The Thinning Zhang-Suen Application Method in the Image of Balinese Scripts on the Papyrus

Made Sudarma
Computer System and Informatics
Department of Electrical Engineering
Faculty of Engineering, Udayana University
Bukit Jimbaran Campus, Bali, Indonesia

Ni Putu Sutramiani
Magister Program of Elctrical Engineering,
Udayana University Graduate Program
Jl. PB Sudirman, Denpasar 80232
Bali - Indonesia

ABSTRACT

One of the cultural heritage which are now starting to be forgotten is the papyrus library. Balinese papyrus is one of the media to write the ideas from minstrels in ancient times. Currently, many of ancient literature that written in papyrus very difficult to identify because the writings were beginning to rot or fade influenced by age. The introduction of Balinese scripts on papyrus can be done first by performing papyrus digitalization. The papyruses are scanned to become one image file. Further the papyrus image is done by Thresholding since prior to the thinning process it is required a binary image. The application of Thinning Zhang-Suen method is very effective because from the original image with 2 sub iteration until yielding in 1 pixel. The benefits of this research is to improve the quality of the image and further segmented to read papyrus making it easier to read text on the papyrus.

Keywords

Balinese scripts, Papyrus digitalization, Thresholding, Binary, image, Thinning Zhang-Suen

1. INTRODUCTION

Papyrus is a literature work (the paper which has good quality, fine, and beautiful) that often teaches us about life philosophy, religion, and many other knowledge which is very useful for us as a provision of life, so that we as young generation it is very important to understand and take care of papyrus. But the problem in understanding papyrus is that papyrus is considered sacred for Hindus in Bali. Not everyone can read and understand papyrus, considering the difficulty to read Balinese scripts on the papyrus. Besides for longer period the papyrus will be susceptible for damage.

The crucial issue which is papyrus damage encourages the author to preserve by way of papyrus digitalization and also wish to improve the quality of papyrus image. In this study, the author wants to conduct a research about the Application of Thinning Zhang-Suen Method for Balinese scripts on Bali's papyrus.

The research regarding papyrus had been existed about clearing up the noise in papyrus [1]. The other research also investigates about morphology operation in hand written from documents of Canada [5]. From those studies the author tries to make research about the Application of Zhang-Suen Method for Balinese scripts on Bali's papyrus.

2. FUNDAMENTAL THEORY

2.1 Papyrus

One of our ancestor's cultural legacies which have important value is ancient manuscripts. All over Indonesia it is known that there are lots of ancient manuscripts written in various scripts and languages. Most of the manuscripts still stored or owned by ordinary people. The others are existed in central and regional agencies, and traditional institutions.

Papyrus is one of the form of ancient manuscripts existed in Indonesia. Papyrus is found abundant in Bali island, but some is found in Java, Sulawesi (called lontara), and in Lombok. Papyrus was used as a media for writing at the time when people had not found paper. Besides the papyrus there was a material resembled with papyrus used as the media for writing, as in Java was using palm leaves (similar with papyrus), *dluwang* (from barks), and *perkamen* (from goatskin), in Sulawesi they were using bamboo (written in a circle) and rattan, meanwhile in Batak besides papyrus also there was tribak (from barks).

Figure 1. *Wariga Palalubangan* Papyrus, Library of Hindu Dharma Institute Denpasar

2.2 Spelling of Balinese Language

According to the decision of Pasamuhan Agung the spelling of Balinese language with Latin alphabets was adjusted with the spelling of Indonesian language [2]. It meant that the spelling was made as simple as possible and the spelling should be phonetic, it meant that it should precise or close to the actual pronunciation. Based on the things mentioned above, then it was stipulated that the letters used to write Balinese language with Latin alphabets were as mentioned below:

a. Voice scripts (vocal): *a, e, i, u, e, o* (six letters, had been changed and pepet and taling is equalized).

