

Writer Identification of Lampung Handwritten Documents Based on Selected Characters

Akmal Junaidi*, Syifa Trianingsih, Muhammad Iqbal

Computer Science Department, FMIPA

Universitas Lampung

Bandar Lampung

*akmal.junaidi@fmipa.unila.ac.id

Abstract-Writer identification is a sub-field in handwriting recognition which its objective is to determine the identity of the writer based on handwriting input. The goal is usually for forensic purposes such as finding the perpetrators of crimes that leave traces of evidence in the form of written messages. In addition, writer identification can also be used to determine the identity of a historical actor if he or she leaves a valuable written artefact. The object of this research is the traditional character of the Lampung region which is so-called Had Lampung by the local community. The traditional character of Lampung consists of 20 main characters and 12 diacritics. Based on selected characters, the writer will be recognized using the Principal Component Analysis (PCA) feature. PCA is one linear feature extraction method of an object in pattern recognition. The PCA algorithm consists of several stages, namely the calculation of the average dataset, the subtraction of the vector dataset with averages, the calculation of covariance, the calculation of eigenvectors and eigenvalues, eigenvector reduction, and the projection of the dataset against reduced eigenvector space. PCA in this paper is used as a feature in image recognition. The dataset utilized in this study is the Lampung Dataset which is a handwritten character recognition (HWCR) dataset. Lampung Dataset consists of 82 Lampung handwritten documents. All Lampung character images in the dataset were extracted from these documents using the connected component extraction algorithm and eventually generated 32,140 images. Furthermore, these images are converted into grayscale images. In this research, as many as 12,500 grayscale images of Lampung handwriting characters were chosen to represent 82 different writers. This data is employed as training and testing data on the proposed method. The highest accuracy of the identification of the writer using this PCA feature is 82.92%, while the lowest accuracy is 28.29%.

Keywords: Lampung Script; Writer Identification; Principal Component Analysis; Lampung Dataset

1. Introduction

Recently, writer identification has become a popular research topic in the area of pattern recognition. An interesting factor in the research topic of writer identification is the handwritten style of each individual who at a glance is similar, but it has its own uniqueness. Handwritten character patterns are an important element for forensic experts to identify the writers. In addition to forensic purposes, writer identification can also be used for the benefit of scientific development. The core process in identifying writers is the process of extracting features from handwritten character image that will be recognized. The object in writer identification can be in the form of a modern or contemporary script as well as a traditional script. This research uses a traditional script originating from the Lampung region, one of a few regions in Indonesia that has a traditional script. The Lampung script, or locally called Had Lampung, has 20 main characters and 12 diacritics. The Lampung handwritten

data compiled in Lampung Dataset consists of 82 raw images of Lampung handwritten documents, 82 text files containing annotations for each document and 32,140 grayscale images of single Lampung character [1]. The grayscale image in the dataset is the result of two stages of preprocessing of these documents. Those stages are connected component extraction followed by converting images into grayscale format. Some character samples of Lampung handwriting in Lampung Dataset can be seen in Figure 1.



Figure 1. Handwritten Character Samples of 4 Different Writers in the Dataset

Character recognition is a research topic that has been developing for more than two decades. In the upstream

side, many researchers provide a dataset to facilitate HWCR research. Some examples are the providing of the Lampung handwritten character dataset [1], historical handwritten digit documents of church records by priests in Sweden [2], and Arabic handwriting from historical manuscripts [3]. The methods and approaches in HWCR have also been applied for various scripts, for instance the Kurdish Text Classification of Sorani dialect [4], Slavic Historical Documents containing Glagolitic and Cyrillic character [5], printed Arabic [6], handwritten Bangla characters from India [7], handwritten Kanji characters [8], offline handwritten Chinese characters [9] and so on. The use of PCA specifically for handwritten character recognition is quite difficult to be found. So far, The research related to PCA has been used for recognizing Urdu characters [10]. The accuracy obtained in the study reached 96.2%. In other research, the use of PCA was not directly applied to handwritten character recognition but was employed as a method of feature space reduction before classification stage [11]. The study only addressed the recognition of digit from the MNIST dataset [12] and CVL Single Digit [13].

