Smart Attend: Face Recognition Attendance System

Abdul Shafiq[1], Dhanush Udaya Acharya[2], Sudithraj S Shetty[3], Sujal M Naik[4], Ms. Abhijnha B N[5]
[1,2,3,4]*Student, Department of Artificial Intelligence and Data Science*
[5]*Assistant Professor, Department of Artificial Intelligence and Data Science*
doi.org/10.64643/IJIRTV12I12-206580-459

Abstract—This smart system based on Automated Facial Recognition (AFC) not only suffers from significant security threats and privacy issues regarding storing facial images but also requires powerful Graphic Processing Units (GPU). Thus, the aim of the present paper is to develop a Smart Classroom Attendance System based on CPU, which will be able to process images with the help of standard hardware while meeting privacy concerns at the same time. Multi-Task Cascaded Convolutional Networks (MTCNN) are used to detect and align the face. After that, FaceNet (InceptionResnetV1) generates 512-dimensions features from the face image. In order to verify possible spoofs of attendance system, a function of liveness check via MediaPipe Face Mesh measures Eye Aspect Ratio (EAR) to establish if the student blinks. The proposed system does not rely on comparing an image to a single picture but utilizes a decision engine, where the cosine similarity threshold is 0.70 or greater for a sequence of at least three consecutive matching images of a person. It is not a video, which is recorded but non-reversible numerical values of the face image are used as embeddings. According to experimental data, processing time of each frame varies from 100 to 200 milliseconds, making detection very accurate.

Index Terms—Biometric authentication, Face embeddings, Facial recognition, Liveness detection, Smart classroom attendance system

I. INTRODUCTION

Tracking student attendance manually in modern educational institutions takes a lot of time and often leads to mistakes, reducing the time available for teaching. To address this, automated attendance systems using facial recognition have been suggested. However, these systems face three key issues: they require powerful hardware, can be tricked using methods like showing a photograph, and raise serious privacy concerns because they store student images or video data. Most advanced systems today need expensive GPUs to work in real time, making them unsuitable for typical classroom computers. Additionally, systems that continuously record classrooms can violate student privacy rules. To overcome these problems, this study presents a CPU-based video attendance system that works within a short time frame, such as the first 45 seconds of a class. It uses FaceNet to generate 512-dimensional feature vectors and compares them using cosine similarity. The system also includes a strong counting method and liveness detection to improve accuracy and reliability.

II. LITERATURE SURVEY

The problem of developing attendance systems using face recognition methods has long been actively discussed in the literature. Several approaches based on classical computer vision methods, deep learning, and intelligent algorithms have been considered and implemented.

The work of Turk and Pentland [1] introduces the concept of Eigenfaces, a PCA-based technique used in face recognition systems. It is known for its high effectiveness but suffers from sensitivity to lighting changes and different facial expressions.

Viola and Jones [2] propose a framework for real-time object detection using Haar Cascade classifier. Effective in controlled environments, this approach performs poorly with variations in lighting and occlusions.

Ahonen et al. [3] introduce the idea of LBPH, which is a significant improvement over previous works. Although better at handling occlusions and lighting changes, LBPH still lacks efficiency when dealing with face variations and large-scale databases.

In recent years, the development of deep learning approaches allowed for great progress in designing face recognition systems. Taigman et al. [4] present

DeepFace, a deep neural network model achieving almost human-like performance in verification tasks.
Schroff et al. [5] propose the embedding-based recognition technique in FaceNet, which consists in mapping facial images into high-dimensional vectors that maximize inter-class differences and minimize intra-class variation.
Deng et al. [6] develop the ArcFace approach, which improves face recognition performance in large-scale databases by using an additive angular margin loss function.

To address the issue of localization, Zhang et al. [7] propose MTCNN, a multi-task cascaded convolutional neural network designed for simultaneous face detection and alignment.
Deng et al. [8] introduce RetinaFace, a fast and efficient single-stage dense face detection algorithm providing excellent face localization and alignment even in uncontrolled environments.
A number of works are dedicated to developing and implementing face recognition–based attendance systems. Thus, Khan et al. [9] introduce a real-time face recognition–based attendance system using Face API and OpenCV. Although it allows for effective automatization of the process, the system requires internet connectivity.

