An Analysis of Ground Truth Binarized Image Variability of Palm Leaf Manuscripts

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

An Analysis of Ground Truth Binarized Image Variability of Palm Leaf Manuscripts

Made Windu Antara Kesiman[1], Sophea Prum[1], I Made Gede Sunarya[2], Jean-Christophe Burie[1] and Jean-Marc Ogier[1]

[1] Laboratoire Informatique Image et Interaction (L3i) University of La Rochelle France

e-mail: {made_windu_antara.kesiman,sophea.prum,jcburie,jean-marc.ogier}@univ-lr.fr

[2] Laboratory of Cultural Informatics (LCI) Ganesha University of Education Bali Indonesia

e-mail: sunarya@undiksha.ac.id

***Abstract*—As a very valuable cultural heritage, palm leaf manuscripts offer a new challenge in document analysis system due to the specific characteristics on physical support of the manuscript. With the aim of finding an optimal binarization method for palm leaf manuscript images, creating a new ground truth binarized image is a necessary step in document analysis of palm leaf manuscript. But, regarding to the human intervention in ground truthing process, an important remark about the subjectivity effect on the construction of ground truth binarized image has been analysed and reported. In this paper, we present an experiment in a real condition to analyse the existance of human subjectivity on the construction of ground truth binarized image of palm leaf manuscript images and to measure quantitatively the ground truth variability with several binarization evaluation metrics.**

***Keywords*—Ground truth image, Binarization, Variability, Palm leaf manuscript.**

I. Introduction

The collection of palm leaf manuscripts in Southeast Asia attracted the attention of researchers in document image analysis, for example some works on document analysis of palm leaf manuscript from Thailand [1][2]. As a very valuable cultural heritage which contains a wide variety of Asian social cultural life aspects, palm leaf manuscripts offer a new challenge in document image analysis system due to the physical characteristics and conditions of the manuscript. Written on a dried palm leaf by using a sharp pen (which looks like a small knife) and colored with natural dyes, palm leaf manuscript images provide a real challenge in the binarization process to separate text from the background, as the early important step in document image analysis. The palm leaf manuscripts contain discoloured parts and artefacts due to aging and low intensity variations or poor contrast, random noises, and fading. Severals deformations in the character shapes are visibles due to the merges and fractures of the use of nonstandard fonts, varying space between letters, and varying space between lines.

With the aim of finding an optimal binarization method for palm leaf manuscript images, some binarization methods which are already proposed and widely used in document image research community have to be tested and evaluated. To evaluate the performance of binarization methods, two approachs are widely used. The first approach evaluates the binarization methods based on the character recognition rate reached by an OCR system applied on those binarized images [3]. But this approach has been criticized that the binarization method is evaluated in their interaction with other process on document analysis pipeline. The second approach evaluates the binarization methods by comparing pixel-by-pixel the difference between binarized image and a ground truthed binarized image [4][5]. In our case where the OCR system for some specific Southeast Asian alphabets are not available yet, the ground truth binarized image of palm leaf manuscripts has to be created to be able to quantitatively measure and compare the performance of all binarization methods. Therefore, creating a new ground truth binarized image of palm leaf manuscripts is a necessary step in our research.

The construction of ground truth binarized image is normally done totally by human, but several frameworks with a specific tool already proposed semi-automatic method to facilitate a much faster and more precise ground truthing process. Regarding to the human intervention in ground truthing process, an important remark about the subjectivity effect on the construction of ground truth binarized image has been analysed and reported. The work of [6] and [7] analysed the binarization ground truthing and the effect of ground truth on image binarization of DIBCO binarized image dataset [4]. The study stated that the different choice of binarization ground truth affects the binarization algorithm design and the performance can vary significantly depending on the choice of ground truth.

In this paper, we present an experiment in a real condition to analyse the human intervention subjectivity on the construction of ground truth binarized image and to measure quantitatively the ground truth variability of palm leaf manuscript images with different binarization evaluation metrics. This experiment measures the difference between two ground truth binarized images from two different ground truthers. Based on the real survey and experiment on this research, we describe some challenges on the construction of ground truth binarized image for palm leaf manuscripts.

