Door Access System Using Face Recognition

J Mae*, F Rabbaanii, and R B Nasza
Computer Engineering Department, Faculty of Engineering, Bina Nusantara University, Jakarta, Indonesia 11480

*Corresponding author: johannes@binus.edu

Abstract. This is a security system to restrict access to rooms using Raspberry Pi as main controller and implementation of Histogram of Oriented Gradients (HOG) feature extraction method. Upon face detection, the system is able to do face recognition for person identification by matching face embedding of the person face with one that already stored in database. This system also presents a random head-pose challenge to identify a real person from an image and then activate a solenoid to unlock the door. In this experiment, if the person is not wearing any face attributes, this system is tested to have 100% successful detection and recognition rate at 60 cm distance. If the person is wearing mask attribute, this system is tested to have 70% successful rate for detection and 45% successful rate for recognition at 60 cm distance.

Keywords : face recognition, Raspberry Pi, Face Features, HOG.

1. Introduction
Personal identification using biometrics is more popular than using access cards, passwords and patterns. As in offices, each workspace has a door that has an RFID security system and access card is required to access it, so that just people cannot enter the room.

The data contained in the RFID card can be duplicated by irresponsible people and people sometimes forget to bring the access card to the office, so this system still has shortcomings. From this case, a system that can solve the problem must be considered. Zufar [1] in his research points out that most access control systems nowadays use a biometric verification. Due to the unique characteristics of biometric features, they are not transferable. Therefore, a door access system using face recognition can solve this problem.

In this research, the Raspberry Pi will be used as the main control system. Images will be provided by a webcam camera, and a solenoid door lock as an actuator.

2. Theoretical Basis

2.1. Microprocessor
Microprocessor is an IC Chip or Integrated Circuit that combines the core functions of the computer's central processing unit (CPU / Central Processing Unit). The IC chip, which is written in English as "microprocessor", is a multipurpose device that can be programmed to receive digital data as input, process it according to the instructions stored in its memory and provide results as output[2].
The microprocessor used in this research is the Raspberry Pi 4 model B. It functions as the main controller on the system that runs a server. When the system recognizes the face and the person is able to complete the challenge, then the server will open a solenoid connected to the GPIO.

2.2. Webcam Logitech C270
A webcam is a video camera that provides an image or video image directly via a computer, such as the internet. A webcam is usually a small camera that is placed on the table, attached to the user's computer, or built inside the user's computer.

The webcam is the primary input device in this research, as it is the main source that continuously capture image frame and then pass it to the main controller to be processed.

2.3. Raspberry Pi LCD Touch Screen 3.5 inch
Raspberry Pi LCD touch screen is an LCD layer that is connected to the Raspberry Pi pin header using the I2C interface. In some situations, it is possible to use the HDMI (High Definition Multimedia Interface) and LCD displays simultaneously.

In this system, the LCD touch screen is used to display output image of the system and also act as feedback for the user.

2.4. Motor Driver L298N
A solenoid is a type of coil made of long wires that are tightly wound and it can be assumed that their length is much larger than their diameter. If there is an iron rod and is placed part of its length in the solenoid, the iron rod will move into the solenoid when the current is applied. This can be used to move levers, open doors, or operate relays.

2.5. Electric Solenoid Bolt EB250
Motor Driver L298N is a type of driver IC that can control the direction of rotation and speed of DC motors and stepper motors. This motor driver can output a voltage of up to 50 volts. This motor driver is used to drive the EB250 bolt electronic solenoid with assistance from an additional power supply because the Raspberry Pi 4 can only provide a maximum output of 5 volts.

2.6. OpenCV
OpenCV (Open Source Computer Vision) is a library of programs primarily intended for real-time computer vision[3]. This library is cross platform, which means it can be used on different operating systems.

2.7. Face Recognition
Face detection is one of the most important initial stages before the face recognition process is carried out. Face recognition is a computer technology that allows us to identify or verify someone's face through a digital image. The trick is to match the texture of our facial curves with facial data stored in a database or a certain file.

