



HAL
open science

Study on Feature Extraction Methods for Character Recognition of Balinese Script on Palm Leaf Manuscript Images

Made Windu Antara Kesiman, Sophea Prum, Jean-Christophe Burie,
Jean-Marc Ogier

► **To cite this version:**

Made Windu Antara Kesiman, Sophea Prum, Jean-Christophe Burie, Jean-Marc Ogier. Study on Feature Extraction Methods for Character Recognition of Balinese Script on Palm Leaf Manuscript Images. 23rd International Conference on Pattern Recognition, Dec 2016, Cancun, Mexico. pp.4006-4011. hal-01422135

HAL Id: hal-01422135

<https://hal.science/hal-01422135v1>

Submitted on 4 Jan 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Study on Feature Extraction Methods for Character Recognition of Balinese Script on Palm Leaf Manuscript Images

Made Windu Antara Kesiman¹, Sophea Prum², Jean-Christophe Burie¹, Jean-Marc Ogier¹

¹Laboratoire Informatique Image Interaction (L3i)

University of La Rochelle, Avenue Michel Crépeau, 17042, La Rochelle Cedex 1, France

{made_windu_antara.kesiman, jcburie, jean-marc.ogier}@univ-lr.fr

²Mimos Berhad, Kuala Lumpur, Malaysia

sopheaprum@gmail.com

Abstract—The complexity of Balinese script and the poor quality of palm leaf manuscripts provide a new challenge for testing and evaluation of robustness of feature extraction methods for character recognition. With the aim of finding the combination of feature extraction methods for character recognition of Balinese script, we present, in this paper, our experimental study on feature extraction methods for character recognition on palm leaf manuscripts. We investigated and evaluated the performance of 10 feature extraction methods and we proposed the proper and robust combination of feature extraction methods to increase the recognition rate.

Keywords—*feature extraction, character recognition, Balinese script, palm leaf manuscript, performance evaluation*

I. INTRODUCTION

Isolated handwritten character recognition (IHCR) has been the subject of intensive research during the last three decades. Some IHCR methods have reached a satisfactory performance especially for Latin script. However, development of IHCR methods for other various new scripts remains a major challenge for researchers. For example, the IHCR challenges for historical documents discovered in the palm leaf manuscripts, written using the script originating from the region of Southeast Asia, such as Indonesia, Cambodia, and Thailand. Ancient palm leaf manuscripts from Southeast Asia have received great attention from historian and also researchers in the field of document image analysis, for example some works on document analysis of palm leaf manuscript from Thailand [1,2] and from Indonesia [3,4]. The physical condition of the palm leaf documents is usually degraded. Therefore, researchers have to deal at the same time with the digitization process of palm leaf manuscripts and with the automatic analysis and indexing system of the manuscripts. The main objective is to bring additional value to the digitized palm leaf manuscripts by developing tools to analyze, to index and to access quickly and efficiently to the content of the manuscripts. It tries to make the manuscripts more accessible, readable and understandable to a wider audience and to scholars all over the world.

The majority of Balinese have never read any palm leaf manuscript because of language obstacles as well as tradition which perceived them as sacrilege. They cannot access to the important information and knowledge in palm leaf manuscript. Therefore, using an IHCR system will help to transcript these ancient documents and translate them to the current language. IHCR system is one of the most demanding systems which has to be developed for the collection of palm leaf manuscript images. Usually, an IHCR system consists of two main steps: feature extraction and classification. The performance of an IHCR system greatly depends on the feature extraction step. The goal of feature extraction is to extract information from raw data which is most suitable for classification purpose [5]. Many feature extraction methods have been proposed to perform the character recognition task [5–13]. These methods have been successfully implemented and evaluated for recognition of Latin, Chinese and Japanese characters as well as digit recognition. However, only a few system are available in the literature for other Asian scripts recognition. For example, some of the works are for Devanagari script [6,14], Gurmukhi script [5,15–17], Bangla script [10], and Malayalam script [18]. Those documents with different scripts and languages surely provide some real challenges, not only because of the different shapes of characters, but also because the writing style for each script differs: shape of the characters, character positions, separation or connection between the characters in a text line.

