Faktor Exacta
Vol. 16, No. 3, Sept 2023, pp. 208~218
eISSN: 2502-339X, pISSN: 1979-276X, DOI: 10.30998/faktorexacta.v16i3.17504

Application of 2DPCA and SOM Algorithms to Identification of Digital Signature Ownership

Norhikmah[1], Rumini[2], Haris Angriawan[3], Acihmah Sidauruk[4]
[1,4.] Department of Information System, Universitas AMIKOM, Yogyakarta, Indonesia
[2, 3] Department of Informatic, Universitas AMIKOM, Yogyakarta, Indonesia

Article Info

Article history:

Received May 20, 2023
Revised Sep 26, 2023
Accepted Oct 4, 2023

Keywords:

Identification
Digital Image Signature
2DPCA
SOM

ABSTRACT

A signature is a proof of the validation of the thesis document. When conventional signatures begin to switch to digital signatures, this can provide a gap in signature forgery. Therefore, a system is needed to identify the ownership of a digital signature image with the first research methodology. It is to collect a digital signature dataset as a thesis document signature image. The second stage is processing the image by changing the color of the image to gray first to get additional features. The third stage analyzes other image features using 2DPCA, and the fourth stage identifies the best matching image units using the Single Organizing Maps (SOM) method and ends with the accuracy level. The results of this research use the 2DPCA and SOM algorithms to identify ownership of digital signatures, with accurate and false test results of 84 patterns from a total dataset of 91 patterns. Resulted the highest accuracy value of 92.3% at a rate of 0.9.

Corresponding Author:

Norhikmah,
Department of Information System,
Universitas AMIKOM Yogyakarta,
Jl.Ring Road Utara, Depok Sleman Yogyakarta.
Email: hikmah@amikom.ac.id

1. INTRODUCTION

Conventional signatures with digital signatures can be used to prove the authenticity of a document in digital form [1]. Legally, the laws that protect government digital signatures are regulated in government regulation no.82 of 2012. Concerning the Operation of Electronic Systems and Transactions, and applies to all electronic system operators. An administrator is any person, state administrator, business entity, or society who provides, manages, and operates electronic systems individually or jointly to electronic system users for their own needs and the needs of other parties [2]. In this Government Regulation, every public service must use an electronic certificate. This Government Regulation is strengthened by the ITE (Information and Electronic transactions) Law, which protects digital signatures.

Currently, COVID is still hitting Indonesia, forcing several agencies to change the ratification of a document that was initially affixed with a wet signature to be changed using a digital signature. The digital signature system is implemented to improve the quality of publishing services documents [3] to provide an opening for counterfeiting digital signature validation. The method used for the extraction of digital image signatures is 2DPCA (Two-dimensional Principal Component Analysis) and SOM (Self Organizing Map) as an algorithm that processes the identification of the original signature of the owner, with an accuracy rate of 76.6% in the application of gender identification cases using lip images [4], 2DPCA method uses image projection techniques live. The 2D (Plan, two-dimensional space) signature image matrix does not need to be transformed into a vector image as in PCA. The image covariance matrix can be directly formed using the original image matrix [5]. Method 2DPCA for signature image extraction and using the Euclidean Distance method to search for similarity of signature data [6]. Identification of signature patterns is one technique used to prove valid authentication to ownership based on the signature. For pattern identification, the Principal Component Analysis (PCA) is a classification method proven to be accurate. Still, it needs optimal weaknesses to differentiate one class from another [7]. SOM differs from neural networks; some have two modes: training and mapping to classify new input vectors. SOM is also helpful in grouping data without knowing class membership and input data [8]. Therefore, this study aims to identify the ownership of image signatures. Digitally and to find out the best rates and iterations to get the highest accuracy results

2. METHOD

The research stages used in identifying digital signature ownership are:

2.1 Data collection: It used the documentation method by taking a picture of the lecturer's signature on the approval page for the thesis or final assignment approved and published in the Amikom repository, with a total dataset of 91 patterns from 35 signature owners.