The robustness of the facial recognition system depending on changes in light conditions or expression or even partially blocked faces can be considered. Several papers [4][5] have proposed various techniques for facial recognition in these conditions. Eigenfaces are variant features extracted for the above factors. The facenet uses a deep convolutional neural network with Google's Inception model architecture and uses a new online triplet mining method to train instead of layers of intermediate bottlenecks.

On the widely used Labeled Faces in the Wild (LFW) dataset, the Facenet system achieves a new record accuracy of 99.63%[6]. However, unfortunately, not only did the database size increase but also the computation costs increased and the recognition accuracy decreased. That's why incremental learning is a learning algorithm that comes close to handling large-scale training data for efficiency and accuracy.
Face Recognition is an integral part of biometrics. In biometrics, basic human nature matches existing data. Facial features are extracted and implemented through an efficient algorithm and some modifications are made to improve the algorithm model. Computers that detect and recognize faces can be applied to a wide variety of uses including criminal identification, system security, identity verification etc.

Face Recognition generally involves two stages, which are face detection and face recognition. The difference between face detection and face recognition is that in detection it determines whether there is one or more faces in the image, but in recognition we determine whose face it is. The features extracted from the faces are processed and compared with the faces in the database. In general, face recognition techniques can be divided into two groups[7]:

- Face representation techniques - these techniques use holistic textural features and are applied to specific areas of the face or in a complete image of the face.
- Feature-based techniques - this technique uses geometric facial features (mouth, eyes, eyebrows, etc.) and the geometric relationships between them.

2.8. **Face Detection using Histogram Of Oriented Gradients**

Histogram of Oriented Gradients (HOG) is a method used for object detection, the histogram contains channels of the direction / orientation of the gradient of image pixels. HOGs are feature descriptors that have been used successfully for object and pedestrian detection, representing objects as a single value vector as opposed to a series of feature vectors where each represents an image area, calculated by sliding the window detector over the image. The HOG descriptor is calculated for each position, while the image scale is adjusted to obtain the HOG feature[8].

The first step of face detection is to scan the image at all scales and locations as shown in Figure 1. A sliding window were used to extract features from a certain part of the image at every scales. Figure 2 shows a part of the image that were extracted and calculated using HOG descriptor. The extracted features will then be matched with trained face features with Support Vector Machine (SVM) algorithm as shown in Figure 3. The detected faces area will likely produce multiple overlap region which is caused by multiple feature extraction by the sliding window within the nearby faces area. This problem can be solved by applying non-maxima suppression to remove overlapping bounding boxes that refer to the same object, as shown in Figure 4.

3. **RESEARCH METHODS**

This system uses a web server that runs on the Raspberry Pi 4 Model B for facial recognition process. The webcam is connected to the Raspberry Pi through a USB cable, and is used to capture image frame for facial recognition and registration for the new faces to dataset. The camera resolution is set to 300 x 300 pixels, so that the display looks proportional when using the 3.5 inch LCD. Using this image resolution, an average of 5 frame-per-second (FPS) were achieved, which is faster than using the default higher resolution resulting only 2 FPS in average. The data that has been processed will be uploaded to the database along with the contents of the data such as name, facial feature extraction and other information from the user. Then if the facial recognition process has face extraction similar to
the dataset registered in the database, the server will give a command to unlock the solenoid and display its status on the monitor. The block diagram of the whole system is shown in Figure 5.

When the system is turned on, it tries to connect to MongoDB which functions as the system database, then the system turns on the camera and ready to detect new faces. When a face is detected, the system will process and extract the face embedding and match it with the pickle dataset in the database. If it is not in the database, the system will detect the face again, but if the face is recognized, the system will start a head pose challenge where the person will be given random directions to turn the face left or right. Upon successful, the person will be given access to the room, and the solenoid will open. The whole process is described in flowchart shown in Figure 6.
This system has an additional security feature called the head pose challenge. This feature serves to minimize false identification when recognizing a face picture from digital sources. The detailed process is shown in Figure 7.