Choosing efficient and robust feature extraction methods plays a very important role to achieve high recognition performance in an IHCR and OCR [5]. The performance of the system depends on a proper feature extraction and a correct classifier selection [10]. With the aim of developing an IHCR system for Balinese script on palm leaf manuscript images, we present, in this paper, our experimental study on feature extraction methods for character recognition of Balinese script. It is experimentally reported that to improve the performance of an IHCR system, the combination of multi features is recommended [19]. Our objective is to find the combination of feature extraction methods to recognize the isolated characters of Balinese scripts on palm leaf manuscript images. We

investigated 10 feature extraction methods and based on our preliminary experiment results, we proposed the combination of feature extraction methods to improve the recognition rate. Two classifiers: k-NN (k-Nearest Neighbor) and SVM (Support Vector Machine) are used in our experiments. We also compared the results with a convolutional neural network based system. In such a system, feature extraction is not required.

This paper is organized as follow: section II gives a brief description about Balinese script on the collection of palm leaf manuscripts in Bali and the challenges for IHCR. Section III presents the brief description of feature extraction methods and our proposed combination of features which were evaluated in our experimental study. The palm leaf manuscript image dataset used in this work and the experimental results are presented in Section IV. Conclusions with some prospects for the future works are given in Section V.

II. BALINESE SCRIPT ON PALM LEAF MANUSCRIPTS AND THE CHALLENGES FOR IHCR

A. The Balinese Script on Palm Leaf Manuscripts

The Balinese manuscripts were mostly recorded on dried and treated palm leaves, called *Lontar*. They include texts on important issues such as religion, holy formulae, rituals, family genealogies, law codes, treaties on medicine, arts and architecture, calendars, prose, poems and even magics. Unfortunately, many discovered *Lontars*, part of the collections of some museums or belonging to private families, are in a state of disrepair due to age and due to inadequate storage conditions. *Lontars* were written and inscribed with a small knife-like pen (a special tool called *Pengerupak*). It is made of iron, with its tip sharpened in a triangular shape so it can make both thick and thin inscriptions. The manuscripts were then scrubbed by the natural dyes to leave a black color on the scratched part as text. The Balinese palm leaf manuscripts were written in Balinese script and in Balinese language. Writing in Balinese script, there is no space between words in a text line. Some characters are written on upper baseline or under the baseline of text line. *Lontars* were written in the ancient literary texts composed in the old Javanese language of Kawi and Sanskrit.

B. The IHCR Challenge for Balinese Script in Palm Leaf Manuscripts

For Balinese IHCR, there are 2 challenges: 1) the complexity of Balinese script, and 2) the poor quality of palm leaf manuscripts. Balinese script is considered to be one of the more complex scripts from Southeast Asia. The alphabet and numeral of Balinese script is composed of ± 100 character classes including consonants, vowels, diacritics, and some other special compound characters. Usually, palm leaf manuscripts have poor quality since the documents are degraded over time due to storage conditions. The palm leaf manuscripts contain discolored parts and artefacts due to aging. They have low intensity, variations or poor contrast, random noises, and fading [4]. Several deformations in the character shapes are visible due to the merges and fractures of the characters, varying space between letters, and varying space

between lines (see Fig. 1). It is known that the similarities of distinct character shapes, the overlaps, and interconnection of the neighboring characters further complicate the problem of an OCR system [7]. One of the main problem faced when dealing with segmented handwritten character recognition is the ambiguity and illegibility of the characters [8]. In the scope of this paper, we will focus on scripts written on palm leaf whose characters are even more elaborated in ancient written form. These characteristics provide a suitable challenge for testing and evaluation of robustness for feature extraction methods which were already proposed for character recognition. Balinese scripts on palm leaf manuscripts offer a real new challenge in IHCR development.

