Journal of
Electronic Imaging

JElectronicImaging.org

Southeast Asian palm leaf manuscript images: a review of handwritten text line segmentation methods and new challenges

Made Windu Antara Kesiman
Dona Valy
Jean-Christophe Burie
Erick Paulus
I. Made Gede Sunarya
Setiawan Hadi
Kim Heng Sok
Jean-Marc Ogier

Made Windu Antara Kesiman, Dona Valy, Jean-Christophe Burie, Erick Paulus, I. Made Gede Sunarya, Setiawan Hadi, Kim Heng Sok, Jean-Marc Ogier, "Southeast Asian palm leaf manuscript images: a review of handwritten text line segmentation methods and new challenges," *J. Electron. Imaging* **26**(1), 011011 (2016), doi: 10.1117/1.JEI.26.1.011011.

Southeast Asian palm leaf manuscript images: a review of handwritten text line segmentation methods and new challenges

Made Windu Antara Kesiman,[a,b,*] Dona Valy,[c,d] Jean-Christophe Burie,[a] Erick Paulus,[e] I. Made Gede Sunarya,[b] Setiawan Hadi,[e] Kim Heng Sok,[c] and Jean-Marc Ogier[a]
[a]University of La Rochelle, Laboratoire Informatique Image Interaction (L3i), Avenue Michel Crépeau 17042, La Rochelle Cedex 1, France
[b]University of Pendidikan Ganesha, Laboratory of Cultural Informatics, Jalan Udayana No. 11, Singaraja, Bali, Indonesia
[c]Institute of Technology of Cambodia, Department of Information and Communication Engineering, Pochentong Boulevard, BP 86, Phnom Penh, Cambodia
[d]Université Catholique de Louvain, Institute of Information and Communication Technologies, Electronic, and Applied Mathematics, Place du Levant 3, Louvain-la-Neuve 1348, Belgium
[e]University of Padjadjaran, Robotics, Artificial Intelligence, and Digital Image Laboratory, Jalan Raya Bandung Sumedang, KM 21, Jatinangor 45363, Indonesia

Abstract. Due to their specific characteristics, palm leaf manuscripts provide new challenges for text line segmentation tasks in document analysis. We investigated the performance of six text line segmentation methods by conducting comparative experimental studies for the collection of palm leaf manuscript images. The image corpus used in this study comes from the sample images of palm leaf manuscripts of three different Southeast Asian scripts: Balinese script from Bali and Sundanese script from West Java, both from Indonesia, and Khmer script from Cambodia. For the experiments, four text line segmentation methods that work on binary images are tested: the adaptive partial projection line segmentation approach, the A* path planning approach, the shredding method, and our proposed energy function for shredding method. Two other methods that can be directly applied on grayscale images are also investigated: the adaptive local connectivity map method and the seam carving-based method. The evaluation criteria and tool provided by ICDAR2013 Handwriting Segmentation Contest were used in this experiment. *© 2016 SPIE and IS&T* [DOI: 10.1117/1.JEI.26.1.011011]

Keywords: handwritten document image analysis; text line segmentation; palm leaf manuscript images.
Paper 16563SS received Jul. 1, 2016; accepted for publication Oct. 26, 2016; published online Nov. 17, 2016.

1 Introduction

An important entity in a document image is a text line. A text line is normally composed of words that are arranged in such spatial position to represent the reading order of all words of the document in the horizontal direction. The vertical position of some text lines also gives important information about a paragraph, representing the layout of the document. Segmentation of the document image into physical spatial entities such as text lines, words, and characters is often performed prior to the recognition step of an optical character recognition (OCR) system.[1–7] The segmentation-based text recognition method needs prior segmentation processing of the document image into text line segments, word segments, or character segments. In this case, properly extracting the text lines in a document will make the extraction of smaller size entities of the document, such as the words or the characters, easier. Consequently, the performance of the OCR system is greatly influenced by the result of the segmentation process.