PCA analysis is one of the classic methods that has been widely applied to various researches. The purpose of using PCA analysis is to reduce information features that are redundant (and large) in order to obtain feature components with lower dimensions while still maintaining the values of discriminative features [14]. This analysis has been widely used to reduce feature dimensions in various pattern recognition tasks, especially for object recognition. In the field of pattern recognition for medical data, PCA is used to reduce the dimension of large size features in dynamic contrast enhanced MR imaging (DCE-MRI) data of hypoxia tumors [15]. In the context of the study, the use of PCA is intended to find the number of components that can distinguish the overall variability of data. The results of the study concluded that the 99% level of overall data variability can be described by only the first three principal components obtained by PCA. Another similar study in the medical field also uses PCA for selection of spectral entropy (SE) features from a 64-channel electroencephalogram (EEG) recording for the detection of alcoholics [16]. The role of the PCA in the study was to evaluate the size of the feature-set at the top-rank position before classification. Variations of the various sizes of these top-rank features indicate a ranking of those features after dimensional reduction. The effect of ranking and PCA during classification by k-NN has shown an improvement in accuracy compared to without ranking. Another unique PCA analysis has been applied for pattern recognition in the gym / fitness center [17]. In that study, PCA was used in gesture recognition (Action Recognition) for 770 fitness exercise movements. PCA analysis is used to reduce the feature dimension of the dataset into 3-7 features. Reduction does not decide a single feature because the number of features in that range is under scrutiny to determine the correlation between number of principal components (PCs) and features

which are the most relevant for correction recognition subsets. With the PCA analysis, the best accuracy achieved was $97 \pm 14\%$. The area of pattern recognition that also utilizes PCA a lot is in the face recognition [18], [19]. The study in [18] uses PCA to recognize faces even at different facial poses and orientations. Whereas research in [19] uses a comparison of PCA and Linear Discriminant Analysis (LDA) for each face recognition. The results concluded that the face recognition performance with PCA was superior compared to LDA.

Writer identification as well as pattern recognition, can use various features for the identification process. The use of global features and local features as one characteristic of handwriting is an appropriate feature combination for writers identification [20]. The dataset was distinguished of 650 and 225 writers, respectively. The identification performance with these features achieved 86% accuracy for the dataset of 650 writers and 79% for the dataset of 225 writers. The method proposed in the study is claimed to be applicable toward the writer identification of non-Latin handwriting such as Asian or Arabic. Other study regarding the writer identification was applied for handwritten Chinese character [21]. The research uses 16,000 Chinese words from 40 different writers. PCA in the study was applied to these words to find unique personal handwriting style characteristics and the best discriminative representation to other personal. The identification process did not use the entire writing texts but only a pieces of words. Beside achieving a high identification accuracy of 97.5%, the use of this approach has impact on time reduction during identification process.

Most of the previous scientific studies have been carried out using the entire document or complete words as identification input, while in this study the identification input used is piece of images of the single character image or double or five Lampung character images in grayscale format. The use of the grayscale format in this study is based on two reasons. First, the grayscale image is fairly well to store the characteristics of the Lampung characters. Grayscale image representation consists of intensity values with a range of 0-255 so it is quite significant to represent variations and details of the image. Second, the grayscale image consists of one channel only which will lighten the computational workload compared to the RGB image which consists of three channels. The use of features extracted from RGB images will increase computational workload by threefold.

One important step in pattern recognition is feature extraction which is defined as a process to extract the special value of the object so that the object can be recognized using this feature. This writer identification process also applies a feature extraction process to identify the writer of handwriting. A variety of feature extraction methods can be used to identify writers, such as the histogram method, line-based representation method, to linear transformation [22]. The feature extraction method implemented in this study uses the PCA approach, which is a linear transformation method that works by reducing

the feature dimensionality of the object as a feature extraction stage. PCA is generally applied to data that has very high dimensionality.

Based on the explained introduction, the purpose of this research is to implement PCA as a feature extraction method in identifying the writers of the Lampung handwritten documents. In addition, this study also aims to determine the accuracy of the PCA feature in identifying the writers of handwritten documents based on selected characters. This research is useful to understand the stages of PCA implementation in identifying writers of Lampung handwritten documents. So far, there have been no results of studies on writer identification of the Lampung handwritten character. The result reported in this paper is a novelty that has never been published formerly. Thus, this paper is expected to be useful for reference to other similar studies. In a wider scope, this research is expected to be able to contribute to the development of science in the pattern recognition domain especially in handwritten character objects.