This paper is organized as follow: section II briefly presents the digitization process and sample images of palm leaf manuscript used in this experiment. Section III describes the semi-automatic framework for the construction of ground truth binarized image and also the binarization evaluation metrics. The experimental results are presented in Section IV and conclusions with some prospects for the future work are given in Section V.

II. Digitization process

Our first initial campaign on the digitization of palm leaf manuscript in Southeast Asia is collecting the sample images of the manuscript from Bali, Indonesia. In order to obtain the variety of the manuscript images, the sample images used in this experiment come from 9 different collections (contents), randomly selected from all collections found in Museum of Gedong Kertya, City of Singaraja, Regency of Buleleng, North Bali, Indonesia. A summary of the collection characteristics is listed in Table 1.

For capturing process, we use a Canon EOS 5D Mark III, and set the camera setting as follows: F-stop: f/22 (diafragma), exposure time: 1/50 sec, ISO speed: ISO-6400, focal length: 70 mm, flash: On - 1/64, distance to object: 76 cm, focus: Quick mode - Auto selection On. We designed a black box camera support by wood (see Fig.1) to avoid the irregular lighting/luminance condition and to fits our semi outdoor capturing location. Two additional light are added inside the black box support with White Neon 50 cm 20 watt. From the 9 collections, we

captured 47 pages of palm leaf manuscript. Thumbnail samples of these images are showed in Fig. 2.

Fig. 1. Black box camera support for digitizing process of palm leaf manuscripts.

Table 1. Collection of palm leaf manuscripts from Museum Gedong Kertya, Singaraja Bali, Indonesia

No	Content	Reference Code in Museum	Physical size of manuscript	Nb of total pages	Nb of captured pages
1.	Sima Desa Tejakula	IIA-5-789	50 x 4 cm	27	7
2.	Awig-awig Desa Tunju	IIA-10-1534	50 x 4 cm	6	2
3.	Dewa Sasana	IIB-2-180	50 x 4 cm	39	4
4.	Panugrahan Bhatara Ring Pura Pulaki	IIIB-12-306	50 x 4 cm	39	8
5.	Buwana	IIIB-42-1526	50 x 4 cm	18	2
6.	Pambadah	IIIB-45-2296	50 x 4 cm	12	5
7.	Krakah Sang Graha	IIIC-19-1293	50 x 4 cm	22	6
8.	Taru Pramana	IIIC-20-1397	50 x 4 cm	5	7
9.	Siwa Kreket	IIIC-23-1506	50 x 4 cm	8	6
			TOTAL		47 pages

Fig. 2. Sample images of palm leaf manuscripts from Museum Gedong Kertya, Singaraja Bali, Indonesia

III. Construction of Ground Truth Binarized Image and Binarization Evaluation Metrics

The first part of this experiment is the construction of ground truth binarized image of palm leaf manuscripts. The variability of the ground truth binarized image is then measured by several binarization evaluation metrics.

A. Framework for construction of ground truth binarized image

In this experiment, we adopted a semi-automatics framework for construction of ground truth binarized image which were used to build the database in DIBCO competition series, as illustrated in Fig. 3 [3]. It consists of several steps: initial binarization process, skeletonization of the characters, character edge detection, manual correction of skeleton, and estimation of ground truth image by a conditional dilation. This framework needs a first initial binarization method to optimally separate text from the background, and it provides a first binary image for the skeletonizing process of the characters on manuscript. The human intervention is done by a user interaction process on manual correcting procedure based on character edges (detected using Canny algorithm [8]) and character skeletons. After the manual correction process, in this framework, the final estimated ground truth binarized image is then automatically constructed by dilating the corrected skeleton image, constrained by the character edges. The skeleton is dilated until half of the Canny edges intersect each binarized component of the dilated skeleton.

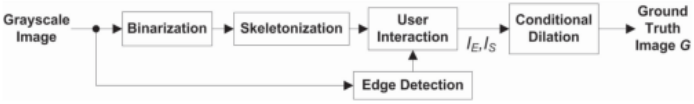

Fig. 3. Ground truth construction procedure used for DIBCO series [3].