The system will load the face landmark predictor on the face recognized by the system, where the system will place the essential points on the person face. Then the system will calculate the euler angle. There are three euler angle points, namely: point X, Y, and Z. In this system only point Y is used, because point Y is the point that projects the position of the head when turning. The formula used to calculate the euler angle is:

\[
\begin{bmatrix}
X \\
Y \\
Z
\end{bmatrix} =
\begin{bmatrix}
r_{00} & r_{01} & r_{02} & t_x \\
r_{10} & r_{11} & r_{12} & t_y \\
r_{20} & r_{21} & r_{22} & t_z
\end{bmatrix}
\begin{bmatrix}
U \\
V \\
W \\
1
\end{bmatrix}
\]

where:
- X, Y, and Z are euler angles
- \( r_{ij} \) is a rotation matrix
- \( t_x, t_y, t_z \) are translation vectors
- U, V, and W are the three-dimentional points of the face detected by the face landmark predictor

![Figure 7. Head Pose Challenge Flowchart](image)

4. Results and Discussion

4.1. Experimentation of Face Detection and Recognition Rate with Various Distance from Camera

This experiment is conducted to find the optimum face detection and recognition rate in various object distances from the camera. Each volunteer in this experiment will take 5 measurements data for each distance specified. The success rate in detecting faces up to a distance of 90cm is 99.16%. While at a distance of 120cm the system starts to show some detection failure because the face images are
smaller. At 150cm distance, the system has difficulties in detecting the user's face and the detection rate has dropped significantly. The success rate of face recognition will always be lower than the face detection rate, because when the system fails to detect the face, it will not proceed to face recognition phase. The testing result shows that the optimal distance for the system to recognize the person face is at a distance of 60cm. At the distance above 90cm the system began to have difficulty determining the person face and at a distance of 150 cm the face recognition result dropped significantly. The experiment result is presented in Table 1.

| Person | Detection Rate (%)  | Recognition Rate (%) |
|--------|---------------------|----------------------|
|        | 30cm | 60cm | 90cm | 120cm | 150cm | 30cm | 60cm | 90cm | 120cm | 150cm |
| 1      | 100  | 100  | 100  | 100   | 20    | 100  | 100  | 100  | 100   | 20    |
| 2      | 100  | 100  | 100  | 100   | 40    | 100  | 100  | 100  | 100   | 20    |
| 3      | 80   | 100  | 100  | 100   | 0     | 80   | 100  | 100  | 100   | 0     |
| 4      | 100  | 100  | 100  | 100   | 0     | 100  | 100  | 100  | 100   | 0     |
| 5      | 100  | 100  | 100  | 60    | 40    | 100  | 100  | 100  | 100   | 40    |
| 6      | 100  | 100  | 100  | 0     | 0     | 100  | 100  | 100  | 0     | 0     |
| 7      | 100  | 100  | 100  | 100   | 0     | 100  | 100  | 100  | 100   | 0     |
| 8      | 100  | 100  | 100  | 0     | 0     | 100  | 100  | 100  | 0     | 0     |
| Avg    | 97.5 | 100  | 100  | 70    | 12.5  | 97.5 | 100  | 97.5 | 62.5  | 7.5   |

4.2. Experimentation of Face Recognition Processing Speed with Various Distances from Camera

The face recognition time measurement begins from the moment the webcam takes a picture, creates a face bounding box, until the system can identify the detected face embedding with the one stored in the database. Each time cell is taken from the calculation of five experiments per user. The system can recognizes the person face within 0.420117 - 0.484913 seconds time range. Almost all distances have a close average time, so it can be concluded that difference in distance does not affect the processing time of the system. The experiment result is presented in Table 2.

