



Face Identification Based on K-Nearest Neighbor

Ni Kadek Ayu Wirdiani, Praba Hridayami, Ni Putu Ayu Widiari, Komang Diva Rismawan, Putu Bagus Candradinatha, I Putu Deva Jayantha

¹Information Technology Department, Faculty of Engineering
Universitas Udayana, Indonesia

Email: ¹ayu_wirdi@yahoo.com, ²prabahridayami@student.unud.ac.id, ³widiariayyu@gmail.com,
⁴divarismawan.20@gmail.com, ⁵putubagus32@gmail.com, ⁶devajayantha870@gmail.com

Abstract

At this moment, face identification has been widely applied for security on gadgets, smart home security, and others. Face dominates as a biometric which is most increase in the next few years. Face is used for biometric identification which is considered successful among several other types of biometrics and accurate results. Face recognition utilizes facial features for security purposes. The classification method in this paper is K-nearest Neighbor (KNN). The K-Nearest Neighbor algorithm uses neighborhood classification as the predictive value of a good instance value. K-NN includes an instance-based learning group. This paper developed face identification using Principal Component Analysis (PCA). The preprocessing method that used in this research is contrast stretching, grayscale and haar cascade segmentation. This research is registered 30 people, each person had 3 images used for training and 2 images used for testing. The result obtained from several test of k value gives the best accuracy is 81% with k=1.

Keywords: Convolutional Neural Network, K-Nearest Neighbor, Principal Component Analysis, Haar Cascade

1. INTRODUCTION

Along with the rapid development of technology, it needs to be balanced with appropriate security enhancements so that users are comfortable with the personal information on the technology. Face is one of the most popular types of biometrics at the moment. The face dominates as biometrics which is most increase in the next few years, which is 38%, followed by multimodal (22%), iris (11%), and fingerprint 9% [1].

Facial biometrics is a personal identification system that is very different from one person to another. This difference makes face as a biometric that is widely applied besides is easily acquired. Research related to face identification had previously been carried out by Changxing Ding, Chang Xu, and Dacheng Tao with his research entitled Multi-task Pose-Invariant Face Recognition. This paper of shooting faces in unrestricted environments usually contains significant variations in poses, which dramatically reduce the performance of algorithms designed to recognize the front face. This study developed a face verification algorithm for the variations in significant facial poses. The best results were

obtained using the High-dim LBP and Joint Bayesian methods with an accuracy of 93.18% [2].

The study entitled Real-Time Face Detection and Recognition in Complex Back-ground has researched facial biometrics. This study developed an algorithm for real-time face detection and recognition with a complex background that is efficient and resilient. Ada Boost, cascade classifier, Local Binary Patent (LBP), Haar-like features, face image pre-processing and Principal Component Analysis (PCA) are a series of signal processing methods. The PCA algorithm is used to recognize faces efficiently. This rhythm reaches 99.2% for correct facial recognition and a true positive level of 98.8% for face detection [3].

There are three stages carried out in this study to identify face, that is face detection, feature extraction and classification. Face detection is a face recognition to find the position of the face from an image that will be extracted later. Face detection displays the location of all faces in the input image given, usually in the form of a box divider. Feature extraction is a step to determine the natural characteristics of a face which will then be classified or recognized. Whereas, classification is the process of matching input with data in a database [4].

The face is one of the biometrics that is very easily acquired, namely using a camera. This paper develops face identification using the K-Nearest Neighbor classification method and using feature extraction Principal Component Analysis (PCA). The classification method used is K-nearest Neighbor (KNN). This paper will produce a program using the python programming language, which is used for the purpose of identifying faces.

2. RELATED WORK

The application of face identification using the PCA (Principal Component Analysis) method as feature extraction and KNN as a classification has been done before, so that it becomes a reference for making a system that can do more activities than the previous system.

One of the studies that has been conducted is stated in the journal entitled "Multi-Faces Recognition Process Using Haar Cascades and Eigenfaces Methods". The proposed facial recognition process is carried out using the Haar Cascades and Eigenface method hybrid process, which can detect many faces (55 faces) in one detection process. The image pre-processing process consists of several stages, namely training data, grayscale conversion, and preprocessing with Haar Cascade. The feature extraction process consists of two stages, namely the stages of training and testing, where the data being tested will be divided into two namely training data and testing data which will be processed using the PCA (Principal Component Analysis) method, better known as eigenfaces method. The face identification process uses the Euclidean Distance similarity method. This enhanced face recognition approach is able to recognize many faces with an accuracy rate of 91.67 [5].

