CTU Journal of Innovation and Sustainable Development
ISSN 2615-9422 | e-ISSN 2815-5602

DOI: 10.22144/ctu.jen.2023.017

Design of a smart doorbell for a leader's office with availability status notification and visitor recognition features

Nguyen Khac-Nguyen, Nguyen Chanh-Nghiem*, Bui Huu-Danh, Tran Minh-Tuong, Lam Chanh-Nghia, Truong Quoc Bao, Nguyen Hoang Dung, and Nguyen Chi-Ngon

College of Engineering Technology, Can Tho University, Viet Nam

**Correspondence: Nguyen Chanh-Nghiem (email: ncnghiem@ctu.edu.vn)*

Article info.

Received 11 Dec 2022
Revised 16 Dec 2022
Accepted 16 Dec 2022

Keywords

Audio communication, face recognition, MobileNet, office doorbell

ABSTRACT

Smart doorbells have become a critical component of smart homes and modern offices. However, a smart doorbell, particularly designed for a leader's office, has not been introduced. In this study, a smart doorbell is developed for a leader's office. The system includes an application that allows availability status notification on the doorbell module and voice communication with the visitor from inside the office based on a private Wi-Fi network without an Internet connection to prevent the leader from potential privacy and security issues. It also features a live video capture of the visitor with face recognition by implementing a MobileNet model. In training and testing this model, 1,549 free face images of 125 people were augmented to generate training, validation, and testing datasets of 9,185, 2,500, and 5,000 face images, respectively. An additional authentication testing dataset of 1,068 AI-generated face images was also used to evaluate the system's False Acceptance Rate (FAR). A high confidence level of 0.945 was selected for the developed MobileNet model to obtain zero FAR and high accuracy, recall, and F-score values of 0.960, 0.960, and 0.978, respectively. Therefore, the proposed doorbell could be used for an office leader, showing potential use for biometric authentication.

1. INTRODUCTION

Recently, many smart doorbells have become commercially available. Most of these doorbells have integrated camera and audio peripherals, allowing visual surveillance and audio communication with the visitor from the inside (Panasonic, 2022). Some smart doorbells also feature video recording (Panasonic, 2022), 3D motion detection (Yaro, 2020), and smartphone application support (Delaney, 2021). High specifications with more security features are also available for smart home systems (BKAV SMARTHOME, 2021).

Although there are smart doorbells particularly developed for office users, their main use is to notify the user inside the office who is requesting an entrance by video communication (Omnicore, 2022). Additionally, they may have the smart feature of face recognition based on artificial intelligence and cloud computation technologies (Omnicore, 2022; Patel et al., 2021). However, such doorbell systems are still commercially limited in Vietnam.

So far, an office doorbell system that allows the user inside to update their availability status, such as being busy, available for visitors, available after some time, and visitor recognition for special alerts, has not been developed. Such a feature is important

for a leader's office, helping the leader concentrate on tasks of greater importance and higher priority. Therefore, this study aims to develop a smart doorbell system for a leader's office that allows the leader to notify an availability status on the doorbell and communicate with the visitor outside by voice. Moreover, it also provides a live view of the visitor with a face recognition feature. This paper is organized as follows. Section 2 describes the materials and methods, followed by the results and discussion in section 3. Finally, section 4 provides the conclusion of this study.

2. MATERIALS AND METHOD

2.1. Design constraints

Because the smart doorbell is for a leader's office, three design constraints because of security issues were considered:

i) Private wireless connection

ii) No cloud computing for face recognition

iii) No Internet connection should be allowed

2.2. System overview

An overview of the smart doorbell system is depicted in Figure 1. The system comprised two modules, i.e., the doorbell module on the office door and the user module on the office desk inside. Communication between the two modules was performed via a wireless connection for which a Wi-Fi hotspot had been created by the user module. However, no Internet connection was established for the user module, hence the Wi-Fi hotspot to prevent potential network attacks. Because of privacy issues, face recognition is performed in the user module, and any face image database created during system implementation is saved in the user module. As a result, the user module of the designed doorbell system should comprise a wireless Wi-Fi module, a user interface with face recognition capability, and audio peripherals for voice communication. Considering the main computation load on the user module and the ergonomic aspects such as comfort, ease of use, and productivity, a surface laptop was chosen to develop the user module. The design challenges for the user module are thus only limited to developing the user application without focusing on the mechanical design and integrating the required hardware. Developing the doorbell module involves the utilization of a microcontroller, a webcam, a speaker, a microphone, and a touch display. The specification of the system is summarized in Table 1.

