 8(4):e1249, 2018.

[5] Suleman Khan, M. Hammad Javed, Ehtasham Ahmed, Syed A A Shah, and Syed Umaid Ali. Facial recognition using convolutional neural networks and implementation on smart glasses. In *2019 International Conference on Information Science and Communication Technology (ICISCT)*, pages 1–6, 2019.

[6] Sudha Sharma, Alpesh Soni, and Vijay Malviya. Face recognition based on convolution neural network (cnn) applications in image processing: a survey. *Proceedings of Recent Advances in Interdisciplinary Trends in Engineering & Applications (RAITEA)*, 2019.

[7] Khin San Myint and Chan Mya Mya Nyein. Fingerprint based attendance system using arduino. *International Journal of Scientific and Research Publications*, 8(7):422–426, 2018.

[8] Unnati Koppikar, Shobha Hiremath, Akshata Shiralkar, Akshata Rajoor, and V. P. Baligar. Iot based smart attendance monitoring system using rfid. In *2019 1st International Conference on Advances in Information Technology (ICAIT)*, pages 193–197, 2019.

[9] Kennedy O. Okokpujie, Etinosa Noma-Osaghae, Olatunji J. Okesola, Samuel N. John, and Okonigene Robert. Design and implementation of a student attendance system using iris biometric recognition. In *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*, pages 563–567, 2017.

[10] Shrija Madhu, A Adapa, Ms Visrutha Vatsavaya, and Ms Padmini Pothula. Face recognition based attendance system using machine learning. *Int J Mang, Tech and Engr*, 9, 2019.

[11] Shubhobrata Bhattacharya, Gowtham Sandeep Nainala, Prosenjit Das, and Aurobinda Routray. Smart attendance monitoring system (sams): A face recognition based attendance system for classroom environment. In *2018 IEEE 18th International Conference on Advanced Learning Technologies (ICALT)*, pages 358–360, 2018.

[12] Shreyak Sawhney, Karan Kacker, Samyak Jain, Shailendra Narayan Singh, and Rakesh Garg. Real-time smart attendance system using face recognition techniques. In *2019 9th International Conference on Cloud Computing, Data Science Engineering (Confluence)*, pages 522–525, 2019.

[13] Xudong Sun, Pengcheng Wu, and Steven CH Hoi. Face detection using deep learning: An improved faster rcnn approach. *Neurocomputing*, 299:42–50, 2018.

[14] Changzheng Zhang, Xiang Xu, and Dandan Tu. Face detection using improved faster rcnn. *arXiv preprint arXiv:1802.02142*, 2018.

[15] Neethu A., Athi Narayanan S., and Kamal Bijlani. People count estimation using hybrid face detection method. In *2016 International Conference on Information Science (ICIS)*, pages 144–148, 2016.

[16] Kennedy Chengeta and Serestina Viriri. A survey on facial recognition based on local directional and local binary patterns. In *2018 Conference on Information Communications Technology and Society (ICTAS)*, pages 1–6, 2018.

[17] Madan Lal, Kamlesh Kumar, Rafaqat Hussain Arain, Abdullah Maitlo, Sadaquat Ali Ruk, and Hidayatullah Shaikh. Study of face recognition techniques: a survey. *International Journal of Advanced Computer Science and Applications*, 9(6):42–49, 2018.

[18] Nataliya Boyko, Oleg Basystiuk, and Nataliya Shakhovska. Performance evaluation and comparison of software for face recognition, based on dlib and opencv library. In *2018 IEEE Second International Conference on Data Stream Mining & Processing (DSMP)*, pages 478–482. IEEE, 2018.

[19] Mangala BS Divya and NB Prajwala. Facial expression recognition by calculating euclidian distance for eigen faces using pca. In *2018 International Conference on Communication and Signal Processing (ICCSP)*, pages 0244–0248. IEEE, 2018.

[20] Aloysius Neena and M Geetha. Image classification using an ensemble-based deep cnn. In *Recent Findings in Intelligent Computing Techniques*, pages 445–456. Springer, 2018.