Even though some of the text line segmentation methods are already performed very well in a printed document, segmenting the text lines in a handwritten document is obviously challenging. In a handwritten document, the spatial positions of the words and the characters that compose the text lines are typically not in straight horizontal and vertical positions. They are often arranged in a skewed medial axis, and in some documents they form a curved baseline. These irregular conditions of the medial axis and baselines directly and greatly increase the challenge of detecting the separating paths between the text lines in the text line segmentation process. The size variation of the characters and the different spaces between text lines further complicate the text line segmentation task by presenting challenges of touching characters and oversized characters that cover two consecutive text lines. For historical documents that were written in Asian-type scripts, the existence of many diacritics or other smaller sized characters that were written separately above or under the main text line is another challenge. A new collection of handwritten documents that attracts the attention of the researcher in document analysis is the collection of palm leaf manuscripts from Southeast Asia. Some preliminary studies on these collections describe challenges in document analysis for palm leaf manuscript images. This collection includes the primary characteristics of degraded historical documents, such as the low intensity and low contrast of the document, the varying space between letters, and the varying space between lines, the merges, fractures, and other deformations ofcharacter shapes.

Many methods of text line segmentation for handwritten document image have already been proposed.[1,8–12] Some

*Address all correspondence to: Made Windu Antara Kesiman, E-mail: made_windu_antara.kesiman@univ-lr.fr

1017-9909/2016/$25.00 © 2016 SPIE and IS&T

works deal directly with the text line and character segmentation and recognition.[2–4] But most of those methods basically still depend on the binary image of the document. Some other methods use combined information from both binary and grayscale images.[2,3] In this case, a good initial binarization process is required. Unfortunately, for some types of historical document images, e.g., the palm leaf manuscript images from Southeast Asia, the binarization process to separate the ancient text from the background is a challenge.[13–15] A review of the evaluation of optimal binarization techniques for character segmentation in historical manuscripts was presented in Ref. 13. In our previous work,[15] we experimented and compared several alternative well-known binarization algorithms on the palm leaf manuscript images. We showed that those binarization methods do not give a good binary image for palm leaf manuscript images. All methods extract unrecognizable characters on palm leaf manuscripts with noise. Consequently, the text line and character segmentation methods that are based on the binary image will not provide good results for this kind of document image. Some methods for text line or character segmentation directly applied to grayscale images have already been proposed.[4,7,16,17] A survey of text line segmentation methods for historical documents is given in Ref. 18. In this paper, we investigated the performance of six promising text line segmentation methods by conducting comparative experimental studies on the collection of palm leaf manuscript images.

This paper is organized as follows: Sec. 2 gives a brief description of the collection of palm leaf manuscripts, the ground truthing process, and the challenges for text line segmentation. Section 3 presents the detailed description of the text line segmentation methods that are investigated in our experimental studies. The results and evaluations of the experimental studies are presented in Sec. 4. Conclusions with some prospects for future works are given in Sec. 5.

2 Palm Leaf Manuscripts

This section describes the corpus collection of palm leaf manuscript images that is used in our experimental studies, the ground truth construction process, and the challenges in text line segmentation tasks.

2.1 *Collection of Palm Leaf Manuscript Images*

Our corpus is comprised of sample images of palm leaf manuscripts from three different scripts: Balinese script from Bali, Sundanese script from West Java, both from Indonesia, and Khmer script from Cambodia. For the collection from Bali, to obtain a large variety of manuscript images, the sample images have been collected from 23 different collections (contents), which come from five different locations (regions): two museums and three private families. It consists of ten randomly selected collections from Museum Gedong Kertya, City of Singaraja, Regency of Buleleng, North Bali, Indonesia, four collections from manuscript collections of Museum Bali, City of Denpasar, South Bali, seven collections from the private family collection from Village of Jagaraga, Regency of Buleleng, and two collections from private family collections from Village of Susut, Regency of Bangli, and Village of Rendang, Regency of Karangasem. From those 23 collections, we selected 35 pages of palm leaf manuscript. For the collection from West Java, 12 images were randomly selected from a collection of 64 pages manuscripts about the story of Ramayana, found in Situs Kabuyutan Ciburuy, region of Garut. These manuscripts are estimated to have been written in the 15th century. Most of the manuscripts consist of four text lines. For the collection from Cambodia, 43 original document images of palm leaf manuscripts were randomly selected from Ecole française d'Extrême-Orient database. A summary of the collection is listed in Table 1. Figure 1 shows the sample images from the three different palm leaf manuscripts.