Figure 1. Example of a digital signature

2.2 Signature Image Processing: The image preprocessing process converts the intensity value of RGB (Red, Green, and Blue) images into a gray image with an intensity value of 8 bits (0-255). The conversion process of an RGB color image with a size of 24bit (8bit red, 8bit green, 8bit blue) to a gray image can be done by summing the RGB image and dividing it by a value of 3. The final result of the process is a gray image with a size of 8 bits (0- 255).

Figure 2. Example of digital signatures converting to greyscale

Convert a color image to a gray image using the equation formula:

$$s = (r + g + b)/3 \tag{1}$$

2.3 2DPCA analysis: The feature extraction method used in this study is Two-Dimensional Principal Component Analysis (2DPCA). The 2DPCA feature extraction method is a development of the classical PCA method. The 2DPCA method uses direct image projection techniques. 2D image matrices do not need to be transformed into vector images as in PCA. The image covariance matrix can be directly formed using the original image matrix. The final result of the 2DPCA feature extraction process is a feature vector. The steps taken to produce feature vectors through the 2DPCA method are as follows [1]:

2.3.1 Matrix Analysis of Signature Image

The image has been converted to a gray appearance and Two-Dimensional Principal Component Analysis (2DPCA). 2D image matrices do not need to be transformed into vector images as in PCA. The image covariance matrix can be directly formed using the original image matrix. Where the idea Aj is the original image matrix mxn, Aj = [A1, A2, .., Am], (j = 1, 2,…, M) with the image dimensions (210 x 280) projected onto in a 2-dimensional matrix with equation (2):

$$A_j = \begin{bmatrix} X_{11} & X_{21} & \dots & \dots & X_{M1} \\ X_{12} & X_{22} & \dots & \dots & X_{M2} \\ \dots & \dots & \dots & \dots & \dots \\ X_{1N} & X_{2N} & \dots & \dots & X_{MN} \end{bmatrix} \tag{2}$$

2.3.2 Analysis of the Mean Value of Signature Images (Mean)

The following 2DPCA feature extraction process calculates the average value of the signature image matrix. The average value is obtained by adding up the pixel values of each signature image and then dividing equally by the total of all data. Calculate the average of the total set matrix with equation (3):

$$\bar{A} = \frac{(x_{1,j} + x_{2,j} + x_{3,j} + \cdots + x_{m,j})}{M} \tag{3}$$

2.3.3 Analysis of Zero Mean Signature Images

Next, the calculation is carried out with zero mean, or the adjusted image is a reduction in the value of the t image matrix you hand with the average image value described in equation (4):

$$\theta = Aj - \bar{A} \tag{4}$$

2.3.4 Analysis of Covariance Matrix Signature Image

The calculation results are zero means used to obtain the value covariance matrix. The calculation process to get the value covariance matrix can be generated by all the multiplication between the zero mean (θ) and the value zero mean that has been transposed (θT). The result of the multiplication process is divided by the number of pixels in the image. The final result is a covariance matrix obtained using the formula (5):

$$G_t = \frac{1}{M}\sum_{j=1}^{M}\left(A_j - \bar{A}\right)^T \left(A_j - \bar{A}\right)$$

Or obtained using the formula:

$$G_t = \frac{\text{Image Adjusted X Adjusted Image Transpose}}{\text{image pixel count}} \tag{5}$$

Image adjusted x Adjusted image transpose) / (Image pixel count)

2.3.5 Analysis of EigenValue and EigenVector Image Signature

The 2DPCA feature extraction stage is obtaining eigenvalues from the values in the covariance matrix. Here are the steps: The eigenvalues obtained are sorted descending from the most significant value to the smallest value in the form of a matrix. An eigenvector with a large eigenvalue is taken to get the optimal value of X. In this case, the X value is accepted in the first four columns because the 2DPCA method tends to place more optimal values to the left or in the initial columns.

2.3.6 Analysis of Value FeatureImage Image Signature

After obtaining the eigenvalues or vector projection value X from the image, the feature extraction from image A is in accordance with equation (6):

$$Yk = Axk \tag{6}$$

2.4 The final result of the 2DPCA process is called a Feature Image, where the value of the image will be trained with SOM.