2. METHOD

a. Writer identification

Writer identification is the recognition of the writer based on handwritten-character by matching of an unknown handwriting sample to the sample of data for which the writers have been known previously. The identification process is done by counting and comparing feature of handwritten samples with a feature database that has been stored. The results of identification of the most similar writers will have the highest level of similarity [23]. Two approaches for writer identification are to analyze based on characters and the approach by using textures from documents [24]. In this study, the proper approach is to use Principal Component Analysis (PCA) analysis to character images.

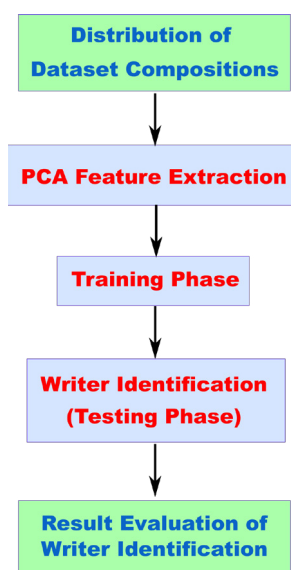


Figure 2. Research Stage of Writer Identification

PCA is a linear transformation method which is also known as the Karhunen-Loeve Transform (KLT) method. The feature of the PCA approach is the result of feature extraction which has been reduced in a simpler form. Dimensionality reduction is done by compressing the specific information that characterizes the object. This special feature set is represented as an eigenvector and the writer identification is evaluated from projected eigenvector of training and testing images.

The steps of the research on the writer identification on the Lampung handwritten documents is illustrated in Figure 2.

The procedure as shown in Figure 2 is in principle a pattern recognition framework, but it is adapted for the purpose of writer identification. Detailed descriptions of each stage are given in the following sub-sections.

b. Extracting PCA Features

The PCA feature extraction process is carried out in several steps. The calculation of average, subtraction, eigenvectors, eigenvalues, elimination of eigenvalues, and its projections in this study refers to the steps of the standard PCA algorithm [18]. The algorithm is explained briefly in the following.

- 1) Step 1: Prepare the image objects.
The object image is the image with representations as $I_1, I_2, I_3, I_4, \dots, I_M$. These image objects must have the same dimension.
- 2) Step 2: Prepare the dataset.
Each image object I_i is transformed into a vector and used as a training set S , where $S = \{\Gamma_1, \Gamma_2, \Gamma_3, \Gamma_4, \dots, \Gamma_M\}$.
- 3) Step 3: Calculate the average of dataset.
The average vector of dataset (Ψ) can be calculated using formulas:

$$\Psi = \frac{1}{M} \sum_{n=1}^M \Gamma_n \tag{1}$$

where:

Ψ : average vector of dataset

M : number of data

n : index of data, n lies from 1 to M

Γ_n : n^{th} training image vector

- 4) Step 4: Subtract of dataset vector and its average.
The dataset vector (Γ_i) is subtracted from average vector of dataset (Ψ) and stored in the Φ_i variable. The formula is given below:

$$\Phi_i = \Gamma_i - \Psi \tag{2}$$

where:

Φ_i : subtraction vector

Γ_i : i^{th} image vector

Ψ : average vector of dataset

- 5) Step 5: Calculate the covariance matrix.
The formula to compute the covariance matrix C is denoted in the following:

$$C = \frac{1}{M} \sum_{n=1}^M \Phi_n \Phi_n^T \quad (3)$$

where:

C : covariance matrix

M : number of data

Φ_n : n^{th} subtraction vector

Φ_n^T : transpose of n^{th} subtraction vector

Equation (3) can be simplified into $C = A \times A^T$ with $A = \{\Phi_1, \Phi_2, \Phi_3, \dots, \Phi_M\}$.