Other research examples are listed in the journal entitled "Handwriting Recognition using Eccentricity and Metric Feature Extraction based on K-Nearest Neighbors". This journal proposed a recognition process that consists of several

stages such as thresholding, noise removal, and cropping before feature extraction and classification. The dataset will be divided into training data and testing data. The feature extraction method used is eccentricity and metrics. Eccentricity is obtained between the value determining between the small elliptical focal distance and the main focus of the ellipse of an object. While the metric is the ratio between the area and circumference of the object. For the classification used the KNN method is used to classify objects based on training data with the distance nearest to the object, where the formula for calculating the distance used in this paper is the Euclidean distance formula. Based on the results of the testing obtained accuracy of 85.38% for the Handwriting Recognition using Eccentricity and Metric Feature Extraction based on K-Nearest Neighbors [6].

Examples of other studies are listed in the title "Voice Recognition using K Nearest Neighbor and Double Distance Method". This journal developed a new method to improve the accuracy of using data outliers, namely double distance method. This doubled distance method will be combined with the KNN method with $k=1$ as the center of the voice recognition. Frame work consist two stages are training and testing process. The training process is feature extraction using Mel Frequency Cepstrum Coefficients (MFCC). While the testing process through the introduction stage using the KNN method. Testing process is divided into two parts, the first part used the KNN method and the second used the doubled distance method. The similarity between testing data and training data is calculated by the Euclidean distance formula. Based on the results of testing, the method of KNN with one data center is 84.85% and the accuracy of the doubled distance method is 96.97%. From the result, we know double distance method improve the accuracy of voice recognition [7].

3. METHODS

Face identification Using the K-Nearest Neighbor Method consists of two phases namely the training phase and the testing phase. The dataset used in the training phase are 790 images consisting of 158 classes with each class consisting of three training images and two for testing images. Figure 1 shows the training phase in detail.

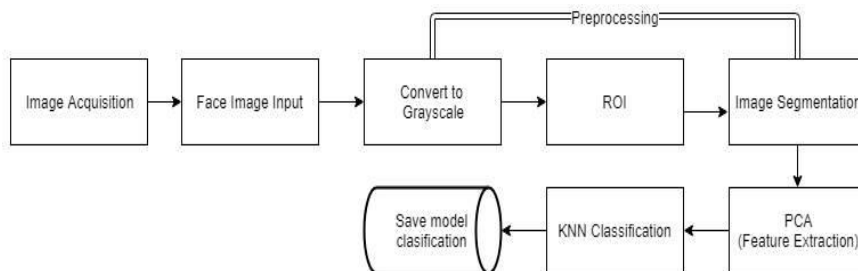


Figure 1. Training Phase of Face identification

While the testing or testing phase involves a database of features that have been obtained from the results of training. Figure 2 shows the testing phase in detail.

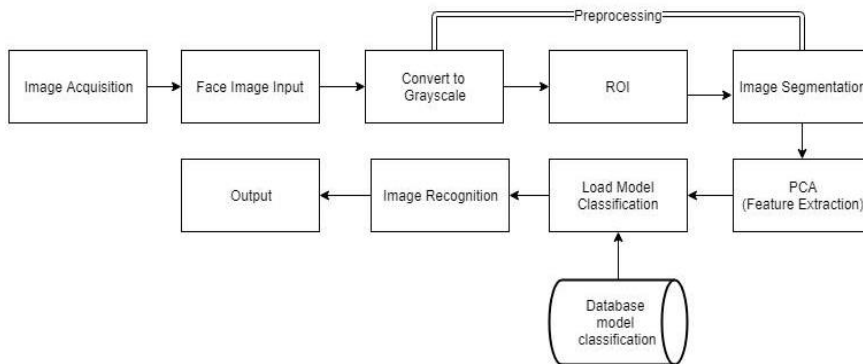


Figure 2. Testing Phase of Face Identification

3.1 Data Source

Data source used comes from manual shoots. This dataset research contains 790 faces from 158 people taken from several angles. The dataset has not gone through the segmentation process so that there is still a background that is quite significant outside the face object. The dataset has been separated into each folder containing the name of the individual owner of the face.



