Development Geofencing Process and Face Recognition Design Using Haversine Formula and the K-Nearest Neighbor Algorithm in the Employee Attendance Application

Fahrurrozi Lubis[1,a)], Ifan Prihandi [2,b)], Wendi Usino[2,d)], Nor Alina Binti Ismail[3,c)]

[1] *Faculty of Computer Science and Information Technology, University of North Sumatra, Medan, Indonesia*
[2] *Faculty of Computer Science, Mercu Buana University, Jakarta, Indonesia*
[3] *Faculty of Data Science and Computing, Universiti Malaysia Kelantan, Malaysia*

[a)] *Corresponding author: fahrurrozi.lubis@usu.ac.id*
[b)] *ifan.prihandi@mercubuana.ac.id*
[c)] *alina.i@umk.edu.my*
[d)] *wendy.usino@budiluhur.ac.id*

Abstract. Attendance is an important factor in assessing employee discipline for a company. Discipline issues in recording employee attendance are an important concern for companies that have many branches or outlets because the company cannot directly monitor their employees. Some companies usually use finger print machines and some even use manual methods. This method is a problem for several companies that have several outlets or branches to record attendance data centrally and accurately. As a solution that is with the application of employee attendance that applies geofencing and face recognition as security in recording the attendance of employees who are at the company's outlets or branches. Geofencing is a technology for conducting remote surveillance of a predetermined area. While face recognition is a technology from a computer to recognize a person's face. In this study using the haversine formula, Euclidean distance, and KNN (K-Nearest Neighbor) algorithm in determining the last location and face recognition of employees by comparing face data that has been registered in the database. Users who record attendance outside the radius cannot record attendance. The face recognition accuracy rate has a percentage of 98%.

Keywords: Geofencing, K-Nearest Neighbor, Face Recognition

INTRODUCTION

The attendance system is one of the important components for the company in managing employee attendance data. Attendance management is an act of managing attendance [1]. Attendance management is one of the basic and important processes in any organization [2]. Currently the growth of technology is much better into all existing sector, attendance system is one of them. The system update is expected to be able to support the need for faster data information. Monitoring attendance and counting hours worked is very important for almost every institution or organization [3]. Despite this there are many industries, companies and institutions that take attendance manually in attendance registers by calling out their names [4]. The attendance system is intended so that employee attendance data can be recapitulated and recorded in real-time on the system to be able to be evaluated [5]. The attendance system not only requires a lot of time but is also less effective if manual calculations and information are slow [6]. With so many obstacles encountered when using the manual method of recording attendance. then the employee attendance system application is made.

Geofencing is a feature in a software program that is used in conjunction with a global positioning system (GPS) in determining geographic boundaries or virtual parameters of a map. The Geofence definition of the area may be a polygon or a radius around a specified point [7]. Global Positioning Service System or GPS is relative new technology. The GPS receiver calculates the exact position of the device and the speed at which the device moves [8]. Even though it's mainly used for the military the goal at the time of discovery, this technology then used in many consumer applications [9]. Geofencing is a technique that allows applications to provide information in a more appropriate way, at the right time and in the right place [10]. In cellular-based geofencing systems, the positioning of devices, determined by satellite-GPS technology, together by matching positions with a set of geofences executed on a mobile

Proceedings of the 6th International Conference on Computing and Applied Informatics 2022
AIP Conf. Proc. 2987, 020007-1–020007-9; https://doi.org/10.1063/5.0200763
Published by AIP Publishing. 978-0-7354-4900-8/$30.00

device [11]. GPS navigation allows many people who are not related to Geomatics to use it in many parts of their lives [12].

In addition to geofencing this application also applies Face Recognition technology, which is one of the biometric technologies that have been widely applied in security systems in addition to recognition of the retina of the eye, fingerprint recognition and iris. The first stage is face detection, where photos are sought to find any face. The second stage is face recognition, where faces are detected and processed, and compared with a database of known faces to determine who the person is [13].

The algorithm used to get the closest distance of vector or face image comparison with data stored in the database is using the Euclidean distance algorithm, and to classify the results of Euclidean distance calculations using the KNN (K-Nearest Neighbor) algorithm.