- 6) Step 6: Calculate eigenvectors and eigenvalues of covariance matrix.
Calculating the eigenvectors and eigenvalues can consider the formula $L = A^T \times A$ for reasons of efficiency and reducing the dimensions of the matrix during the computation process.
- 7) Step 7: Eliminate eigenvectors.
Eigenvalues obtained from previous step are then eliminated partially and hold the most relevant values that can significantly represent the objects.
- 8) Step 8: Calculate the dataset projection into the eigenvector space.
The next step is to calculate the dataset projections to eigenvector space using the formula below:

$$\omega_i = L^T \times (\Gamma_i - \Psi) \quad (4)$$

where:

ω_i : image projection

Γ_i : i^{th} training image vector

L^T : transposed eigenvector

Ψ : average vector of dataset

These steps have been implemented in a computer program to carry out the two processes in this study. The first process is training data as an effort to learn the characteristics of handwriting by the system. The next process is identification as a system decision in recognizing the writer of the document from the handwriting contained in the document.

c. Training phase

After all steps of feature extraction have been carried out, the next stage is the training stage. This stage is the training of available features obtained from former stage to be used at the writer identification stage, i.e. by projecting each eigenvector on the Lampung feature vector. The results of this projection will be used as a decision reference at the writer identification stage.

Data samples in the training process are arranged into 3 different configuration schemes. These configuration differences are arranged such that the use of sample data

represents units of data consisting of 1, 2 and 5 single character images. The selection of this configuration is decided based on trial and error. A detailed explanation of this configuration can be found in subsection III.

d. Writer Identification Matching Scheme (Phase Testing)

A number of images of which the writer(s) to be identified must go through a matching scheme as described in Figure 3 [25]. Stages of writer identification aim to predict writer from input sample of handwritten images. The decision of the identification result is determined by comparing the value of the projected training image to the input image which is a sample of the testing image based on a minimum value of Euclidean distance.

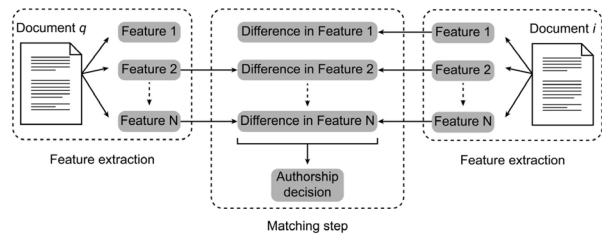


Figure 3. Matching Scheme on the Writer Identification System

If all decisions on the results of the matching steps have been obtained, the next step is to measure their accuracy using equation (5).

$$\text{Accuracy} = \left(\frac{\text{Amount of Correct Identification}}{\text{The amount of data}} \right) \times 100\% \quad (5)$$

The level of accuracy of identification is calculated after the overall results of the matching decisions are obtained completely. The counting of the matching is done with the aim to know the number of documents that are recognized correctly as an indicator of the accuracy on the writer identification process. The accuracy can be used as an evaluation whether the proposed method and the features indicate a good performance or not.

3. Results

Lampung dataset in this study is image collection of Lampung handwritten character in grayscale format with size 32x32 pixels. This dataset is distributed into two parts, one group as a training set for development of an identification model and another as a testing set for writer identification matching scheme. The character image sample is randomly selected as many as 12,424 images out of the total 32,140 character images in the dataset. The selected characters are further divided into two parts, 11,768 character images as the training set and the remaining 656 character images as the testing set. The characters from each part are then randomly selected to be divided into three sample groups for writer identification. Details of these sample distributions are listed in Table 1.