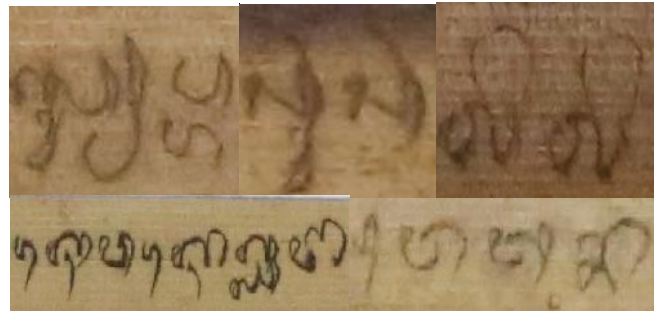


Fig. 1 Balinese script on palm leaf manuscripts

III. FEATURE EXTRACTION METHODS AND OUR PROPOSED COMBINATION OF FEATURES

Many feature extractions methods have been presented in the literature. Each method has its own advantages or disadvantages over other methods. In addition, each method may be specifically designed for some specific problem. Most of feature extraction methods, extract the information from binary image or grayscale image [6]. Some surveys and reviews on features extraction methods for character recognition were already reported [19–24]. In this work, first, we investigated and evaluated some most commonly used features for character recognition: histogram projection [6,10,18], celled projection [10], distance profile [6,18], crossing [6,10], zoning [6,8,9,18], moments [14,18], some directional gradient based features [5,11], Kirsch Directional Edges [6], and Neighborhood Pixels Weights (NPW) [6]. Secondly, based on our preliminary experiment results, we proposed and evaluated the combination of NPW features applied on Kirsch Directional Edges images, with Histogram of Gradient (HoG) features and zoning method. This section will only briefly describe the feature extraction methods which were used in our proposed combination of features and the convolutional neural network. For more detail description of other commonly used feature extraction methods which were also evaluated in this experimental study, please refer to references mentioned above.

A. Kirsch Directional Edges

Kirsch edges is a non-linear edge enhancement [6]. Let A_i ($i=0,1,2,\dots,7$) be the eight neighbors of the pixel (x,y) , i is

taken as modulo 8, starting from top left pixel at the moving clock-wise direction.

Four directional edge images are generated (Fig. 2) by computing the edge strength at pixel (x,y) in four (horizontal, vertical, left diagonal, right diagonal) direction, defined as G_H , G_V , G_L , G_R , respectively [6]. They can be denoted as bellow:

$$G_H(x,y) = \max(|5S_0 - 3T_0|, |5S_4 - 3T_4|) \quad (1)$$

$$G_V(x,y) = \max(|5S_2 - 3T_2|, |5S_6 - 3T_6|) \quad (2)$$

$$G_R(x,y) = \max(|5S_1 - 3T_1|, |5S_5 - 3T_5|) \quad (3)$$

$$G_L(x,y) = \max(|5S_3 - 3T_3|, |5S_7 - 3T_7|) \quad (4)$$

Where S_i and T_i can be computed by:

$$S_i = A_i + A_{i+1} + A_{i+2} \quad (5)$$

$$T_i = A_{i+3} + A_{i+4} + A_{i+5} + A_{i+6} + A_{i+7} \quad (6)$$

Each directional edge image is thresholded to produce a binary edge image. The binary edge image is then partitioned into N smaller regions. Then, edge pixel frequency in each region is computed to produce the feature vector. In our experiments, we computed Kirsch feature from grey scale image with 25 smaller regions to produce a 100 dimensions feature vector. Based on the empirical tests for our dataset, the Kirsch edge image can be optimally thresholded with a threshold value of 128. The feature value is then normalized by the maximum value of edge pixel frequency from all regions.

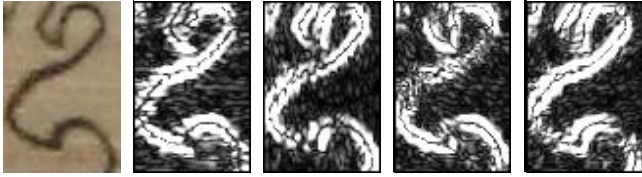


Fig. 2 Four directional Kirsch edge images

B. Neighborhood Pixels Weights (NPW)

Neighborhood Pixels Weight (NPW) was proposed by Satish Kumar [6]. This feature may work on binary as well as on gray images. NPW considers four corners of neighborhood for each pixel: top left, top right, bottom left, and bottom right corner. The number of neighbors considered on each corner is defined by value of layer level (see Fig. 3). Level 1 considers only pixels in layer 1 on each corner (1 pixel), level 2 considers pixels in layer 1 and 2 (4 pixels), and level 3 considers pixels in all layers (9 pixels). In the case of binary image, the weight value on each corner is obtained by counting the number of pixel character, divided by total number of neighborhood pixels on that corner. For grayscale image, the weight value on each corner is obtained by summing the gray level of all neighborhood pixels, divided by maximum possible weight due to all neighborhood pixels on that corner ($nb_neighborhood_pixels \times 255$). Four weighted plans are constructed for each corner from the weighted value of all pixels on the image. Each plane is divided into N smaller regions, and the average weight of each region is computed. The feature vector is finally constructed from the average weight of each region from each plane ($N \times 4$ vector dimension).