| Person | Average Recognition speed (s) |
|--------|------------------------------|
|        | 30 cm | 60 cm | 90 cm | 120 cm | 150 cm |
| 1      | 0.428708 | 0.44235 | 0.431383 | 0.425838 | 0.431505 |
| 2      | 0.427042 | 0.42513 | 0.437887 | 0.421523 | 0.423708 |
| 3      | 0.420345 | 0.420117 | 0.42515 | 0.422834 | N/A |
| 4      | 0.424046 | 0.425549 | 0.422847 | 0.429225 | N/A |
| 5      | 0.430367 | 0.436426 | 0.484913 | 0.426283 | 0.424533 |
| 6      | 0.435847 | 0.455515 | 0.458466 | N/A | N/A |
| 7      | 0.447804 | 0.442572 | 0.447753 | 0.452324 | N/A |
| 8      | 0.430307 | 0.433663 | 0.484814 | N/A | N/A |
| Avg    | 0.430558 | 0.435165 | 0.449152 | 0.429671 | 0.426582 |

4.3. Experimentation of Challenge Verification Process with Various Distances from Camera

The person distance from the camera greatly affects the challenge successful rate. When the person is too close to the camera, it is difficult to see the face movement. While at the greater distance, the system will have difficulty to detect the person face. So, it is important to specify the ideal distance as part of the system specification. Experiment result shows that a person at a distance of 60-90 cm has the higher rate to complete the challenges that the system presents. While at the distance above 90 cm
the result shows that successful rate significantly decrease. The experiment result is presented in Table 3.

| Person | Challenge successful rate (%) |
|-------|-------------------------------|
|       | 30 CM | 60 CM | 90 CM | 120 CM | 150 CM |
| 1     | 100   | 100   | 100   | 80     | 20     |
| 2     | 100   | 100   | 100   | 80     | 40     |
| 3     | 80    | 100   | 100   | 80     | 0      |
| 4     | 100   | 100   | 100   | 100    | 0      |
| 5     | 60    | 80    | 100   | 0      | 20     |
| 6     | 100   | 100   | 100   | 0      | 0      |
| 7     | 100   | 100   | 100   | 40     | 0      |
| 8     | 100   | 100   | 100   | 0      | 0      |
| Avg   | 92.5  | 97.5  | 100   | 76     | 26.3   |

4.4. Experimentation of Additional Face Attributes to the Successful Detection Rate
The experiment is conducted when the person uses the attributes on the face with a distance of 60 cm from the camera, and the position of the person face is perpendicular to the camera, and the lighting condition is such in a room condition that has no windows and only uses ordinary light bulb. Each person performs each experiment for five times. The attributes used for this test are reading glasses, sunglasses, masks, hats and a combination of these attributes. An example of these attributes are shown in Figure 8.

![Figure 8. Example of Face Attributes](image_url)

Table 4 provides a comparison of successful detection and recognition rate between multiple face attributes combinations. The data shown in Table 4 is an average data from 8 persons for each type of attribute combination. If the person uses a mask that covers almost 50% of the person face, the system will have difficulty detecting and recognizing the face. Hat attribute has minimum effect for the detection and recognition process, but have some difficulty while doing the head-pose challenge. When the person uses a combination of attributes that almost cover all faces, the system will find it difficult to detect and recognize the person. When the mouth, forehead and eyes of the person is obstructed, the system can no longer detect the face.
Table 4. Experiment Data of Multiple Face Attributes Combination

| Attribute                  | Average Detection Rate (%) | Average Recognition Rate (%) | Average Detection Speed (s) | Average Recognition Speed (s) | Challenge Successful Rate (%) |
|----------------------------|----------------------------|-----------------------------|---------------------------|-------------------------------|-----------------------------|
| Glasses                    | 100                        | 100                         | 0.068                     | 0.372                         | 97.5                        |
| Sunglasses                 | 100                        | 98                          | 0.0678                    | 0.375                         | 70                          |
| Mask                       | 70                         | 45                          | 0.0578                    | 0.362                         | 15                          |
| Hat                        | 100                        | 98                          | 0.0844                    | 0.358                         | 65                          |
| Hat, Mask, Glasses         | 53                         | 0                           | 0.0675                    | N/A                           | 0                           |
| Hat, Mask                  | 27                         | 20                          | 0.071                     | 0.387                         | 0                           |
| Hat, Mask, Sunglasses      | 20                         | 0                           | 0.0647                    | N/A                           | 0                           |