Figure 3. Example dataset model

3.2 Image Enhancement

Image Enhancement is an accentuation or sharpening of the elements of an image such as edge and boundaries or contrast levels that can make the graphic display of the image more useful for analysis and display [8]. Stages of image enhancement in the face identification system using the K-Nearest Neighbor (KNN) method consists of ROI (Region of Image), image conversion to grayscale and contrast stretching. An image enhancement technique increase contrast in the image by stretching called Contrast stretching. The concept of contrast stretching is to maintain the range of values of intensity it contains to reach the desired range of values. This is used to enhance the information in the image and maintain other details [9]. The following is the image enhancement result of face identification research using the KNN method.



Figure 4. Image Enhancement

3.3 Feature Extraction

Feature extraction is a stage to find the characteristic features of an inputted image. Feature extraction is a process which is extracted features to encourage the classifier to make decisions when classifying.

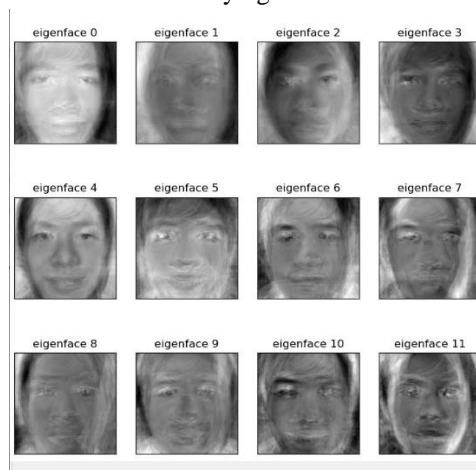


Figure 5. Feature Extraction using PCA

The Principal Component Analysis feature extraction method used in this paper is one of the popular extraction methods [10]. The PCA method reduces the dimensions of data with the least amount of information loss [11]. This method is used in many fields, such as biometrics, feature extraction, image processing, data compression, etc. In the PCA method, faces are described as linear combinations of eigenvector weights called Eigenfaces. This eigenvector is a covariance matrix from the image database. The number of images in the database will be the same as the number of Eigenfaces received [12]. Furthermore, is an example of feature extraction using the face identification system using the Principal Component Analysis (PCA) method.

3.4 KNN Classification

Face classification is a stage for the process of matching testing data and training data from face datasets. KNN is one of the simple algorithms that can be

used for classification. Regardless of its simplicity, this method is quite effective as a classification. This method was first proposed by T. M. Cover and P. E. Hart in 1967 [13] but then modified to improve the performance of the KNN. The basic concept of KNN is to have several training samples and testing samples determined by members. If $k=1$, the testing sample is assigned to the nearest single neighbor class. However, finding the right k value for a particular problem is a problem that affects the performance of the KNN [14]. The classification stages in face identification systems use the K-Nearest Neighbor (KNN) method where the eigen image of the feature extraction process used as input is as follows.



Figure 6. The results of identification using KNN

4. RESULT AND DISCUSSION

Tests were carried out using face datasets manually for research material. The following is a face identification system test using the K-Nearest Neighbor method with a change in parameter k value which is neighbor points of each class.

Table 1. Testing with $k = 1$

Name	Precision	Recall	F1-score	Support
Adhi Pramudya	1.00	0.50	0.67	2
Agung Meidi	0.67	1.00	0.80	2
Agus Hendra	1.00	1.00	1.00	2
Agus Wiratha	0.67	1.00	0.80	2
Alit D	1.00	0.50	0.67	2
Alit Sudawiaya	0.67	1.00	0.80	2
Aniek Laksmidewi	1.00	1.00	1.00	2
Ariana	1.00	0.50	0.67	2
Arik	1.00	1.00	1.00	2
Aris Eko P	0.50	1.00	0.67	2

Arsa Suteja	1.00	1.00	1.00	2
Arsika	1.00	1.00	1.00	2
Artajaya	1.00	0.50	0.67	2
Arya D	1.00	1.00	1.00	2
Aryana	1.00	0.50	0.67	2
Aryawan	1.00	1.00	1.00	2
Ayu Puspa Dewi	1.00	0.50	0.67	2
Bawa Arnawa	1.00	0.50	0.67	2
Bawa P	1.00	0.50	0.67	2
Benediltus	1.00	1.00	1.00	2
Budiasa	0.50	0.50	0.50	2
Citra Aprilia	1.00	1.00	1.00	2
Deazy	1.00	1.00	1.00	2
Deny Satriawan	0.33	0.50	0.40	2
Eka Cahyadi	0.67	1.00	0.80	2
Gita Indrawan	1.00	1.00	1.00	2
Gusur	1.00	1.00	1.00	2
Ibu Budi	0.50	1.00	0.67	2
Iwan Tofani	1.00	1.00	1.00	2
Maya Ritasari	1.00	0.50	0.67	2
avg / total	0.88	0.82	0.81	2

Table 1 is a testing process used the K-Nearest Neighbor (KNN) method with parameters $k=1$ using face training data with 30 classes consist of 2 testing data and 3 training data, which obtained an accuracy or F1-score of 81% . All persons can be identified but there are a few testing data cannot be identified. We suspect that the testing data face position is not so clear therefore it is not identified.