Table 1 Summary of the palm leaf manuscript collection.

Collection	Number of pages	Total number of text lines
Collection from Bali, Indonesia	35 pages	140 text lines
Collection from Sunda, West Java, Indonesia	12 pages	46 text lines
Collection from Cambodia	43 pages	191 text lines
Total	90 pages	377 text lines

2.2 *Ground Truth Construction*

For the manuscripts from West Java, the binary ground truth images were manually generated using PixLabeler.[19] The text line segmentation ground truth data were then generated by hand based on the binary ground truth images. For the manuscripts from Bali, the binary ground truth images were created with a semiautomatic scheme.[14,15] It used a specific semilocal binarization scheme to overcome the ground truth creation difficulty on degraded and low quality palm leaf manuscript images. This scheme is used as the initial binarization process for the semiautomatic framework for the construction of ground truth binarized images. This framework is based on the one used to build the database in DIBCO competition series.[20] Figure 2 shows some samples of binary ground truth images of our palm leaf manuscript collections. The text line segmentation ground truth data were also generated by hand based on the binary ground truth images. For the manuscript from Cambodia, a local thresholding method is applied, and a median filter is used to remove noises caused by isolated pixel. Then the resulted image is superimposed on the original image, and strokes are traced manually one character at a time. A stylus with a pressure sensitive tip is used to maintain the variation of width of each stroke (Fig. 3). A semiautomatic tool is used to construct line segmentation ground truths from the binary ground truth images. A set of initial text line midpoints is generated and positioned at the estimated locations on the binary images (Fig. 4). Those points are then moved manually to separate and assign connected components to their correct text lines. Touching components that spreads over multiple lines are marked, and the vertical cutting locations are also noted.

2.3 *Challenges in Text Line Segmentation Tasks*

Due to its specific characteristics, palm leaf manuscripts provide new challenges in document analysis. Usually, palm leaf manuscripts are of poor quality since the documents have degraded over time due to bad storage conditions.

Fig. 1 From top to bottom: Sample images of palm leaf manuscripts from Khmer, Sunda, and two images from Bali.