4.5. Experimentation of System Robustness to Two-Dimensional Images

In this experiment, a two-dimensional image is placed perpendicular to the camera at a distance of 60 cm. Each registered person image will be presented to the camera for five attempts and will be tilted left and right as if the user is looking left and right. The expected output is that this attempt should not pass the challenges given by the system. All the presented images can be detected and recognized properly, and the average image level that can pass this challenge is 27.5% as shown in Table 5. The hardware of the system being used is presented in Figure 9.

Table 5. System Testing With Two-Dimensional Images To The Camera

| Person | Detection Rate (%) | Recognition Rate (%) | Challenge Rate (%) |
|--------|--------------------|----------------------|-------------------|
| 1      | 100                | 100                  | 40                |
| 2      | 100                | 100                  | 20                |
| 3      | 100                | 100                  | 40                |
| 4      | 100                | 100                  | 60                |
| 5      | 100                | 100                  | 20                |
| 6      | 100                | 100                  | 20                |
| 7      | 100                | 100                  | 20                |
| 8      | 100                | 100                  | 0                 |
| Avg    | 100                | 100                  | 27.5              |

Figure 9. Hardware of The System
5. Conclusion
This research demonstrate the use of low-cost microprocessor to make a simple door security system. The main feature of this research is the implementation of HOG feature extraction method to detect and recognize faces. Throughout the experiment, the optimum distance of a person is obtained to be 60 cm ahead from the camera. By using HOG feature extraction method and the optimum distance, the system can achieve successful detection and recognition rate of 100% with processing time of 0.435 seconds. There are some limitations occur in this implementation, such as the ability to detect and recognize faces that heavily obstructed with face attributes, and a considerably high false detection rate when a two-dimensional images were presented to the camera. While the latter can easily solved by using a dual-camera setup, this experiment proves the ability of additional head-pose challenge to minimize false detection rate in a single camera setup.

6. References

[1] M. Zufar, "Convolutional Neural Networks Untuk Pengenalan Wajah Secara Real-Time," *Jurnal Sains dan Seni ITS*, vol. 5, no. 2, pp. 2337-3520, 2016.

[2] K. D., "Pengertian Mikroprosesor (Microprocessor) dan Cara Kerja Mikroprosesor," 3 January 2020. [Online]. Available: https://teknikelektronika.com/pengertian-mikroprosesor-cara-kerja-mikroprosesor- adalah/. [Accessed 24 September 2020].

[3] K. Pulli, A. Baksheev, K. Kornyakov and V. Eruhimov, "Real-time computer vision with OpenCV," *Communications of the ACM*, vol. 55, no. 6, pp. 61-69, 2012.

[4] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of cognitive neuroscience*, vol. 3, no. 1, pp. 71-86, 1991.

[5] F. Schroff, D. Kalenichenko and J. Philbin, "Facenet: A unified embedding for face recognition and clustering," *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815-823, 2015.

[6] T. Nguyen, B. Lakshmanan and W. Sheng, "A smart security system with face recognition," *arXiv preprint*, 2018.

[7] P. Kamencay, M. Benco, T. Mizdos and R. Radil, "A new method for face recognition using convolutional neural network," *Digital Image Processing and Computer Graphics*, vol. 15, no. 4, pp. 663-672, 2017.

[8] C. Rahmad, R. A. Asmara, D. R. H. Putra, I. Dharma, H. Darmono and I. Muhiqqin, "Comparison of Viola-Jones Haar Cascade Classifier and Histogram of Oriented Gradients (HOG) for face detection," *IOP Conference Series: Materials Science and Engineering*, vol. 732, no. 1, pp. 1-8, 2020.