For our experiment, in order to measure the variability of human subjectivity in our ground truth creation, we did not apply any initial binarization and skeletonization methods. The skeletonization process is completely performed by human. The skeleton drawn manually by user is dilated until Canny edges intersect each binarized component of the dilated skeleton in a ratio of 0.1. This value of minimal ratio between number of pixels in intersection of Canny edge and number of pixels of the dilated skeleton is found based on our empirical experiment and observation on the thickness of the character stroke in our manuscripts.

B. Binarization evaluation metrics

As presented in [6], three metrics of binarization evaluation which are used in the DIBCO 2009 contest [4] are used in this analysis to measure the difference between two ground truth binarized images from two different ground truthers.

F-Measure (FM)

FM is defined from Recall and Precision.

$$\mathrm{Re}call = \frac{TP}{FN+TP} * 100 \quad (1)$$

$$\mathrm{Pr}ecision = \frac{TP}{FP+TP} * 100 \quad (2)$$

TP is defined as true positive, occurs when the image pixel is labeled as foreground and the ground truth is also. FP is defined as false positive, occurs when the image pixel is labeled as foreground but the ground truth is labeled as background. FN is defined as false negative, occurs when the image pixel is labeled as background but the ground truth is labeled as foreground.

$$FM = \frac{2 * \mathrm{Re}call * \mathrm{Pr}ecision}{\mathrm{Re}call + \mathrm{Pr}ecision} \quad (3)$$

A higher F-measure indicates a better match.

Peak SNR (PSNR)

PSNR is calculated from Mean Square Eror (MSE).

$$MSE = \sum_{x=1}^{M} \sum_{y=1}^{N} \frac{(I_1(x,y) - I_2(x,y))^2}{M*N} \quad (4)$$

$$PSNR = 10 * \log_{10}\left(\frac{C^2}{MSE}\right) \quad (5)$$

where C is defined as 1, the difference between foreground and background colors in the case of binary image. A higher PSNR indicates a better match.

The value of FM and PSNR when we assumed the image drawn by the first ground truther as ground truth image will be the same with

the value of FM and PSNR when we assumed the vice versa, the image drawn by the second ground truther as ground truth image.

Negative Rate Metric (NRM)

NRM is defined from negative rate of false negative (NRFN) and negative rate of false positive (NRFP).

$$NR_{FN} = \frac{FN}{FN+TP} \quad (6)$$

$$NR_{FP} = \frac{FP}{FP+TN} \quad (7)$$

TN is defined as true negative, occurs when both the image pixel and ground truth are labeled as background. TP, FN, and FP are the same definitions as in F-Measure.

$$NRM = \frac{NR_{FN}+NR_{FP}}{2} \quad (8)$$

The value of NRM when we assumed the image drawn by the first ground truther as ground truth image will not be the same with the value of NRM when we assumed the image drawn by the second ground truther as ground truth image. In this case, we calculated two value of NRM: NRM1 and NRM2. A lower NRM indicates a better match.

IV. Experiments and Results

A. Scheme diagram of our experiment

For this experiment, we worked with 70 students from Department of Informatics Education, Ganesha University of Education, Singaraja Bali. Each student was asked to trace manually the skeleton of the Balinese character found in palm leaf manuscript image with PixLabeler tool [9]. One student worked with two different images, and one image was ground truthed by two different students. These two manually skeletonized image will be re-skeletonized with Matlab function bwmorph[1] to make sure that the skeleton is one pixel wide for the next process of automatic ground truth estimation with conditional dilation and Canny edge constraint. As mention in sub-section III.A, Figure 4 shows the scheme diagram of our experiment. Figure 5 shows some samples images as the result example of this experiment.