Table 1. Voice Scripts

Number	Balinese script	Latin script
ꦏꦿꦴ	ꦲ ꦲꦴ ꦄ ꦌꦴ	a
ꦌꦴ	ꦈꦴ	ê
ꦇꦴ	ꦈꦴ ꦈꦴ ꦶ ꦷ	i
ꦈꦴ	ꦈꦴ ꦈꦴ ꦈ ꦈ ꦲ	u
ꦲꦴ	ꦩꦲ ꦩꦲ ꦫ ꦫꦴ	e
ꦫꦴ	ꦩꦲꦴ ꦈꦴ ꦈꦴ	o

If Balinese letters in this case are Voice Scripts which exist now is written with Balinese Latin writing it will be as follows:

Description:

- *Wisarga ha* (with the verb) is written the same with *a kara*.

- *Suara dirga (ulu sari, suku ilut, and so on)* is equalized and written with hrasua scripts (short).

- The writing between pepet and taling is equalized (has been adjusted with *EYD*).

- The form of double voice scripts: *ai / ia = e (taling)* and *au / ua = o* (taken the form of code only).

According to the experts, it is stated that the words written with *wisarga ha*, in general are the words originated from Bali (considered as Balinese), meanwhile the words written with a kara are the words not originated from Bali (Ancient Java or Sanskrit). In connection with that, so:

Table 2. The Substitution of Voice Scripts.

Voice script	Replaced
ꦄ	ꦲꦴ
ꦇ	ꦈꦴ
ꦈ	ꦈꦴ
ꦲ	ꦩꦲ
ꦇꦴ	ꦩꦲꦴ

b. *Wianjana* Scripts (consonants): *h, n, c, r, k, g, t, m, ng, h, s, w, l, p, d, j, y, ny (18 letters).*

- The number of Balinese Latin scripts is equal to the number of *ha, na, ca, ra, ka, Bali (18 letters).*

- Writing

(Da madu, murdania) does not have a separate form, in contrast to Ancient Javanese Latin scripts. The naming of *wianjana* scripts with the term of dantia and murdania is stated incorrect. The correct one is besides the common names also it is expected to use the term in accordance with the name of *pakadangan aksara* (articulation area).

Table 3. *Wianjana* Scripts

Number	Balinese script	Latin script	Number	Balinese script	Latin script
ꦏ	ꦲ	h / a	ꦏꦴ	ꦭ	l
ꦤ	ꦤ ꦤ	n	ꦏꦏ	ꦩ	m
ꦕ	ꦕ ꦕ	c	ꦏꦒ	ꦒ ꦒ	g
ꦭ	ꦫ	r	ꦏꦧ	ꦧ ꦧ	b
ꦏ	ꦏ ꦏ	k	ꦏꦔ	ꦔ	ng
ꦢ	ꦢ ꦢ ꦢ	d	ꦏꦥ	ꦥ	p
ꦠ	ꦠ ꦠ ꦠ	t	ꦏꦗ	ꦗ	j
ꦱ	ꦱ ꦱ ꦱ	s	ꦏꦪ	ꦪ	y
ꦮ	ꦮ	w	ꦏꦚ	ꦚ	ny

Table 4. The Writing of *Wianjana* Scripts

Script	Meanig	Script name	Script	Meanig	Script name
ꦤ	na-rambat	murdania	ꦧ	ba-kembang	ostia mahaprana
ꦕ	ca-laca	talawia mahaprana	ꦱ	sa-sapa	Murdania
ꦒ	ga-gora	kantia mahaprana	ꦱ	sa-saga	Talawia
ꦠ	ta-tawa	dantia mahaprana	ꦥ	pa-kapal	ostia mahaprana
ꦠ	ta-latik	Murdania	ꦢ	da-madu	dantia mahaprana

c. Pangangge

Pangangge is the symbol which can not stand alone, written by attaching a wianjana script or voice script and influences the way of reading and writing Balinese scripts. There are various types of *pangangge*, among others are voice *pangangge*, *pangangge tengenan* and script *pangangge*.

