1. Mohammed GHAZAL[1], 2. Roaa ALBASRAWI[2], 3. Najwan WAISI[3], 4. Mayyada AL HAMMOSHI[4]

Northern Technical University (1,2,3), Fairfax University of America (4)
ORCID: 1. 0000-0003-3159-4983, 2. 0000-0001-6071-3967, 3. 0000-0003-4259-1294, 4. 0000-0001-5818-420x

doi:10.15199/48.2022.03.21

Smart Meeting Attendance Checking Based on A multi-biometric Recognition System

Abstract. *Multimodal biometric can address some of the restrictions of the unimodal biometric by the combination of multi-biometric information for the same person in the decision-making operation. In this regard, the development in deep learning technologies has been employed in the multimodal biometric system. The deep learning techniques in object detection, such as face recognition and voice identification, are become more popular. Meeting Attendance checking carry out a very important role in meeting management. The manual checking attendance such as calling names or sign-in sheets is time-consuming. Face recognition and voice identification can be applied for attendance checks based on deep learning techniques. This paper presents an automatic multimodal biometric attendance checking system using Convolutional Neural Networks (CNN). The system uses a known dataset for the meeting participants, to train the CNN algorithm with a known set of input data. A computer with a high-quality webcam is used during the meeting attendance check, the system detects the attender face and voice then compares it with the known dataset, whenever matched, the attendee's name will be recorded in an excel file. The final result is an excel file with all attendance names. The result of the system shows that the proposed CNN architectures attained a high accuracy. Furthermore, this result could be beneficial in student attendance records, particularly in surveillance and person identification systems.*

Streszczenie. *Biometria multimodalna może rozwiązać niektóre ograniczenia biometrii unimodalnej poprzez połączenie informacji multi-biometrycznych dotyczących tej samej osoby w operacji podejmowania decyzji. W związku z tym rozwój technologii głębokiego uczenia się został wykorzystany w multimodalnym systemie biometrycznym. Coraz popularniejsze stają się techniki uczenia głębokiego w wykrywaniu obiektów, takie jak rozpoznawanie twarzy i identyfikacja głosu. Sprawdzanie obecności na spotkaniach pełni bardzo ważną rolę w zarządzaniu spotkaniami. Ręczne sprawdzanie obecności, takie jak wywoływanie nazwisk lub arkusze logowania, jest czasochłonne. Rozpoznawanie twarzy i identyfikacja głosu mogą być stosowane do sprawdzania obecności w oparciu o techniki głębokiego uczenia się. W artykule przedstawiono automatyczny multimodalny biometryczny system sprawdzania obecności z wykorzystaniem Convolutional Neural Networks (CNN). System wykorzystuje znany zbiór danych dla uczestników spotkania, aby wytrenować algorytm CNN ze znanym zbiorem danych wejściowych. Podczas sprawdzania obecności na spotkaniu używany jest komputer z wysokiej jakości kamerą internetową, system wykrywa twarz i głos uczestnika, a następnie porównuje je ze znanym zestawem danych, po dopasowaniu nazwisko uczestnika zostanie zapisane w pliku Excel. Ostatecznym wynikiem jest plik Excela ze wszystkimi nazwami obecności. Wynik działania systemu pokazuje, że proponowane architektury CNN osiągnęły wysoką dokładność. Ponadto wynik ten może być korzystny w rejestrach obecności uczniów, zwłaszcza w systemach nadzoru i identyfikacji osób.* (***Inteligentne sprawdzanie obecności na spotkaniach w oparciu o multi-biometryczny system rozpoznawania***)

Keywords: Multi-biometric, Deep Learning, Face Recognition, Speaker identification.
Słowa kluczowe: biometria, Deep Learning, rozpoznawanie twarzy.

Introduction

Nowadays, the need for online video conferencing platforms such as Google Meet and Zoom have been increased due to the rapid spread of the Coronavirus. These platforms play a pivotal role in eLearning, business, and press. In addition, it can accommodate a large number of attendees, which makes the attendance registration process difficult. Many ways for attendance check have been used, such as recording attendance manually by the instructor as mentioned in [1], and use Google Forms and Google Spreadsheets for meeting attendance record [2], by using Google Spreadsheets the instructor can create a new subject with classroom enrollment data, then create Google Forms, assigning topics, time to attendance, attendee ID, and attender name. Both methods are time-consuming because manual recording takes a lot of time and for Google Form, the instructor needs to create a Form and send the link of the Google Form to the attendees to fill out.