In our experiments, we computed NPW features in level 3 neighborhood with 25 smaller regions ($N=25$) to produce a 100 dimensions feature vector. The feature value is normalized by the maximum value of average weight from all regions. We tested the performance of NPW feature for both binary and grayscale image.

3	3	3		3	3	3
3	2	2		2	2	3
3	2	1		1	2	3
			P			
3	2	1		1	2	3
3	2	2		2	2	3
3	3	3		3	3	3

Fig. 3 Neighborhood pixels for NPW features

C. Histogram of Gradient (HoG)

The gradient is a vector quantity comprising of magnitude as well as directional component computed by applying its derivatives in both horizontal and vertical directions [5]. The gradient of an image can be computed either by using for example Sobel, Roberts or Prewitt operator. The gradient strength and direction can be computed from the gradient vector. Gradient feature vector used in [5] is formed by accumulating the gradient strength separately along different directions.

To compute HoG, first, we compute the gradient magnitude and gradient direction of each pixel of the input image. The gradient image is then divided into some smaller cells, and in each cell we generate the histogram of directed gradient by assigning the gradient direction of each pixel into certain range of orientation bin which are evenly spread over 0 to 180 degrees or 0 to 360 degrees (Fig. 4 & 5). The histogram cells are then normalized with a larger overlap connected blocks.

*	*	*	*	*	×	*	*
*	*	*	×	*	*	*	*
*	*	*	*	*	*	*	*
*	*	*	*	×	*	×	*

Fig. 4 An image with 4x4 oriented histogram cells and 2x2 descriptor blocks overlapped on 2x1 cells

The final HoG descriptor is then generated from all concatenated vector of the histogram after the block normalization process. For our experiments, we used the HoG implementation of VLFeat¹. We computed HoG feature from gray scale image with cell size of 6 pixels and with 9 different orientations to produce a 1984 dimensions feature vector.



Fig. 5 The representation of the array of cells HoG

¹ <http://www.vlfeat.org/api/hog.html>

D. Zoning

Zoning is computed by dividing the image into N smaller zones: vertical, horizontal, square, diagonal left and right, radial or circular zone (see Fig. 6). The local properties of image are extracted on each zone. Zoning can be implemented for binary image and grayscale image [6]. For example, in binary image, the percentage density of character pixels in each zone is computed as local feature [9]. In grayscale image, the average of gray value in each zone is considered as local feature [18]. Zoning can be easily combined with other feature extraction methods [10], for example in [8]. In our experiments, we computed zoning with 7 zone types (zone width or zone size = 5 pixels) and combined them into a 205 dimensions feature vector. We also tested the zoning feature on the image of the skeleton.

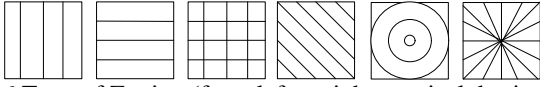


Fig. 6 Type of Zoning (from left to right: vertical, horizontal, block, diagonal, circular, and radial zoning)

E. Our proposed combination of features

After evaluating the performance of 10 individual feature extraction methods, we found that the HoG features, NPW features, Kirsch features and Zoning method give a very promising result (see Table I in Section IV). We obtained 62,45% of recognition rate only by using Kirsch features. It means that the four directional Kirsch edge images already serve as a good feature discriminants for our dataset. The shape of Balinese characters are naturally composed by some curves. We can notice that Kirsch edge image is able to give the initial directional curve features for each character. On the other hand, NPW features have an advantage that it can be applied directly to gray level images. Our hypothesis is the four directional Kirsch edge images will provide a better feature discriminants for NPW features. Based on this hypothesis, we proposed a new feature extraction method by applying NPW on Kirsch edge images. We call this new method as NPW-Kirsch (see Fig. 7). Finally, we concatenate NPW-Kirsch with two other features, HoG and Zoning method.