2.3 Local Adaptive Thresholding

Thresholding process or also called threshold will result in binary image, namely the image which has two grayscale values that is black and white [3]. The threshold process of grayscale image to generate a binary image in general is as follows:

$$G(x,y) = \begin{cases} 1 \ if \ f(x,y) \geq T \\ 0 \ if \ f(x,y) < T \end{cases} \quad\ldots\ldots\ldots\ldots\ldots\ldots (1)$$

With g(x,y) is a binary image of grayscale f(x,y), and T represents threshold value. T value holds a very important role in threshold process. The result quality of binary image is very much dependent on T value used. The method used in this research is Local Adaptive Thresholding In this method, the value of local threshold can be calculated with one of three ways that is:

$$T = \frac{\Sigma_{(x,y)} \ \Sigma_{\in W} \ f(x,y)}{N_W} - C, \ atau,$$

or…………………. (2)

$$T = median \{f(x,y), (x,y) \in W\}$$ ………….. (3)

$$T = \frac{\max \{f(x,y),(x,y)\in W\}+\min \{f(x,y),(x,y)\in W\}}{2}.$$ (4)

With *W* represents the processed blocks, *Nw* represents the amount of pixels at each *W* block, and *C* represents a constant which can be defined freely. If *C = 0*, it means threshold value is equal to the average value of each pixel in the block concerned. The three ways above respectively calculate *T* values by calculating the average values, median, and maximum and minimum average values of the pixels in the window.

2.4 Thinning Zhang-Suen

The simplified morphology can be defined as the form and structure of an object or in another description is also called the setting and relation between parts of an object [4]. In morphology operation using two sets of input namely an image and a kernel. Particularly in morphology, the term of kernel is also called Structuring Elements (SE). SE is a matrix and generally small in size. Structuring elements can be random in size and having axis point (also called origin point/reference point). In this research, the morphology operation used is Thinning Zhang-Suen [4].

Thinning Zhang-Suen is a parallel method consists of two sub-iterations. First iteration serves to erase boundary point of south-east and vertex of north-west while the other one serves to erase boundary point of north-west and vertex of south-east. The rule in Zhang-Suen algorithm is inside of the iteration there are two sub-iterations. The following is sub-iterations presence in Zhang-Suen method:

    1. Sub-iteration 1

        a.   2 < N (P1) < 6

        b.   S(P1) = 1

        c.   P2 * P4 * P6 = 0

        d.   P4 * P6 * P8 = 0

    2. Sub-iteration 2

        a.   2 < N (P1) < 6

        b.   S(P1) = 1

        c.   P2 * P4 * P8 = 0

        d.   P2 * P6 * P8 = 0

3. RESEARCH METHOD

3.1 Data Acquisition

Data acquisition is a data acquisition process from analog data to digital that is from papyrus to become papyrus image by using scanner. Balinese papyrus is scanned first. After yielding a Balinese papyrus image as the result of scanning then the image is stored in the computer and will be processed into preprocessing stage.

3.2 Preprocessing

a. Thresholding

Data acquisition which results in Balinese papyrus image prior to thinning will be followed by thresholding process first. In this process the image will change to be a binary image. Binary image will facilitate morphology process since Balinese scripts presence in papyrus image will be clearer. The method used in the research is Local Adaptive Thresholding. In this method, local threshold value can be calculated with one of three ways which can be seen on (2), (3), and (4). With *W* represents blocks being processed, *Nw* represents the amount of pixels in each of W's blocks and C represents a constant which can be defined freely. If *C = 0*, it means threshold value is equal to the average value of each pixel in the block concerned. All the three ways above respectively calculate *T* value by calculating the average values, median, and maximum and minimum average values from the pixels in the window.