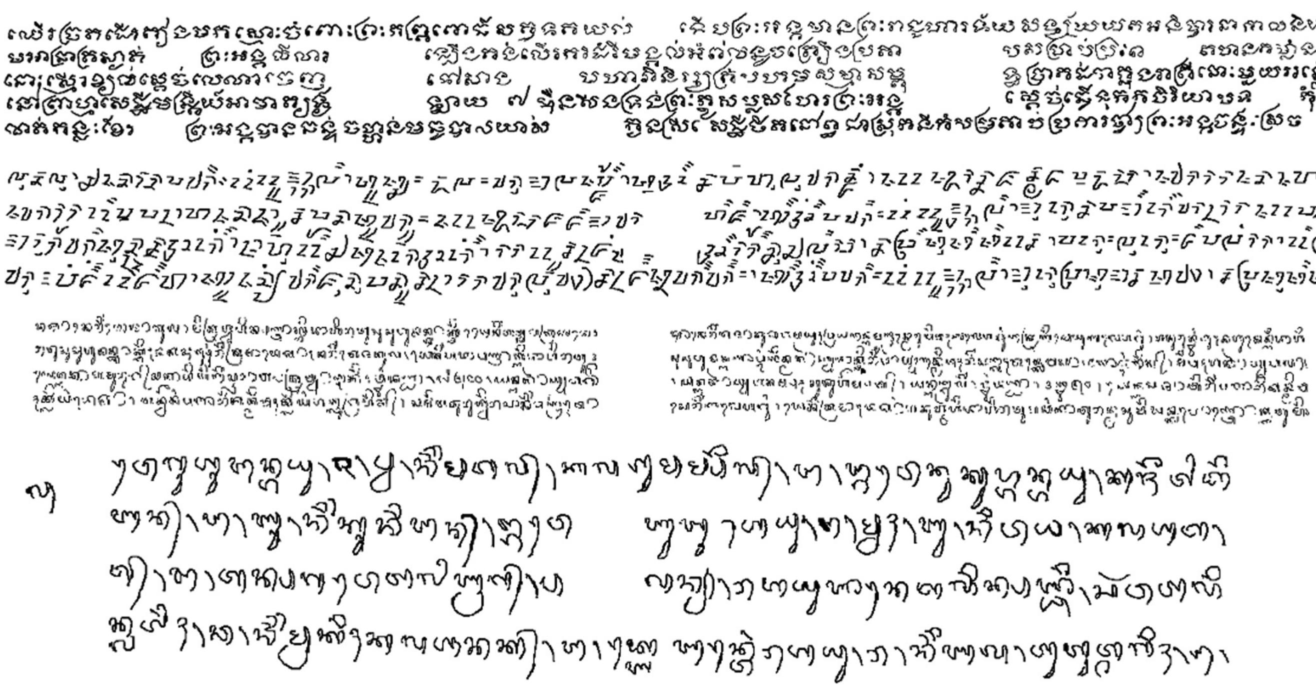

Fig. 2 From top to bottom: Ground truth binary images of palm leaf manuscript of Fig. 1.

Natural materials from palm leaves certainly cannot fight against time; therefore, the process of digitizing and indexing palm leaf manuscripts is very important. In general, the palm leaf manuscripts contain discolored parts and artifacts due to aging and low intensity variations or poor contrast, random noises, and fading.[15] 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. The palm leaf manuscripts contain some obstacles for line segmentation, e.g., skewed and fluctuating text lines and irregularity in geometrical properties of the line, such as line width, height, and distance between lines.[1] These characteristics provide a suitable challenge for text line segmentation.

3 Methods Investigated

In this section, we describe six promising text line segmentation methods that are used in our experimental studies. Two methods work on the binary images: the adaptive partial projection (APP) line segmentation approach and the shredding

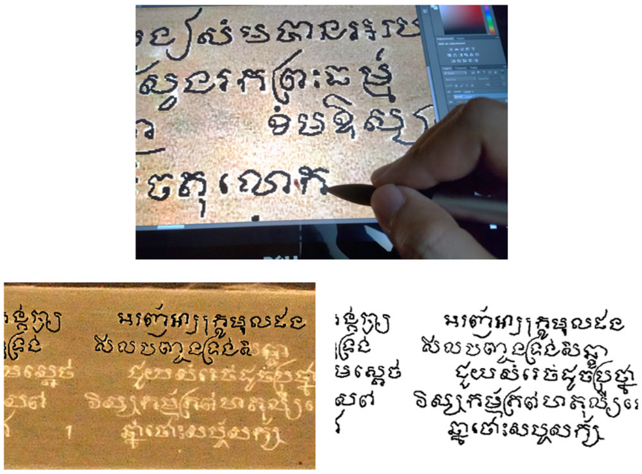

Fig. 3 Example of binary ground truthing for the Khmer manuscripts.

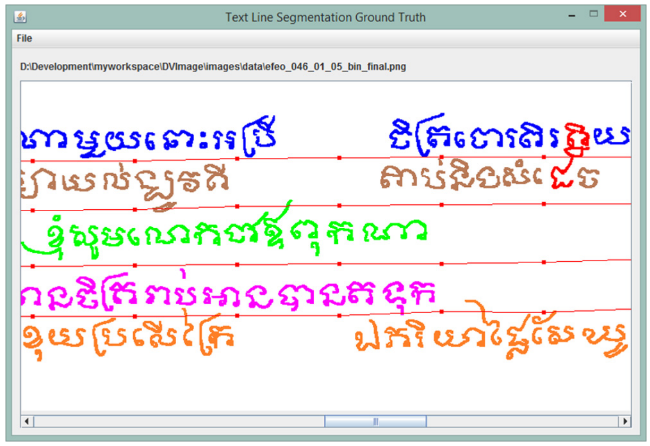

Fig. 4 Example of line segmentation ground truthing.

method. Two other methods can be directly applied on gray-scale images: the adaptive local connectivity map (ALCM) method and the seam carving-based method.