Dang [10] introduces an improved smart attendance system that implements advanced face recognition techniques. As a result, it provides higher recognition accuracy through improved preprocessing and classification processes.
Bairagi et al. [11] implement a real-time attendance system using Haar Cascade classifiers and machine learning techniques, illustrating how attendance systems can be effectively automated in classroom settings. Still, the system may experience some problems due to its dependence on specific hardware devices.

Nandhini et al. [12] develop a face recognition–based attendance system that helps to automate the attendance process while suffering from problems associated with real-time operations.
Some researchers explore the idea of implementing IoT-based attendance systems. For example, Yusof et al. [13] propose an internet-based attendance system providing real-time data and centralized access to it. However, such systems suffer from high dependency on internet access and associated privacy concerns.
Another approach that has been considered in recent years is cloud-based attendance systems that provide great storage capabilities and scalability but lead to increased latency and increased risks of data breaches due to its necessity to store information outside of the system.
The recent trend in developing face recognition–based systems is the use of embedding-based approaches supplemented with classifiers, such as SVM and cosine similarity measures, which allow improving identity discrimination and reducing the rate of false acceptances. Anti-spoofing measures such as liveness detection through texture and blink analysis may be used to increase system security. At the same time, these approaches add considerable computational overhead.
Many existing systems face several challenges, including high computational costs, GPU dependence, lack of privacy protection measures, and inability to resist spoofing attacks. Besides, the absence of temporal verification causes unstable recognition results in real-time environments. Differently from existing approaches, the proposed system combines RetinaFace face detector, InsightFace embedding generator, and a Support Vector Machine (SVM) classifier. Moreover, it uses cosine similarity verification and temporal face tracking. Most importantly, the system operates offline, which improves privacy and eliminates dependency on internet connection while maintaining efficient performance.

III. PROPOSED SYSTEM

The suggested system will provide an efficient and accurate solution for automated attendance management through face recognition technology. Specifically, the Smart Attendance System will be focused on face detection and identity verification with subsequent secure offline storage of attendance data, without requiring expensive computational power.
Mainly, the system's objective will be the elimination of conventional manual attendance tracking, prevention of proxy attendance issues, and increased accuracy of attendance registration. Attendance will be registered in real time automatically via webcam

feed detection and aligned with the structured attendance register stored digitally.

There will be several functions that would make the Smart Attendance System run efficiently. First, face detection and alignment will occur using MTCNN; second, liveness detection will be applied based on facial landmarks to avoid any fraud attempts; third, facial embeddings will be produced using FaceNet technology; fourth, identity verification will be done through calculating cosine similarity and applying classification procedures; fifth, attendance records will be logged automatically into the storage database.

Temporal verification is used to increase reliability of the system and avoid false positive identification in case the system identifies wrong persons due to various factors, including lighting and partial face occlusions. Furthermore, the system is intended to work offline only, meaning there will not be any data transferred to or accessed through the internet.

In comparison to both conventional methods of attendance tracking and available biometric solutions, the proposed system would minimize manual intervention and physical interactions significantly, providing improved reliability and efficiency in operation.

IV. OBJECTIVES

The key goals of the Smart Classroom Attendance System include:

1) The creation of a CPU efficient attendance system, which will be able to perform well on standard computing devices without requiring any GPUs, hence making the system affordable and accessible to many educational institutions.
2) The development of a reliable face recognition pipeline which makes use of state-of-the-art models for accurate face detection, face alignment and face recognition in challenging and real-world classroom conditions.
3) Integration of anti-spoofing features, such as liveness detection, which is capable of preventing any spoofing attacks such as photo/spoof attacks, video spoofs and still images/3D spoofs.
4) Ensuring the privacy of users' information by eliminating the requirement of storing actual facial images or video frames and making use of numeric embedding vectors.
5) Implementation of a reliable decision-making model using counter-driven aggregation in order to improve accuracy in attendance registration and minimize any possible false positives.
6) Ability to rapidly capture the attendance of students within the allocated time frame.
7) Reducing storage requirements and computation overhead.
8) Making the system scalable and practical by finding the right balance between the different components of the pipeline.