b. Thinning Zhang-Suen

The thresholding result which yielding a binary image, further will be performed a morphology process. In this research it is using morphology operation of Thinning Zhang-Suen. This Zhang-Suen morphology is selected in order to make clearer an object which in this case Balinese scripts with its background and clear up the noises in the papyrus. In this morphology process, the papyrus image will be conducted the iteration process which inside there are two sub-iterations. This process is done pixel by pixel until the end and obtains the result which is needed, the result of this morphology will look clearer because Balinese scripts will look very clear and will be very helpful in segmentation process.

c. The Application of Thinning Zhang-Suen Method

The application of Thinning Zhang-Suen can be exemplified as follows:
- The Matrix of Balinese Script Image
The matrix of script image " " which can be seen in Figure 1. That image which will be Thinning by using Zhang-Suen algorithm.
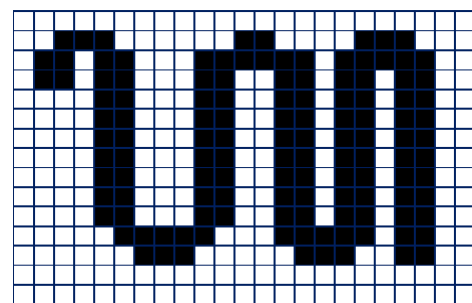

Figure 2. The Matrix of Balinese Script Image

- Zhang-Suen Algorithm
Zhang-Suen algorithm consists of 2 sub-iterations namely as follows:
1. Sub-iteration 1

        a.   2 < N (P1) < 6

        b.   S(P1) = 1

        c.   P2 * P4 * P6 = 0

        d.   P4 * P6 * P8 = 0

2. Sub-iteration 2

a.   2 < N (P1) < 6

b.   S(P1) = 1

c.   P2 * P4 * P8 = 0

d.   P2 * P6 * P8 = 0

First Step:

It provides a flag for the pixels that want to be deleted. The conditions are as follows:

a.   2 <= N(p1) <= 6

b.   S(p1) = 1

c.   p2.p4.p6 = 0

d.   p4.p6.p8 = 0

N(p1)=p2+p3+…+p8+p9

S(p1)=transition 0 to 1

P9	P2	P3
P8	P1	P4
P7	P6	P5

Second Step
a.   For the first step, all the conditions must be fulfilled. If one is not qualified then it should not given a sign / flag. First step is done for all the points in the image. Further the signed pixels are changed to 0.

b.   Delete all the pixels that have been signed.

c. Thinning using Zhang-Suen Algorithm
The image in Figure 2 will be thinning by using Zhang-Suen algorithm. In iteration process, we can take 1 sample to describe iteration process using Zhang-Suen algorithm. Iteration process in the first block, it is not yet getting any change since it is not yet qualified as shown in Figure 3.
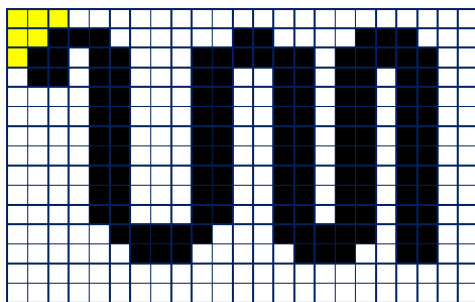
Figure 3. Iteration on the first block

In the first iteration process, it will be getting any change when it is already fulfilling the condition of Zhang-Suen method. As well as the next process so that yielding an image as the thinning result shown in Figure 4.
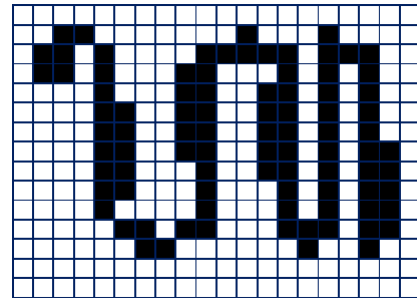
Figure 4. The Image as the result of First Iteration