3.1 *Adaptive Partial Projection Line Segmentation Approach*

The APP line segmentation approach was proposed by Chamchong and Fung.[21] It is an improved technique from their previous work[22] in which they adapted modified partial projection and smoothed the histogram with recursion. The technique first constructs the global horizontal projection of the text image to determine the number and average positions of text lines and the average distance between two adjacent lines. These details will be used throughout as reference values. The whole image is then divided into vertical columns. The column size is estimated to be "3*average_char_width" as it is normally the size of a word. The average character width and height are automatically calculated from connected component analysis of each binary image of the manuscript. The smoothed horizontal projection profile is extracted from each column, and the valleys of the profile are considered to be the baselines of that column. For each column, incorrect baselines are removed, and new baselines are inserted based on the referenced values mentioned above. The approach also deals with connected components that spread over multiple lines by recursively dividing the column in which those components belong into two and by traversing up and down from the old baselines until it reaches a more appropriate position. The baselines of all columns are joined together to form separating lines (Fig. 5).

In our experiments, we first calculate the number of lines based on the number of peaks of the global horizontal projection profile. To find the number of peaks, we smooth the projection profile using a moving average filter to remove spurious peaks. The window size of the filter depends on the average height of all connected components in the page. For Khmer and Balinese scripts, we used "avg_char_height/2" as the window size. However, for Sundanese scripts, since they contain lots of small connected components, we increase the window size to "avg_char_height" to improve the result.

3.2 *Shredding Method*

The shredding method was proposed by Nicolaou and Gatos.[1] This technique tries to shred the image into text lines from one side of the image to the other by following the white-most and black-most paths. This approach considers a topological assumption that for each text line, there exists a path from one side of the image to the other that traverses only one text line. The shredding method is applied on the binary image.

In the preprocessing stage, the binary image is blurred with a blurring filter that is based on the size of the estimated letter height from all connected component heights on the binary image of the manuscript. The most frequent letter height found in all connected component heights on the binary image of the manuscript is used. Let I be the binary image, and LH be the estimated letter height. The width of the blurring window is defined as $BW = LH * 8$, and the height of the blurring window is defined as $BH = LH * 0.8$. The blurring image B is defined as

$$B(x, y) = \sum_{i=-BW/2}^{i=BW/2} \sum_{k=-BH/2}^{k=BH/2} I(x + i, y + k).$$

The size of BW and BH is defined in such a way that this operation blurs out the intracharacter and intraword spaces while keeping the spaces between text lines (Fig. 6). A recursive tracer function Tr is finally applied on the blurred image.

$$Tr_{k,B}(1) = k,$$

$$Tr_{k,B}(n+1) = \begin{cases} Tr_{k,B}(n) - 1 \Rightarrow \text{if}: B(n, Tr_{k,B}(n) + BH/2) > B(n, Tr_{k,B}(n) - BH/2) \\ Tr_{k,B}(n) \Rightarrow \text{if}: B(n, Tr_{k,B}(n) + BH/2) = B(n, Tr_{k,B}(n) - BH/2) \\ Tr_{k,B}(n) + 1 \Rightarrow \text{if}: B(n, Tr_{k,B}(n) + BH/2) < B(n, Tr_{k,B}(n) - BH/2) \end{cases}.$$

