P-ISSN : 2715-2448 | E-ISSSN : 2715-7199
Vol.3 No.2 July 2022
Buana Information Tchnology and Computer Sciences (BIT and CS)

Signature Verification Using The K-Nearest Neighbor (KNN) Algorithm and Using the Harris Corner Detector Feature Extraction Method

Aang Alim Murtopo [1]
Study Program
Technical Information
STMIK YMI TEGAL
aang.alim@gmail.com

Bayu Priyatna [2]
Study Program
Information System
Universitas Buana Perjuangan Karawang
bayu.priyatna@ubpkarawang.ac.id

Rini Mayasari [3]
Study Program
Information System
Universitas Singaperbangsa Karawang
Rini.mayasari@staff.unsika.ac.id

Abstrak— Keamanan proses transaksi sangat penting di zaman sekarang ini. Tanda tangan dapat digunakan sebagai sarana penjamin keamanan suatu transaksi selain sidik jari. Namun, ancaman pemalsuan tanda tangan bagi mereka yang menggunakan tanda tangan sebagai keamanan masih sangat tinggi dan sering terjadi. Dalam penelitian ini, kami akan memverifikasi keaslian tanda tangan dan mengujinya menggunakan algoritma K-Nearest Neighbor (KNN) dan metode ekstraksi fitur Harris Corner. Ada dua macam perhitungan jarak yang akan digunakan pada algoritma K-NN yaitu dengan menghitung jarak dari Euclidean Distance dan Manhattan Distance.

Kata kunci— Tanda tangan, verifikasi, KNN (K-Nearest Neighbour), Harris Corner Detector, Euclidean Distance, Manhattan Distance.

Abstract — The security of the transaction process is very important in this day and age. Signatures can be used as a means of guaranteeing the security of a transaction other than fingerprints. However, the threat of signature forgery for those who use signatures as security is still very high and frequent. In this research, we will verify the authenticity of a signature and test it using the K-Nearest Neighbour (KNN) algorithm and the Harris Corner feature extraction method. There are two kinds of distance calculations that will be used in the K-NN algorithm, namely by calculating the distance from Euclidean Distance and Manhattan Distance. The k value at KNN taken is at k = 1, k = 3, and k = 5.

Keywords— signature, verification, KNN (K-Nearest Neighbour), Harris Corner Detector, Euclidean Distance, Manhattan Distance.

I. INTRODUCTION

Security in transactions needs to be given important attention today. There are several ways that can be done for transaction security measures such as fingerprints and PINs are examples of frequently used security. One of the well-known safeguards is to use a signature. Signatures are considered easier to use, cheaper, and quite effective. However, signature forgery is still common and poses a security threat to signature users. Signature verification is a way to find out if the signature is genuine or fake. Signature verification is divided into two forms, namely on-line and off-line signature verification. On-line means, when the signature is taken, recording of the time, pressure, and others. In off-line verification, the signature is only taken from a photo or image of that signature. The problem that often occurs in signature verification is the difference between the tools used to write signatures and the image retrieval process, image resolution, images that contain noise, and others.

This signature itself is a handwriting that has a unique character and each person must have a difference. We often encounter unreadable signatures, however, signatures can be read as images and recognized by computers [1]. Because of its uniqueness, the signature can be used as a security system and identifier of a person's identity.

Various studies that have been done before, have used various methods and feature extraction in this signature verification. Among them, research [2] was conducted to verify signatures online using the Feed Forward Back Propagation Error Neural Network and the Discrete Wavelet Transform feature extraction method. Research [3] used the Euclidean Distance Model and Geometric Centre method for feature extraction. Research [4] with K-Nearest Neighbour and Gabor Wavelet as characteristic extraction. Research [5] used PCA classification and Multilayer Feed Forward Artificial Neural Network and for its extraction using Fourier Descriptor and Chain Codes. Research [6] used the Euclidean Distance classification and a new extraction method that is dividing into several images.

