Indonesian Journal of Physics
Vol. 34. No.2, 25 December 2023
Journal homepage: http://ijphysics.com

\

Utilization and Optimization of Histogram of Oriented Gradients and Machine Learning in Face Recognition System

M. Ervandy Rachmat.[1*], Irfan Dwi A.[1], Fahdzi Muttaqien[1]
[1]*Physics Department, Bandung Institute of Technology, Ganesa St. No.10, Bandung, 40132, Indonesia*

(Received: 29 September 2023, Revised: 03 January 2024, Accepted: 05 January 2024)

Abstract

Computer science and technology development in recent years has experienced great developments. This time, some types of technology digitise almost everything related to human life, including facial recognition. In recent years, various methods for recognising human faces have developed. One of them is using the Histogram of Oriented Gradients (HOG). On this occasion, an image processing system will be designed to recognise human faces using Histograms of Oriented Gradients (HOG) and machine learning such as Convolutional Neural Networks (CNN) and Support Vector Machines (SVM). This system detects the winking of the face by using computer-recognisable points in the eye area from 68 facial landmarks. From these results, the distance between the upper and lower eyelids can be measured. If the distance (in pixels) is small enough, it can be interpreted as a wink. In addition, it is also limited by the distance of faces that can be detected to blink. In the end, if a recognised face blinks are detected, the time and date will be recorded. It will then open a solenoid lock using serial communication via Arduino Uno to become a security system. From 100 facial photos and 207 blink tests, 89.86% found that the computer could detect a "True Positive" wink. Besides, the recommended tolerance parameter value for this facial recognition system is between 0.42 and 0.48.

Keywords: face, histogram of oriented gradients, image processing, machine learning

INTRODUCTION

The human face, with its unique characteristics, has become a focal point in the development of machine learning and security systems. As technology evolves, computers are increasingly capable of performing tasks traditionally done by humans, including those related to security. Modern security systems are now integrating with computers, gradually replacing conventional methods. These systems often utilize distinct human features for verification purposes.

This paper will delve into the intricacies of integrating facial recognition systems with computers. The system under discussion is designed to differentiate between registered and unregistered faces in its database. It is a sophisticated piece of technology that must discern between a real face, and one represented in a photograph. This is achieved by instructing the detected face to blink. If a blink is detected, the system concludes that the face is genuine and not a photographic representation. This paper also discusses about the best parameter used in this facial recognition system.

The history of face recognition is the result of research that has been running for more than 50 years, it is started between year's 1964 and 1966 when Woodrow W Bledsoe tried to determine how computers recognize current human faces. That system is simply measuring the hairline, eyes, and nose. However, this did not work successfully, this is because there are many variations in parts of the face, for example the tilt of the face towards the camera, facial expressions, age of the face, light intensity, etc., which are very unlikely to be the same every time. A major source of difficulty in many real-world artificial intelligence applications is that many factors of variation influence every single piece of data such that it is very difficult to extract abstract feature from raw data using traditional system [7]. This problem began to be overcome due to improvements in

[1*] Corresponding author.

E-mail address: rachmatervandy@gmail.com

technology such as cameras, computer processes, and the emergence of machine learning [1]. With the computer's ability to recognize a person's face, this ability can be used into a system that is useful for society. Applications of this technology include facial recognition for employee access to banks [2], facial recognition for student lecture attendance [3], burglary detection system [6], and various other systems that use facial recognition as the foundation of their operations. Other countries like in UK, USA, and Australia has been installed and operated this system in various types of (quasi)public space – including factories, cafes, airports, shopping areas, and government buildings [10].

The facial recognition system, despite the sophistication of this system, it turns out that it still has weaknesses that have the potential to be exploited by irresponsible parties. Based on the Usenix security conference, security and computer vision specialists from the University of North Carolina said that the system uses 3D (3 Dimensional) digital facial models based on available photos which are then displayed using virtual reality to penetrate facial recognition security systems [4]. Apart from that, it was also added that photos originating from social media such as Facebook, LinkedIn, Instagram, and even Google can be a source for getting photos of someone's face. Even though the photos obtained are of poor quality, they can still be used to create a 3D model system that can be used to penetrate facial recognition systems. This is dangerous because if the uniqueness of our face which is used as a biometric can easily spread/distributed on social media, this is the same as giving away our security system password for free. To avoid this, it is necessary to improve the facial recognition system by adding features that can strengthen the security system.