Figure 1. Overview of the smart doorbell system

Table 1. Hardware information

Module	Specification
Doorbell module	+ Microcontroller board Raspberry PI zero 2W (1GHz quad-core 64-bit Arm Cortex-A53 CPU; 512MB SDRAM; 2.4GHz 802.11 b/g/n wireless LAN) + C270 HD Logitech webcam with microphone (Max Resolution: 720p/30fps; camera mega pixel:0.9) + Speaker + 2.4-inch RPi Touch Display (320x240 pixel)
User module	+ Surface laptop (4 GB RAM, i3-8130U CPU, Win10 64bit)

The working principle of the developed doorbell system is summarized as follows.

i) A guest touches the call button on the touch screen of the doorbell module.

ii) The user application alerts the user and shows a snapshot of the visitor's image with a face recognition result and live video capture.

iii) The user can enable voice communication, notify an entrance invitation, or update an availability status (Figure 2c-d).

iv) The doorbell module resets to its previous availability status when voice communication is stopped or when instructed by the user from the user module.

Like most video doorbells commercially available for a smart home, video capture is only available for the leader, and voice communication is only enabled from inside, i.e., by the leader in the office room. The user can also update the availability status in real-time. Specifically, when the system cannot recognize a new visitor, the user can save the new visitor's images so that retraining the face recognition model can be triggered later. The main window of the user application is briefly introduced in Figure 3.

Figure 2. Notifications on the doorbell module's screen for (a) default availability status, (b) entrance invitation, (c) visit request declination, (d) temporarily unavailable