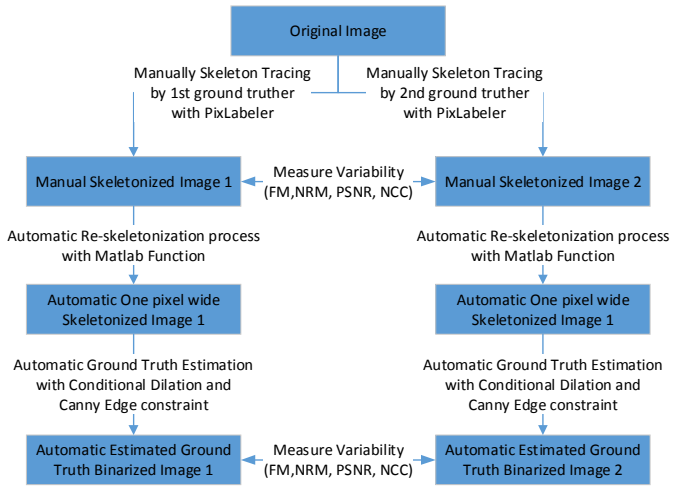

Fig. 4. Scheme diagram of experiment.

[1] http://fr.mathworks.com/help/images/ref/bwmorph.html

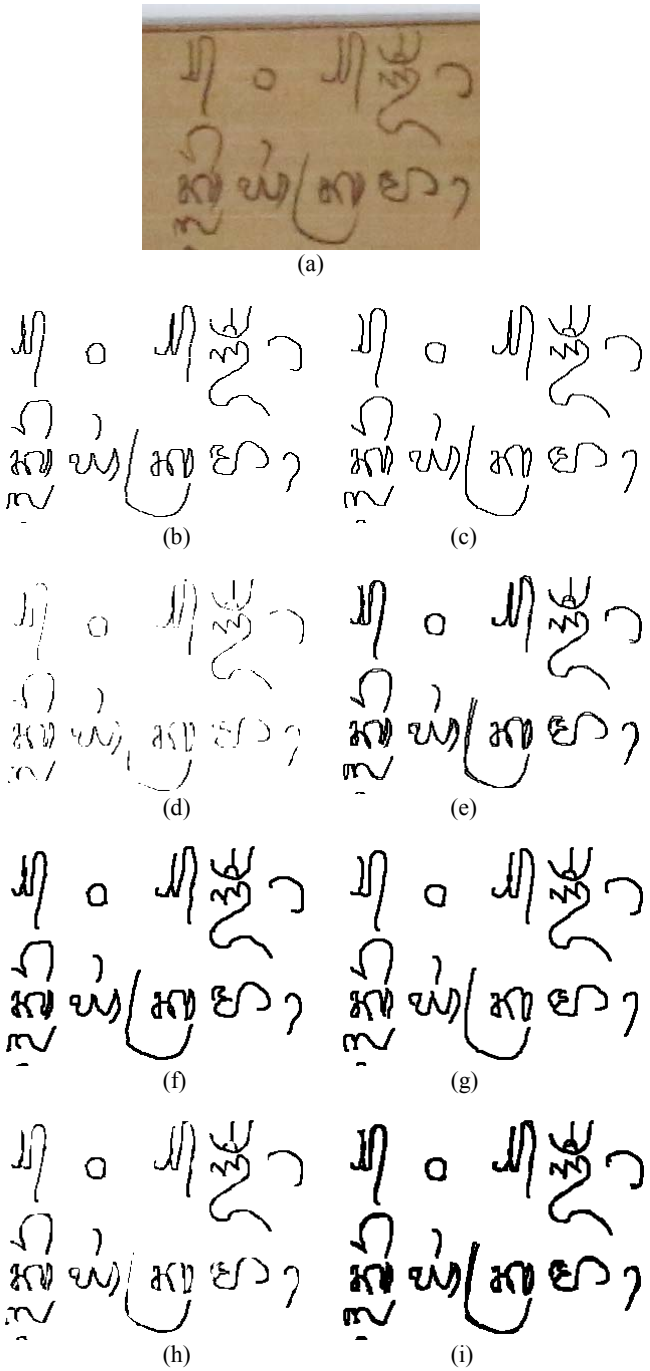

Fig. 5. Example of ground truth binarized image from the experiment: (a) original image, (b) skeletonized image by 1st ground truther, (c) skeletonized image by 2nd ground truther, (d) image intersection between (b) and (c), (e) image union between (b) and (c), (f) estimated ground truth binarized image from (b), (g) estimated ground truth binarized image from (c), (h) image intersection between (f) and (g), (i) image union between (f) and (g)

By observing visually the two skeletonized image created by two different ground truthers, we can see how different the two ground truthers in choosing the trace of the character skeleton. All the broken parts of in image of intersection between two skeletonized images show the different skeleton traces between two ground truthers. And all the double-lined parts in image of union between two skeletonized images show how far the different position of the skeleton traced by two ground truthers.