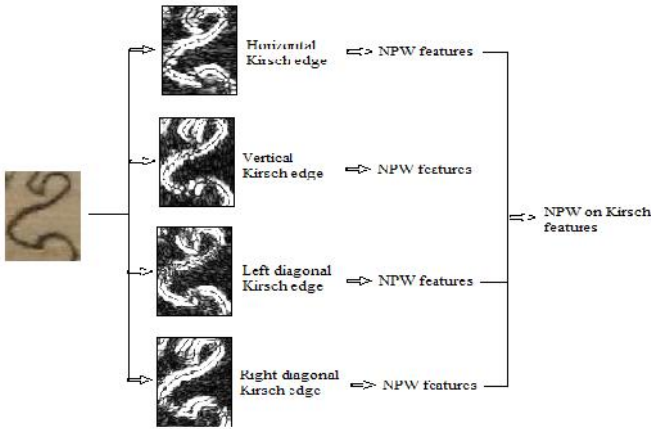


Fig. 7 Scheme of NPW on Kirsch features

F. Convolutional neural network

In this experiment, we also use convolutional neural network where feature extraction step is not required. The network considers the value of each pixel of the input image as the input layer. In this experiment, we used Tensorflow library².

More specifically, multilayer convolutional neural network is used. The architecture of our network is illustrated in Fig. 8. Given an input grayscale image of 28x28 pixels, the first convolutional layer (C1) is computed by using a sliding window of 5x5 pixels. We obtain C1 layer containing 28x28x32 neurons, where 32 is the number of features maps chosen. For each sliding window on the neuron (i, j), the convolutional $C(i, j)$ output can be computed by:

$$C(i, j) = \sum_{k=0}^4 \sum_{l=0}^4 (W_{(k,l)} I_{(i+k, j+l)}) \quad (7)$$

Where W is a 5x5 matrix to be used as the shared weights matrix, and I is the input neurons. Rectified linear unit is then applied. We obtain: $C = \text{ReLU}(b+C)$, where b is the bias. Then, we apply max-pooling using a window of 2x2 which consists in choosing the maximum activation in the selected region as the output. We obtain P2 layer consists of 14x14x32 neurons. After computing the second convolutional layer (C3) and second max-pooling (P4), we obtain a layer (P4) of 7x7x64 neurons. We add a fully-connected layer (F5) of 1024 neurons, where

$$F5 = \text{ReLU}(P_4 W_4 + b_4) \quad (8)$$

P_4 is a one dimension matrix containing the 3136 neurons in P4. W_4 is a 3136x1024 shared weight matrix, and b_4 is the bias.

Finally, we fully connect this F5 layer to the output layer, where the number of neurons equals the number of classes by using the equation 8.

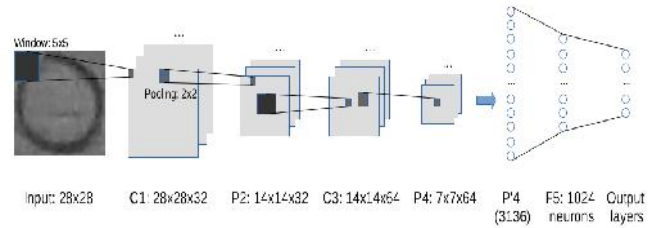


Fig. 8 Architecture of multilayer convolutional neural network

IV. DATASET, EXPERIMENTS AND RESULTS

We present our experimental study on feature extraction methods for character recognition of Balinese script on palm leaf manuscripts.