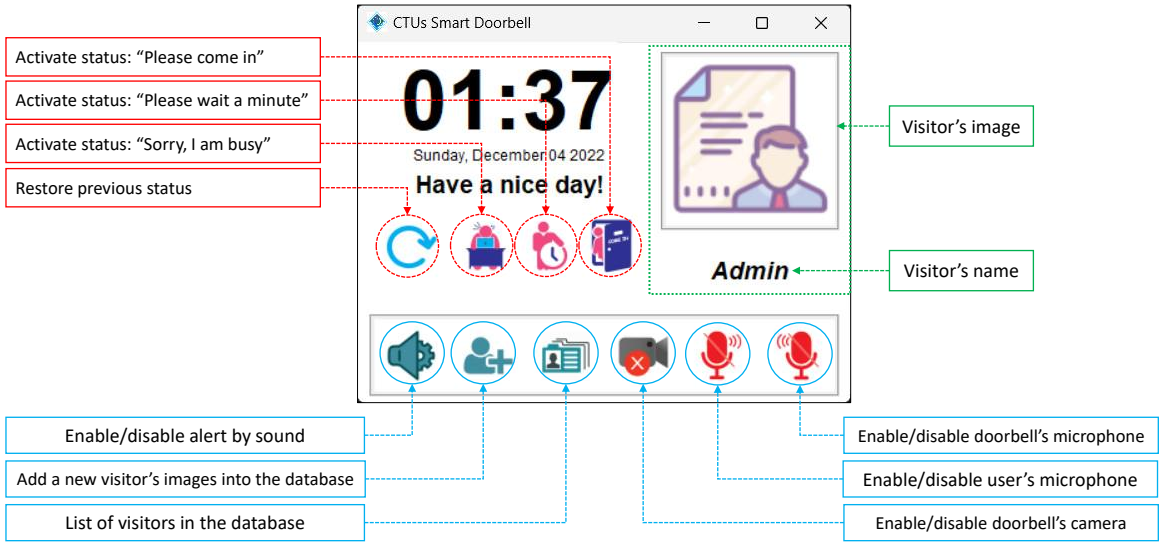

Figure 3. An overview of the main window of the user application

2.3. Face recognition

2.3.1. Database preparation

In this study, face images from two free databases were used with general information described in Table 1. The face images used to develop a face recognition model for the smart doorbell system were extracted from LFW face database (Learned-Miller, 2014). A total of 1,549 images of 125 people were randomly selected. For a doorbell system particularly designed for the leader's office, the successful recognition of 125 people would satisfy the system's requirements. Data augmentation was implemented using Keras library with a horizontal flip; a rotation range of 10 degrees; a width shift, a height shift, and a zoom range of 10% (Figure 4). As a result, 16,685 images were obtained and randomly divided into the training, validation, and testing datasets of 9,185, 2,500, and 5,000 images, respectively.

Another database of 1,068 selected face images was also obtained from Generated Media (Media, 2019). These Artificial-Intelligence generated faces images were used to evaluate the model's performance with unknown faces (Table 2).

2.3.2. Training configuration

Transfer learning was applied in this study with the pre-trained MobileNet model (Howard et al., 2017). This model originally had 1,000 output classes. Therefore, a new densely connected neural network replaced the last five layers of the original MobileNet model. This network was initially created with 125 classes, which was the number of people in the augmented images.

As a transfer learning strategy, only the first 22 layers of the modified MobileNet model were not subjected to training. These layers were selected based on empirical suggestions (DeepLizard, 2020). Should a new person be added to the system for future recognition, the detected person's images will be included in the training dataset. The model can be retrained following the aforementioned strategy, i.e., a new densely-connected neural network with an incremented number of output classes will be substituted for the old one before training is performed for all network layers, excluding the first 22 layers.

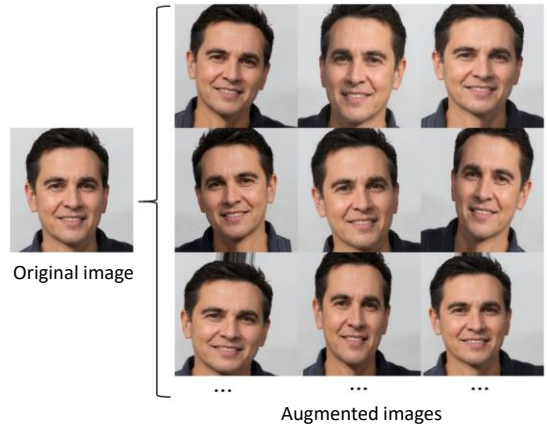

Figure 4. Examples of augmented images

Table 2. Datasets information

Dataset source	Description
LFW database	+ Augmented face images of 125 people + Training dataset: 9,185 images + Validation dataset: 2,500 images + Testing dataset: 5,000 images (40 images/person), from which 4,971 correctly recognized face images were used to calculate False Rejection Rate (FRR)
Generated Media	+ 1,068 AI-generated face images + Used to calculate False Acceptance Rate (FAR)

2.3.3. Performance evaluation

Accuracy, *Precision*, *Recall*, and *F-score* were calculated to evaluate the model performance with the testing datasets. These metrics are commonly used to evaluate the model performance (Afreen & Bajwa, 2021; Li et al., 2021). For a two-class classification problem, *Accuracy*, *Precision*, *Recall*, and *F-score* are formulated, respectively as

$$Accuracy = \frac{TP + TN}{TP + FP + TN + TN}, \quad (1)$$

$$Precision = \frac{TP}{TP + FP}, \quad (2)$$

$$Recall = \frac{TP}{TP + FN}, \quad (3)$$

$$F\text{-}score = 2\frac{Precision \times Recall}{Precision + Recall}, \quad (4)$$

or

$$F\text{-}score = \frac{2TP}{2TP + FP + FN}, \quad (5)$$

where *TP*, *FP*, and *FN* are the number of true positives, false positives, and false negatives of the predicted instances, respectively. However, these metrics can also be extended to multiple class problems. Because of a multi-class recognition problem in this study, the macro averaged *Precision*, *Recall*, and *F-score* values were calculated based on the model's evaluation results with the testing dataset. These are, respectively, the average *Precision*, *Recall*, and *F-score* values calculated for all available classes.