B. Measure of ground truth variability

First, we measured the variability between two skeletonized ground truthed images manually drawn by two different ground truthers with the three Binarization Evaluation Metrics as presented in sub-section III.B. The results of these 3 measures are presented in Table 2. The wide range between the maximum and the minimum value and also the mean and variance value of all three binarization evaluation metrics from 47 images show that there is a large variability between the ground truthers for each image.

Table 2. Variability between two manually skeletonized ground truthed image.

Comparison metric	FM1	NRM1	NRM2	PSNR
Maximum	58,945	0,371	0,458	60,166
Minimum	14,058	0,209	0,209	26,882
Mean	41,459	0,302	0,303	33,196
Variance	77,764	0,002	0,003	60,083

We then measured the variability between the two ground truth binarized images automatically estimated from two different manually skeletonized image for each image of manuscript (see sub-section III.A). Table 3 illustrates this variability. The wide range between the maximum and the minimum value and also the mean and variance value of all three binarization evaluation metrics from 47 images show that there is still a large variability between the estimated ground truth images for each image.

Table 3. Variability between two ground truthed image automatically estimated from two different manually skeletonized image.

Comparison metric	FM1	NRM1	NRM2	PSNR
Maximum	74,731	0,309	0,446	59,196
Minimum	18,615	0,128	0,130	23,961
Mean	59,556	0,214	0,215	31,110
Variance	89,880	0,002	0,003	61,383

By comparing the value of binarization evaluation metrics between the two manually skeletonized ground truth images (see Table 2) and between the two automatic estimated ground truth images (see Table 3), we can see that the variability of two ground truth images in F-Measure and NRM for all 47 images decreases after the estimation ground truth process. The value of PSNR decreases because the number of different foreground-background pixels between the two estimated ground truth images also increases after the automatic estimation process, not only the number of common foreground pixels from the two estimated ground truth images. Figures 6 to 9 show that the ground truth estimation process tends to decrease the variability between two ground truthers to produce a better match between two ground truth images.

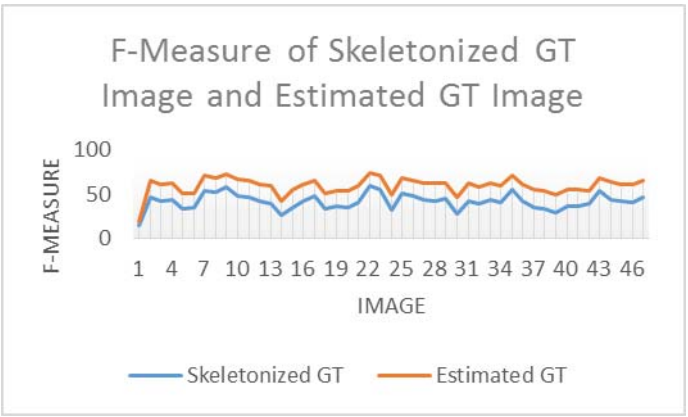

Fig. 6. Comparison of F-Measure between two skeletonized ground truth image and between two estimated ground truth image