2.5 Analysis Self Organizing Maps (SOM): The training process Self Organizing Maps (SOM) is used to classify image patterns from the feature extraction results using the 2DPCA method. According to Song dan Resman (1999) The following are the steps for the SOM algorithm [1]:

2.5.1 Initialize the input data from the feature extraction to determine the start learning rate value and the maximum iteration. The values are the start learning rate and maximum iteration obtained based on input from the user because there are no absolute rules in determining the start learning rate value and the maximum iteration of the SOM algorithm. A small start learning rate will result in insignificant changes to the weight when it is updated. Meanwhile, suppose the value of the start learning rate given is large enough but not followed by a sizeable maximum iteration value. In that case, this will not be effective so that the determination of the start learning rate and the maximum iteration must be adjusted accordingly.

2.5.2 Initializing the initial weight, determining the initial weight randomly with a value between 0 and 1. The initial weight is generated by as many as the exact dimensions from the feature extraction matrix.

2.5.3 Input data, namely the training data attribute that, affects the weight changes during the computation process training data.

2.5.4 The calculation of the closest distance uses the method Euclidean Distance, which is between the input data (vector) with weights, and the node with the minimum distance between the input data and the weight node is declared the winner. The search for the closest distance can be expressed mathematically in equation (7):

a.

$$d_j = \sum_{j=0}^{i=1} \left(X_i(t) - W_{ij}\right)^2 \quad (7)$$

b. Perform weights updates. The winning weight node will then be updated with equation (8):

$$w_{ij}(t+1) = w_{ij}(t) + d.\alpha(t).\left(x_i(t) - w_{ij}(t)\right), j \in N_e \quad (8)$$

c. Calculation of *update learning rate* with geometric subtraction (9):

$$\alpha(t+1) = \alpha(t) * -exp(i\,t + 1/\,i) \quad (9)$$

d. When the computation process is complete, save the computed weight as a reference used for classification.

2.6 Implementation of the 2DPCA and SOM algorithms in digital signature recognition applications using the desktop-based Java programming language.

2.7 Testing: a system test was conducted to test the accuracy of the rate values and iteration results to obtain the highest accuracy in identifying digital signature ownership.

3. RESULT AND DISCUSSION

3.1 Feature Extraction Using 2DPCA: Steps in 2DPCA Calculation for feature extraction of digital signature images as follows (Yang, 2014):

3.1.1 Reading image and converting it into a 4x4 matrix, Suppose an example of a digital signature image matrix.

Table 1. Reading and converting image

25	24	24	24
24	25	25	24
25	25	25	24
25	25	25	24

3.1.2 Finding the Average Value of the Image, according to equation (3)

$$\bar{A} = \frac{AM}{M}$$

$$\bar{A} = \frac{\begin{array}{c}25 + 24 + 24 + 24 + 24 + 24 + 25 + 25 + 24 + 25 \\ +25 + 25 + 25 + 24 + 25 + 25 + 25 + 24\end{array}}{16}$$

$$\bar{A} = 24{,}5625$$

3.1.3 Adjusting Data (Data Adjustment) The initial image matrix pixel value is reduced by the average value. Hereinafter referred to as Adjusted Image, according to equation (4).

Table 2. Adjusted Image

5.35	4.35	4.35	4.35
4.35	5.35	5.35	4.35
5.35	5.35	5.35	4.35
5.35	5.35	5.35	4.35

3.1.4 Calculating the Covariance Matrix, according to equation (5)

a The adjusted image that was exposed to the image

Table 3. Adjusted Image

5.35	4.35	5.35	5.35
4.35	5.35	5.35	5.35
4.35	5.35	5.35	5.35
4.35	4.35	4.35	4.35

b That was adjusted was multiplied by the image that was transposed

Table 4. Transposed adjusted Image

85.39	88.74	94.09	94.09
88.74	95.09	99.44	99.44
94.09	99.44	104.79	104.79
94.09	99.44	104.79	104.79

c The above results should be divided by 16 (Is the number of pixels in the image)

Table 5. Result Image

4.2695	4.437	4.7045	4.7045
4.437	4.7545	4.972	4.972
4.7045	4.972	5.2395	5.2395
4.7045	4.972	5.2395	5.2395

3.1.5 Counting Eigen Vectors and Eigen Value

Using Library

3.1.6 Sorting eigenvector in ascending and grab some parts of the course which is hereinafter called Feature Vector

a Eigenvector and Eigenvalue features have been obtained and taken 2 feature columns.