Problems that often arise in the attendance system using finger print or manual method generally when the process of recording attendance data that takes longer than using an online attendance system because the data is centralized on one server and also a problem of the accuracy of data for those who use the manual method.

The purpose of this research is to produce an Android-based employee attendance system application that is fast and accurate in the process of recording attendance data for companies that have many branches of the company by applying geofencing and face recognition technology to support the accuracy of the attendance data.

RESEARCH METHOD

Software Development Method

A software development project is a structure imposed on the development of a software product [14]. Model Application design in this system, the authors use the process method SDLC (System Development Life Cycle) or the Waterfall Method. Waterfall model is imposed as a software methodology [15]. The Waterfall method is a classic life cycle, which implies a systematic and sequential approach. Consist of Requirements, Implementation , Verification and Maintenance

Algorithm

In this study using the Haversine Formula, Euclidean Distance, and KNN (K-Nearest Neighbor) algorithm which is applied to the Android-based employee attendance system application to validate employee attendance data [16].

The Haversine formula is a method that calculates the considered distance appropriately and accurately [17]. The Haversine formula is an important equation in navigation, giving the distance of a large circle between two points on the surface of the earth based on longitude and latitude [18]. The Haversine formula is used to calculate the distance between a user location and a point [19].

$$D = (sin\ sin\ (lat1)\ .sin\ sin\ (lat2)\ + cos\ cos\ (lat1)\ .cos\ cos\ (lat2)\ cos\ cos\ (long2 - long1)\ )\ .R \quad (1)$$

Information :
R = the radius of the earth 6371(km)
d = distance(km).

The Euclidean distance is an accurate technique even in conditions such as sudden, temporary change replacement operations [20]. The Euclidean distance formula takes the square root of the overall difference [21]. Euclidean distance calculates the distance between two points.

$$d(p,q) = \sqrt{\sum_{i=1}^{n} (q_i - p_i)^2} \quad (2)$$

The formula can be implemented on the system. Equation d is the result of calculating the closest distance, equation q is the result of encoding data or the results of solving face images whileequation p is vector data or face image stored

in a database. After getting the closest distance from the vector or face image comparison, the system will calculate into a percentage.

The K-Nearest neighbor algorithm is a simple artificial one-based learning model [22]. K-Nearest Neighbor algorithm classifies based on distance learning data closest to the object [23]. The KNN algorithm is used because it is a simple machine learning algorithm [24]. In this application KNN is used to convert the Euclidean distance calculation results into percentages [25]

The Confusion Matrix is a method for calculating the correct number and wrong predictions which are further summarized by the number of calculated values and descriptions into each class [26]

Predicted values	Actual Values		
		Positive (1)	Negative (0)
	Positive (1)	TP	FP
	Negative (0)	FN	TN

FIGURE 1. Confusion Matrix Table

Then the accuracy value can be calculated by using the following formula:

$$Accuracy = \frac{TP + TN}{TP + FP + FN + TN} \qquad (3)$$

RESULTS AND DISCUSSION

Requirement Analysis

After analysing the data, the needs of the Android-based employee attendance system application are as follows:

1. **User Needs**

   Here are some user requirements analysis:

- Can see their last location
- Can take face photos
- Can review their attendance data before sending to the server
- Can see the results of attendance records on the same day

2. **Admin Needs**

   Here are some admin requirements analysis:

- Can manage employee data
- Can manage employee work location data
- Can manage employee work shift data
- Can download report attendance

Geofencing

In conducting geofencing validation, this study uses the haversine formula algorithm. After doing the calculation process using the haversine formula algorithm, the next step is to convert the results of the haversine formula algorithm into meters by multiplying by one thousand. Validation of the radius in this system is that if the haversine formula calculation distance is more than the radius data stored in the database, the user will get a message that the person is outside the specified radius. Geofencing validation process flowchart as follows:

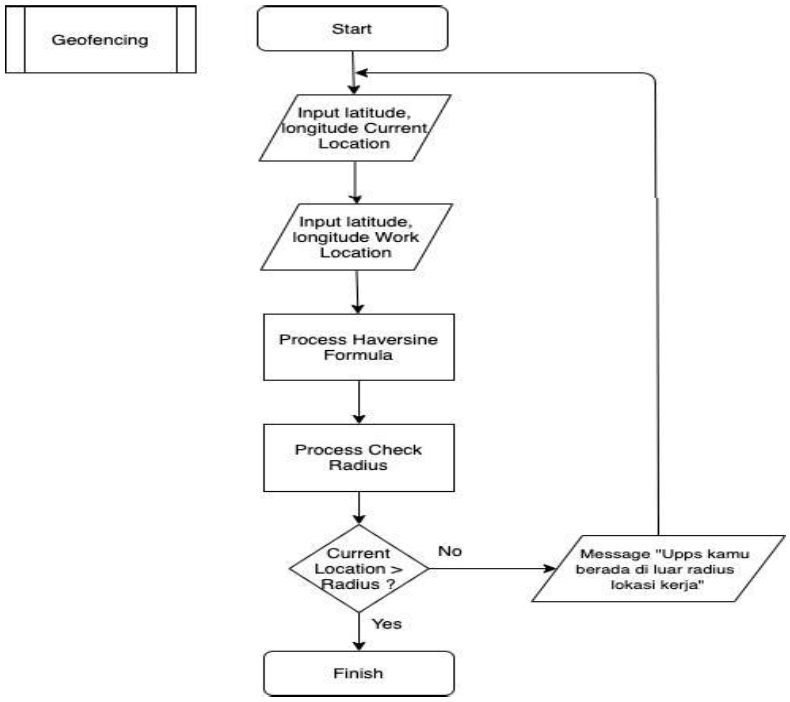

FIGURE 2. Process Geofencing

The accuracy of taking the user last location by the system can be influenced by user smartphone device factors. The better the quality of GPS in a smartphone hardware, the accuracy will also be higher. Users can refresh the page to get the most accurate location if the location obtained by the system was not yet right. In this test the radius is set to 50 meters the maximum distance from the center point of a predetermined location. The results are as follows:

The first experiment is to record attendance outside the specified radius and the results are as follows

FIGURE 3. The user outside radius
Translation:
Apakah lokasi kamu sudah sesuai? = Is your location appropriate?

From Define the user at the outer radius of the specified work location. Because it is outside the radius, the system will display messages that are claimed to be outside the radius.

a. The second experiment is to record attendance within a predetermined radius. When the geofencing validation process is successful, the system will switch pages to face capture page. The results are as follows:

FIGURE 4. The user inside radius
Translation:
Apakah lokasi kamu sudah sesuai? = Is your location appropriate?

From the results of this geofencing test show that the results are in accordance with the expected. The system successfully applied geofencing technology to validate employee attendance locations using the haversine formula algorithm.

Face Recognition

The application of face recognition on the Android-based employee attendance system application is a technology to validate the accuracy of the user's or employee's face when doing attendance. In this study using the face-api.js library to get face images. face recognition is beneficial for passive and non-intrusive systems in verifying personal identity [27].

This library represents each face as a point in an imaginary 128-dimensional space. To check if two faces match, it checks if the distance between those two points is less than 0.6 (the default threshold). Using a lower threshold than 0.6 makes the face comparison more strict [25].

After the next process of solving face images is classified using the KNN (K-Nearest Neighbor) algorithm. At value k=1, and check if matches are thereby fixing a threshold value as 0.6 for distance metric (Sinha, Ghosh, and Manju 2020). The higher the threshold value, the chance of misclassification of data will be more inaccurate. In this study using a threshold value of 0.4 because the accuracy is higher than the threshold value of 0.6. The following are the steps of the face recognition process in the system