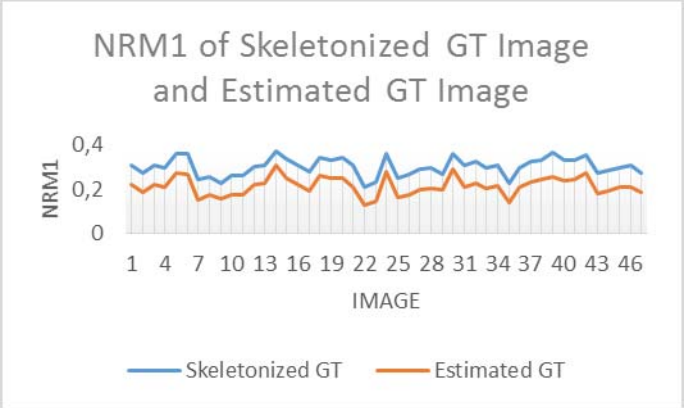

Fig. 7. Comparison of NRM1 between two skeletonized ground truth image and between two estimated ground truth image

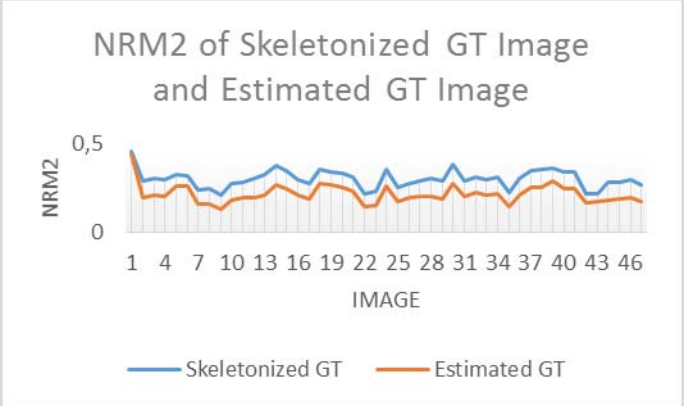

Fig. 8. Comparison of NRM2 between two skeletonized ground truth image and between two estimated ground truth image

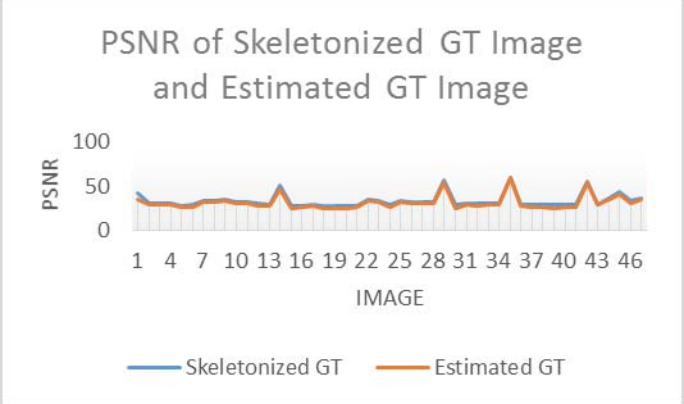

Fig. 9. Comparison of PSNR between two skeletonized ground truth image and between two estimated ground truth image

We also tested and estimated the ground truth binarized image from the union of two skeleton images manually drawn by two differents ground truthers (see exemple in Fig. 5(e)). The variability between this estimated union ground truth image with two other estimated ground truth images from each ground truther is then measured. Tabel 4 and 5 illustrate the results of the comparison metric for all 47 images in experiment. The ground truth image estimated from union of two skeleton images indicates a better match with two other ground truth images from two different ground truthers.

Tabel 4. Variability between ground truth image estimated from union of two skeleton images with ground truth image estimated from the first ground truther

Comparison metric	FM	NRM1	NRM2	PSNR
Maximum	89,758	0,076	0,418	67,095
Minimum	27,823	0,038	0,064	29,854
Mean	80,539	0,066	0,132	37,759
Variance	71,677	0,000	0,003	70,775

Tabel 5. Variability between ground truth image estimated from union of two skeleton images with ground truth image estimated from the second ground truther

Comparison metric	FM	NRM1	NRM2	PSNR
Maximum	94,182	0,090	0,227	65,188
Minimum	66,806	0,025	0,035	30,815
Mean	81,155	0,067	0,129	37,816
Variance	17,054	0,000	0,001	63,464