Table 6. 2 feature columns

15	17
-3	24
7	20
6	-2
-4	12

b Multiplying the initial image times the feature

vector Feature Vector x, Adjusted Matrix transposed

Table 7. Result Image transposed

15	25	24	76	76
25	87	98	128	23
16	54	56	123	23
43	76	89	3	66

Multiplication Result Matrix

Table 8. Multiplication Result Matrix

470	2095
1476	4493
1116	2718
794	5121

3.1.7 The final result of the 2DPCA process is a feature image matrix from the training image, according to equation (6).

If used as a final feature, it can be made into an array form 1D

Table 6. Result Image

470	1476	1116	794	2095	4493	2718	5121

3.2 Identify Digital Signature Image Using SOM

3.2.1 Input Vector PCA 4 Elements with 2 Identification Ownership of digital image signatures,

Table 7. Input Vector PCA

No	T1	T2	T3	T3	Name
1	12	80	10	31	Winda
2	3	4	2	2	Andi

3.2.2 Map SOM (SOM Map for Clustering) Measuring 2x2

a Weight Position in MAP

Table 8. Weight Position

N1	N2
N3	N4

b Cluster Label (Not yet known classification)

Table 9. Cluster Label

?	?
?	?

c The long weight is the length of the feature, in this example it is 4, Generated Randomly

Table 10. Weight Generated random

Bobot	T1	T2	T3	T3
N1	6	70	20	10
N2	3	3	4	4
N3	7	50	25	21
N4	6	4	9	2

3.2.3 Calculating the Euclidean Distance between Data and Weight, according to the equation (7)

a Data 1

Table 11. Data 1

D1_vs_N1	26.01922366	Best Matching Unit and D1 (winda)
D1_vs_N2	82.31038817	
D1_vs_N3	35.35533906	
D1_vs_N4	81.57205404	

b MAP Update

Table 12. MAP Update

Winda	?
?	?

c Update Weight N1 (because N1 is selected as the most suitable weight with feature D1), according to equations (8 and 9).

Table 13. Update Weight N1

Bobot	T1	T2	T3	T3
N1	0.6	61	29	-8.9
N2	2	2	3	3
N3	6	60	20	11
N4	5	3	8	1

d Data 2

Table 14. Data 2

D2_vs_N1	64.05130756	
D2_vs_N2	2.645751311	Best Matching Unit and D2 (Andi)
D2_vs_N3	59.58187644	
D2_vs_N4	6.480740698	

e Update Map

Table 15. Update Map

Winda	Andi
?	?

f Update Weight N2 (because N2 was chosen as the most suitable weight with feature D2), according to equations (8 and 9).

Table 16. Update Weight N2

Bobot	T1	T2	T3	T3
N1	5	75	25	8
N2	1.1	0.2	3.9	3.9
N3	6	60	20	11
N4	5	3	8	1

3.2.4 Identification Phase

Table 17. Identification Phase

Data Uji	3	60	24	5

a Testing SOM Weight

Table 18. Testing SOM Weight

Data Uji Vs N1	15.45962483	BMU pada N1
Data Uji Vs N2	61.09574453	
Data Uji Vs N3	7.810249676	
Data Uji Vs N4	59.37171044	

b Map Label

Table 18. Update Weight N1

Winda	Andi	N1	N2
?	?	N3	N4

Identification name resulting Nama Winda

3.3 Implementation of the 2DPCA AND SOM Algorithm.

Interface design is done at the 2DPCA and SOM Algorithm Implementation stage using a desktop-based Java programming language, as shown below 3. There are two features in this application, namely, training and testing features.