From some of the studies above, these researchers have attempted to find a method of classification and feature extraction that can produce the highest level of accuracy. This level of accuracy is very important to avoid signature forgery so as to provide a sense of security in transactions and everything related to the economic sector.

This research will use the K-Nearest Neighbour (KNN) algorithm and the Harris Corner method for its feature extraction. From several previous studies, KNN is considered capable of classifying well for difficult images, such as in the study [7] "Automatic Medical Image Classification and Abnormality using K-Nearest Neighbour", KNN classifies medical images with an accuracy of 80% and greater when compared to SVM linear and RBF kernel. Harris Corner can be used for grayscale images and produces a more consistent extraction value from distorted images. as in research [8] "Harris Operator Corner Detection using Sliding Window Method", with the Harris Corner method palms can be detected with an accuracy of 97.5%.

II. METHOD

2.1 . Study of Literature

Research [2], which has been done, namely online signature verification, the classification algorithm used is Feed Forward Back Propagation Error Neural Network and feature extraction of Discrete Wavelet Transform produces an accuracy of 95%. Using a sample of 100 signatures consisting of 10 original signatures and 10 fake signatures for each person. This sample was taken from 5 people who gave signatures.

Research [3] with the Euclidean Distance Model classification algorithm and geometric center feature extraction on signature verification, resulted in a random FAR of 2.08%, simple 9.75% and 16.36% skilled forgeries. Meanwhile, the FRR was 14.58%. Signatures tested were as many as 21 original signatures and 30 fake signatures. From these signatures, 9 original signatures were found.

Research [4] carried out from signature verification, using the Nearest Neighbour classification algorithm and using Gabor Wavelet feature extraction. This study resulted in verification with accuracy ≥ human accuracy in carrying out signature verification with the smallest FAR and FRR values in this study were 22.5% and 15.5%.

Research [5] identified signatures based on Fourier Descriptor and Chain Codes and produced FAR = 2.6% and FRR = 1.6% in the verification process. This study uses the PCA algorithm and ANN Feed Forward for classification. Meanwhile, for feature extraction using Fourier Descriptor and Chain Codes.

Research [6] the approach used for feature extraction is to divide the image into rectangles based on the midpoint of gravity of the signature. The classification uses Euclidean Distance. The result is 0% random FAR, 0% simple and 1% skilled. Meanwhile, the FRR is 0.5%.

a) ***K-Nearest Neighbour* (KNN)**

K-Nearest Neighbour is a non-parametric classification even though it is simple but it is one of the popular and well-known and often used. The key to this method is the user-defined k parameter. When k is selected and given the x pattern, assign the pattern to the class that has the greatest number of K- Nearest Neighbour (a calculation to measure the distance to the nearest neighbour), because K-NN uses a calculation by determining the closest distance.

The calculation of the distance used in the K-NN [9]:

I.Euclidean Distance

$$j(v1,v2) = \sqrt{\sum_{k=1}^{N}(v1(k) - v2(k))^2}$$

II.Manhattan Distance

$$d(v1, v2) = \sum_{i=1}^{N}|v1(k) = v2(k)|$$

b) ***Harris Corner***

Harris corner detector (Harris angle detector) is a point (angle) detector that is often used because it is able to produce consistent values despite rotation, scale, lighting variations and noise. Harris angle detector based on the autocorrelation function of the local signal which calculates the local change of the signal. This detector also functions to detect local gradients in the horizontal and vertical directions at each surrounding point, the aim is to find the image value whose intensity varies from the two directions.