Several methods of automatic attendance checking systems have been developed to overcome the wasted time of the traditional methods with different technologies such as fingerprint [3], RFID cards [4], Iris scan [5], and recently face recognition [6]. Nowadays, the method of face recognition and voice identification has more interest to researchers due to the great development in deep learning techniques. Human voice identification is the system's ability to distinguish the speaker from the set of speaker samples available in the system [7]. There are two types of speaker voice identification, the first is called text-dependent systems, and the second is text-independent speaker recognition which can distinguish the voice in any language or text [8].

In this paper, a meeting attendance system using a face and voice identification technique based on a deep learning algorithm has been proposed. Different deep learning algorithms are used for face recognition which results in great improvements in the system performance. The deep learning algorithms are based on the training dataset machine gets to learn and thus automatically extract features. Different issues concerning computer vision easily handled by using deep learning such as image classification [9], handwriting recognition [10], object detection [11,12], and face recognition [13]. Neural networks make up the backbone of deep learning algorithms and therefore different neural network topologies are produced in deep learning thus each network has different functionality [14]. This network's development process involves the addition of multiple layers thus more weights and represents the relationship between the layers [15].

The rest of this paper is divided into 3 sections: Section 2, clarify the implementation of the proposed attendance checking system. Our experimental work with results is presented in Section 3. While the conclusions and future work directions are mentioned in Section 4

The proposed approach of attendance system

1. System Overview

The main tasks of the proposed attendance checking system based on multimodal biometric traits (face and voice) are discussed as shown in Fig.1. Firstly, creating and training the datasets for both traits. Secondly, collecting features by applying different deep learning models for each trait. Finally, the attendee's identity can be

assigned after comparing the query and training traits. The required database design process is subject to circumstances, according to the current health conditions due to the Coronavirus, which caused a general lockdown to contain the outbreak of the virus, we have used a Google Form to collect the required data as shown in Fig. 2. All the participants in the meet are required to take ten pictures, record a voice message, write their full name and email address to generate the required database. The custom database is used as a reference for real-time attendance recognition. To check the attendance of a participant to the meet, the computer takes the face image and voice of the participant through the real-time video stream and employs deep learning neural networks to predict whether the participant matches anyone in the database, and (if yes) further identifies the name of the participant. The result of this system will be used to update the attendance record in the format of an excel file.

2.Face Recognition System

Recently, VGGFace [16] and VGGFace2 [17] models have achieved high accuracy results in face recognition in comparison with the classical methods [18,19]. In the proposed system VGGFace2 deep learning model has been used for face recognition. The model has been included in Keras library; it can be used freely with the available pre-trained models and third-party open-source libraries [20]. VGGFace2 is a very large dataset that contains more than 3 million images belong to 9131 persons, with about 360 images for each person [21]. The images are collected from the Google search engine of images and they have large variations in pose, age, and illumination. Detect faces process is needed before recognition. The detection process has two purposes: the first is to extract faces from images and exclude the background, and the second is an automatic drawing of bounding boxes around their extent. The Multi-Task Cascaded Convolutional Neural Network (MTCNN) [22] has been used to find and extract faces from images. The pre-trained model ResNet-50 [23] is used for the face recognition process; it's trained on the VGGFace2 dataset. The face identification method is based on cosine similarity measure learning [24]. In this method we can calculate the similarity between the features samples, then compare the similarity (S) with a preset threshold value (Sth); If (S> Sth) then the face image is for the same person, otherwise, the face image is not for the same person. Before starting the Meet, attendance is required to show their faces at the front of the web camera one at a time. The program will automatically capture a face image and then the searching algorithm is performed to find the similarity with the faces in the dataset.

3.Speaker Recognition System

Person identification by voice recognition is the second option for the proposed system; it has been combined with the face recognition trait to compose the multimodal biometric attendance checking system. speaker recognition depends on two factors; the first is voice features extraction method such as Mel Frequency Cepstral Coefficients (MFCC), and the second is the modeling technique such as Artificial Neural Networks (ANN) [25,26]. The expected voice signal is in the range of (8 - 44) KHz, of the uncompressed audio file in Wav format. In this work, a preprocessing stage to decrease the noise content and remove the silent parts in the signal is required. Followed by the feature extraction stage using the MFCC method, and then resize the dimension of the feature vectors to fit with the modeling input stage. The final stage is related to the probability of matching the speaker voice with the trained model to recognize the speaker and returns it as output.

