LONTAR KOMPUTER VOL. 10, NO. 1 APRIL 2019 **p-ISSN 2088-1541**
DOI : 10.24843/LKJITI.2019.v10.i01.p02 **e-ISSN 2541-5832**
Accredited B by RISTEKDIKTI Decree No. 51/E/KPT/2017

Implementation of Zoning and K-Nearest Neighbors in Character Recognition of Wrésastra Script

I Wayan Agus Surya Darma

Informatics Engineering, STMIK STIKOM Indonesia
Tukad Pakerisan 97th Denpasar, Bali, Indonesia
surya@stiki-indonesia.ac.id

Abstract

Balinese script is an important aspect that packs the Balinese culture from time to time which continues to experience development along with technological advances. Balinese script consists of three types (1) Wrésastra, (2) Swalalita and (3) Modre which have different types of characters. The Wrésastra and Swalalita script are Balinese scripts which grouped into the script criteria that are used to write in the field of everyday life. In this research, the zoning method will be implemented in the feature extraction process to produce special features owned by Balinese script. The results of the feature extraction process will produce special features owned by Balinese script which will be used in the classification process to recognize the character of Balinese script. Special features are produced using the zoning method, it will divide the image characters area of Balinese scripts into several regions, to enrich the features of each Balinese script. The result of feature extractions is stored as training data that will be used in the classification process. K-Nearest Neighbors is implemented in the special feature classification process that is owned by the character of Balinese script. Based on the results of the test, the highest level of accuracy was obtained using the value K=3 and reference=10 with the accuracy of Balinese script recognition 97.5%.

Keywords*: Balinese Script, Feature Extraction, Classification, Zoning, KNN*

1. Introduction

Bali is one of the regions in Indonesia that has diverse cultural wealth. This culture has become a tourist attraction in Bali even famous throughout the world. It is an obligation to preserve the culture that is owned by the people in Bali. One of the cultural heritages in Bali is Balinese script which has been used by ancestors in Bali in writing in the lontar script.

Balinese script is an important aspect that packs the Balinese culture from time to time which continues to experience development along with technological advances [1]. Balinese script consists of three types (1) *Wrésastra*, (2) *Swalalita* and (3) *Modre* which have different types of characters. The *Wrésastra* and *Swalalita* script are Balinese scripts which grouped into the script criteria that are used to write in the field of everyday life. The *Modre* script is a sacred Balinese script that is used in the writing of sacred spells in Balinese culture. *Wrésastra* script is a Balinese script that popularly referred to as *anacaraka* consisting of 18 characters as shown in Figure 1.

Figure 1. The characters of Balinese *Wrésastra* script

Wrésastra script is a Balinese script that is used as a guide for writing Balinese latin script. *Wrésastra* script also has a *Pengangge Suara* which is a vocal component in writing Balinese script. It consists *of Tedong, Ulu, Suku, Pepet, Taling, Surang, Adeg, Wisah* and *Cecek*. *Pengangge* characters in Balinese script has a function as a component of character combined with *Anacaraka* character in writing Balinese latin text.

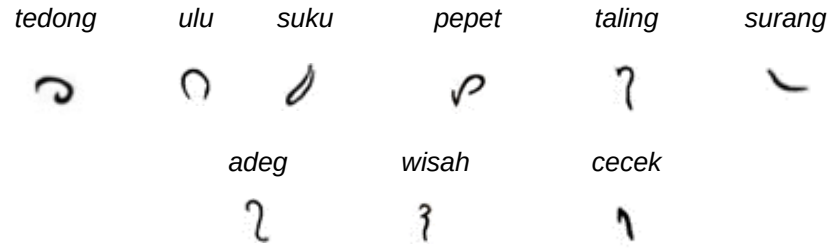

Figure 2. The characters of *Pengangge* script

Some researchers have implemented the K-Nearest Neighbors (KNN) method in identifying the character of Balinese script [2] based on the special characteristics of Balinese script. A study that provides an overview of the special feature extraction of Balinese script using the *zoning* method also has been carried out [3]. The results of this study can be used as a basis for the development of a Balinese script characters recognition system. Other researches on the implementation of KNN have also been carried out [4] for the classification of Batik Parang motifs so that batik types can be identified based on the motif features produced. Other research on KNN for the introduction of Arabic characters has also been done [5]. This study implements KNN in the classification of Arabic characters.