Harris Corner based on:

$$S_{ij} = \sum_{m=i-D}^{i+D} \sum_{m=j-d}^{j+D} W_{mn} \begin{bmatrix} h_{mn}^2 & h_{mn}v_{mn} \\ h_{mn}v_{mn} & v_{mn}^2 \end{bmatrix}$$

where is S_{ij} calculated in the area of measure (2D + 1) x (2D + 1) around position (i, j). h_{mn} represents the derived filter response horizontally, W_{mn} on vertical, and W_{mn} is the weight that reduces the impact of the position. The following is the Harris Corner detection algorithm [10]:

1. Calculate the x and y derivatives of the figure.

$$I_x = G_\sigma^x * I$$
$$I_y = G_\sigma^y * I$$

2. Calculate the derivative of each pixel

$$I_{x2} = I_x . I_x \quad I_{y2} = I_y . I_y \quad I_{xy} = I_x . I_y$$

3. Calculate the product of the derivative of each pixel.

$$S_{x2} = G_{\sigma'} * I_{x2} \quad S_{y2} = G_{\sigma'} * I_{y2} \quad S_{xy} = G_{\sigma'} * I_{xy}$$

4. Matrix form.

$$H(x,y) = \begin{bmatrix} S_{x2}(x,y) & S_{xy}(x,y) \\ S_{xy}(x,y) & S_{y2}(x,y) \end{bmatrix}$$

5. Calculate the detection response in each pixel.

$$R = Det(H) - k(Trace(H))^2$$

6. Threshold response value.

c) *Recall*, Precision, True Negative Rate and Accuracy

Recall to find out the answers to the system obtained from:

$$Recall = \frac{tp}{tp + fn}$$

Precision to determine the accuracy of the system to recognize the authenticity of signatures obtained from:

$$Precision = \frac{tp}{tp+fn}$$

Accuracy to determine the performance of the feature extraction and classification used.

$$Accuracy = \frac{tp+tn}{tp+fp+tn+fn}$$

Table 1. Confusion Matrix

	Total Population	True Condition	
		Positive Condition	Negative Condition
Prediction Condition	Predicted Condition Positive	True Positive	False Positive
	Predicted Condition Negative	False Negative	True Negative

True positive (tp) is when the program recognizes the authenticity of a signature image, it indicates that the signature image is genuine (true). False positives (fp) are when the program mistakenly recognizes the fake signature image as the original signature image. False negative (fn) is when the program mistakenly recognizes the original signature image as a fake signature image. True negative (tn) is when the program recognizes a fake signature image as a fake (true) signature image.

2.2 Data Collection

Signature image data, obtained from the scanning process and measuring 400x400 pixels totalling 300 images consisting of 150 original signatures and 150 fake signatures. These signatures were obtained from 10 different people.

2.3 Design

This research step is depicted in Figure 1. Verification of the signature starts from inputting the image of the signature to be tested, in the form of a grayscale image measuring 400x400 pixels, then extracting it to get the characteristics of the image. The results of this feature extraction in the form of coordinates that show the location of the angle are then stored with the KNN model of the grayscale image feature extraction to be an example. The extraction results are then calculated the distance with the extracted samples one by one.

After the distance calculation results are known, they are sorted from smallest to largest. The signature to be verified by category/label (fake or real), selected from a predetermined radius, will be tested.

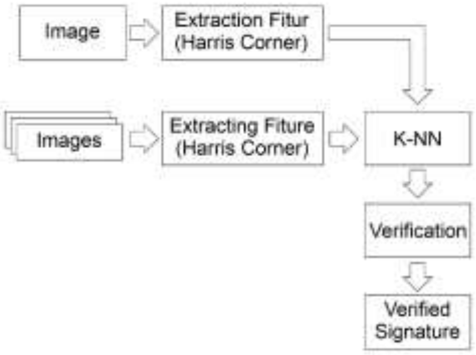
Image
Extraction Fitur (Harris Corner)
Images
Extracting Fiture (Harris Corner)
K-NN
Verification
Verified Signature

Figure 1. Design Diagram

III. RESULTS AND DISCUSSION

This signature verification system, its appearance can be seen in Figure 2 below. This signature verification system will be carried out in two ways for calculating the distance, namely Euclidean Distance and Manhattan Distance. In Figure 2, to verify the signature by pressing the test image button to input the image to be tested. Then press the original image button or the fake image button to input the image that will be used as training data. Original and fake training data signature image is required for verification. After that the training data image is entered, press the verification button, the results will appear in the massage box.