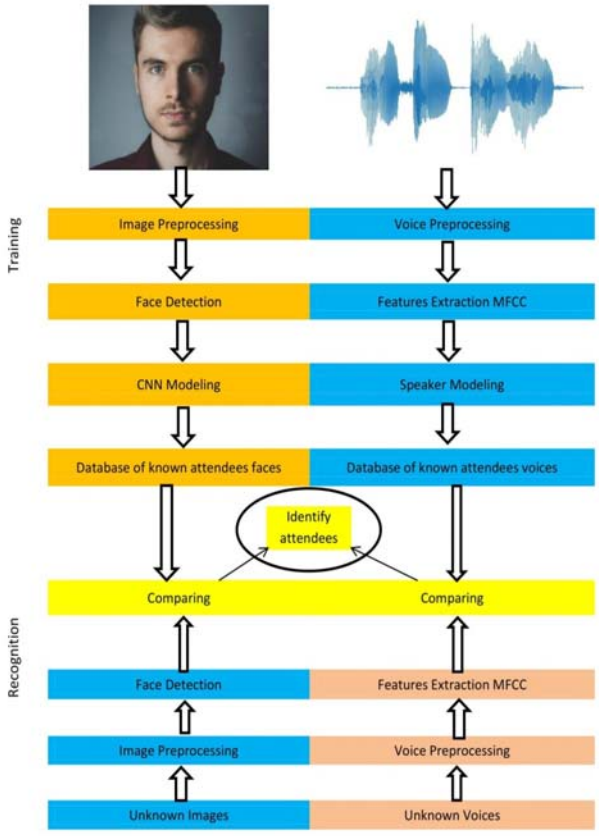

Fig.1. The proposed meeting Attendance model architecture.

4.Sign in Attendance System

The attendance information such as the name, login time, and email address will be saved in an excel file as shown in Fig. 3. The system has the property of determining the number of user recognition attempts. In case of a failure of attendance identification check, the participant should notify the instructor of the situation for a manual attendance check-in in the excel file.

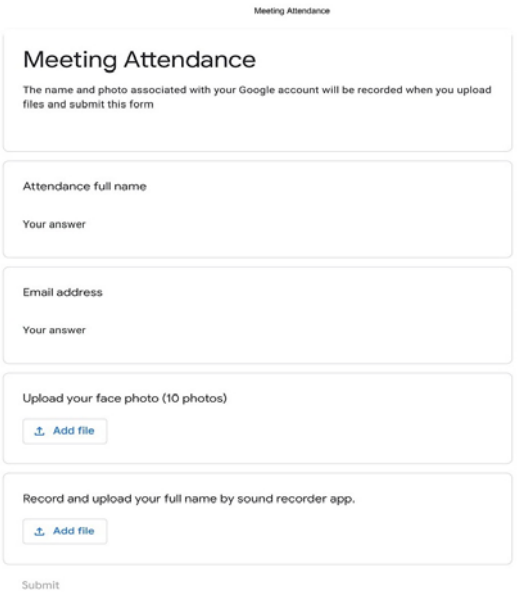

Fig.2. Google Form for collecting the data.

Clipboard Font

B16

	A	B	C
1	name	attenddance time	
2	karam	10:23 AM	
3	ali	10:24 AM	
4	mohammed	10:24 AM	
5	sara	10:25 AM	
6	muna	10:30 AM	
7	wesam	10:31 AM	
8	zaid	10:35 AM	
9	yaseen	10:40 AM	
10	mustafa	10:40 AM	
11	yasmeen	10:42 AM	
12	rana	10:43 AM	
13	maryam	10:45 AM	
14	ahmed	10:46 AM	
15	belal	11:00 AM	

Fig.3. Example of attendance Sign-in.

Experimental results

In this section, we discussed the experimental result of the attendance check system based on two biometric traits; facial recognition and speaker identification. The system is implemented using a computer from ASUS TUF with the specification as follows, Intel Core i7-11370H 11th Gen, 16GB of RAM, GeForce RTX 3070 8GB Graphics, 1 TB SSD Hard disk, and Ubuntu 21 operating system with python version 3.8. Despite the high-quality hardware and software used in the proposed multimodal biometric system, we must observe that the execution time of the multimodal biometric system has a proportional relation with many factors: the number of biometric traits, the number of subjects, and their images in the database. In the experimental work, we created a dataset of 1000 images for 40 different persons with about 25 images per person. In our proposed system, a pre-trained model of MTCNN has been used for face detection and the ResNet-50 module is used for face recognition. The system was able to recognize 39/40 persons accurately on normal conditions. The proposed face recognition system performance measurement in terms of benchmark evaluation metrics; Accuracy 96.54 %, Precision 96.8 %, and Recall/Sensitivity 99.45 %. And for the speaker recognition, we created a dataset of 10 voice samples of 10 persons are recorded and processed as mentioned in the proposed method. For each audio sample, 60 frames are extracted using the MFCC method and all the frames are tested on the neural net. We observed that the system was able to recognize 7/10 of voice recorded under lab conditions. The speaker recognition system accuracy is 70%. We observed that the measurement of the system performance can be different because of many factors such as Brightness, Hue, Saturation, and Noise. An example of the system is shown in Fig.4, a meeting with unknown person, both faces have been detected but only one is recognized. The recognized person name is (Karam) and his name is shown over the bounding box. For the unrecognized person, he is named by (Unknown). The unknown person did not fill out the Google Form in Fig. 3. Thus, his information did not include in the known database.