V. SYSTEM ARCHITECTURE

The smart classroom attendance system architecture comprises several functional modules, which include the Input Acquisition Module, Face Detection and Alignment Module, Liveness Detection Module, Feature Extraction Module, Similarity Matching Module, Decision Engine, and Database Module. These modules work sequentially to allow accurate and efficient attendance management the Input Acquisition Module captures video footage via a standard webcam within a predefined time interval. Thus, it does not increase the computational complexity because there is no constant capturing of data; it is done for a specified period only.

The Face Detection and Alignment Module detects facial regions and calculates their landmarks based on the obtained images' analysis. The landmarks help in normalizing the facial region's orientation and scale. In other words, it helps to achieve the same input in further processing despite possible face pose differences due to lighting changes and others to reduce the risk of spoofing attacks, the system incorporates the Liveness Detection Module that checks whether there is an alive individual in front of the webcam or not. It works by evaluating facial dynamics that should correspond to those of a live person.

The Feature Extraction Module uses a deep learning algorithm to convert facial regions from the previous step into embeddings that are numerical vectors of features of each face. Such embeddings do not allow to restore any personal information related to users, which protects their privacy.

Next, the Similarity Matching Module compares generated embeddings and compares them with those previously stored in the database according to a specific similarity criterion. When a certain similarity threshold is exceeded, it considers that a match occurs. The Decision Engine Module improves system reliability by performing the matching procedure across several consecutive video frames, and if a specific number of frames contains matches, then it records attendance.

Lastly, the Database Module stores embeddings and records attendance. Since it only contains embeddings and does not store video streams or even images, it makes data management more efficient and increases privacy.

Overall, the interaction between modules makes it possible to manage attendance accurately and efficiently, provides real-time monitoring capabilities, and supports fast processing on CPU-based devices. Besides, the system can easily be enhanced and expanded further.

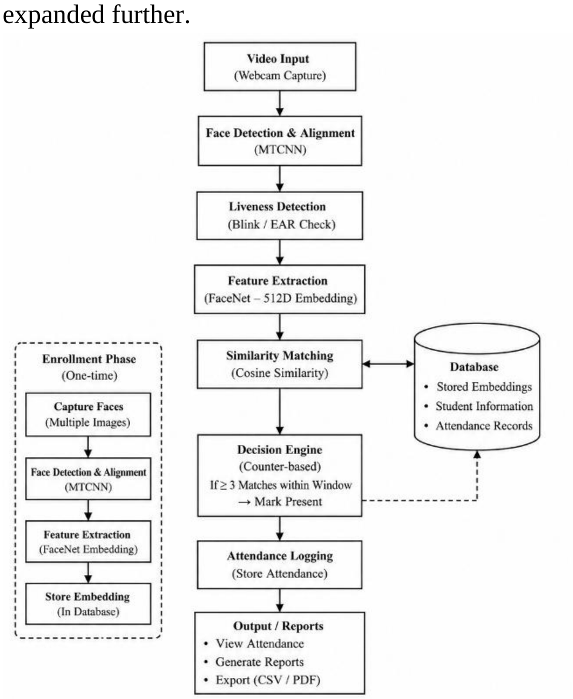

Fig. 1. Functional Block Architecture of the Smart Classroom Attendance System

VI. PROPOSED METHODOLOGY

The suggested methodology is grounded on Chaudhry et al. and refers to a framework of designing an attendance system with face recognition in real time. The basic characteristics of such framework involve a sequential approach to face detection and processing, including feature extraction and comparison to achieve lower computational complexity. Furthermore, the proposed methodology is based on CPU-based architecture, and thus no expensive hardware and high computational power are required.

First, the suggested methodology implies creating a database with all facial embeddings of users. Every entry includes a unique user identifier and corresponding embeddings acquired during enrollment stage. The system stores only embeddings rather than faces to protect personal information reliably.

During enrollment procedure, the suggested methodology proposes obtaining a video stream with a certain duration. After that, faces will be identified, aligned, and feature vectors will be calculated for them. Then, averaging several vectors will help obtain one vector for detecting presence.