A. Palm leaf manuscript images dataset

Our Balinese palm leaf manuscript images dataset come from 5 different locations in Bali, Indonesia: 2 museums and 3 private families. In order to obtain the variety of the manuscript images, the sample images used in this experiment are randomly selected from 23 different collections (contents),

² <http://arxiv.org/abs/1603.04467>

with the total of 393 pages. By using the collection of word-level annotated patch images from our palm leaf manuscript collection, which were produced manually using Aletheia³ [25] in our previous ground truthing process, we collected our isolated Balinese character dataset. We applied binarization process to automatically extract all connected component found on the word patch images. Our Balinese philologists manually annotated the segment of connected component that represent a correct character in Balinese script. All patch images that have been segmented and annotated at character level will serve as our isolated character dataset. It consists of 133 character classes, with a total number of 19,383 character samples and 11,710 samples are randomly selected and used as train set. The remaining samples are used as test set. The number of sample images for each class is different. Some classes are frequently found in our collection of palm leaf manuscripts, but some others are rarely used. Thumbnail samples of these images are showed in Fig. 9.

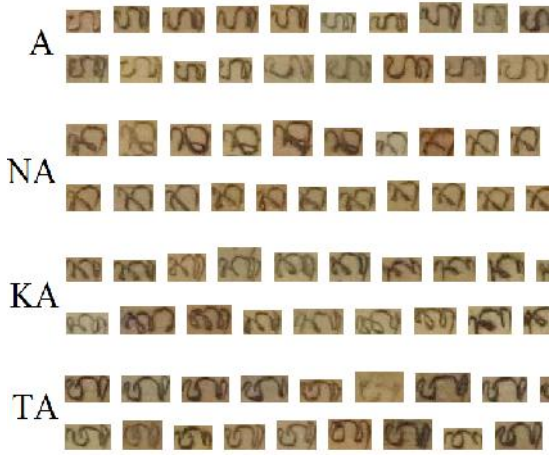


Fig. 9 Samples of character-level annotated patch images of Balinese script on palm leaf manuscripts

B. Experiments of character recognition

We investigated and evaluated the performance of 10 feature extraction methods and the proposed combination of features in 29 different schemes. We also compared the experimental result with the convolutional neural network. For all experiments, a set of image patches containing Balinese characters from the original manuscripts will be used as input, and a correct class of each character should be identified as a result. We used $k=5$ for k-NN classifier, and all images are resized to 50x50 pixels (the approximate average size of character in collection), except for Gradient features where images are resized to 81x81 pixels to get evenly 81 blocks of 9x9 pixels, as described in [11].

The results (Table I) show that the recognition rate of NPW features can be significantly increased (up to 10%) by applying it on the four directional Kirsch edge images (NPW-Kirsch method). Then, by combining this NPW-Kirsch features, HoG features and Zoning method can increase the recognition rate up to 85%. This result is slightly better than

using the convolutional neural network. In our experiments, the number of training dataset for each classes is not balance, and it influences the performance of convolutional neural network. But this condition was already clearly stated and can not be avoided as real challenge of our IHCR development for Balinese script on palm leaf manuscripts. Some ancient characters are not frequently found in our collection of palm leaf manuscripts.

Table I. Recognition rate from all schemes of experiment

No	Method	Feature Dim.	Classifier	Recog. Rate %
1.	Histogram Projection (Binary)	100	SVM	26,313
2.	Celled Projection (Binary)	500	SVM	49,9414
3.	Celled Projection (Binary)	500	k-NN	76,1632
4.	Distance Profile (Binary)	200	SVM	40,1277
5.	Distance Profile (Binary)	200	k-NN	58,947
6.	Distance Profile (Skeleton)	200	SVM	36,7653
7.	Crossing (Binary)	100	SVM	15,0007
8.	Zoning (Binary)	205	SVM	50,6451
9.	Zoning (Binary)	205	k-NN	78,5351
10.	Zoning (Skeleton)	205	SVM	41,848
11.	Zoning (Grayscale)	205	SVM	52,4176
12.	Zoning (Grayscale)	205	k-NN	66,128
13.	Gradient Feature (Gray)	400	SVM	60,0417
14.	Gradient Feature (Gray)	400	k-NN	72,5792
15.	Moment Hu (Gray)	56	SVM	33,481
16.	Moment Hu (Gray)	56	k-NN	33,481
17.	HoG (Gray)	1984	SVM	71,2759
18.	HoG (Gray)	1984	k-NN	84,3477
19.	NPW (Binary)	100	SVM	51,388
20.	NPW (Gray)	100	SVM	54,1249
21.	Kirsch (Gray)	100	SVM	62,4528
22.	HoG with Zoning (Gray)	1984	SVM	69,6859
23.	HoG with Zoning (Gray)	1984	k-NN	83,5006
24.	NPW-Kirsch (Gray)	400	SVM	63,5736
25.	NPW-Kirsch (Gray)	400	k-NN	76,7105
26.	HoG on Kirsch edge (Gray)	1984*4	k-NN	82,0931
27.	HoG + NPW-Kirsch (Gray)	1984+400	k-NN	84,7517
28.	Zoning + Celled Projection (Binary)	205+500	k-NN	77,701
29.	HoG + NPW-Kirsch (Gray) + Zoning (Binary)	1984+400+205	k-NN	85,1557
30.	Convolutional Neural Network			84,3086