Fig.4. Experimental result of face recognition system.

Comparison with related works and discussion

The proposed Meeting attendance system is a result of combining a deep learning model trained with face images dataset and a voice recognition model. This combined model has a good performance in Meeting attendee detection system. To improve the model Adam optimizer has been used. Indeed, most of the related work concentrate on the classification of face recognition only. The performance of the proposed model in compare with different methods is shown in Table 1. According to the results of our work the accuracy is equal to 96.54 %.

Table 1.

Reference	Method	Classifi ca-tion	Detectio n	Result %
Ref [1]	DNN-FaceNet	Yes	Yes	AC= 95
Ref [27]	Inception-ResNetV1	Yes	Yes	AC=96.02
Ref [28]	CNN	Yes	Yes	AC=92
Proposed	ResNet-50	Yes	Yes	AC=96.54

Conclusion

Meeting attendance has become one of the most important requirements of life at present. Therefore, the participants' attendance record is important in the Meeting management. The traditional method of manual-record considered as a wasting time process. This research presents a meeting attendance checking system, based on multimodal biometric traits identification. The use of deep learning techniques enhanced the system performance in the face and speaker recognition. The system has been successfully tested in a meeting of 40 persons. The practical results show that the proposed system reduced the utilization time in compare with the traditional system. In the future, we will improve the system by adding more traits such as ear recognition and design a Graphical User Interface (GUI).

***Authors**: Mohammed Talal Ghazal obtained his M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2016, E-mail: mohammed.ghazal@ntu.edu.iq; Roaa Albasrawi her M.Sc. degree from Computer Engineering Technology, Northern Technical University, Mosul, Iraq in 2011, E-mail: roaa.albasrawi@ntu.edu.iq;Najwan Zuhair Waisi obtained her M.Sc. degree from Computer Science Department, Mosul University, Mosul, Iraq in 2014, E-mail: najwan.tuhafi@ntu.edu.iq; Mayyada Al Hammoshi obtained her PhD. Degree from Computer Science Department, Mosul University, Mosul, Iraq, E-mail mhammoshi@fxua.edu Mayyada is working as a Computer Science program chair and associate professor at Fairfax University of America since 2016;*

REFERENCES

[1] W. Kuang and A. Baul, "A Real-time Attendance System Using Deep-learning Face Recognition," in *2020 ASEE Virtual Annual Conference Experience*, 2020.

[2] A. Z. Mansor, “Managing student's grades and attendance records using google forms and google spreadsheets,” UKM Teaching and Learning Congress 2011., vol.59, 2012.

[3] B.K. Mohamed and C. Raghu, “Fingerprint attendance system for classroom needs,” India Conference (INDICON), Annual IEEE, 2012, pp. 433-438.

[4] S. N. Shah and A. Abuzneid, “IoT based smart attendance system (SAS) using RFID,” IEEE Long Island Systems, Applications and Technology Conference (LISAT), 2019.

[5] A. Khatun, A.K.M.Fazlul Haque, S. Ahmed, and M. M. Rahman, “Design and implementation of Iris recognition based attendance management system,” 2nd Int’l Conf. on Electrical Engineering and Information & communication Technology (ICEEICT) 2015, Bangladesh.

[6] S.Sawhney, K.Kacker, S. Jain, S. N.Singh, and R.Garg, “Real-time smart attendance system using face recognition techniques,”, 9th Int’l Conf. on Cloud Computing, Data Science & Engineering, 2019, pp. 522-525.

[7] R. Jagiasi, S. Ghosalkar, P. Kulal, and A. Bharambe, "CNN based speaker recognition in language and text-independent small scale system," in *2019 Third International conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud)(I-SMAC)*, 2019, pp. 176-179: IEEE.