The research that related to the introduction system has been carried out in the introduction of the palm entitled "*High Performance Palm print Identification System Based on Two Dimensional Gabor*" [6] resulting in a success rate of 98.7% aims to introduce the palm ROI segmentation method at the center of the two-stage moment and apply the two-dimensional Gabor method to produce the palm code as a palm feature and use the *hamming* distance method to measure the similarity level of two palm vectors. The similar research also has been conducted to recognize android-based faces with the Eigen face Method that is used to extract relevant information from a face image, then change it into the most efficient set of codes and the code is compared with the code of the face image that has been stored in the database [7]. Previous research related to KNN which uses a custom image to train the classifier [8] to recognize handwritten or printed text. Previous research on remote sensing image of SVM and KNN classification [9] was also conducted to calculate accuracy in classification. Previous research about feature extraction and classification of pollen species and types is an important task in many areas like forensic palynology, archaeological palynology and melissopalynology [10]. Other research about feature extraction to extract color image using texture methods [11] which these features can be used as a primary key to identify and recognize the image. A research about feature extraction and classification for X-Ray Medical Image [12] which proposed pertinent feature extraction algorithm for X-ray medical images and determined machine learning methods for automatic X-ray medical image classification. Previous research on Japanese Hiragana character recognition using Euclidean distance [13] obtained 94.1% average accuracy. Other research related to character recognition on Hindi Optical Character Recognition for Printed Documents using KNN [14] and a research about Javanese handwritten character classification [15].

The character of Balinese script that is used as primary data is the image characters of Balinese script from handwriting. In this research, the zoning method will be implemented in the feature extraction process to produce the special feature of Balinese script. The result of feature

extraction process will produce special features of Balinese script that will be used in the classification process to recognize the characters of Balinese script.

Special features is produced using the zoning method, it will divide the image characters area of Balinese scripts into several regions, to enrich the features of each Balinese script. The result of feature extractions are stored as training data that will be used in the classification process. KNN is implemented in the special feature classification process that is owned by the character of Balinese script. In the classification process, the character feature of Balinese script will be classified and matched with the Balinese script features stored in the training data. The results of the introduction will be presented on the percentage of the nearest neighbors, in accordance with the similarity of the characteristics of the Balinese script with training data that has been through the classification process using KNN.

2. Research Methods

The implementation of zoning and KNN in the introduction of *Wrésastra* script through two main stages, namely the feature extraction stage for the formation of training data and the introduction stage. This stage is shown in Figure 3.

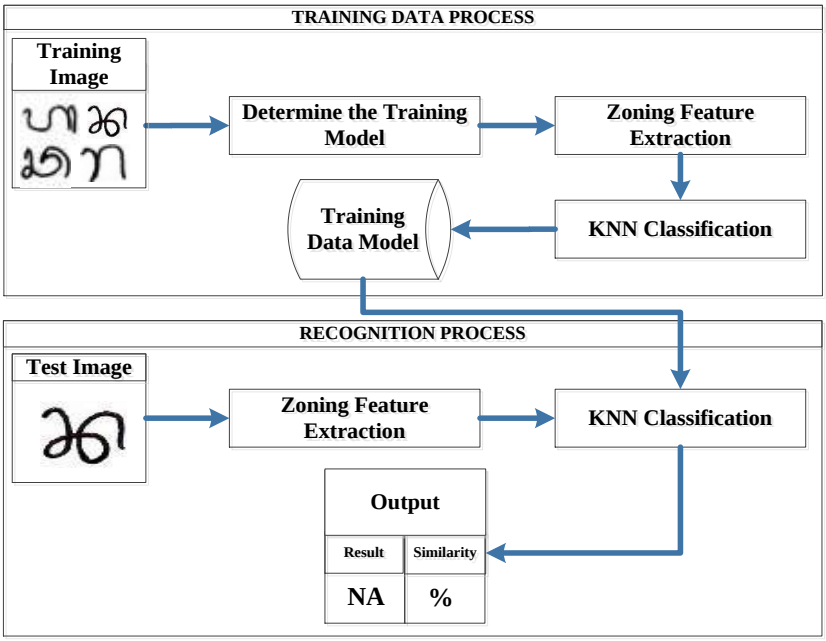

Figure 3. The overview of zoning and KNN implementation in the *Wrésastra* script recognition system.