Table 2. Testing with $k = 2$

Name	Precision	Recall	F1-score	support
Adhi Pramudya	0.33	0.50	0.40	2
Agung Meidi	0.50	0.50	0.50	2
Agus Hendra	0.50	1.00	0.67	2
Agus Wiratha	0.33	0.50	0.40	2
Alit D	0.33	0.50	0.40	2
Alit Sudawiaya	0.50	1.00	0.67	2
Aniek Laksmidewi	1.00	1.00	1.00	2
Ariana	1.00	1.00	1.00	2
Arik	0.67	1.00	0.80	2
Aris Eko P	0.50	1.00	0.67	2
Arsa Suteja	0.67	1.00	0.80	2
Arsika	1.00	1.00	1.00	2
Artajaya	1.00	1.00	1.00	2
Arya D	1.00	1.00	1.00	2
Aryana	0.50	0.50	0.50	2
Aryawan	0.50	1.00	0.67	2
Ayu Puspa Dewi	0.00	0.00	0.00	2

Bawa Arnawa	0.00	0.00	0.00	2
Bawa P	0.00	0.00	0.00	2
Benediltus	0.25	0.50	0.33	2
Budiasa	1.00	0.50	0.67	2
Citra Aprilia	1.00	0.50	0.67	2
Deazy	1.00	0.50	0.67	2
Deny Satriawan	0.50	1.00	0.67	2
Eka Cahyadi	0.67	1.00	0.80	2
Gita Indrawan	0.00	0.00	0.00	2
Gusur	0.00	0.00	0.00	2
Ibu Budi	0.00	0.00	0.00	2
Iwan Tofani	1.00	0.50	0.67	2
Maya Ritasari	0.00	0.00	0.00	2
avg / total	0.52	0.60	0.53	60

Table 2 a testing process used the K-Nearest Neighbor (KNN) method with parameters $k=2$ using face training data with 30 classes, which obtained accuracy or F1-score of 53%. There are seven people who cannot be identified at all with a zero F1-score. The zero F1-score obtained because of the precision and recall is also scored zero, that means there is no neighbor who close by can be a recommendation for the identification.

Table 3. Testing with $k = 3$

Name	Precision	Recall	F1-score	Support
Adhi Pramudya	0.67	1.00	0.80	2
Agung Meidi	1.00	1.00	1.00	2
Agus Hendra	0.50	1.00	0.67	2
Agus Wiratha	0.00	0.00	0.00	2
Alit D	0.33	0.50	0.40	2
Alit Sudawiaya	0.33	1.00	0.50	2
Aniek Laksmidewi	1.00	0.50	0.67	2
Ariana	1.00	0.50	0.67	2
Arik	0.50	1.00	0.67	2
Aris Eko P	0.25	0.50	0.33	2
Arsa Suteja	1.00	0.50	0.67	2
Arsika	0.33	0.50	0.40	2
Artajaya	0.50	0.50	0.50	2
Arya D	1.00	1.00	1.00	2
Aryana	1.00	0.50	0.67	2
Aryawan	1.00	0.50	0.67	2
Ayu Puspa Dewi	0.00	0.00	0.00	2
Bawa Arnawa	0.00	0.00	0.00	2
Bawa P	0.00	0.00	0.00	2
Benediltus	0.33	0.50	0.40	2
Budiasa	1.00	0.50	0.67	2
Citra Aprilia	1.00	0.50	0.67	2
Deazy	1.00	1.00	1.00	2

Deny Satriawan	0.33	0.50	0.40	2
Eka Cahyadi	0.33	0.50	0.40	2
Gita Indrawan	0.00	0.00	0.00	2
Gusur	0.00	0.00	0.00	2
Ibu Budi	0.50	0.50	0.50	2
Iwan Tofani	0.50	0.50	0.50	2
Maya Ritasari	0.00	0.00	0.00	2
avg / total	0.51	0.50	0.47	60

Table 3 a testing process used the K-Nearest Neighbor (KNN) method with parameters $k=3$, using face training data with 30 classes where accuracy or F1-score is 47%.