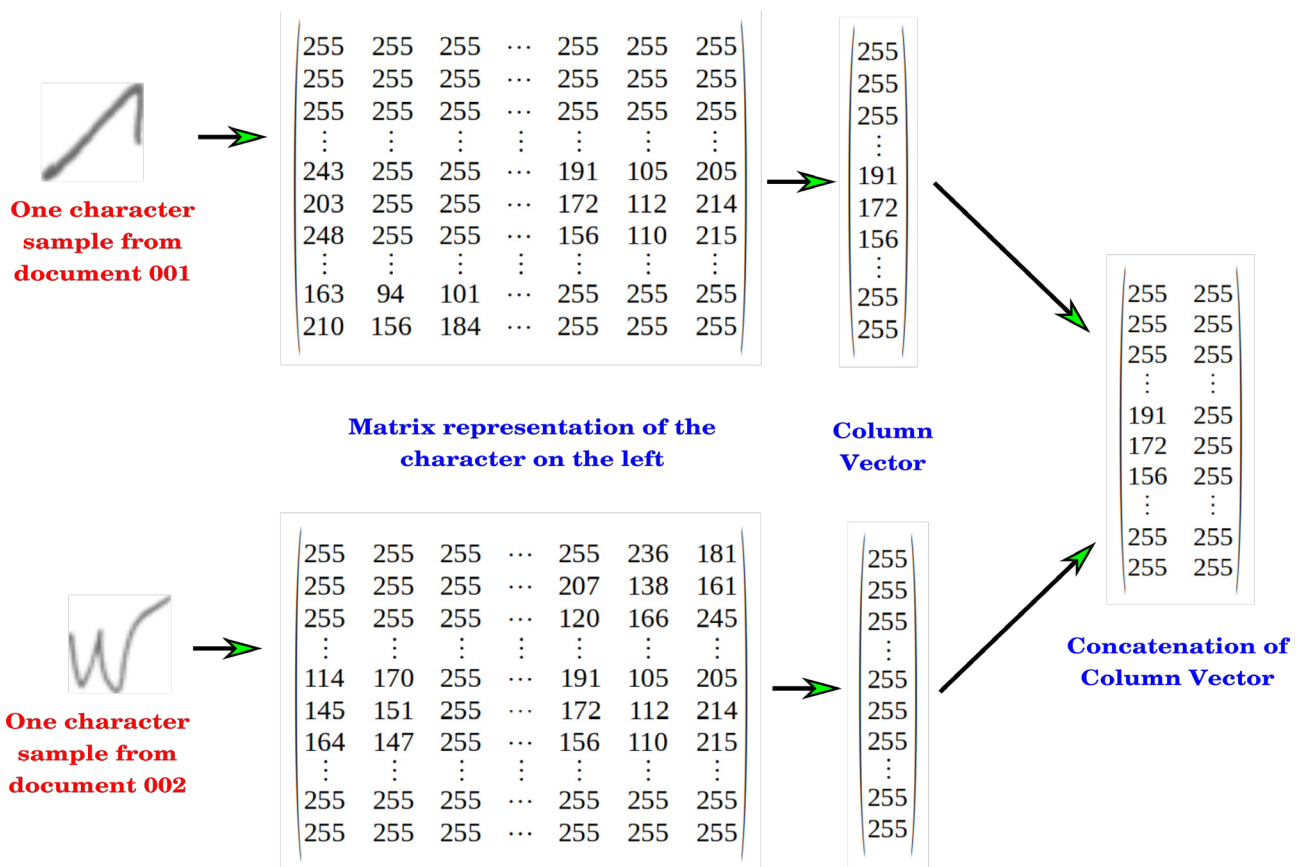


Figure 4. Transforming Process of Lampung Handwriting Sample into Column Vector.

Table 1. Distribution of Training and Testing Set of Writer Identification

No	Sample Name	Number of Training Set	Number of Testing Set
1	Sample I	1.548	82
2	Sample II	4.920	164
3	Sample III	5.300	410
Total sample group		11.768	656
Total amount of sample		12.424	

Sample I is a sample group consisting of one character image as one unit of data. The second sample group is a sample with 2 character images as one unit of data. The last sample group uses 5 character images as one unit of data. The elements of one unit of data in sample II and III are also randomly assigned.

Before the feature extraction stage using PCA is performed, the entire image sample is converted into a column vector. This is conducted for the sake of efficiency and convenience during the computational process as well as dimensionality reduction. An example of transforming process of two image samples of Lampung handwritten character into a column vector is illustrated in Figure 4.

After the entire selected character in each sample is converted into a column vector, the next step is feature extraction using PCA. Each sample is processed by following the algorithm and steps described in section II.B. These steps are calculating the average, then subtraction

of the training set and its average, followed by calculating the covariance of subtraction and finally calculating the character projection. The final step is conducted as a reference at the writer identification stage. All these steps are the training phase on the writer identification system.

