Advances and Applications in Mathematical Sciences
Volume 20, Issue 3, January 2021, Pages 459-469

SMART ATTENDANCE PORTAL USING FACIAL RECOGNITION

HIMANSHU TRIPATHI, AKHILESH KUMAR SRIVASTAVA and ATUL BISHT

Computer Science and Engineering Department
ABES Engineering College
Ghaziabad, U.P. 201009, India
E-mail: himanshu.16bcs1193@abes.ac.in
akhilesh.srivastava@abes.ac.in
atul.16bcs1085@abes.ac.in

Abstract

Automated face recognition technology (AFR) has made a lot of improvements in this evolving world. Smart Attendance with the help of Live Face Recognition is a real-world solution that comes with the everyday tasks of managing attendance of people. Attendance system based on facial recognition is the process of marking the presence of students with the help of live data form the high definition video and other information sensors. In this facial recognition-based attendance project, the system will detect and recognize the human faces and mark their attendance with surveillance video camera. Many types of techniques and algorithms are developed for improving accuracy and performance of facial recognition technique, the technique which is used here is Deep Learning. It works by converting video frames to still images, then the face is detected and recognized from this image and attendance is marked in the database automatically.

I. Introduction

Here a method is proposed which marks the attendance of student automatically which it obtains by monitoring the classroom with camera. Continuous monitoring helps in good estimation and increases the performance of the attendance system. Attendance is marked by monitoring the positions and image of face. By monitoring the classroom, the system

2010 Mathematics Subject Classification: 62H35, 68T10, 94A08.
Keywords: Facial Recognition, Facial detection, Spoof Detection, Convolutional neural networks.
Received May 25, 2020; Accepted August 5, 2020

computes the position where the student is seated and location with respect to classroom which is used to mark the attendance. This system will continuously take images of students sitting in the classroom and mark the attendance of them. The main differentiating feature of this system is its reliable spoof detection which will reduce the cases of false attendance.

A. Objective

It is proposed that the faces of students will be fed into the database. Now when the student is attending the class, the face of the student needs to be captured in high-resolution so that maximum number of features of the face of each student can be captured. This cut short the need of manual attendance because the system continuously sends the image frames and after further processing, the faces are matched to the database and the attendance is marked. This attendance is assessable over web portal and will have proper management and maintenance apparatus.

B. Problem Identification & Definition

The authors identified that there is an immense scope of application of Facial recognition and Deep learning in the product based market where we can use it to solve many problems. It was observed that maintaining attendance was the sort of job that could be automated efficiently using facial recognition. Therefore the problem was defined to detect and identify faces for the students entering the classroom and then mark their attendance. The attendance would be managed by the online portal, which would make it very easy for students and teachers to manage it.

II. Literature Survey

In Paper [1], The Authors are focusing on detecting faces from surveillance cameras, when a large number of faces will be in single frame. Especially because in this use case the camera will be installed in classrooms, where a large number of students will be present in single frame, due to which resolution for a single face will be very small.

A critical point is that the database typically contains features extracted from high resolution images while the probes, taken by surveillance cameras, can be at a very low resolution. A breakthrough occurred in 2012 when AlexNet was created and won the Image Net Large Scale Visual Recognition Competition (ILSRVC) improving upon the state-of-the-art by a noticeable margin. Since then, CNN-based methods for image retrieval received considerably more attention from the scientific community From a theoretical perspective, the inner layers of a CNN realize an abstraction of the input that describes specific concepts contained inside the data. Moreover, due to the typical structure of deep models architecture, inner layers combine the information available from previous layers thus achieving a higher level of abstraction that summarizes the overall content of the input data.

Common Space Projection techniques concern the ability of a neural networks to minimize the distance among deep features, in a common space, extracted from a low resolution image and its high resolution counterpart.

The use of Multi-task Cascaded Convolutional Neural Networks (MTCNN). This step is performed once for each input frame. After all the faces have been identified from the picture, they are cropped, pre-processed and then used as input for the feature extractor. The preprocess step includes the rescaling of the image IJB-B dataset [17]. Specifically, the creators tried the model against the 1:1 confirmation convention intending to evaluate its capacity to extricate discriminative highlights.

In Paper [2] The creators talk about the 3 layered system model that was utilized to advance and improve the face location system. For face confirmation, given a couple of countenances, the two face highlights are looked at utilizing a closeness metric. L2 distance and cosine similarity are the two most usually utilized measurements for contrasting two face feature portrayals.

A single shot detector (SSD) is made from the truncated version of VGG-16 model for object detection. The SSD networks will generate the fixed number of image scores and bounding boxes. Over these extra layers were added to perform multi-scale detection.