After going through second iteration process then it generates the image of the thinning result as shown in Figure 5. From the result of Thinning Zhang-Suen operation it can be seen that this research is very suitable to use Zhang-Suen algorithm since it facilitates image segmentation process.
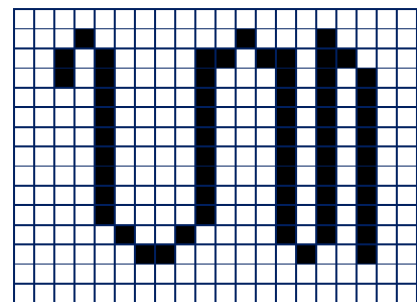
Figure 5. The Image as the result of Second Iteration

4.   RESULTS AND ANALYSIS

The experiments in this study is the process of Zhang-Suen algorithm in image script bali. Here is the process of thinning the Zhang Suen method. The first stage, is a process, where the pixels of the processed part, by using a sub-iteration condition on the first iteration, and the result of the experiment:

-    N(P1) = 3 $(2 \leq N (P1) \leq 6)$ (requirements are met)

-    S (P1) = 1 (condition met)

-    P1 * P4 * P6 = 0 * 1 * 1 = 0 (Terms fulfilled)

-    P4 * P6 * P8 = 1 * 1 * 0 = 0 (Terms fulfilled)

-    Since all conditions are met then be marked on the marked pixels.

The next step performed a similar process, followed by moving the SE pixel by pixel in the input image to get the first iteration final results.

The results of the first iteration of the sub-iteration, the next iteration in the second sub-process done by using the rule of sub-iterations to 2. with the results of iteration:

-    N(P1) = 4 $(2 \leq N (P1) \leq 6)$ (Conditions are met)

-    S (P1) = 1 (Conditions are met)

-    P2 * P4 * P8 = 0 * 1 * 0 = 0 (Terms fulfilled)

-    P2 * P6 * P8 = 0 * 1 * 0 = 0 (Terms fulfilled)

-    Since all conditions are met then the pixels will be marked.

Furthermore, a similar process is done repeatedly until the last pixel that matches only one pixel only.The trial test in this

research is to conduct Zhang-Suen algorithm process in Balinese script image. The following is several results of thinning Zhang-Suen:


Table 5. The Result of Thinning Zhang-Suen

Iteration will continue so that all parts of the script bali into one pixel. Here are the results of Zhang Suen thinning that results in the framework script bali. Based on the data of testing 54 characters bali, Zhang-Suen thinning method successfully processed 47 characters bali image. There are 7 image of Balinese script that did not make it through the process of thining and considered broken, this is because the script is not a good image because of the difficulty of writing letters in the palm of Bali. The results of this test success percentage of 87.04% gain. Here is the test result data Zhang-Suen thinning method on 54 images bali script.

Based on the above framework Balinese script shown that the method is very effective for Zhang Suen morphological processes and generate success percentage of 87.04%. This method can produce a good bali literacy framework that can be used in the image segmentation stage script bali.

5. CONCLUSION

Based on the test result then it can be concluded that the utilization of Zhang-Suen method is able to clarify the curvature of Balinese scripts. The result of this Thinning process is very useful for segmentation process. The further research can be used for papyrus segmentation. From the result of the thinning process it facilitates to papyrus segmentation process

6. ACKNOWLEDGEMENTS

7. REFERENCES

[1] Prapitasari. Luh Putu Ayu, Noise Removal for Ancient Palm Leaf Manuscript of Bali, Konferensi Nasional Sistem Informasi (KNSI) No ISBN 9786029876802, 2012

[2] http://www.babadbali.com/aksarabali/books/ppebb.htm

[3] Putra. IKG Darma, Pengolahan Citra Digital, Andi No ISBN 9789792914436, 2010

[4] Zhang. T. Y., Suen. C. Y., A Fast Parallel Algorithm for Thinning Digital Patterns, ACM Vol 27 No. 3, 1984

[5] H R. Mamatha, K. Srikantamurthy, Morphological Operations and Projection Profiles based Segmentation of Handwritten Kannada Document, International Journal of Applied Information Systems (IJAIS) Vol. 4 No. 5, 2013