As soon as input acquisition module works for some period of time to activate attendance detection procedure, faces will be recognized and aligned for normalization. Then, a face will be recognized and the behavior of eyes will be analyzed for identifying whether the face belongs to a live person – i.e., liveness detection will take place. For instance, if a person blinks, it means that he/she could be recognized.

The next step in the suggested methodology is to perform feature extraction for the recognized face. Afterwards, the obtained vector will be compared to embeddings of people kept in the database and the score of similarity will be obtained.

In contrast to traditional procedures when feature vectors from one frame are being compared, the suggested methodology will use a counter-based decision algorithm. According to this principle, the number of times when a particular face will be identified will be counted and the score will be compared with threshold. If the similarity score is higher than the threshold during a certain period of time, a person will be recognized.

Finally, the system requires implementing a real-time data update process, which is provided by batch-processing of new frames and accumulation of intermediate results. The suggested methodology will not employ any storage except RAM to secure sensitive information.

The workflow of the suggested methodology can be described as follows: enrollment → face detection →

liveness verification → feature extraction → similarity comparison → decision aggregation → attendance detection.
The suggested methodology is scalable because face detection algorithm can be improved to become more advanced.

VII. TECHNOLOGY USED

The Smart Classroom Attendance System that is being developed relies on modern deep learning algorithms, light weight computer vision technologies and smart software optimization strategies to provide fast and highly accurate performance in CPU-based devices, guaranteeing reliable protection of the user's privacy.

A. Programming Environment and Frameworks –

The development process makes use of Python 3.10+, which is a flexible programming language with vast support for both machine learning and computer vision. The core of deep learning algorithms uses the powerful framework PyTorch, which allows efficient processing of tensors and provides pre-trained model architectures. OpenCV library is used to perform real time video capture and image preprocessing.

B. Face Detection and Alignment –

For the purpose of face localization and landmark extraction Multi-Task Cascaded Convolutional Network (MTCNN) model is used, which is able to accurately detect and align faces under various illumination and orientation conditions. At the same time, it is suitable for deployment on CPUs only.

C. Facial Feature Extraction –

Features of the face are extracted through the use of pre-trained FaceNet (InceptionResnetV1) architecture. Using it, each detected face can be represented by the 512-dimensional vector. Pretrained weights of InceptionResnetV1 trained on large-scale VGGFace2 dataset are used to provide more accurate recognition results.

D. Liveness Detection Algorithm –

To prevent potential spoof attacks, MediaPip Face Mesh technology is used to detect detailed facial features. Based on the information from detected landmarks, Eye Aspect Ratio (EAR) metric is calculated to detect eye blinks.

E. Face Embedding and Similarity Comparison –

Cosine similarity measure is used to compare live embeddings with previously captured ones from the database, and thus detect the identity of person in the camera view. Identity match criterion is set to be equal or higher than 0.70.

F. Storage –

All data (name and unique ID) about users will be stored in SQLite database along with their facial embeddings, which are represented by 512-dimensional numeric vectors.

G. Optimization Strategies –

To provide efficient CPU-only processing speed several optimization techniques will be used: processing of only 5-7 frames per second, usage of counter-based method to avoid false positives and discarding of raw video frames.

VIII. RESULTS AND DISCUSSION

Experimental validation of the proposed Smart Classroom Attendance System was conducted using a regular computer system with an Intel Core i5/7 processor without the assistance of GPU acceleration. The objective is to analyze the performance of the system with respect to its accuracy, efficiency, and robustness when deployed in actual classroom settings. The results demonstrate that the system operates effectively without using many resources.

A. Evaluation Metrics

Evaluation was done by using some classification measures, such as accuracy, precision, and recall that are recognized to be used. Such criteria were measured based on the total number of positive results detected correctly.

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN}$$

$$Precision = \frac{TP}{TP+FP}$$

$$Recall = \frac{TP}{TP+FN}$$

(TP), True Negative (TN), False Positive (FP), and False Negative (FN) values obtained during experimentation.
Accuracy is determined by the overall performance accuracy of the model, whereas Precision focuses on the ability to prevent the marking of incorrect attendance. Recall indicates how good the model is at

identifying all the students that are present. It can be seen that Precision has increased drastically using the counter-based model.