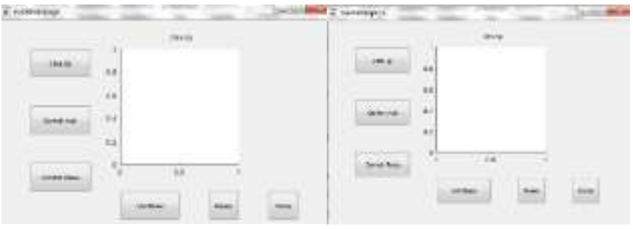

Figure 2. Display of the Verification System

Test Result

The test carried out aims to determine the results of the level of accuracy. In Figure 3 below, an example of a signature that will be used in research:

Figure 3. Sample Signature

In Figure 4 below, are the results of the signature test image. The signature image that will be used is obtained from the results of the scan process without resizing, totalling 300 image data. The data is divided into 200 images of training data and 100 images as test data. The signature image which is the training data will not be used in the test data and vice versa. When testing, training data and test data will go through a feature extraction process using the Harris Corner method. The results of the coordinates generated in the feature extraction process will be calculated the distance to K-Nearest Neighbour. Testing on the same signature image model using three kinds of variable k at K-Nearest Neighbour, namely k = 1, k = 3, k = 5 for the calculation of Euclidean Distance.

Figure 4. Image Testing Results

Table 2. Signature Testing Accuracy K = 1 Euclidean Distance

Model Type	Precision (%)	Recall (%)	True Negative Rate (%)	Accuracy (%)
0	0	0	100	50
1	100	80	100	90
2	50	100	0	50
3	50	100	0	50
4	50	100	0	50
5	50	100	0	50
6	50	100	0	50
7	0	0	100	50
8	0	0	100	50
9	50	100	0	50

Testing with k = 1 in Table 2 using Euclidean distance calculations produces an average precision of 40%, an average recall of 68%, an average true negative rate of 40%, and an average accuracy of 54%.

Table 3. Signature Testing Accuracy K = 3 Euclidean Distance

Model Type	Precision (%)	Recall (%)	True Negative Rate (%)	Accuracy (%)
0	0	0	100	50
1	50	40	60	50
2	100	20	100	60
3	50	100	0	50
4	50	100	0	50
5	50	100	0	50
6	50	100	0	50
7	0	0	100	50
8	0	0	100	50
9	50	100	0	50

Testing with k = 3 in Table 3 using Euclidean distance calculations produces an average precision of 40%, an average recall of 56%, an average true negative rate of 46%, and an average accuracy of 51%.

Table 4. Signature Testing Accuracy K = 5 Euclidean Distance

Model Type	Precision (%)	Recall (%)	True Negative Rate (%)	Accuracy (%)
0	0	0	100	50
1	66.66	40	80	60
2	0	0	100	50
3	50	100	0	50
4	50	100	0	50
5	50	100	0	50
6	50	100	0	50
7	0	0	100	50
8	0	0	100	50
9	50	100	0	50

Testing with k = 5 using the Euclidean distance calculation in Table 4 produces an average precision of 31.66%, an average recall of 54%, an average true negative rate of 48%, and an average accuracy of 51%.

Table 5. Signature Testing Accuracy K = 1 Manhattan Distance

Model Type	Precision (%)	Recall (%)	True Negative Rate (%)	Accuracy (%)
0	0	0	100	50
1	0	0	100	50
2	50	100	0	50
3	50	100	0	50
4	50	100	0	50
5	50	100	0	50
6	50	100	0	50
7	0	0	100	50
8	0	0	100	50
9	50	100	0	50

Testing with k = 1 using the Manhattan distance calculation in Table 5 produces an average precision of 30%, an average recall of 60%, an average true negative rate of 40%, and an average accuracy of 50%.