The creators, resize the information picture to such an extent that the side with least length has an element of 512. After each convolutional square, to reduce the feature maps by 1/2 and increase the stride by 2 times max-pooling is applied. For example, conv3 layer gives out the feature maps which have the min- spatial size of 128. Moreover, 4 pixel stride in the original picture is equal to 1 pixel stride in the particular layer. As appeared in paper, beginning layers of a DCNN have feature maps with low stride, which is advantageous for identifying modest countenances since little size faces can be coordinated with 0.5 a high Jaacard overlap.

Paper [3] presents a brought together framework for face confirmation (is this a similar individual), acknowledgment (who is this individual) and bunching (discover ordinary citizens among these appearances). This technique depends on learning an Euclidean type embedding per picture utilizing a deep CNN. The system is prepared with the end goal that similarity between the faces is indicated by L2 distance. Appearances of a similar individual have little separations and countenances of unmistakable individuals have enormous separations.

$$d(p, q) = d(q, p) = \sqrt{(q_1 - p_1)^2 + (q_2 - p_2)^2 + \ldots + (q_n - p_n)^2}$$

$$= \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2}.$$

After the creation of embedding, at that point the previously mentioned errands become straight-forward: face check essentially includes thresholding the separation between the two embeddings; acknowledgment turns into a k-NN arrangement issue; and bunching can be accomplished utilizing off-the rack procedures, like agglomerative clustering or kmeans algorithm.

The most significant piece of this methodology lies at start-to-end learning of the entire framework. To this end the creators utilize the triplet loss that straightforwardly reflects what we need to accomplish in face confirmation, acknowledgment and bunching. To be specific, it makes progress toward an implanting $g(t)$, from a picture t into a component space Fs, with the end goal that the squared separation between all faces, free of imaging conditions, of a similar personality is little, though the squared separation between a couple of face pictures from various characters is huge. Despite the fact that the authors didn't directly contrast with different losses, for example the one utilizing sets of positives and negatives, they accept that the triplet loss is progressively appropriate for face confirmation. The inspiration is that the loss puts all the face data of one character to be anticipated onto a solitary point in the embedding space. The triplet loss, in any case, attempts to authorize an edge between each pair of face data from one individual to every other face. This permits the appearances from one personality to live on a complex, while as yet authorizing the separation and subsequently discriminability to different characters.

In Paper [4], Broad investigation on face spoof analysis is introduced. There are 2 major fundamental parts that is descriptors and the other one is classifiers. This organized study additionally brings a similar presentation investigation of the works considering the most significant open informational collections in the field. The philosophy followed right now especially pertinent to watch transient advancement of the field, drifts in the current methodologies, to talk about still opened issues, and to propose new viewpoints for the fate of face spoof recognition.

A. Types of face spoofing attacks:

Most people have their photo available online. Therefore, Printed photo is the most common with a lot of possibilities that it can sabotage the system.

People try to mimic blinking behavior with eye-cut photo attacks. Here, eye regions are removed.

Warped photo attacks here the photograph is printed and then bent and to any direction to pass spoof.

Video playback attack is targeted so that behavior of a real face can be mimicked and fools the system by detailing the intrinsic features which would be of some valid user.

There are 2 types of Mask attacks-one where is exactly as Humans face, Other where mask is made of paper. The target of these are spoof detection systems as these analyze 3-D face details, this makes the attack very difficult to identify.

B. Overview of Descriptors:

Texture of the given face: Textual features are extracted and because the printed material portray different textures that real image photograph, it can be detected by algorithms.

Motions of face-

Such as eye blinking, facial expressions, mouth development, head rotation etc.

Analysis of how the face interact with the environment.

Frequencies in face: as the image is not a real photograph, in the extensive analysis it can be observed that micro frequencies are not the same.

Color of the face. In the maximum number of scena5rios, the color of the face is not consistent due to changing light intensity of the environment. But there are some parameters which can be analyzed to differentiate between real face and imposter face.

Shape of the face: It is observed that the plainer surface geometry is different in a printed photograph from the original image set.

Reflectance of face: We can observe in the real life that on capturing the image of a printed photograph, we find that the light is reflected in a different way than the light from the original image of the face. This analysis can be used to distinguish between images.

Paper [5] proposes to solve the problem of spoof detection from a texture investigation perspective. To be sure, the texture analysis of the printed photograph can be done which helps us to find out defects in printing. The suggested methodology breaks down the details of surface of the photographs with the help of technique called as LBP or multi-scale local binary pattern. Besides, human faces and printed materials reflect light in various manners on the grounds that are-

A face is rigid object it does not changes and the texture defects are different.

A photograph of the face photo has a different texture structure which can be observed a 2d unbending article.