False Acceptance Rate (FAR) and False Rejection Rate (FRR) were also calculated to evaluate the model's performance. These are among the critical performance measures for a biometric security system that can be calculated as

$$FAR = \frac{n_{unauthorized\_access}}{N_{unauthorized\_access}}, \quad (6)$$

$$FRR = \frac{n_{rejected\_authorized\_access}}{N_{authorized\_access}}, \quad (7)$$

Where,

$n_{unauthorized\_access}$ is the number of unauthorized user instances that were incorrectly granted access;

$N_{unauthorized\_access}$ is the total number of unauthorized user instances;

$n_{rejected\_authorized\_access}$ is the number of authorized user instances that were incorrectly rejected access;

$N_{authorized\_access}$ is the total number of authorized user instances.

Thus, FAR can be regarded as the likelihood of a false positive when the biometric authentication system grants access to an unauthorized user (Devi et al., 2022). FRR, on the other hand, represents the probability that the biometric authentication system will incorrectly reject an authorized user's access attempt. FAR will be calculated based on untrained face images from Generated Media (Media, 2019). In contrast, FRR will be computed based on the correctly detected images from the testing face images from the augmented images from the selected LFW database (Table 2).

3. RESULTS AND DISCUSSION

3.1. Model training

Figure 5 shows the training and validation results. The optimal model was obtained after six epochs with the training accuracy and validation accuracy of 100% and 99.2%, respectively. The model was evaluated with the testing dataset, and good results were obtained with more than 99% for all performance metrics (Table 3). The confidence level was set relatively high at 0.805 to ensure an accurate prediction of the obtained model. For each output class, the model outputs a predicted value. If the maximum predicted value for all classes is smaller than the predefined confidence level, a classification failure occurs; otherwise, the model outputs the class corresponding to the maximum predicted value.

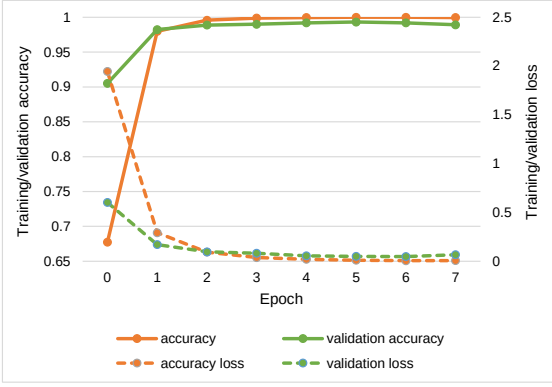

Figure 5. Training results

Table 3. Testing performance with a confidence level of 0.805

Metrics	Macro average value
Precision	0.994
Recall	0.994
F-score	0.992

3.2. Selecting the model's confidence level

As previously discussed, the final predicted class depends on the model's maximum predicted value and the predefined confidence level. Increasing the confidence level might decrease the model's accuracy; however, it will prevent the model from outputting false positives, which should be a severe problem for a biometric security system. In this study, the possibility of accepting false positives of the system is evaluated by calculating FAR using a different face image dataset obtained from Generated Media (Media, 2019). A few images from this dataset are shown in Figure 6.

A large confidence level might also lead to a significant classification failure rate. This effect was evaluated based on the FRR that was calculated using the 4,971 correctly recognized face images in the testing dataset with a confidence level of 0.805.

Figure 6. Some AI-generated face images from used for calculating the false acceptance rate

Table 4. FAR and FRR

Confidence level	FAR (%)	FRR (%)
0.805	0.28	1.43
0.825	0.28	1.55
0.845	0.28	1.67
0.865	0.28	1.89
0.885	0.28	2.09
0.905	0.19	2.31
0.925	0.09	2.76
0.945	0	3.44
0.965	0	5.47
0.985	0	11.21
0.995	0	26.49

Table 4 summarizes the calculated FAR and FRR when increasing the confidence level from 0.805 to an extreme of 0.995. FAR dropped significantly when the confidence level exceeded 0.905. A smaller FAR is more favourable for a biometric security system than a smaller FRR. To promote the potential development of the proposed doorbell system into a smart door lock, the confidence level of 0.945 was selected to achieve zero FAR. As a result, a precision of 1.0 was obtained with reduced *Accuracy*, *Recall*, and *F-score* values of 0.960, 0.960, and 0.978, respectively (Table 5).