Figure 3. Main Menu

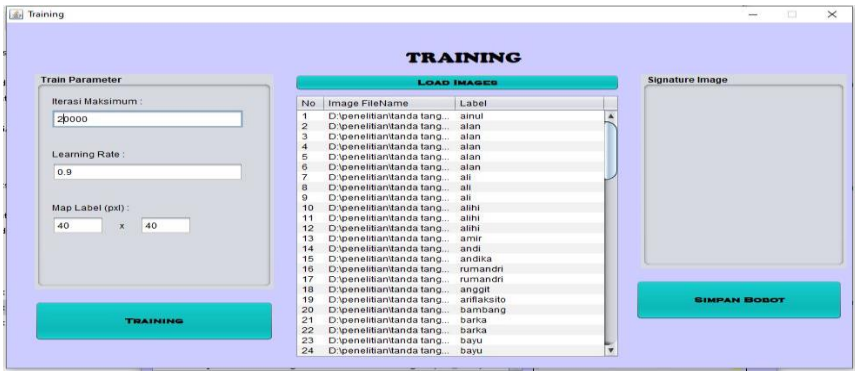

Figure 4. Data Training

In Figure 4, there is a menu for iteration and learning rate settings, as well as a menu for taking the signature image and reviewing the image that has been taken. If the signature image data is ready for training, click the training menu, then the menu for saving weights will be active. The processing time in training depends on the number of data sets rate settings, and iterations

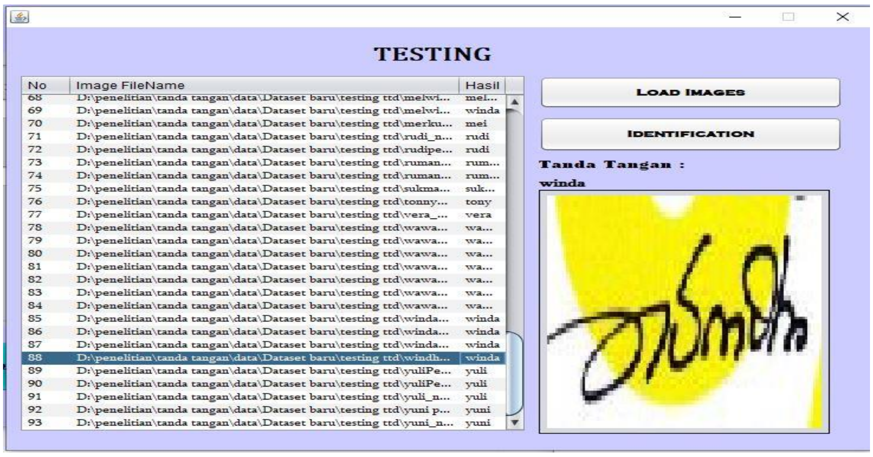

Figure 5. Data Testing

Figure 5 shows the form for testing a digital signature image. If it has finished loading the image to be tried, click the identification menu, and we will see whether the identification results are appropriate or not.

3.4 Application Testing

Results software testing results with a percentage of accuracy level identification of signature image are correctly classed. Testing done to 91 pattern data images of the 35's name signature owner can be seen on digital with the following Table 19.

Table 19. Result Testing

Testing	Iteration	Rate	FALSE	TRUE	Accuracy	Time (Minutes)
1	10000	0.9	23	68	74,4%	80
2	15000	0.9	15	76	83,5%	127
3	10000	0.6	35	56	61,5%	85
4	20000	0.9	7	84	92,3%	174
5	20000	0.6	17	74	81,3%	173
6	25000	0.9	16	75	82,4%	190
7	30000	0.9	8	83	91,2%	203