The authors have proposed a powerful methodology which is more efficient and does not involve any interaction from the client. This technique embraces the local binary patterns, an incredible texture analyzer, for depicting the small microsurfaces as well as their planer data. The vectors in the feature space are then given as input to a SVM classifier which decides if the micro-scale surface examples portray a individual or a phony picture. Figure 1 displays instances of 2 pictures (real face and a photograph of photograph) in the first one we have real photograph and its LBP analysis and then we have printed (fake) photograph with LBP analysis. Though the faces look same their corresponding LBP analysis show a lot of difference.

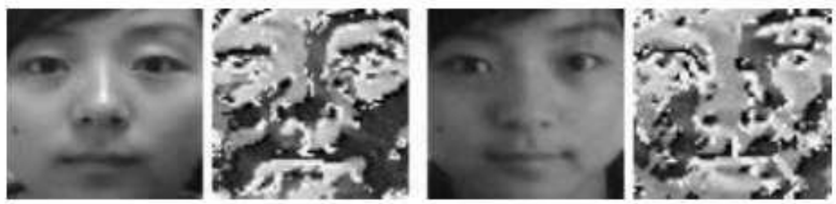

Figure 1. 2 same faces with one real photograph and other a spoof with their corresponding LBP analysis.

The face is first distinguished, edited and standardized into a pixel image of size 64 × 64. Then, the LBP8, 1u2 administrator on the standardized face picture is applied and the subsequent LBP face picture is isolated into 3×3 covering locales. The nearby 59-receptacle histograms from every locale are processed and gathered into a solitary 531-container histogram. At that point, two different histograms are registered from the entire face picture utilizing LBPu2 8, 2 and LBPu2 16, 2 administrators, yielding the histograms which are of 59 and 243 bin respectively that are added to the histogram of 531-bin recently figured. At long last, a radial basis function kernel and SVM classifier which is nonlinear is used to distinguish between live faces and spoofs.

III. Related Work

Deep learning techniques are currently experiencing a huge expansion in their field of application mainly because the advances in the GPU technology and goal oriented fabrication of chips [6]. Moreover, the existence of big data-sets has made it possible to train neural networks and to let them nearly reach human levels of performance when tested against tasks such as image classification [7].

The Face Recognition technology has been researched extensively along with deep learning in the 20 years. In particular a key role in the context of smart surveillance systems [8] is played by these facial recognition technologies. In such systems, the case is usually that a low resolution face image, taken by a surveillance camera, has to be matched against a database containing deep features extracted from high resolution images. To this end, several techniques have been developed in order to train deep models to deal with the images which are either of low quality or resolution. Some examples are Super Resolution technique and the other is Common Space Projection techniques. Super Resolution is a technique based on the prowess of neural networks that is they can rebuild an image with higher resolution from the lower resolution image. The Recognition task is later fulfilled in the high-resolution domain [9].

There are neural networks which, together with the super resolution task, tries to create a high resolution version of a image which is fed to it in a fairly low resolution. Instead, Common Space Projection techniques concern the ability of a neural networks to minimize the distance among deep features, in a common space, extracted from a low resolution image and its high resolution counterpart.

For example, in [10] they train a two-branch CNN to learn a mapping from high/low resolution domain to a common space. Specifically, given a low- and high-resolution image, the model extracts features vectors of size 2048 and their distance is evaluated.

IV. Proposed Methodology

In order to identify people in a classroom with the help of camera and neural networks we will need to establish a data set for this purpose. This will be used for the training of the model. The trained model will be fed with the live data of the classroom, which will return back the results to the web portal. Now the portal needs to be so that it could be easily accessible over the internet.

To achieve best performance over the complex machine learning model, we will implement it with the c++ language which speeds up the performance.

A connector will be required to connect the internet service to the c++ model and fetch the results. The framework for implementing the internet service is selected to enhance the maintainability and scalability of the service. There is a scope of improvement where on camera machine learning could be employed and performance and scalability can be expanded.

Step 1.

Gaining efficient knowledge about Facial recognition and spoof detection (Research paper, books, conference) and all the technologies that are going to be used in Web portal development.

Step 2.

C++ Facial Recognition Library:

Training and testing Facial Detection and recognition model.

Developing C++ Library for the Backend.

Building interface between Node.js and C++ Library.

NodeJS Backend:

Node.js Backend will serve stateless API's for React based frontend.

Frontend:

React, HTML, CSS, JavaScript

Step 3.

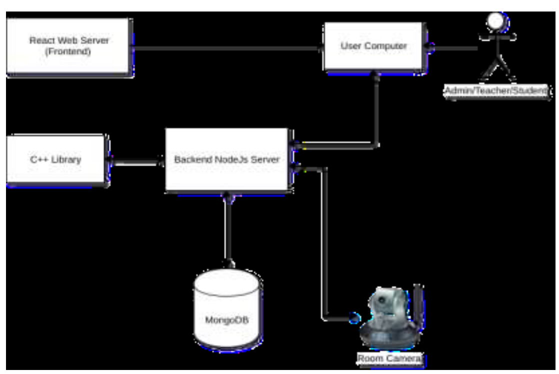

Figure 2. Block diagram of Proposed Methodology.