The development and implementation of such a system require a deep understanding of both computer science and human physiology. It involves complex algorithms and machine learning models that can accurately analyze and interpret human facial features. Moreover, it requires a comprehensive database of faces to ensure the system's accuracy and reliability.

The potential applications of this technology are vast, ranging from personal device security to access control in high-security areas. However, it also raises important questions about privacy and data security that must be addressed.

This innovative approach to security has significant implications for the future of computer-integrated systems. It represents a major step forward in the field of biometric security and opens up new possibilities for research and development.

THEORY

Histogram of Oriented Gradients is a face detection method that uses calculations such as calculating the gradient of a 2-dimensional plane. A photo is a collection of pixels consisting of several colour combinations whose values vary, but for HOG, the colour format used is grayscale. Grayscale is often used because this colour format only consists of one range of values from 0 to 255; 0 represents black, and 255 represents white (this value can be reversed). The image is first divided into small, connected regions, called cells, and for each cell a histogram of edge orientations is computed. The histogram channels are evenly spread over 0–180° or 0–360°, depending on whether the gradient is ‘unsigned’ or ‘signed’. The histogram counts are normalized to compensate for illumination [5].

	90	
50		120
	10	

Fig. 1. Example of block cell

To measure the gradients magnitude, it can be done using this formula.

$$Gradient\ magnitude = \sqrt{\Delta x^2 + \Delta y^2} \qquad (1)$$

And for measure the gradient angle can be done with this formula.

$$Gradient\ angle = \tan^{-1}\left(\frac{\Delta x}{\Delta y}\right) \qquad (2)$$

Δx and Δy represent the different distance values on the same axis, which is the x-axis and y-axis. So, from Fig.1, the Δx is 70 (120-50), and Δy is 80 (90-10), the gradient magnitude is 106.3 and the gradient angle is 48.2^o. The entire process is repeated until all block cell sections are done.

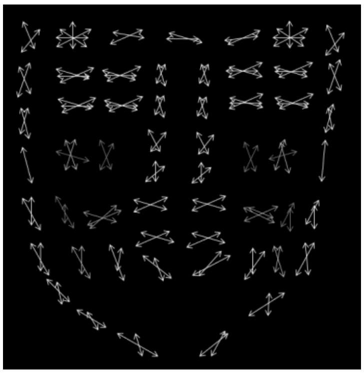

Fig. 2. HOG from human face

The image above results from a facial pattern based on HOG obtained from pictures of human faces

[1]. The HOG results are generated from various human photos, which are then given to the computer to be trained so that the HOG shape for the human face is obtained and can be used as a reference for a face detection system. The image explains that there are shapes of eyes, nose, mouth, and other parts of the face, so now computers can recognise human facial patterns. For practical purposes, the results of the HOG pattern obtained from the image above will be used as a reference for computers to recognise facial patterns.

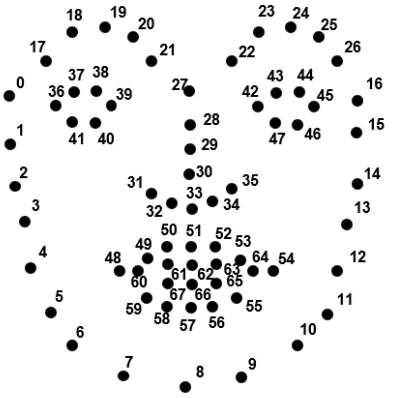

Fig. 3. Facial landmark features

The image above illustrates the application of the HOG method [8] in facial recognition, highlighting 68 distinct points on the face. These points serve various purposes, such as creating facial filters on social media by determining the positions of the eyes, nose, mouth, and other features. In this research, these facial landmarks are specifically used to identify the position of the eyelids.