B. Performance Metrics

The system exhibits effective real-time performance even in CPU-only scenarios. The entire processing pipeline – from face detection, liveness checking, embedding generation, and matching – requires about 100 ms to 200 ms per face. This translates into a frame rate of around 5-7 FPS, which is sufficient given the 45-second attendance recording timeframe.

Table I: System Performance and Resource Utilization

Metric	Measured Value
Processing Time per Face	100 - 200 ms
Operating Framerate	5 - 7 FPS
Enrollment Storage per Student	~8 KB (4 embeddings)
Total Storage for 60 Students	~480 KB (< 1 MB)
RAM Utilization	~400 - 600 MB

In addition to this, effective resource utilization is one more advantage. The maximum amount of memory space required per individual for storing information would be 8 KB. This means that the memory space required by 60 individuals would be less than 1 MB. In addition to this, the amount of RAM used by the program would be 400-600 MB.

C. Impact of Counter-Based Aggregation

One notable enhancement of this proposed system can be seen in the manner the system makes use of a counter-based attendance determination strategy. Contrary to existing techniques where only one frame match is required to confirm attendance, this system requires the existence of three matches for verification. By doing so, false alarms due to environmental noise or partial obstruction are greatly reduced. From the experiment performed, it was observed that although false alarms may occur in a single-frame match, there were virtually no false alarms in the new system despite an enhanced level of robustness.

Table II: Single Frame vs. Counter-Based Accuracy Evaluation

Approach	False Positives	False Negatives	Robustness to Occlusion
Single Frame Match	Moderate	Low	Low
Counter-Based (≥ 3)	**Zero**	Low	**High**

D. Limitations

Nevertheless, some disadvantages of this system should be highlighted. First, due to the fact that it is based on CPU, the efficiency level lowers when there are more than five faces in the frame. Second, while EAR-based face liveness detection proved itself efficient enough in case of static images, it might fail in cases where video replay is used for spoofing. These results prove that an appropriately designed CPU-based model is capable of producing good real-time results, which means it is applicable in smart classrooms.

IX. CONCLUSION

A lightweight Smart Classroom Attendance System was developed to operate effectively on generic CPU computers with no specialized hardware. It is equipped with an efficient face detector, feature extractor, and liveness detector to provide high-quality attendance tracking. Multi-frame validation provides higher accuracy and reliability since fewer mistakes occur while validating attendance during lectures.

According to the experiments, the system has good performance in terms of stability and low processing time, memory consumption, and storage requirements. Therefore, the proposed method can be implemented in real-life scenarios. Moreover, encoding attendance data into vectors rather than saving the actual images of students preserves their anonymity and minimizes the risk of data breaches.

Nevertheless, there are still limitations of the presented approach. First of all, the performance of the system decreases significantly when several faces need to be processed at once. Second, the suggested system cannot cope with advanced face spoofing attacks. Future work will be focused on solving the problem of efficiency and robustness to attacks.

REFERENCES

[1] Turk, M., and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[2] Viola, P., and M. Jones, "Rapid object detection using a boosted cascade of simple features," in Proc. IEEE Comput. Soc. Conf. Computer Vision and Pattern Recognition (CVPR), 2001, vol. 1, pp. I-511–I-518.

[3] Ahonen, T., A. Hadid, and M. Pietikäinen, "Face recognition with local binary patterns," in Proc. European Conf. Computer Vision (ECCV), 2004, pp. 469–481.

[4] Taigman, Y., M. Yang, M. Ranzato, and L. Wolf, "DeepFace: Closing the gap to human-level performance in face verification," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2014, pp. 1701–1708.

[5] Schroff, F., D. Kalenichenko, and J. Philbin, "FaceNet: A unified embedding for face recognition and clustering," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2015, pp. 815–823.

[6] Deng, J., J. Guo, N. Xue, and S. Zafeiriou, "ArcFace: Additive angular margin loss for deep face recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2019, pp. 4690–4699.

[7] Zhang, K., Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," IEEE Signal Processing Letters, vol. 23, no. 10, pp. 1499–1503, 2016.