Table 5. Testing performance with the selected confidence level of 0.945

Metrics	Value
Accuracy	0.960
Precision	1.000
Recall	0.960
F-score	0.978

4. CONCLUSION

This study successfully proposed a design scheme for developing a smart doorbell system for a leader's office that allows voice communication, availability status notification, and live video capture with a visitor recognition feature. As a stand-alone system that does not require cloud computation and an Internet connection, the proposed doorbell met the design constraints that help prevent privacy issues. Preliminary results showed that the developed MobileNet model used for face recognition had zero FAR when evaluated with the authentication testing dataset and high *Accuracy*, *Recall*, and *F-score* values of 0.960, 0.960, and 0.978, respectively, when evaluated with the testing dataset. As a result, the proposed smart doorbell could be used for a leader’s office with promising use for biometric authentication.

ACKNOWLEDGMENT

This study was funded by Can Tho University's scientific project TĐH2021-01.

REFERENCES

Afreen, H., & Bajwa, I. S. (2021). An IoT-Based Real-Time Intelligent Monitoring and Notification System of Cold Storage. *IEEE Access*, *9*, 38236–38253. https://doi.org/10.1109/ACCESS.2021.3056672

BKAV SMARTHOME. (2021). *Nhận dạng khuôn mặt công nghệ trí tuệ nhân tạo*. https://bkavsmarthome.vn/nhan-dang-khuon-mat-cong-nghe-tri-tue-nhan-tao

DeepLizard. (2020). *Fine-Tuning MobileNet on Custom Data Set with TensorFlow’s Keras API - deeplizard*. https://deeplizard.com/learn/video/Zrt76AIbeh4

Delaney, J. R. (2021, May 18). *Ring Video Doorbell Pro 2 Review | PCMag*. https://www.pcmag.com/reviews/ring-video-doorbell-pro-2

Devi, R. M., Keerthika, P., Suresh, P., Sarangi, P. P., Sangeetha, M., Sagana, C., & Devendran, K. (2022). Retina biometrics for personal authentication. *Machine Learning for Biometrics: Concepts, Algorithms and Applications*, 87–104. https://doi.org/10.1016/B978-0-323-85209-8.00005-5

Howard, A. G., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., Andreetto, M., & Adam, H. (2017). MobileNets: Efficient Convolutional Neural

Networks for Mobile Vision Applications. *CoRR, abs/1704.0*. http://arxiv.org/abs/1704.04861

Learned-Miller, G. B. H. E. (2014). *Labeled Faces in the Wild: Updates and New Reporting Procedures* (Issue UM-CS-2014-003).

Li, S., Wu, J., Long, C., & Lin, Y.-B. (2021). A Full-Process Optimization-Based Background Subtraction for Moving Object Detection on General-Purpose Embedded Devices. *IEEE Transactions on Consumer Electronics*, *67*(2), 129–140. https://doi.org/10.1109/TCE.2021.3077241

Media, G. (2019). *Gallery of AI Generated Faces | Generated.photos*. https://generated.photos/faces

Omnicore. (2022). *The 8 Best Smart Doorbell with Video Camera for Office + Home*. https://www.omnicoreagency.com/best-smart-doorbells/

Panasonic. (2022). *Video Intercom*. https://www.panasonic.com/middleeast/en/business/security/video-intercom.html

Patel, V., Kanani, S., Pathak, T., Patel, P., Ali, M. I., & Breslin, J. (2021). An Intelligent Doorbell Design Using Federated Deep Learning. *8th ACM IKDD CODS and 26th COMAD*, 380–384. https://doi.org/10.1145/3430984.3430988

Yaro. (2020, April 16). *Best Wireless Doorbells in 2022*. https://thehousetech.com/best-wireless-doorbells/