V. CONCLUSIONS AND FUTUR WORKS

Based on our data survey after the experiment with all ground truthers, we observed and remarked some challenges on the ground truth creation of palm leaf manuscripts as follows: 1. The balinese aphabets found on the manuscripts are not daily used by the ground truthers. Most of the ground truthers learned those alphabets in their elementary school until their junior or senior high school, but they never re-used those alphabets after the classroom learning process. There are some alphabets that they have never seen before. For those kind of alphabets, the ground truthers could not make a smooth and natural trace of the character skeleton. Regarding the variability of ground truth images produced in this experiment, we suggest that this kind of important fact or condition should be always taking into account in every ground truthing process of ancient manuscripts project. 2. The time needed to semi-manually corrected the skeleton of image from an initial automatic method can be much greater than making the skeleton totally manual started from zero. In our first trial experiment, we need 4 until 6 hours to corrected the semi-automatic generated skeleton. It is due to the physical carateristics of the manuscripts which make the binarizing and skeletonizing method do not tend to produce the optimal good skeleton of the characters. We finally decided to make it totally manual, and it takes between 2 until 3 hours to trace the skeleton started from zero.

From the result of this experiment, we proved that the human subjectivity has a great effect in producing a great variability of ground truth binarized image. This phenomen becomes much more visible when we are working on the binarization process of ancient type document or manuscript where the physical characteristics and conditions of the manuscript are not good enough or it is still hard to be ground truthed even by human. The method of binarization evaluation by comparing and measuring pixel-by-pixel with a ground truth binarized image should be re-evaluated to avoid the great bias from human subjectivity. Some other measures should be proposed to evaluate the binarization process of document image of ancient manuscripts. The choice of ground truth data set and the variability within the ground truth should be analyzed quantitatively before the performance measure of any binarization methods. In the case of a manuscript with specific ancient characters, the qualitative observation and validation should also be made by the philologist to guarantee the correctness of the binarized characters on the manuscripts.

ACKNOWLEDGMENT

The authors would like to thank Museum Gedong Kertya in Singaraja Bali, Indonesia, for providing us the sample of palm leaf manuscripts, and the students from Department of Informatics Education, Ganesha University of Education for helping us in binarization 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, and K. W. Wong, "Comparing Binarisation Techniques for the Processing of Ancient Manuscripts," in *Cultural Computing*, vol. 333, R. Nakatsu, N. Tosa, F. Naghdy, K. W. Wong, and P. Codognet, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, pp. 55–64.

[2] Chun Che Fung and R. Chamchong, "A Review of Evaluation of Optimal Binarization Technique for Character Segmentation in Historical Manuscripts," presented at the Third International Conference on Knowledge Discovery and Data Mining, WKDD 2010, 2010, pp. 236–240.

[3] K. Ntirogiannis, B. Gatos, and I. Pratikakis, "Performance Evaluation Methodology for Historical Document Image Binarization," *IEEE Trans. Image Process.*, vol. 22, no. 2, pp. 595–609, Feb. 2013.

[4] B. Gatos, K. Ntirogiannis, and I. Pratikakis, "DIBCO 2009: document image binarization contest," *Int. J. Doc. Anal. Recognit. IJDAR*, vol. 14, no. 1, pp. 35–44, Mar. 2011.

[5] I. Pratikakis, B. Gatos, and K. Ntirogiannis, "ICDAR 2013 Document Image Binarization Contest (DIBCO 2013)," presented at the ICDAR '13 Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, 2013, pp. 1471–1476.

[6] E. H. B. Smith, "An analysis of binarization ground truthing," presented at the DAS '10 Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, 2010, pp. 27–34.

[7] E. H. B. Smith and C. An, "Effect of 'Ground Truth' on Image Binarization," presented at the DAS '12 Proceedings of the 10th IAPR International Workshop on Document Analysis Systems, 2012, pp. 250–254.

[8] J. Canny, "A Computational Approach to Edge Detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. PAMI-8, no. 6, pp. 679–698, Nov. 1986.

[9] E. Saund, J. Lin, and P. Sarkar, "PixLabeler: User Interface for Pixel-Level Labeling of Elements in Document Images," presented at the ICDAR '09. 10th International Conference on Document Analysis and Recognition, 2009, pp. 646–650.