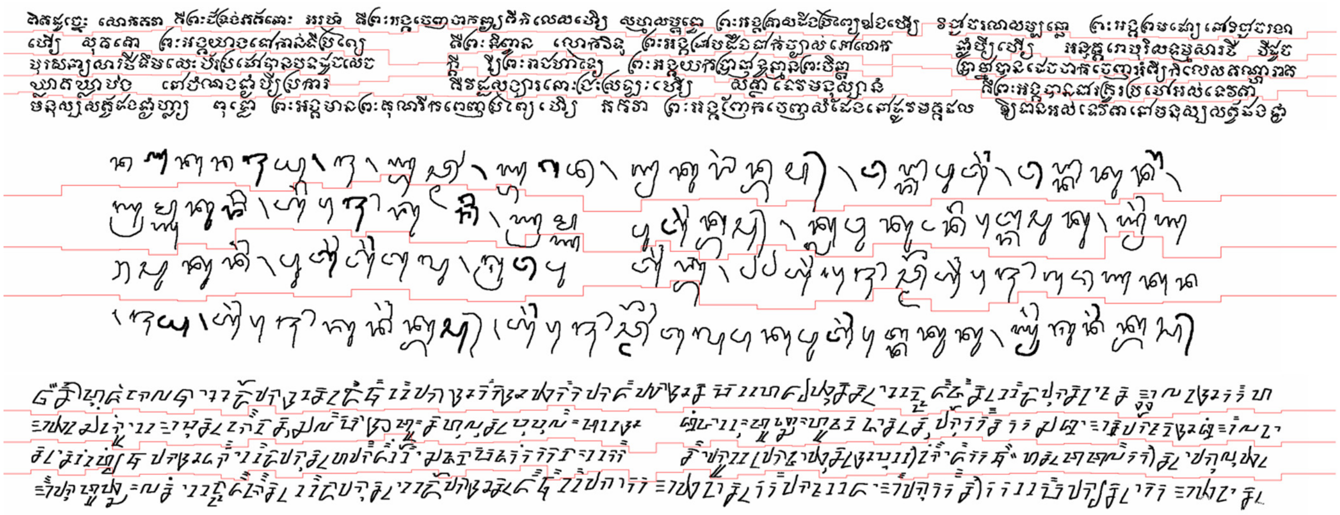

Fig. 5 Some results of the APP approach on Khmer, Balinese, and Sundanese manuscripts.

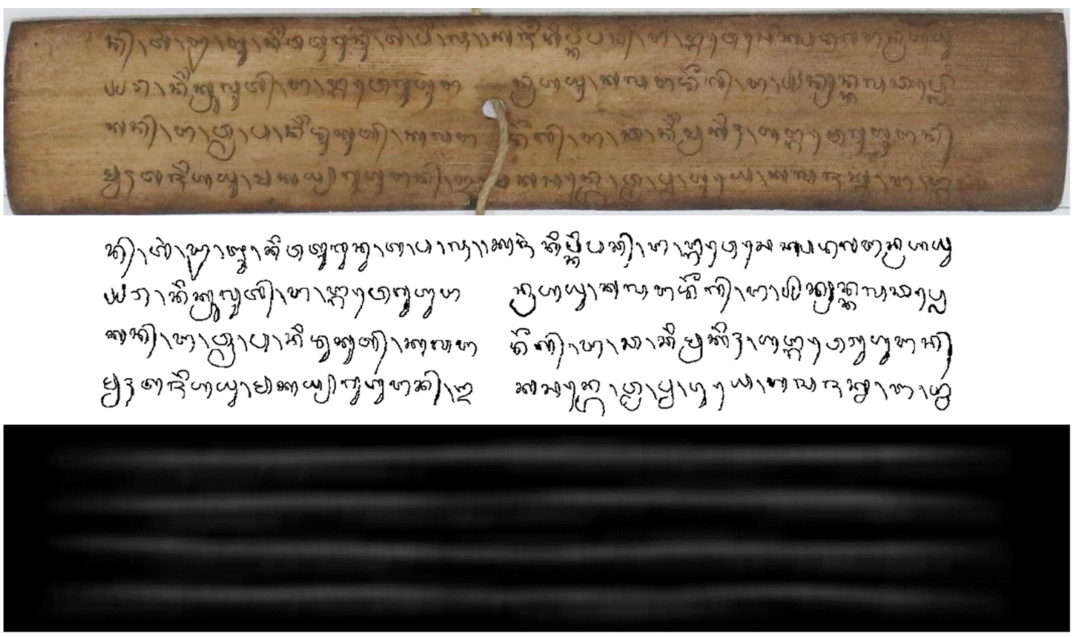

Fig. 6 From top to bottom: original image, binary ground truth image, and blurred image of the shredding technique.

It results in the shredded text line areas (Fig. 7). The text line areas that are smaller than LH^2 were filtered out. To detect the medial axis of text lines (Fig. 8), the same recursive tracer function is applied on the inverted blurred image—$B(x, y)$. The next step consists of assigning each connected component from the binary image input based on the intersection with line areas and line centers or medial axis.