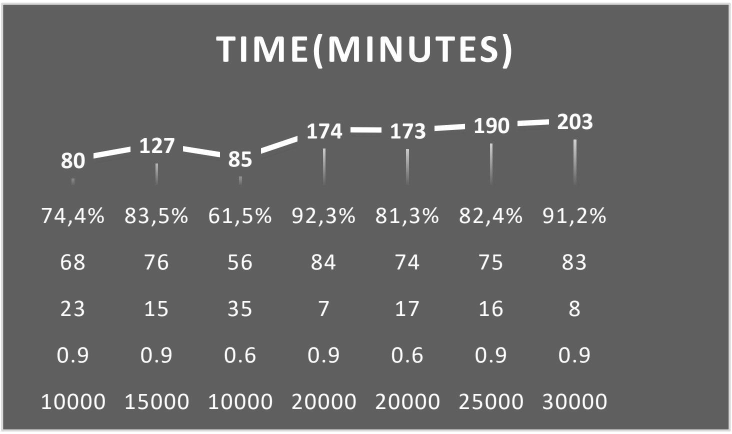

Figure 6. Result Testing

Table 19 and Figure 6 show the highest accuracy results at a rate of 0.9 and 2000 iterations with a duration of 174 minutes or 2 hours 54 minutes with an accuracy value of 92.3%. The data shows that the higher the rate, the higher the accuracy, and for the longer time in testing is 29 minutes, what affects the test besides the iteration and rate determination. One of the datasets used is the number of training and testing datasets.

4. CONCLUSION

The conclusion of this research is that it has succeeded in building a system for identifying the authenticity of digital signature ownership using the 2DPCA and SOM algorithms with an accuracy of 92.3% in 20000 iterations and a rate of 0.9. Suggestions for developing this research system are to add a feature that is connected directly to the camera with a digital image signature identification application and add a dataset of digital image signature patterns.

ACKNOWLEDGEMENT

Thank you to the research directorate of AMIKOM University Yogyakarta for research funding assistance

REFERENCES

[1] R. A. Azdy, "Tanda tangan Digital Menggunakan Algoritme Keccak dan RSA," *Jurnal Nasional Teknik Elektro dan Teknologi Informasi (JNTETI)*, vol. 5, no. 3, pp. 184–191, 2016, doi: 10.22146/jnteti.v5i3.255.

[2] "Aturan pemerintah no.82 tahun 2012. Tentang Penyelenggaraan Sistem dan Transksi Elektronik,"

[3] D. Rositawati and P. Romansyah, "Model Digital Signature Pada Dokumen Formal Akademik," *Cyberpreneurship Innovative and Creative Exact and Social Science*, vol. 6, no. 1, pp. 22–32, 2020.

[4] Norhikmah and H. Angriawan, "Implementation of 2DPCA and SOM Algorithms to Determine Sex According to Lip Shapes," in *ICITISE (IEEE)*, 2019, pp. 101–106. doi: DOI: 10.1109/ICITISEE48480.2019.

[5] C. Liu Y. C., Wu and M. Liu, "Research of fast SOM clustering for text information," *Expert Syst Appl*, vol. 38, no. 8, pp. 9325–9333..

[6] D. P. P. Didik Tri Setiawan, "Mengidentifikasi Citra Tanda Tangan," in *Seminar Nasional Inovasi Teknologi UN PGRI Kediri*, 2017, pp. 465–470.

[7] M. Harahap, A. M. Husein, and A. D. Program, “Signature Identification Based On Som Kohonen With Principal Component,” in *Seminar Nasional Aptikom (SEMNASTIKOM)*, 2017, pp. 0–5.

[8] S. Roy and S. K. Bandyopadhyay, “Gender recognition using Self Organizing Map (SOM) -an unsupervised ANN approach,” *International Journal of Emerging Research in Management &Technology*, vol. 9359, no. 38, pp. 2278–9359, 2014.

[9] Yang *et al.*, “Two- Dimensional PCA: A New Approach to Appearance-Based Face Representation Recognition,” in *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2004, pp. 26 (1), 131-137.

[10] Hidayat, Akik., and R. dan Shofa, N., “Self Organizing Maps (SOM) Suatu Metode Pengenalan Aksara Jawa,” *Jurnal Siliwangi*, vol. 2, no. 1, pp. 64–70, 2016.