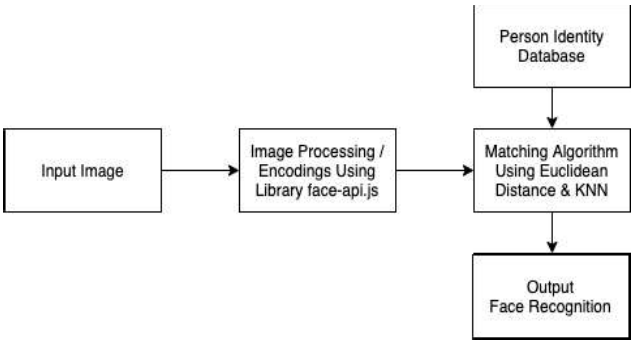

FIGURE 5. Block Diagram Process Face Recognition

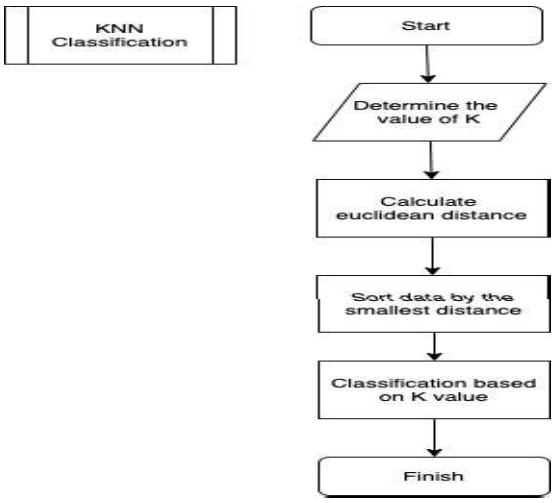

FIGURE 6. KNN Classification

Validating face recognition requires user data that has been stored in the database. In this application as a comparison of data that uses thirty examples of the user's face image stored in the database. When the user is in the attendance process, if the user first records attendance, then the first photo taken will be used as comparison data for the next face photo. The second and subsequent photos only face data whose accuracy is above 75% will be saved to the database as a comparison of the next face data.

The level of accuracy of the face recognition can be influenced by factors including light when taking photos, and blur levels when taking photos. The following are the results of face recognition testing with a value of K = 1 and a

threshold value of 0.4 using euclidean distance and KNN classification. In this face recognition test is tested by logging in using one of the employee accounts that have been registered in the Android based employee attendance application system. Photos will be compared based on the employee account photo training data. The account photos will be compared with photos stored in the database to get the percentage of face recognition results.

FIGURE 7. Photo when recording attendance. Photo is blurred in this manuscript due to privacy reason.

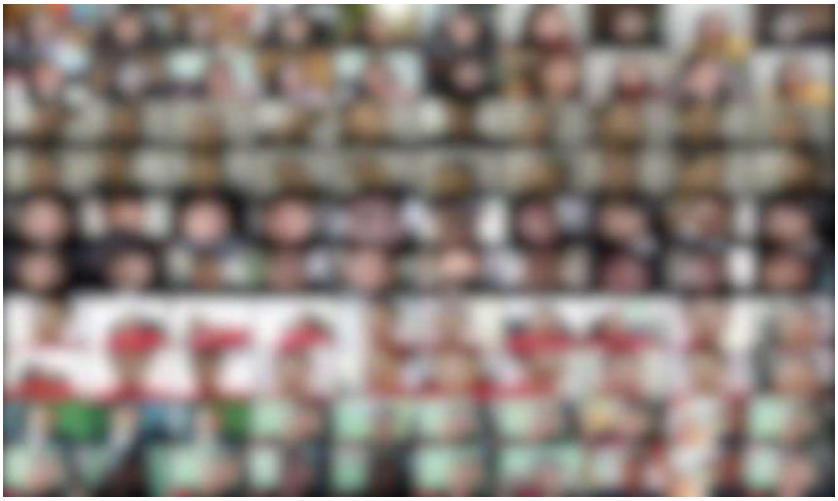

FIGURE 8. Face photo data stored in the database. Photo is blurred in this manuscript due to privacy reason.

In the test scenario, the face photo data will be tested with 100 facial image training data stored in the database and 50 test data. In the test data, there are photos of faces from various positions, including photos of faces facing forward, to the edges and also photos of faces with minimal lighting. In addition, the 50 test data also consisted of photo data registered in the database and face photo data that were not registered in the database. To analyze the results of this scenario, confusion matrix method is used.

TABLE 1. Confusion Matrix

		ACTUAL VALUE	
		TRUE	**FALSE**
PREDICTION VALUE	**TRUE**	TP = 45	FP = 1
	FALSE	FN = 0	TN = 4

Based on table 1, the accuracy value can be calculated using the following formula:

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

$$= \frac{45+4}{45+1+0+4} = \frac{49}{50}\ x\ 100\ \% = \ 98\ \%$$

The results of facial recognition testing, the accuracy rate of facial recognition success is 98% from the test results with training data of 100 photos of face coding results and test data of 50 photos of faces.