The writer identification is decided based on the Euclidean distance between the sample image of the testing set with all the images of the writer in the training set. The smallest distance among pairs indicates a high degree of similarity and implies the writer identity. The formula to compute this Euclidean distance is given in equation (6).

$$Euc = \sqrt{\sum_{n=1}^M (\Gamma_{in} - \Gamma_{jn})^2} \tag{6}$$

where:
 Euc: Euclidean distance
 M: number of data

Γ_{in} : training image vector

Γ_{jn} : testing image vector

By using the formula in equation (6), Euclidean distance is computed for all three sample groups. Then the computation outcome is evaluated to find the minimum Euclidean distance as the closest pair among testing and training data. The final results of the writer identification by this procedure are summarized in Table 2.

Table 2. The Accuracy of Writer Identification on the Lampung Handwritten Document.

Sample Name	Number of Testing Set	Number of Correct Identification	Accuracy (%)
Sample I	82	68	82,92
Sample II	164	82	50,00
Sample III	410	118	28,29

Evaluation of Euclidean distance is a testing phase or matching scheme of writer identification through a series of system processes. The evaluation shows that the highest accuracy obtained from sample I. Two other samples show a significantly decreasing accuracy rate.

4. Discussion

The observations in Table 2 show that the highest accuracy of the writer identification is found in Sample I with total 68 correct identifications, resulting the accuracy of 82.92%. The lowest accuracy occurs in the result of Sample III as many as 118 correct identifications or 29.29%. The research prediction for the best accuracy of proposed method was at least 75%. This means that the target research has met the expectations and the performance of the PCA method has provided adequate results for the writer identification of the Lampung handwritten character. However, the direct implication of this result is that there is still a space for improvement of the accuracy and the writer identification of Lampung handwriting is relatively a new "brand" in the subdomain of writer identification. Some interesting research opportunities are explained in the advice section.

Based on the accuracy noticed in Table 2, observations and analyzing were carried out on the samples that were incorrectly identified. Several possible causes of writer identification failure were successfully observed. Three main reasons of the failure in this identification process are explain in the following:

a. The combinations and permutations of characters for each unit of data in sample III is quite large

The first factor has a significant effect on the accuracy of sample III because of the large difference in the combination of characters in the training set and one unit of data that should be formed. With the total number of Lampung characters as many as 18 characters, the unit data that must be arranged with member of 5 characters

for the training data should be in total 8,568 units of data. This amount is obtained from the calculation on the combination concept of ${}_{18}C_5$. This phenomenon results in a state space explosion of character variations for units of data. While the number of random selected training set is only 5,300 single characters (see Table 1) which must be arranged in a 5-character formation without replacement. With this arrangement, only 1,060 units of data have been formed. The difference as of 7,508 is the minimum number of combination arrangements that are lack in this training set. Consequently, the representation of the model for the writer identification developed during the training process does not reflect all possible combinations. If there is a newly identification data (the one from testing set) belonging to the 7508 group, it is most likely that the accuracy of the writer identification will be biased. The possibility of this bias is quite large because of the number of this group category is also large. Moreover, the arrangement of characters in one unit of data has a lot of permutations. Although there are 5 characters in one unit of data, the total composition of the arrangement of 5 images is 5!. This case doubles the bias which is already large by the former circumstance. Both have a direct impact on the decreasing in accuracy of the writer identification.

b. The similar shapes of some Lampung characters

Another factor leading to an improper prediction of a writer is similarities among of Lampung characters. The basic shape of many characters resembles each other. It is like a quadratic curve and mainly flows to upright. As the results, the Euclidean distance will be small so that two character images with essentially similar in its basic shape will be identified as the same character. Thus, the identification process generates an error for this case. Samples of similar shape of the Lampung character are shown in Figure 5.

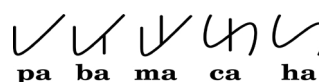


Figure 5. Similar Basic Shape of Some Lampung Characters

The three leftmost characters of Lampung Script in Figure 5 have a basic shape that is similar to the shape of a parabolic curve that opens upward. The basic shape of other similar characters is shown in the last two characters in Figure 5. Both characters have a basic shape like the letter S rotated to 90° clockwise. The main difference between the characters is only the presence or absence of a short line in the middle of the basic shape of the character. Apart from these two basic forms, the Lampung writing system still has some basic forms of the two or more characters.

c. Similarity in writing style among writers

In addition to both cases, writing style among writers who are similar each other is also the most likelihood the evidence for wrong prediction of the writers. Two examples of the Lampung handwritings in the sample

that are similar in appearance but written by two different writers are visualized in Figure 6.