V. CONCLUSIONS AND FUTURE WORKS

Balinese scripts on palm leaf manuscripts offer a real new challenge in IHCR development. We present our experimental study on feature extraction methods for character recognition of Balinese script on palm leaf manuscripts by investigating 10 feature extraction methods for IHCR. We proposed the proper and robust combination of feature extraction methods to increase the recognition rate. Our study shows that the recognition rate can be significantly increased by applying NPW features on four directional Kirsch edge images. And the use of NPW on Kirsch features in combination with HoG features and Zoning method can increase the recognition rate up to 85%, and it still slightly better than using the convolutional neural network. For our future works, we will investigate more specific curve and curvature based features to achieve better recognition rate of Balinese script IHCR.

³ <http://www.primaresearch.org/tools/Aletheia>

ACKNOWLEDGMENT

The authors would like to thank Museum Gedong Kertya, Museum Bali, and all families in Bali, Indonesia, for providing us the samples of palm leaf manuscripts, and the students from Department of Informatics Education and Department of Balinese Literature, Ganesha University of Education for helping us in ground truthing process for this experiment. This work is also supported by the DIKTI BPPLN Indonesian Scholarship Program and the STIC Asia Program implemented by the French Ministry of Foreign Affairs and International Development (MAEDI).

REFERENCES

- [1] R. Chamchong, C.C. Fung, K.W. Wong, Comparing Binarisation Techniques for the Processing of Ancient Manuscripts, in: R. Nakatsu, N. Tosa, F. Naghdy and all, K.W. Wong, P. Codognot (Eds.), *Cult. Comput.*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2010: pp. 55–64.
- [2] Chun Che Fung, R. Chamchong, A Review of Evaluation of Optimal Binarization Technique for Character Segmentation in Historical Manuscripts, in: *IEEE*, 2010: pp. 236–240. doi:10.1109/WKDD.2010.110.
- [3] M.W.A. Kesiman, S. Prum, J.-C. Burie, J.-M. Ogier, An Initial Study On The Construction Of Ground Truth Binarized Images Of Ancient Palm Leaf Manuscripts, in: 13th Int. Conf. Doc. Anal. Recognit. ICDAR, Nancy, France, 2015.
- [4] M.W.A. Kesiman, S. Prum, I.M.G. Sunarya, J.-C. Burie, J.-M. Ogier, An Analysis of Ground Truth Binarized Image Variability of Palm Leaf Manuscripts, in: 5th Int. Conf. Image Process. Theory Tools Appl. IPTA 2015, Orleans, France, 2015: pp. 229–233.
- [5] A. Aggarwal, K. Singh, K. Singh, Use of Gradient Technique for Extracting Features from Handwritten Gurmukhi Characters and Numerals, *Procedia Comput. Sci.* 46 (2015) 1716–1723. doi:10.1016/j.procs.2015.02.116.
- [6] S. Kumar, Neighborhood Pixels Weights-A New Feature Extractor, *Int. J. Comput. Theory Eng.* (2009) 69–77. doi:10.7763/IJCTE.2010.V2.119.
- [7] N. Arica, F.T. Yarman-Vural, Optical character recognition for cursive handwriting, *IEEE Trans. Pattern Anal. Mach. Intell.* 24 (2002) 801–813. doi:10.1109/TPAMI.2002.1008386.
- [8] M. Blumenstein, B. Verma, H. Basli, A novel feature extraction technique for the recognition of segmented handwritten characters, in: *IEEE Comput. Soc.*, 2003: pp. 137–141. doi:10.1109/ICDAR.2003.1227647.