Figure 3 shows an overview of zoning and KNN implementation in the *Wrésastra* script recognition system. The stages in the introduction process are divided into two main stages, first, the training data formation phase and the introduction stage.

2.1 Data Training Process

The implementation of the zoning and KNN methods is carried out at each of the character recognition stages in *Wrésastra* script. In the first stage, there is a process of forming training data that is used as a reference in the introduction stage. During the training data formation phase, the character image of *Wrésastra* script through the feature extraction process. At this stage, the zoning method is implemented to divide the character image of the Balinese script. The division of regions in the image of Balinese script aims to provide special features variations on each character image of Balinese script. The results of this process will go through the classification stage and stored into the training data model. In this research, the formation of

training data models can be determined based on the number of training data for each character image of *Wrésastra* script

a. Training Image

The first process is to register the training image to be formed into a training model. This process aims to determine the character image of Balinese script that will be used as training data. In this study, the variation of training image used for each Balinese script are 5, 7 and 10 images, which means the total number of images used as training data for 27 *Wrésastra* Scripts are variation of 135, 189 and 270 images. Variations in the amount of training data for each *Wrésastra* script will be used as a training model for the testing phase for the recognition of *Wrésastra* script.

b. Determine the Training Model

The second process at this stage is the feature extraction process to acquire the special features of each Balinese script characters that are used as training data model. In this process, it can be created training data model based on the number of Balinese scripts that is used.

c. Zoning Feature Extraction

The third process at this stage is the feature extraction process to acquire the special features of each Balinese script characters that is used as a training data model. The zoning method is implemented to divide the image area of Balinese script, therefore the resulting features will be more varied. The implementation of the zoning method in the feature extraction process is done in order to obtain unique features for each character. The zoning method is implemented by zoning on each character during the feature extraction process. The following is an example of zoning in feature extraction to obtain a stop point in Figure 4.

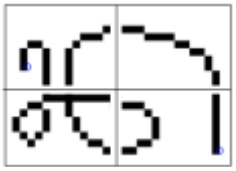

Figure 4. Zoning feature extraction on character

The first feature that is generated based on the regional division of Balinese script characters is the direction feature based on table 1.

Table 1. Features of Balinese Script Directions

Form	Value	Direction
	2	Vertical
	4	Horizontal
	3	Right Diagonal
	5	Left Diagonal

The determination of direction value in the feature extraction of Balinese script through the following steps:

1. The determination on starting pixel point of the character

   The starting point of the character is the first pixel point found in the character image, which is the lower left point. In this process, there will be an iteration of searching the starting point and direction values that is started at the starting pixel point until no more pixels that are characters formation which is has no direction value. After the iteration of

searching the starting point and direction value is complete, the point that has been found is given a temporary value to mark the pixel of the Balinese script for each character.

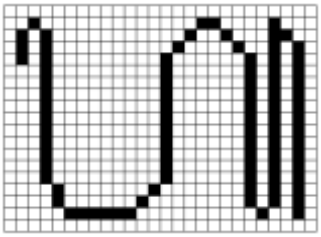

Figure 5. Pixel of Balinese script image

2. Encoding based on direction features

Based on table 1, the character image of Balinese script will be encoded based on the character form of the Balinese script by referring to the table of direction features. There are four directions that are used as guidelines in encoding the features of Balinese script, namely vertical, horizontal, diagonal right and diagonal left. The results of the first step will be encoded based on the direction value shown in table 1. Therefore the character pixel of Balinese script have values based on the direction feature table, Figure 6 shows the results of the encoding of the Balinese script image.

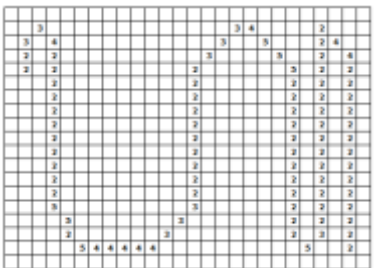

Figure 6. Conversion of feature direction values

The second feature that is the results from the feature extraction process is the semantic feature of Balinese script. Based on Figure 7, the semantic features of Balinese script consist of the number of stop points, character length, character width, loops, horizontal lines, vertical lines, based on the pixels formed into a vector, it can be seen that the strokes of these pixels will form a written *Wrésastra Scripts*.