Figure 6. Two Character “Ba” Written by Different Writers

In this example, the character “ba” in part (a) was written by the first writer whereas part (b) was written by the 19th writer. Both images do not exhibit much significant variation so that they look alike. Therefore, the difference of the Euclidean distance between both samples is reasonably small. As a result, both images are considered as the characters derived from the same person during the writer identification process. This type of mistake occurs quite a lot in the testing sample during this final stage. Consequently the accuracy of the writer identification encounters a considerable degradation even in Sampel I as well. This kind of mistake is not triggered by the system operation but instead it is a purely independent factor.

5. Conclusion

The study on the writer identification of the Lampung handwritten documents based on selected characters notice some conclusion:

- The PCA feature extraction method has been successfully applied to the identification of writers on the Lampung handwritten documents.
- The highest performance was obtained from the evaluation of Sample I containing of 82 testing images with the accuracy of 82.92%. The lowest performance was confirmed from 410 testing images of Sample III with an accuracy of 28.29%. This highest performance is moderately convincing for the Lampung characters as the new initiating characters on the writer identification sub-field.
- The implication of the results shown in Table 2 indicates that taking one character image from a Lampung handwriting document as one unit of data is fairly enough to identify the writer.

Based on the analysis during the study, the research team also managed some interesting challenges from this study. Some prospects can be considered as a new subject in the next study or enhanced the existing approach for future development. The topic can be one of the following:

- Implement the Principal Component Analysis (PCA) feature extraction method as an image-based writer identification extracted from a complete character or line-based handwriting from a document.
- Use the PCA approach to perform features selection that are the most relevant to Lampung handwritten characters and compare the results of some feature configurations of those selected features.
- Promote the writer identification process on the Lampung handwriting document using other

classification methods such as k-Nearest Neighbor (k-NN), Support Vector Machine (SVM), Naïve Bayes, Decision Tree or Hidden Markov Model (HMM).

References

- [1] A. Junaidi, S. Vajda and G. A. Fink, “Lampung - A New Handwritten Character Benchmark: Database, Labeling and Recognition,” dalam *Join Workshop on Multilingual OCR and Analytics for Noisy Unstructured Text Data*, Beijing, 2011.
- [2] H. Kusetogullari, A. Yavariabdi, A. Cheddad, H. Grahn and J. Hall, “ARDIS: A Swedish Historical Handwritten Digit Dataset,” *Neural Computing and Applications*, 2019.
- [3] K. Adam, A. Baig, S. Al-Maadeed, A. Bouridane and S. El-Menshawy, “KERTAS: Dataset for Automatic Dating of Ancient Arabic Manuscripts,” *International Journal on Document Analysis and Recognition (IJ DAR)*, vol. 21, no. 4, p. 283–290, December 2018.
- [4] A. M. Saeed, T. A. Rashid, A. M. Mustafa, R. A. A.-R. Agha, A. S. Shamsaldin and N. K. Al-Salih, “An Evaluation of Reber Stemmer with Longest Match Stemmer Technique in Kurdish Sorani Text Classification,” *Iran Journal of Computer Science*, vol. 1, no. 2, p. 99–107, June 2018.
- [5] D. Brodić, Z. N. Milivojević and Č. A. Maluckov, “Script Characterization in the Old Slavic Documents,” dalam *International Conference on Image and Signal Processing*, Cherboung, France, 2014.
- [6] M. Alghamdi and W. Teahan, “Experimental Evaluation of Arabic OCR Systems,” *PSU Research Review*, vol. 1, no. 3, pp. 229–241, 2017.
- [7] M. Z. Alom, P. Sidike, M. Hasan, T. M. Taha and V. K. Asari, “Handwritten Bangla Character Recognition Using the State-of-the-Art Deep Convolutional Neural Networks,” *Computational Intelligence and Neuroscience*, 2018.
- [8] M. Grębowiec and J. Protasiewicz, “A Neural Framework for Online Recognition of Handwritten Kanji Characters,” dalam *Federated Conference on Computer Science and Information Systems (FEDCSIS)*, Poznań, Poland, 2018.
- [9] Z. Zhong, L. Jin and Z. Xie, “High Performance Offline Handwritten Chinese Character Recognition Using GoogLeNet and Directional Feature Maps,” dalam *International Conference on Document Analysis and Recognition (ICDAR)*, Nancy, France, 2015.
- [10] K. Khan, R. Ullah, N. A. Khan and K. Navid, “Urdu Character Recognition using Principal Component Analysis,” *International Journal of Computer Applications*, vol. 60, no. 11, pp. 1–4, 2012.
- [11] A. Das, T. Kundu and C. Saravanan, “Dimensionality Reduction for Handwritten Digit Recognition,” *EAI Endorsed Transactions on Cloud Systems*, vol. 4, no. 13, 2018.
- [12] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, “Gradient-based Learning Applied to Document Recognition,” dalam *Proceedings of the IEEE*, 1998.
- [13] M. Diem, S. Fiel, A. Garz, M. Keglevic, F. Kleber and R. Sablatnig, “ICDAR2013 Competition on Handwritten