CONCLUSION

Based on various tests that have been done, it can be concluded that :

1. Users who record attendance outside the radius cannot record attendance.
2. This process of solving face images will produce each face as a point in 128 imaginary dimensions.
3. Validating face recognition requires user data that has been stored in the database.
4. Face accuracy above 75% will be saved to the database as the next face comparison data.
5. The level of accuracy of face recognition has an average percentage above 98%.
6. The higher the face matching threshold the chance of misclassification of data will be more inaccurate.
7. Based on testing the system is running well, geofencing validation and face recognition are proceeding as expected.

REFERENCES

[1] B. Nii, W. O., and K. Andoh, "AttendanceBot Location based Attendance Tracker for Workers," *International Journal of Computer Applications*, vol. 178, no. 51, pp. 12–18, Sep. 2019, doi: 10.5120/ijca2019919414.

[2] T. Sutabri, P. Pamungkur, A. Kurniawan, and R. E. Saragih, "Automatic Attendance System for University Student Using Face Recognition Based on Deep Learning," *International Journal of Machine Learning and Computing*, vol. 9, no. 5, pp. 668–674, Oct. 2019, doi: 10.18178/ijmlc.2019.9.5.856.

[3] S. Sultana, A. Enayet, and I. J. Mouri, "A Smart, Location Based Time and Attendance Tracking System using Android Application," *International Journal of Computer Science, Engineering and Information Technology*, vol. 5, no. 1, pp. 01–05, Feb. 2015, doi: 10.5121/ijcseit.2015.5101.

[4] M. N. Kallimani Veena G S, Divakar Harekal, Jagadish S, "Biometric Based Attendance Management System using Cloud," *International Journal of Advanced Science and Technology*, vol. 29, no. 3s, Mar. 2020, Accessed: Jan. 30, 2023. [Online]. Available: http://sersc.org/journals/index.php/IJAST/article/view/5551

[5] Y. Faridah, H. Nasir, A. K. Kushsairy, S. I. Safie, S. Khan, and T. S. Gunawan, "Fingerprint Biometric Systems," *Trends in Bioinformatics*, vol. 9, no. 2, pp. 52–58, Sep. 2016, doi: 10.3923/tb.2016.52.58.

[6] R. Singh, J. Kumar, "Wireless LAN based Mobile Attendance System (WL-MAS)," *International Journal of Recent Technology and Engineering*, vol. 8, no. 2, pp. 4555–4564, Jul. 2019, doi: 10.35940/ijrte.b3456.078219.

[7] M. Alsaqer, B. Hilton, T. Horan, and O. Aboulola, "Performance Assessment of Geo-triggering in Small Geo-fences: Accuracy, Reliability, and Battery Drain in Different Tracking Profiles and Trigger Directions," *Procedia Engineering*, vol. 107, pp. 337–348, 2015, doi: 10.1016/j.proeng.2015.06.090.

[8] M. T. U. Scott, T. J. Scott, and V. G. Kelly, "The Validity and Reliability of Global Positioning Systems in Team Sport," *Journal of Strength and Conditioning Research*, vol. 30, no. 5, pp. 1470–1490, May 2016, doi: 10.1519/jsc.0000000000001221.

[9] N. Chan Soe and T. Lai Lai Thein, "Haversine Formula and RPA Algorithm for Navigation System," *International Journal of Data Science and Analysis*, vol. 6, no. 1, p. 32, 2020, doi: 10.11648/j.ijdsa.20200601.14.

[10] A. M. Muriach, "*Geofencing Event-Based System*," 2015.

[11] V. Carchiolo, P. W. Modica, M. P. Loria, M. Toja, and M. Malgeri, "A Geofencing Algorithm Fit for Supply Chain Management," Proceedings of the 2018 Federated Conference on Computer Science and Information Systems, Sep. 2018, doi: 10.15439/2018f238.

[12] H. Abbas Neam Al_ dami and I. Riyadh Abdulridha, "The Possibility of Using Navigation GPS Devices in the Survey Engineering," *International Journal of Engineering & Technology*, vol. 7, no. 4.20, p. 293, Nov. 2018, doi: 10.14419/ijet.v7i4.20.25942.