From the results of research by applying the facial identification system using K-Nearest Neighbor (KNN) or eigenface divided into two parts, namely the training process and testing process. Testing scenario in this research by using parameter $k=1$, $k=2$, and $k=3$, give the best accuracy is on $k=1$ with 81%, while $k=2$ give 53%, and $k=3$ is 47%. It shows the value of k affected to the accuracy. From this research we can tell that the higher k value is the smaller accuracy that we get for face identification using KNN.

5. CONCLUSION

In this paper we have presented an experiment for face identification using the KNN method. KNN is one of the simplest algorithms that can be used for classification. The sources of data come from manual shooting which divided into 30 classes. The face identification using the KNN method consists of two stages, such as the training phase and the testing phase. Based on the results by changing the parameter k value obtained results are different for each parameter. The results give accuracy 81% for $k=1$, give accuracy 53% for $k=2$, give accuracy 47% for $k=3$. From the results, it shows that the value of k greatly affects the level of accuracy of the system. The parameter k value and accuracy are inversely proportional, the greater k value gives the smaller accuracy of the identification system. From this research can be concluded that the higher k value is the smaller accuracy that we get for face identification using KNN.

6. REFERENCES

- [1] Biometrics Institute. (2017). *Kairos- Protect your business and create frictionless customer experiences with face recognition*. Retrieved from <https://www.kairos.com/face-recognition-for-security>
- [2] Ding, C., Xu, C., & Tao, D. (2015). Multi-Task Pose-Invariant Face Recognition. *IEEE Transactions on Image Processing*, 24(3), 980 - 993.
- [3] Zhang, X., Gonnot, T., & Sanjie, J. (2017). Real-Time Face Detection and Recognition in Complex Background. *Journal of Signal and Information Processing*, 8(2), 99-112.

- [4] Ranjan , R., Bansal, A., Zheng, J., Xu, H., Gleason, J., Lu, B., et al. (2015). A Fast and Accurate System for Face Detection., *JOURNAL OF LATEX CLASS*, 14(8).
- [5] Mantoro, T., Ayu, M. A., & Suhendi. (2018). Multi-Faces Recognition Process Using Haar Cascades and Eigenface Methods. *2018 6th International Conference on Multimedia Computing and Systems (ICMCS)*.
- [6] Rachmawanto, E. H., Anarqi, R. G., Setiadi, D. I., & Sari, A. (2018). Handwriting Recognition Using Eccentricity and Metric Feature Extraction Based on K-Nearest Neighbors. *2018 International Seminar on Application for Technology of Information and Communication (iSemantic)*.
- [7] Fransiska, R. (2016). Voice Recognition Using k Nearest Neighbor and Double Distance Method. *2016 International Conference on Industrial Engineering, Management Science and Application (ICIMSA)*.
- [8] Asana, I., Widyantara, I., Wirastuti, N., & Adnyana, I. (2017). Metode Contrast Stretching untuk Perbaikan Kualitas Citra pada Proses Segmentasi Video. *OJS Unud*, 16(2).
- [9] Natesan, M., & Bhuvaneshwaran, K. (2014). Image Enhancement Using Multilevel Contrast Stretching and Noise Smoothing Technique for CT Images. *International Journal of Scientific and Engineering Research*, 5(5), 713-718.
- [10] Mahmud , F., Khatun, T., Zuhori, S., Afroge, S., Aktar, M., & Pal, B. (2015). Face recognition using Principle Component Analysis and Linear Discriminant Analysis. *International Conference on Electrical Engineering and Information Communication Technology (ICEEICT)*.
- [11] Paul , L., & Sumam, A. (2012). Face Recognition Using Principal Component Analysis Method. *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, 1(9).
- [12] Min, W., Romanova, E., Lisovec, Y., & San, A. (2019). Application of Statistical Data Processing for Solving the Problem of Face Recognition by Using Principal Components Analysis Method. *IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus)*. Saint Petersburg and Moscow, Russia, Russia.
- [13] Cover, T., & Hart, P. (1967). Nearest Neighbor Pattern Classification. *IEEE Transactions on Information Theory*, 13(1).
- [14] Mahmud, F., Islam, B., Hossain, A., & Goala, B. P. (2018). Facial Region Segmentation Based Emotion Recognition Using K-Nearest Neighbors. *International Conference on Innovation in Engineering and Technology (ICIET)*. Dhaka, Bangladesh, Bangladesh.