3.3 *Adaptive Local Connectivity Map Method*

The ALCM method was proposed by Shi et al.[7] This method is considered a transform-based method and can be applied directly on grayscale images. It consists of the following steps. First, an ALCM map is generated from a grayscale document using the following transform:[7]

$$\mathrm{ALCM}: f \rightarrow A$$

$$A(x, y) = \int_R f(x, y) G_c(t - x, y) \mathrm{d}t$$

where

$$G_c(x, y) = \begin{cases} 1 \Rightarrow \text{if}: |x| < c \\ 0 \Rightarrow \text{otherwise} \end{cases}.$$

The ALCM transform computes the cumulative intensity by adding up all the intensity values in a certain size of neighborhood of each pixel. The ALCM transformed image is a grayscale image that gives preliminary information about the possible locations of text lines. The second step is the binarization of the ALCM transformed image. It was considered

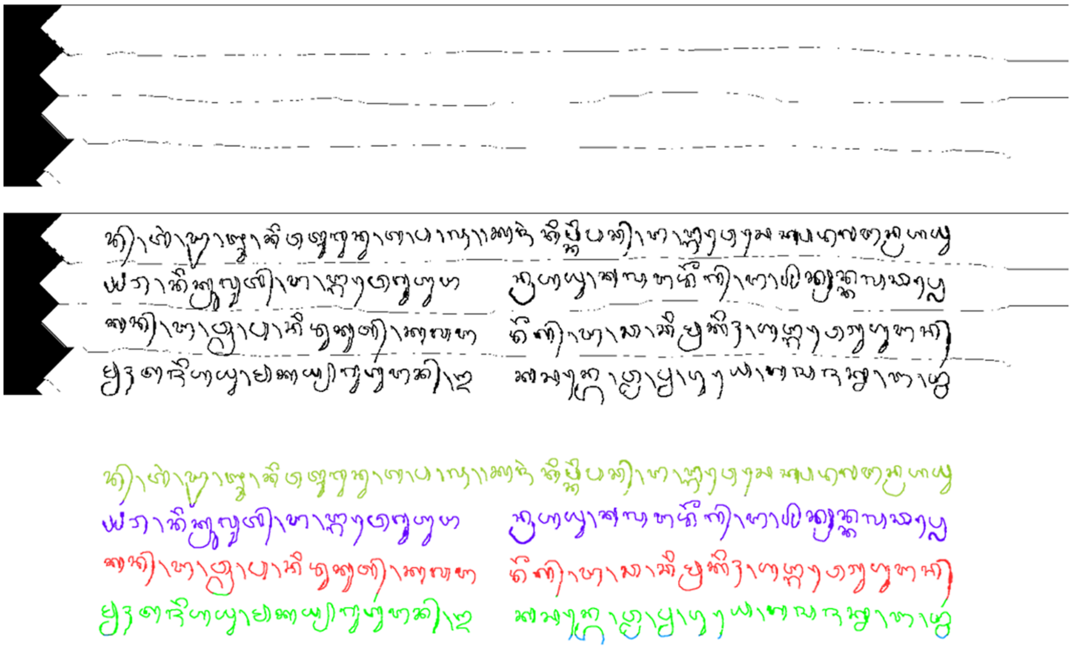

Fig. 7 The shredded text line areas and the text line segments.

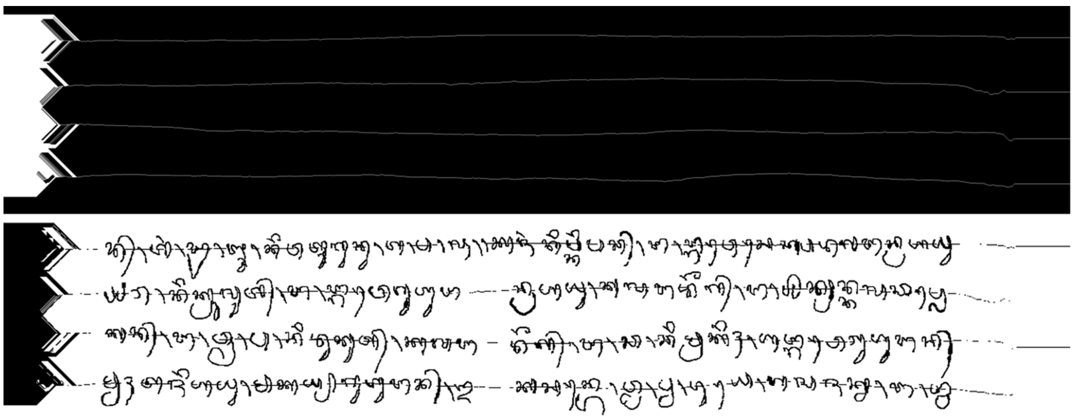

Fig. 8 The detected medial axis of text lines.

that the binarization of the ALCM transformed image is easier than the original grayscale image because it consists of a clear bimodal pixel distribution. The ALCM transformed image is more tolerable for the different binarization algorithms.