Table 6. Signature Testing

Method	Average precision (%)	Recall average (%)	Average True Negative (%)	Average accuracy (%)
Euclidean k = 1	40	68	40	54
Euclidean k = 3	40	56	46	51
Euclidean k = 5	31.66	54	48	51
Manhattan k = 1	30	60	40	50

From the results of tests carried out in Table 6, it is known that the k = 1 value of the Euclidean distance calculation is the best with an accuracy of 54%. However, the value of k = 1 is prone to noise. With a value of k = 1, if the distance to the noise image is the smallest, then the image that is verified will be considered the same type of image as the noise image. This is indicated by the smallest true negative rate with a value of 40% (equal to k = 1 Manhattan distance calculation). A small true negative rate indicates the system has probably made a mistake by accepting a fake signature as the original signature is large enough. The value of true negative rate depends on the similarity of the fake signature and image noise in the sample image. The greater the similarity of fake signature and image noise in the sample image,

From the above test, it can be seen that the Euclidean distance calculation is better than Manhattan distance. This can be seen from the greater precision, recall, and accuracy Euclidean distance values compared to Manhattan distance.

IV.CONCLUSION

Signature verification can be done to determine the authenticity of the signature. Signature verification is done based on the angle found and then applies the K-Nearest Neighbour algorithm. The small K value in the K-Nearest Neighbour algorithm has a tendency to accept the fake signature image as the original signature image is greater than the larger K value. Euclidean Distance calculation is better used for signature verification than Manhattan Distance calculation. At K = 1, the accuracy of the Euclidean Distance is 54% while the Manhattan Distance is 50%.

V.Suggestion

As for suggestions that are useful for further research, namely:

1. Further research can take a signature image using a pen tablet to reduce noise in the signature image.
2. Future research can reproduce the signature image that will be used as an example.
3. Signature verification can be done using other algorithms.

[a]REFERENCES

b.

[1] C. Oz, "Signature Recognition and Verification with Artificial Neural Network Using Moment Invariant Method," in *International Symposium on Neural Networks*, China, 2005.

[2] M. M. Fahmy, "Online handwritten signature verification system based on DWT features extraction and neural network classification," *Ain Shams Engineering Journal,* vol. 1, p. 59–70, 2010.

[3] Y. S. R. D. P. B. Banshider Majhi, "Novel Features for Off-line Signature Verification," *International Journal of Computers, Communications & Control (IJCCC),* vol. 1, pp. 17-24, 2006.

[4] M. R. P. H. R. P. Mohamad-Hoseyn Sigari, "Offline Handwritten Signature Identification and Verification Using Multi-Resolution Gabor Wavelet," *CiiT International Journal of Biometrics and Bioinformatics,* vol. 5, pp. 234-248, 2011.

[5] M. A. R. T. E. D. A. S. Ismail A. Ismail, "An efficient offline signature identification method based on Fourier Descriptor and chain codes," *International Journal of Biomedical Engineering and Technology,* vol. 5, pp. 1-10, 2011.

[6] P. I. S. S.A. Daramola, "Novel Feature Extraction Technique for Off-line Signature Verification System," *International Journal of Engineering Science and Technology*, vol. 2, pp. 3137-3143, 2010.

[7] M. K. Rakesh Ramteke, "Automatic Medical Image Classification and Abnormality Detection Using K-Nearest Neighbour," *International Journal of Advanced Computer Research*, vol. 2, pp. 190-196, 2012.

[8] R. D. G. S. Jyoti Malik, "Harris Operator Corner Detection using Sliding Windows Method," *International Journal of Computer Applications*, vol. 22, 2011.

[9] A. Kadir, Image Processing Theory and Application, Yogyakarta: Andi Offset, 2013.

[10] S. J. Prince, Computer Vision: Models, Learning and Inference, Cambridge: Cambridge University Press, 2012.