Once computers can detect human faces using HOG, the subsequent challenge lies in enabling computers to distinguish one face from another. This involves recognizing faces that belong to specific individuals. To address this challenge, machine learning techniques are employed. In this context, a Convolutional Neural Network (CNN) is utilized due to its proficiency in handling image processing problems. CNNs are a kind of artificial neural networks (ANNs) that use convolution operations in at least one of their layers [8]. A CNN usually takes an order 3 tensor as its input, e.g., an image with H rows, W columns, and 3 channels (R, G, B color channels). Higher order tensors inputs, however, can be handled by CNN in a similar fashion. The input then sequentially goes through a series of processing. One processing step is usually called a layer, which could be a convolution layer, a pooling layer, a normalization layer, a fully connected layer, a loss layer, etc [8]. Multilayer networks can learn complex and high dimensional patterns from large datasets, making them obvious candidates for images recognition task [9].

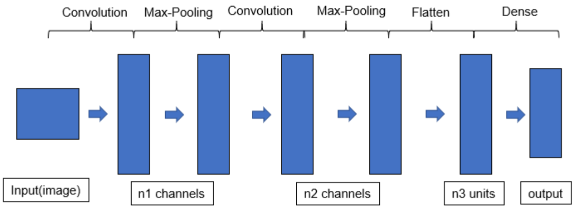

Fig. 4. Illustration of CNN process

In the realm of facial recognition, the Histogram of Oriented Gradients (HOG) method is utilized to detect faces, while a Convolutional Neural Network (CNN) is employed to distinguish one face from another. In this context, the CNN evolves into a Deep Convolutional Neural Network (DCNN), tasked with identifying crucial characteristics of the human face.

The process of facial feature extraction conducted by the DCNN yields 128 measurement points, referred to as face embeddings. These points represent the faces of each individual in the database. Although the physical significance of these 128 measurement points remains unclear, it is not a critical concern for this research. The primary objective here is to devise a method to differentiate one face from another effectively.

This innovative approach, combining HOG and DCNN, offers a robust solution for facial recognition. It not only enhances the accuracy of distinguishing faces but also contributes significantly to advancing machine learning and image processing techniques.

EXPERIMENTAL METHOD

In this research, the computer leverages the Histogram of Oriented Gradients (HOG) algorithm to detect facial patterns. Following this, a Deep Convolutional Neural Network (DCNN) is employed to extract facial features, resulting in 128 calculations known as "face embeddings". Upon face detection, the computation of 68 facial landmarks is carried out to ascertain the position of the upper and lower eyelids of both eyes.

A Support Vector Machine (SVM) is utilized to classify the 128 DCNN output calculations against the HOG and DCNN calculations from the reference face photo (training). The closest result captured by the camera to the face from the image (database) will be identified as the face in question. If the calculation deviates significantly from that in the database, it will be labeled as 'Unknown'. However, if the results align closely with those in the training database, then the person's name will appear, and they will be asked to blink for verification of authenticity (not a photo).

To enable the computer to detect whether an eye is blinking, 68 facial landmarks are used. These landmarks measure the distance (in pixel size) between the upper and lower eyelids of both eyes. Subsequently, the computer measures the distance between the left and right eyes, calculated from the midpoint between the upper and lower eyelids for both eyes. If this measured distance falls between 100 and 200 pixels, a blink will be requested. Upon detection of a blink, real-time data will be captured, including the name, time, and date of blinking. This data is recorded in an Excel file format.

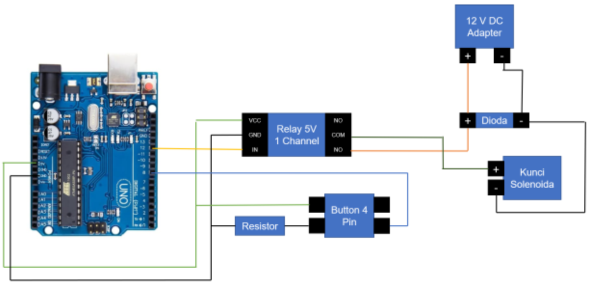

Fig. 5 Hardware scheme

This research introduces an additional component to the facial recognition system - a solenoid lock. The lock is designed to open when a person, who is recognized in the database, blinks. This feature simulates a facial recognition system that can be applied in home security systems or other security systems requiring facial verification to unlock.

The inclusion of a blink as a form of verification enhances the security system's robustness, reducing breaches where previous systems failed to differentiate between real faces and faces from photos. To evaluate the accuracy of facial recognition, facial data from 100 individuals were used, including 33 from ITB Physics undergraduate students and the rest from various public figures such as artists, researchers, athletes, whose photos were sourced from the internet.