V. Implementation and Results

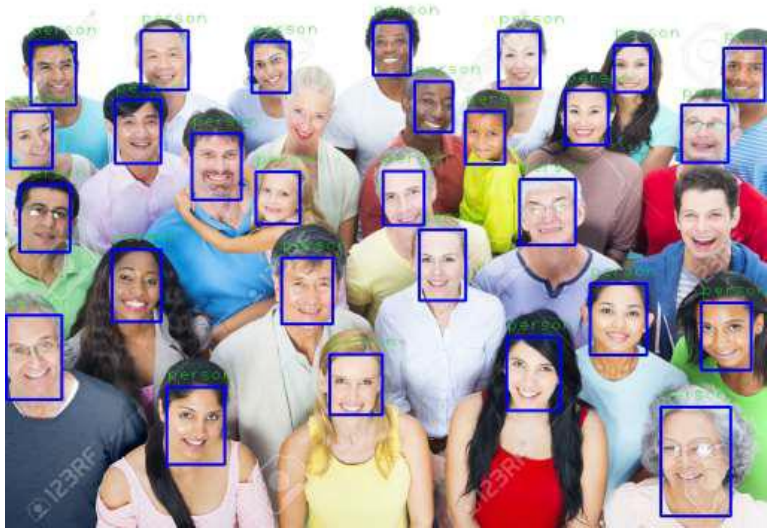

Figure 3. Snapshot of Face Detection and L2 Distance calculation for recognition.

The implementation of the facial recognition gives a good accuracy and the spoof detection is able to weed out the maximum spoof attacks which can be used to trick the system. The real-world performance will have several key factors into play such as performance of the facial recognition unit, no of points for feeding attendance, connection speed, and quality of camera.

On the testing the whole system should process images at speed of 30-50 ms per picture on our current facial recognition algorithm setup.

VI. Conclusion

This project displays the immense progress achieved by the deep learning technologies. The main part of the project is CNN based facial recognition and detection, and its application over making the daily life simple is the example how far these technologies have come. There is a lot of work desirable over hardware and software in these areas which will in turn make these technologies more accessible to the common people and everyone will be able to reap the benefits of machine learning as a whole. Also, the research over the scalability of these solutions must be done. As there is a limit on the scalability of this solution.

References

[1] F. V. Massoli, G. Amato, F. Falchi, C. Gennaro and C. Vairo, CNN-based System for Low Resolution Face Recognition Conference: 27th Italian Symposium on Advanced Database Systems, AIMIR Research Activities 2019. (2019)

[2] R. Ranjan, A. Bansal, J. Zheng, H. Xu, J. Gleason, B. Lu, A. Nanduri, J. C. Chen, C. D. Castillo and R. Chellappa, A Fast and Accurate System for Face Detection, Identification, and Verification. arXiv preprint arXiv:1809.07586v1

[3] F. Schroff, D. Kalenichenko and J. Philbin, FaceNet: A unified embedding for face recognition and clustering, 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Boston, MA, (2015), 815-823.

[4] L. Souza, L. Oliveira and M. Pamplona, How far did we get in face spoofing detection? Engineering Applications of Artificial Intelligence, 72 (2018), 368-381.

[5] J. Atta, A. Hadid and M. Pietika Inen, Face Spoofing Detection From Single Images Using Micro-Texture Analysis, International Joint Conference on Biometrics. 10.1109/IJCB.2011.6117510. (2015)

[6] Miller, Erik, Huang, Gary, Roy Chowdhury, Aruni, Li, Haoxiang and Hua, Gang, Labeled Faces in the Wild: A Survey. 10.1007/978-3-319- 25958-1_8. (2016).

[7] K. He, X. Zhang, S. Ren and J. Sun, Deep Residual Learning for Image Recognition, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, (2016), 770-778.

[8] B Lavi , M.F. Serj and I. Ullah, Survey on deep learning techniques for person re-identification task. arXiv preprint arXiv:1807.05284 (2018).

[9] K. Grm, M. Pernuš, L. Cluzel, W. Scheirer, S. Dobrišek and V. Štruc: Face hallucination revisited: An exploratory study on dataset bias, arXiv preprintarXiv:1812.09010 (2018).

[10] E. Zangeneh, M. Rahmati and Y. Mohsenzadeh, Low resolution face recognition using a two-branch deep convolutional neural network architecture, arXiv preprintarXiv:1706.06247 (2017).