[8] S. Kinkiri and S. Keates, "Speaker Identification: Variations of a Human voice," in *2020 International Conference on Advances in Computing and Communication Engineering (ICACCE)*, 2020, pp. 1-4: IEEE.

[9] T.-H. Chan, K. Jia, S. Gao, J. Lu, Z. Zeng, and Y. J. I. t. o. i. p. Ma, "PCANet: A simple deep learning baseline for image classification?," vol. 24, no. 12, pp. 5017-5032, 2015.

[10] O. Sudana, I. W. Gunaya, and I. K. G. D. J. T. Putra, "Handwriting identification using deep convolutional neural network method," vol. 18, no. 4, pp. 1934-1941, 2020.

[11] P. Patel, A. J. I. J. o. E. Thakkar, and C. Engineering, "The upsurge of deep learning for computer vision applications," vol. 10, no. 1, p. 538, 2020.

[12] M. Ghazal, N. Waisi, and N. J. T. Abdullah, "The detection of handguns from live-video in real-time based on deep learning," vol. 18, no. 6, pp. 3026-3032, 2020.

[13] G. Hu *et al.*, "When face recognition meets with deep learning: an evaluation of convolutional neural networks for face recognition," in *Proceedings of the IEEE international conference on computer vision workshops*, 2015, pp. 142-150.

[14] N. Mathur, S. Mathur, D. Mathur, and P. Dadheech, "A Brief Survey of Deep Learning Techniques for Person Re-identification," in *2020 3rd International Conference on Emerging Technologies in Computer Engineering: Machine Learning and Internet of Things (ICETCE)*, 2020, pp. 129-138: IEEE.

[15] A. Garcia-Garcia, S. Orts-Escolano, S. Oprea, V. Villena-Martinez, P. Martinez-Gonzalez, and J. J. A. S. C. Garcia-Rodriguez, "A survey on deep learning techniques for image and video semantic segmentation," vol. 70, pp. 41-65, 2018.

[16] O. M. Parkhi, A. Vedaldi, and A. Zisserman, "Deep face recognition," 2015.

[17] Q. Cao, L. Shen, W. Xie, O. M. Parkhi, and A. Zisserman, "Vggface2: A dataset for recognising faces across pose and age," in *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, 2018, pp. 67-74: IEEE.

[18] M. T. Ghazal and K. J. T. Abdullah, "Face recognition based on curvelets, invariant moments features and SVM," vol. 18, no. 2, pp. 733-739, 2020.

[19] N. Waisi, N. J. T. Abdullah and M. Ghazal,"The Automatic Detection of Underage Troopers from Live- Videos Based on Deep Learning," vol. 2021, no. 9, pp. 85-88, 2021.

[20] Ö. Sahin, "Integrating Keras Models," in *Develop Intelligent iOS Apps with Swift*: Springer, 2021, pp. 137-164.

[21] N. J. T. Abdullah, M. Ghazal, and N. Waisi, “Pedestrian age estimation based on deep learning," vol. 22, no. 3, 2021.

[22] K. Zhang, Z. Zhang, Z. Li, and Y. J. I. S. P. L. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," vol. 23, no. 10, pp. 1499-1503, 2016.

[23] C. Zhenzhou and D. Pengcheng, "Face recognition based on improved residual neural network," in *2019 Chinese Control And Decision Conference (CCDC)*, 2019, pp. 4626-4629: IEEE.

[24] H. V. Nguyen and L. Bai, "Cosine similarity metric learning for face verification," in *Asian conference on computer vision*, 2010, pp. 709-720: Springer.

[25] Athira Aroon, S.B. Dhonde, 2015, “Speaker Recognition System using Gaussian Mixture Model”, International Journal of Computer Applications (0975 – 8887), Volume 130 – No.14

[26] Zhang, Chunlei, and Kazuhito Koishida. "End-to-End TextIndependent Speaker Verification with Triplet Loss on Short Utterances." In Interspeech, pp. 1487-1491. 2017

[27] A. Rai, R. Karnani, V. Chudasama and K. Upla, "An End-to-End Real-Time Face Identification and Attendance System using Convolutional Neural Networks," *2019 IEEE 16th India Council International Conference (INDICON)*, 2019, pp. 1-4, doi: 10.1109/INDICON47234.2019.9029001.

[28] S. Chowdhury, S. Nath, A. Dey and A. Das, "Development of an Automatic Class Attendance System using CNN-based Face Recognition," *2020 Emerging Technology in Computing, Communication and Electronics (ETCCE)*, 2020, pp. 1-5, doi: 10.1109/ETCCE51779.2020.9350904.