These photos are crucial for testing the accuracy level of this facial recognition system, determining whether faces not in the database will be detected as other people in the database or vice versa. It's important to note that the database photos used for training consist of only one photo per person, with varying sizes, backgrounds, and positions in the photo, not all directly facing the camera.

RESULTS AND DISCUSSION

In this study, facial recognition trials were conducted using facial data from 100 individuals. Of these, 38 were ITB Physics undergraduate students, 25 of whom participated in the blink test. The remaining 75 faces comprised 13 Physics undergraduate students who did not participate in the blink test and 62 other individuals, including artists, researchers, athletes, whose photos were sourced from the internet.

These photos play a vital role in evaluating the precision of the facial recognition system, helping to ascertain if a person in the database is mistakenly identified as another individual or not detected entirely. This rigorous testing process ensures the reliability and effectiveness of the facial recognition system, contributing significantly to its potential applications in various fields.

Table. 1. Summary face and blink detection

	Positive	Negative
True	186 (89.86%)	-
False	20 (9.66%)	1 (0.48%)

The facial recognition system under study was evaluated using data from 100 individuals. This included 38 ITB Physics undergraduates, 25 of whom participated in the blink test, and 62 individuals from diverse backgrounds such as artists, researchers, and athletes, whose photos were obtained from the internet.

From 100 individuals then 207 blink tests conducted, which each person can tried the blink test more than once, the result from total blink test are 186 result test were detected as "True Positive" blinks, where the individual blinked and the computer correctly identified the blink. There were 20 instances of "False Positive" where one person blinked but the computer detected it as another person blinking. There was one instance of "False Negative", where an individual was in front of the camera but was detected as "Unknown" despite having their photo in the database.

The system's accuracy is influenced by various factors such as the brightness level of the environment, the angle of the face towards the camera, and the background captured by the camera. The program predicts the output of the training results closest to the data in the database, so there may be several errors in the investigation.

A parameter called "tolerance parameter" is used to measure faces against data in the database. The smaller this value, the stricter the computer program will be in determining the face detected by the camera. However, this value can be adjusted freely, but it will also affect the level of accuracy.

In one case of "False Negative", despite using the same percentage parameter value, the computer did not detect a person when testing was carried out. This could be due to factors such as noise in the background of the photo, position of face not directly facing camera, size of face relatively far from camera and attributes like glasses which cover eyebrow line and create shadows on face.

This research demonstrates that while facial recognition systems have made significant strides in accuracy and reliability, there are still challenges to overcome. The findings underscore the importance of continuous improvement and adaptation in response to these challenges.

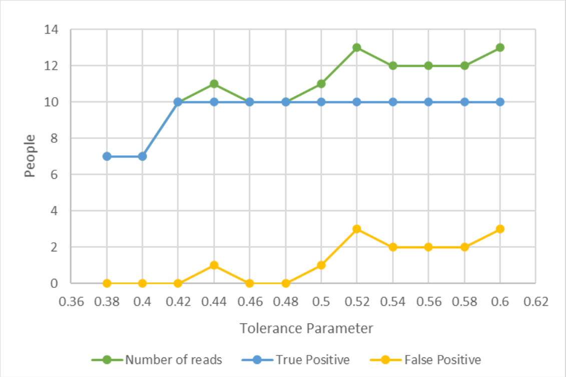

Fig. 6. Number of readable people against the tolerance parameter

The tolerance parameter value in this program must be carefully calibrated to suit the system. If the tolerance parameter value is too small, it can render the faces in the database undetectable during testing. Conversely, if the tolerance parameter value is too large, it can result in reading errors for people who don't have data in the database and are detected as someone else or the person in the database is read as someone else (False Positive).

To address this, a test was carried out using ten people, photos of their faces were entered into a database containing 100 other people's faces, and then the tolerance value test was carried out from a tolerance value range of 0.38 to 0.6. The results were represented graphically, with the x-axis representing the tolerance parameter value used and the y-axis representing the number of readable people when the test was carried out for ten people.