- Digit Recognition (HDRC 2013),” dalam *International Conference on Document Analysis and Recognition*, Washington DC, 2013.
- [14] X. Cui, P. Zhou and W. Yang, “Local Dominant Orientation Feature Histograms (LDOFH) for Face Recognition,” *Applied Informatics*, vol. 5, no. 1, December 2017.
- [15] M. Venianaki, O. Salvetti, E. de Bree, T. Maris, A. Karantanas, . E. Kontopodis, K. Nikiforaki and K. Marias, “Pattern Recognition and Pharmacokinetic Methods on DCE-MRI Data for Tumor Hypoxia Mapping in Sarcoma,” *Multimedia Tools and Applications*, vol. 77, no. 8, p. 9417–9439, April 2018.
- [16] T. K. Padma Shri and N. Sriraam, “Pattern Recognition of Spectral Entropy Features for Detection of Alcoholic and Control Visual ERP’s in Multichannel EEGs,” *Brain Informatics*, vol. 4, p. 147–158, January 2017.
- [17] T. Hachaj and M. R. Ogiela, “Human actions recognition on multimedia hardware using angle-based and coordinate-based features and multivariate continuous hidden Markov model classifier,” *Multimedia Tools and Applications*, vol. 75, p. 16265–16285, 2016.
- [18] L. C. Paul and A. Al-Sumam, “Face Recognition Using Principal Component Analysis Method,” *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, vol. 1, no. 9, pp. 135-139, 2012.
- [19] A. Kaur, S. Singh and Taqdir, “Face Recognition Using PCA (Principal Component Analysis) and LDA (Linear Discriminant Analysis) Techniques,” *International Journal of Advanced Research in Computer and Communication Engineering*, vol. 4, no. 3, pp. 308-310, 2015.
- [20] I. Siddiqi and N. Vincent, “A Set of Chain Code Based Features for Writer Recognition,” dalam *10th International Conference on Document Analysis and Recognition*, Barcelona, 2009.
- [21] L. Zuo, Y. Wang and T. Tan, “Personal Handwriting Identification Based on PCA,” dalam *Second International Conference on Image and Graphics*, Hefei, 2002.
- [22] M. Cheriet, N. Kharma, C.-L. Liu and C. Y. Suen, *Character Recognition Systems A Guide for Students and Practitioners*, New Jersey: John Wiley & Sons, Inc., 2007.
- [23] G. Louloudis, N. Stamatopoulos and B. Gatos., “ICDAR 2011 Writer Identification Contest,” dalam *11th International Conference on Document Analysis and Recognition*, Beijing, 2011.
- [24] S. Fiel and R. Sablatnig, “Writer Retrieval and Writer Identificatin using Local Features,” dalam *10th LAPR International Workshop on Document Analysis Systems*, Queensland, 2012.
- [25] S. Al-Maadeed, A. Hassaine, A. Bouridane and M. A. Tahir, “Novel Geometric Features for Off-line Writer Identification,” *Pattern Analysis and Applications*, vol. 19, no. 3, pp. 699-708, 2016.