- [9] M. Bokser, Omnidocument technologies, *Proc. IEEE.* 80 (1992) 1066–1078. doi:10.1109/5.156470.
- [10] M. Zahid Hossain, M. Ashraf Amin, Hong Yan, Rapid Feature Extraction for Optical Character Recognition, *CoRR*. abs/1206.0238 (2012). <http://arxiv.org/abs/1206.0238>.
- [11] Y. Fujisawa, Meng Shi, T. Wakabayashi, F. Kimura, Handwritten numeral recognition using gradient and curvature of gray scale image, in: *IEEE*, 1999: pp. 277–280. doi:10.1109/ICDAR.1999.791778.
- [12] Z. Jin, K. Qi, Y. Zhou, K. Chen, J. Chen, H. Guan, SSIFT: An Improved SIFT Descriptor for Chinese Character Recognition in Complex Images, in: *IEEE*, 2009: pp. 1–5. doi:10.1109/CNMT.2009.5374825.
- [13] M. Rani, Y.K. Meena, An Efficient Feature Extraction Method for Handwritten Character Recognition, in: B.K. Panigrahi, P.N. Suganthan, S. Das, S.C. Satapathy (Eds.), *Swarm Evol. Memetic Comput.*, Springer Berlin Heidelberg, Berlin, Heidelberg, 2011: pp. 302–309. http://link.springer.com/10.1007/978-3-642-27242-4_35 (accessed March 21, 2016).
- [14] R.J. Ramteke, Invariant Moments Based Feature Extraction for Handwritten Devanagari Vowels Recognition, *Int. J. Comput. Appl.* 1 (2010) 1–5. doi:10.5120/392-585.
- [15] G.S. Lehal, C. Singh, A Gurmukhi script recognition system, in: *IEEE Comput. Soc.*, 2000: pp. 557–560. doi:10.1109/ICPR.2000.906135.
- [16] D. Sharma, P. Jhaji, Recognition of Isolated Handwritten Characters in Gurmukhi Script, *Int. J. Comput. Appl.* 4 (2010) 9–17. doi:10.5120/850-1188.
- [17] K. Singh Siddharth, R. Dhir, R. Rani, Handwritten Gurmukhi Numeral Recognition using Different Feature Sets, *Int. J. Comput. Appl.* 28 (2011) 20–24. doi:10.5120/3361-4640.
- [18] Ashlin Deepa R.N, R.Rajeswara Rao, Feature Extraction Techniques for Recognition of Malayalam Handwritten Characters: Review, *Int. J. Adv. Trends Comput. Sci. Eng. IJATCSE.* 3 (2014) 481–485.
- [19] Ø. Due Trier, A.K. Jain, T. Taxt, Feature extraction methods for character recognition-A survey, *Pattern Recognit.* 29 (1996) 641–662. doi:10.1016/0031-3203(95)00118-2.
- [20] Satish Kumar, Study of Features for Hand-printed Recognition, *Int. J. Comput. Electr. Autom. Control Inf. Eng.* 5 (2011).
- [21] Neha J. Pithadia, Dr. Vishal D. Nimavat, A Review on Feature Extraction Techniques for Optical Character Recognition, *Int. J. Innov. Res. Comput. Commun. Eng.* 3 (2015).
- [22] U. Pal, R. Jayadevan, N. Sharma, Handwriting Recognition in Indian Regional Scripts: A Survey of Offline Techniques, *ACM Trans. Asian Lang. Inf. Process.* 11 (2012) 1–35. doi:10.1145/2090176.2090177.
- [23] U. Pal, B.B. Chaudhuri, Indian script character recognition: a survey, *Pattern Recognition.* 37 (2004) 1887–1899. doi:10.1016/j.patcog.2004.02.003.
- [24] V.. Govindan, A.. Shivaprasad, Character recognition — A review, *Pattern Recognit.* 23 (1990) 671–683. doi:10.1016/0031-3203(90)90091-X.
- [25] C. Clausner, S. Pletschacher, A. Antonacopoulos, Aletheia - An Advanced Document Layout and Text Ground-Truthing System for Production Environments, in: *IEEE*, 2011: pp. 48–52. doi:10.1109/ICDAR.2011.19.