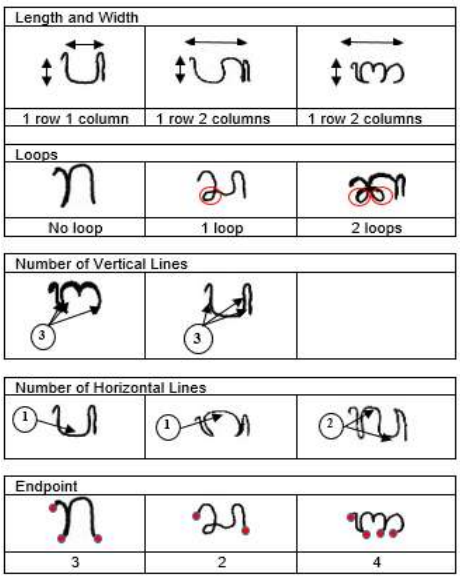

Figure 7. Semantic Feature

Figure 7 shows the length and width of Balinese characters is obtained by search the black pixels in the *Wrésastra script* from the left then moving to the right and character width is obtained by searching black pixels from top to bottom. The loop feature is a feature in the Balinese script that has pixels that make up the loop path. The process of searching for horizontal lines starts with finding the connection to the neighboring right pixel of *Wrésastra script*. The search process for vertical line features begins with searching for pixels that have a connection between the neighbor pixels below. The search for endpoint on *Wrésastra* characters is done by checking pixels that are searched for by neighbors whose pixels are black too.

d. KNN Classification

The classification process aims to classify data from feature extraction results. KNN will classify features of *Wrésastra* script based on the nearest neighbors. The value of the feature extraction results through the KNN classification process, therefore the *Wrésastra* script feature data will be grouped. The classification results are stored as a training data model.

2.2 Recognition Stage

The second stage is the introduction stage of *Wrésastra* script by implementing KNN. At this stage, the character image of Wrésastra script is tested, through the feature extraction process first to get the special feature values of the *Wrésastra* script character image tested. After the special features of *Wrésastra* script are obtained, this feature will be classified with the training data model that has been formed using KNN.

a. Image Acquisition

The image that is used in the introduction process is the handwritten of *Wrésastra* script characters. The number of *Wrésastra* script image is tested in this research is 81 handwritten of *Wrésastra* script character image.

b. Zoning Feature Extraction

Feature extraction process to acquire specific features of each character of *Wrésastra* script that is used as a training data model. The zoning method is implemented to divide the image area of *Wrésastra* script, therefore the resulting features will be more varied. The implementation of zoning method in the feature extraction process is done in order to obtain unique features for each character. The stages in the feature extraction process at the introduction stage are the same as the Data Training Process stage.

c. KNN Classification

The classification process aims to classify features of *Wrésastra* script. The KNN method classifies *Wrésastra* script feature data based on the closest neighbors. The feature value of the *Wrésastra* script test will be classified with feature values in the training data model that has been formed previously at the training process data stage.

3. Result and Discussion

The test results on the *Wrésastra* image recognition process using the Zoning and KNN methods, which are tested with some data. The parameters used in this test are the neighbor value (K) and the number of references used in the introduction process. The K value is the closest number of neighbors used in the KNN classification process. The number of references is a model of training data that has been created at the training image stage. The following is the test results using the value of K = 1 to K = 10, reference = 5 and 81 sample data can be seen in table 2.

Table 2. Testing Result with reference=5

Amount of "K"	Results of Recognition		
Value	Success	Failed	Percentage
1	75	6	92.5%
2	74	7	91.3%
3	76	5	93.8%
4	72	9	88.8%
5	70	11	86.4%
6	73	8	90.1%
7	70	11	86.4%
8	67	14	82.7%
9	72	9	88.8%
10	70	11	86.4%

Based on the results of testing using reference = 5, the highest recognition accuracy is obtained by using the value K = 3 resulting in an accuracy rate of 93.8%. Testing is also done using reference = 7 with 81 sample data, the test results can be seen in table 3.