[8] Deng, J., J. Guo, and S. Zafeiriou, "RetinaFace: Single-stage dense face localisation in the wild," arXiv preprint arXiv:1905.00641, 2019.

[9] Khan, S., A. Akram, and N. Usman, "Automated attendance system using face recognition with Face API and OpenCV," Wireless Personal Communications, vol. 113, no. 4, pp. 2725–2748, 2020.

[10] Dang, T.-V., "Attendance system based on improved facial recognition," Journal of Robotics and Control (JRC), vol. 4, no. 2, pp. 220–226, 2023.

[11] Bairagi, R., S. A. Tisha, M. S. Islam, R. Ahmed, M. S. Sarder, and M. A. Islam, "A real-time face recognition smart attendance system with Haar cascade classifiers," in Proc. IEEE Int. Conf. Inventive Research in Computing Applications (ICIRCA), 2021, pp. 1064–1069.

[12] Nandhini, R., N. Duraimurugan, and S. P. Chokkalingam, "Face recognition-based attendance system," International Journal of Engineering and Advanced Technology, vol. 8, no. 3, pp. 574–577, 2019.

[13] Yusof, Y. W. M., M. A. M. Nasir, K. A. Othman, S. I. Suliman, S. Shahbudin, and R. Mohamad, "Real-time internet-based attendance using face recognition system," International Journal of Engineering & Technology, vol. 7, no. 4, pp. 118–122, 2018.

[14] Zhao, W., R. Chellappa, P. J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," ACM Computing Surveys, vol. 35, no. 4, pp. 399–458, 2003.

[15] Li, S. Z., and A. K. Jain, Handbook of Face Recognition, 2nd ed. London, U.K.: Springer, 2011.

[16] Goodfellow, I., Y. Bengio, and A. Courville, Deep Learning. Cambridge, MA, USA: MIT Press, 2016.

[17] Krizhevsky, A., I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2012, pp. 1097–1105.

[18] Parkhi, O. M., A. Vedaldi, and A. Zisserman, "Deep face recognition," in Proc. British Machine Vision Conference (BMVC), 2015.

[19] Huang, G. B., M. Ramesh, T. Berg, and E. Learned-Miller, "Labeled Faces in the Wild: A database for studying face recognition in unconstrained environments," Univ. Massachusetts Amherst, Tech. Rep. 07-49, 2007.

[20] Wang, H., Y. Wang, Z. Zhou, X. Ji, D. Gong, J. Zhou, and W. Liu, "CosFace: Large margin cosine loss for deep face recognition," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2018, pp. 5265–5274.

[21] Cao, Q., L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "VGGFace2: A dataset for recognising faces across pose and age," in Proc. IEEE Int. Conf. Automatic Face & Gesture Recognition (FG), 2018, pp. 67–74.

[22] Redmon, J., S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, real-time object detection," in Proc. IEEE Conf. Computer Vision and Pattern Recognition (CVPR), 2016, pp. 779–788.

[23] Geitgey, A., Face Recognition Library. GitHub Repository, 2018.

[24] Paszke, A., et al., "PyTorch: An imperative style, high-performance deep learning library," in Proc. Advances in Neural Information Processing Systems (NeurIPS), 2019, pp. 8024–8035.

[25] Chollet, F., Keras. GitHub Repository, 2015.

[26] Szeliski, R., Computer Vision: Algorithms and Applications. London, U.K.: Springer, 2010.
[27] Jain, A. K., A. Ross, and S. Prabhakar, "An introduction to biometric recognition," IEEE Transactions on Circuits and Systems for Video Technology, vol. 14, no. 1, pp. 4–20, 2004.
[28] Zafeiriou, S., C. Zhang, and Z. Zhang, "A survey on face detection in the wild: Past, present and future," International Journal of Computer Vision, vol. 126, no. 9, pp. 935–964, 2018.
[29] Wright, J., A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 31, no. 2, pp. 210–227, 2009.
[30] Sandberg, M., J. Rasmusson, and L. Söderberg, "Face liveness detection using motion and texture analysis," in Proc. IEEE Int. Conf. Image Processing (ICIP), 2020.