This parameter tolerance set by measuring the different between data face from the database and the data during while testing, the parameter tolerance set the highest value that can be accepted by computer. If the measuring result is higher than this tolerance, then the computer will not make any decision in face recognition system like saving the blinking time or whom face is detected from camera. These numbers in Figure 6 got by several experiment, initially the tolerance is set between 0.1 to 0.8, then the numbers was cut between 0.36 and 0.6 because it is good enough to detecting human face correctly but sadly the accuracy to recognizing each person is still low, so these numbers was processed again by testing to get best tolerance parameter value.

In the graph, there are three lines:

- The green line indicates the total number of people read by a computer.
- The blue line shows the number of people correctly detected as "True Positive".
- The yellow line shows the number of people detected as "False Positive".

It was observed that when the tolerance parameter values were 0.38 and 0.4, only seven people were detected, all of them being "True Positive". For larger parameter values in the range 0.52 to 0.6, there was "False Positive" data.

According to this experiment, a suitable range for the tolerance parameter value is around 0.42 to 0.48. However, at a tolerance value of 0.44, there was an additional one person as "False Positive", which occurred due to non-perpendicular and non-parallel face orientation towards the camera during testing.

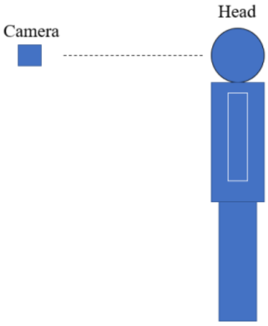

Fig. 7. Best camera position to this system

To make this system work properly, the camera must be placed in front of the face that want to be detected just like figure 7 above. Also make sure that the face must be exposed with good lighting.

CONCLUSION

In conclusion, the research methodology employed in this study yielded facial recognition results with an accuracy level of 89.86% from 100 facial photos in the database, using a tolerance parameter value of 0.47 and conducting 207 tests. During face testing, individuals whose photos are not in the database were consistently recognized as "Unknown", while those in the database were accurately recognized.

The system prompts individuals to blink as a form of verification. If a blink from a recognized person is detected, the computer records the time of the blink in an Excel file named Attendance1.xlsx and saves it locally on the computer. The computer then commands the Arduino Uno to unlock the solenoid for 6 seconds.

It's important to note that only one photo is used for each person in the database. Therefore, when testing, the conditions should be similar to those in the database photos (for example, lighting, face position, and face direction when facing the camera).

Finally, it is recommended that the tolerance parameter value should be within the range of 0.42 to 0.48. In this range, the system demonstrates a higher level of accuracy in recognizing individuals compared to other parameter values.

REFERENCES

[1] THALES. 2023. Facial Recognition History. www.thalesgroup.com. Accessed in August 2023.

[2] Arsal, M., Wardijono, B. A., and Anggraini, D., Face Recognition Untuk Akses Pegawai Bank Menggunakan Deep Learning Dengan Metode CNN. *J. Nas. Teknol. dan Sist. Inf,* **6**(1), 55, 2020.

[3] Santoso, B., & Kristianto, R. P., Implementasi Penggunaan Opencv Pada Face Recognition Untuk Sistem Presensi Perkuliahan Mahasiswa. *Sistemasi: Jurnal Sistem Informasi*, **9**(2), 352, 2020.

[4] Newman, Lily H. (2016). Hacker Trick Facial-Recognition Login with Photos from Facebook (What Else?). https://www.wired.com/2016/08/hackers-trick-facial-recognition-logins-photos-facebook-thanks-zuck/. Accessed in June 2023

[5] Déniz, O., Bueno, G., Salido, J., & De la Torre, F., Face recognition using histograms of oriented gradients. *Pattern recognition letters*, **32**(12), 1598, 2011

[6] Ng, Ri & Lim, Kian & Lee, Chin-Poo & Abdul Razak, Siti Fatimah, Surveillance system with motion and face detection using histograms of oriented gradients. *Indonesian Journal of Electrical Engineering and Computer Science*, **14**, 869, 2019.

[7] Goodfellow, I., Bengio, Y., & Courville, A., Deep learning. MIT press, 2016.

[8] Wu, J., Introduction to convolutional neural networks. National Key Lab for Novel Software Technology. Nanjing University. China, **5**(23), 495, 2017

[9] LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P., Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, **86**(11), 2278, 1998.

[10] Mark Andrejevic & Neil Selwyn (2020) Facial recognition technology in schools: critical questions and concerns, *Learning, Media and Technology*, **45**:2, 115-128, 2020.