Table 3. Testing Result with reference = 7

Amount of "K"	Results of Recognition		
Value	Success	Failed	Percentage
1	77	4	95.0%
2	74	7	91.3%
3	78	3	96.3%
4	74	7	91.3%
5	74	7	91.3%
6	73	8	90.1%
7	74	7	91.3%
8	74	7	91.3%
9	68	13	83.9%
10	70	11	86.4%

Based on the test results using reference = 7, the highest recognition accuracy is also obtained by using the value K = 3 resulting in an accuracy rate of 96.3%. In this second test, the accuracy rate is higher compared to the first test. The higher result is due to the increasing number of references used in the classification process. The next test is done using reference=10 with 81 sample data. Test results can be seen in table 4.

Table 4. Testing Result with reference = 10

Amount of "K"	Results of Recognition		
Value	Success	Failed	Percentage
1	78	3	96.3%
2	74	7	91.3%
3	79	2	97.5%
4	75	6	92.5%
5	76	5	93.8%
6	75	6	92.5%
7	75	6	92.5%
8	76	5	93.8%
9	74	7	91.3%
10	73	8	90.1%

The third test using reference = 10 and 81 sample data resulted in the highest accuracy rate of 97.5% using K = 3. The image of the *Wrésastra* script character that is successfully identified is 79 out of 81 *Wrésastra* script image characters tested.

Based on figure 8 can be seen the results of testing the Implementation of Zoning and KNN applications on *Wrésastra* Script Recognition. The tests are using 81 *Wrésastra* script images data samples of. In the first test using reference = 5, the best accuracy is obtained by using the value K = 3 with 76 data samples of *Wrésastra* script images that were correctly recognized. Furthermore, testing using reference = 7 on 81 *Wrésastra* script sample data, produces the best level of accuracy by using K = 3 with 78 data that is correctly recognized. The third test, using reference = 10 on 81 *Wrésastra* script sample data, produced the best level of accuracy by using K = 3 with 79 data successfully identified correctly. The Graphs of previous tests research can be seen in Figure 8 below.

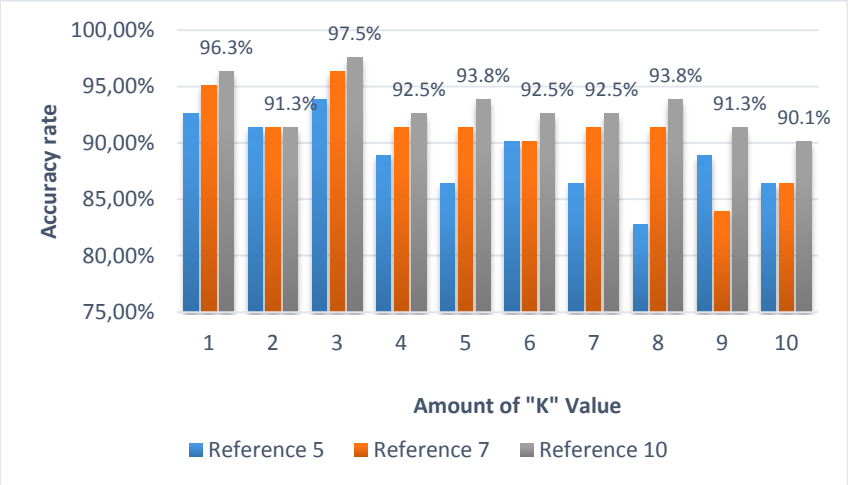

Figure 8. Results of Testing Implementation of Zoning and KNN on *Wresastra* Script Recognition

Based on the results of testing, the highest accuracy rate was obtained by a combination of K=3 and reference=10 which resulted in a value of 97.5%. These results indicate that the implementation of the KNN in the recognition of *Wrésastra* script was able to produce a high accuracy rate. In the previous study, the implementation of KNN on the introduction of Arabic

Character [5] also resulted in 87% accuracy rate. The recognition results are influenced by the value of K. The greater the value of K will affect the accuracy rate. The selection of K values will greatly affect the performance of the KNN. Based on the test results shown in figure 8, the trend of accuracy rate decreases if the K value is higher, because the higher the K value makes the boundary between each classification becomes increasingly blurred so the accuracy rate decreases. The higher result is due to the increasing number of references used in the classification process.