[13] R. G. Radityatama, *Rancang Bangun Aplikasi Mobile Android Sistem Kehadiran Mahasiswa Melalui Pencocokan Wajah Dengan Menggunakan Library Android Face Recognition With Deep Learning Studi*. 2017.

[14] N. Redhal, "*A Detailed Software Model Analysis*," *Int. J. Res.*, vol. 1, no. 10, pp. 918–922, 2018, [Online].

Available: http://edupediapublications.org/journals/index.php/ijr/article/view/910.
[15] N. A. Rahman *et al.*, "*International Journal of Advanced and Applied Sciences Incorporating teamwork in waterfall model-based project*," 2018, *Int. J. Adv. Appl. Sci.*, vol. 5, no. 12, pp. 126–135.
[16] I. Setyorini and D. Ramayanti, "*Finding Nearest Mosque Using Haversine Formula on Android Platform*," *J. Online Inform.*, vol. 4, no. 1, p. 57, 2019, doi: 10.15575/join.v4i1.267.
[17] M. Basyir, M. Nasir, S. Suryati, and W. Mellyssa, "*Determination of Nearest Emergency Service Office using Haversine Formula Based on Android Platform*," *Emit. Int. J. Eng. Technol.*, vol. 5, no. 2, pp. 270–278, 2018, doi: 10.24003/emitter.v5i2.220.
[18] P. Deshmukh, A. Bhajibhakre, S. Gambhire, A. Channe, and N. Deshpande, "*Survey of Geofencing Algorithms*," 2018, *Int. J. Comput. Sci. Eng. Tech.*, vol. 3, no. 2, pp. 1–5, [Online]. Available: http://www.ijcsejournal.org.
[19] R. Purbaningtyas, A. Arizal, and M. Sholehuddin, "*Regional Leading Potential Recommendations: Implementation of Haversine Formula in Sidoarjo on Hands Mobile Applications*," *Probl. Inf. Technol.*, vol. 10, no. 2, pp. 70–76, 2019, doi: 10.25045/jpit.v10.i2.10.
[20] M. Kintali, S. R. Sura, N. R. Pulivarthi, and A. K. Kondaji, "*Euclidean Metric based Fault Diagnosis in Power Transmission Line*," *Int. J. Innov. Technol. Explor. Eng.*, vol. 9, no. 2S3, pp. 226–231, 2019, doi: 10.35940/ijitee.b1057.1292s319.
[21] S. Beugelsdijk, B. Ambos, and P. C. Nell, "*Conceptualizing and measuring distance in international business research: Recurring questions and best practice guidelines*," *J. Int. Bus. Stud.*, vol. 49, no. 9, pp. 1113–1137, 2018, doi: 10.1057/s41267-018-0182-4.
[22] J. Diz, G. Marreiros, and A. Freitas, "*Applying Data Mining Techniques to Improve Breast Cancer Diagnosis*," *J. Med. Syst.*, vol. 40, Sep. 2016, doi: 10.1007/s10916-016-0561-y.
[23] S. Sumarlin, "*Implementasi Algoritma K-Nearest Neighbor Sebagai Pendukung Keputusan Klasifikasi Penerima Beasiswa PPA dan BBM*," *J. Sist. Inf. Bisnis*, vol. 5, no. 1, pp. 52–62, 2015, doi: 10.21456/vol5iss1pp52-62.
[24] T. Sinha, A. Ghosh, and G. Manju, "*Attendance System with Face Recognition*," no. 4, pp. 799–804, 2020, doi: 10.35940/ijeat.D7710.049420.
[25] Adam Geitgey, "*Calculating Accuracy as a Percentage*," 2019. https://github.com/ageitgey/face_recognition/wiki/Calculating-Accuracy-as-a-Percentage (accessed Mar. 10, 2020).
[26] K. J. Bhanushree and M. B. Meenavathi, "*Feature Based Face Recognition using Machine Learning Techniques*," *Int. J. Recent Technol. Eng.*, vol. 8, no. 6, pp. 1313–1317, 2020, doi: 10.35940/ijrte.f7497.038620.
[27] P. Sukhija, S. Behal, and P. Singh, "*Face Recognition System Using Genetic Algorithm*," *Procedia Comput. Sci.*, vol. 85, no. Cms, pp. 410–417, 2016, doi: 10.1016/j.procs.2016.05.183.