4. Conclusion

This paper presented an implementation of zoning and KNN in the recognition of *Wrésastra* characters to test the accuracy rates based on variations in the number of references used as training data. The accuracy rate obtained was 97.5% with a combination of K=3 and reference=10 values against 81 test images. The recognition results are influenced by the selection of values of K and references. Based on the test results, the trend of accuracy rate decreases if the K value is higher, because the higher the K value makes the boundary between each classification becomes increasingly blurred so the accuracy rate is decreasing. Based on the third test using 10 references, indicating the higher result is also due to the increasing number of references used in the classification process.

References

[1] I. N. Duija, "Keberadaan Aksara Wrésastra Dalam Aksara Bali the Existenace of Wrésastra in Balinese Script," *Kajian Budaya, Institut Hindu Dharma Negeri Denpasar*, vol. 29, no. 51, hal. 1, 2017.

[2] M. Sudarma dan S. Darma, "The Identification of Balinese Scripts Characters based on Semantic Feature and K Nearest Neighbor," *International Journal of Computer Applications*, vol. 91, no. 1, hal. 14–18, Apr 2014.

[3] S. Darma, D. Putra, dan M. Sudarma, "Ekstraksi Fitur Aksara Bali Menggunakan Metode Zoning," *Majalah Ilmiah Teknologi Elektro*, vol. 14, no. 2, hal. 44–49, 2015.

[4] W. Ginantra, "Deteksi Batik Parang Menggunakan Fitur Co-Occurrence Matrix dan Geometric Moment Invariant dengan Klasifikasi KNN," *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 7, no. 1, hal. 715–725, 2016.

[5] M. Rashad dan N. Semary, "Isolated Printed Arabic Character Recognition Using KNN and Random Forest Tree Classifiers," *Advance Machine Learning Technologies and Application*, vol. 488, hal. 11–17, 2014.

[6] I. K. G. D. Putra dan Erdiawan, "High Performance Palmprint Identification System Based On Two Dimensional Gabor," *TELKOMNIKA*, vol. 8, no. 3, hal. 309–318, 2010.

[7] I. K. S. Widiakumara, I. K. Gede, D. Putra, dan K. S. Wibawa, "Aplikasi Identifikasi Wajah Berbasis Android," *Lontar Komputer: Jurnal Ilmiah Teknologi Informasi*, vol. 8, no. 3, hal. 200–207, 2017.

[8] T. K. Hazra, D. P. Singh, and N. Daga, "Optical character recognition using KNN on custom image dataset," in *2017 8th Annual Industrial Automation and Electromechanical Engineering Conference (IEMECON)*, 2017.

[10] G. Alimjan, T. Sun, Y. Liang, H. Jumahun, and Y. Guan, "A New Technique for Remote Sensing Image Classification Based on Combinatorial Algorithm of SVM and KNN," *International Journal of Pattern Recognition and Artificial Intelligence*, vol. 32, no. 07, p. 1859012, Jul. 2018.

[11] A. AlQaisi, M. AlTarawneh, Z. A. Alqadi, and A. A. Sharadqah, "Analysis of Color Image Features Extraction using Texture Methods," *TELKOMNIKA (Telecommunication Comput. Electron. Control.*, vol. 17, no. 3, Jun. 2019.

[12] M. M. Abdulrazzaq, I. FT Yaseen, S. Noah, and M. A. Fadhil, "Multi-Level of Feature Extraction and Classification for X-Ray Medical Image," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 10, no. 1, p. 154, Apr. 2018.

[13] S. Das and S. Banerjee, "An Algorithm for Japanese Character Recognition," International Journal of Image, Graphics and Signal Process., vol. 7, no. 1, pp. 9–15, 2015.

[14] P. A. Choksi, K. Kumari, S. Kanojiya, P. Sahu, and N. Rindani, "Hindi Optical Character

Recognition For Printed Documents Using Fuzzy K-Nearest Neighbor Algorithm : A Problem Approach In Character Segmentation," vol. 8, no. 1, pp. 25–34, 2018.

[15] D. R. I. M. Setiadi, A. Susanto, C. A. Sari, E. H. Rachmawanto, and D. Sinaga, "A High Performace of Local Binary Pattern on Classify Javanese Character Classification," *Scientific Journal of Informatics*, vol. 5, no. 1, p. 8, 2018.