After the binarization of the ALCM transformed image, a procedure to filter out the small pieces area is performed. It is based on the statistical size from all areaa in the binary image of the ALCM. The full area of a text line is then generated by filling the preliminary text line area based on their upper and lower profile points. The binarization is finally performed by locally focusing only on areas of text lines. The last step is the connected component mapping and text line component collection. All connected components found in the binarized version of the document are mapped based on the locations of text line areas to make up the text line segments.

In our experimental studies, we computed the cumulative intensity in a neighborhood of size $2c$, where $c = 100$, as it was suggested to approximate c by the value of three times the average height of text. The scanning process to add up all intensity values was done twice, from left to right and right to left, and all cumulative intensities were rescaled to range from 0 to 255 to produce a grayscale ALCM transformed image. For the next step of binarization of the ALCM transformed image, we found that it is still difficult to binarize the ALCM transformed image from our manuscripts. The local adaptive binarization method of Sauvola and Pietikäinen[23] was applied with default values of $k = 0.1$, $R = 128$, and $\text{block\_size} = 50 \times 50$. In most cases, the ascender and descender parts of characters make it difficult to separate two consecutive text line areas. Consequently, the preliminary position of text lines on the document was still hardly extracted (Fig. 9). We finally filtered out the small areas whose the height was less than 20 pixels (the half of the estimated text height) or with a width less than 100 pixels (half of the estimated word width). To make up the text line segments, we mapped all connected components found in the binary ground truth image of the document, based on the intersection of the text line areas. If a connected component intersected more than one text line area, it was assigned to

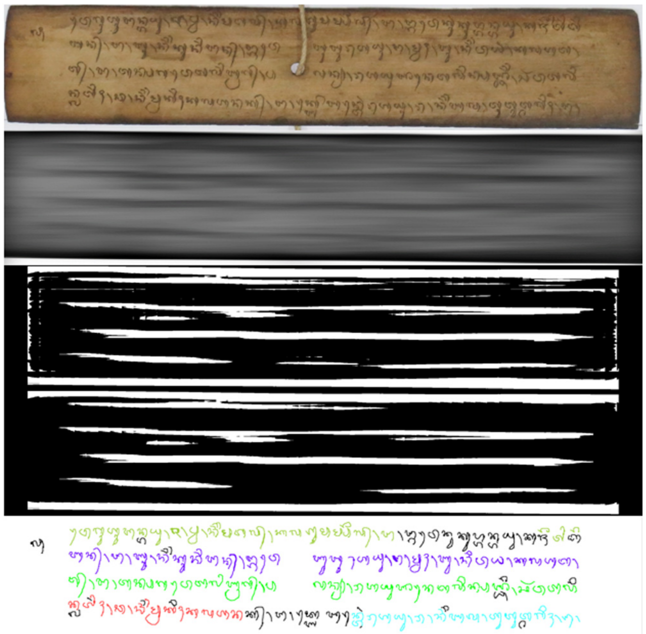

Fig. 9 From top to bottom: original image, ALCM transformed image, the binarized image of ALCM transformed image, the final binarized image of ALCM transformed image after filtering process of small areas, and the text line segmentation.

the text line with the most intersection area. We did not perform any further postprocessing task to investigate the performance of ALCM transform.

3.4 *Seam Carving-Based Method*

The seam carving-based method determines the segmentation path based on a defined seam map that is generated from a given energy function. Some schemes for text line segmentation based on the seam carving method have already been proposed.[10,24–27] By using the same basic idea for the seam carving method, the different schemes differ only in the choice of the energy function and their proposed preprocessing and postprocessing steps.

Saabni and El-Sana[25] proposed a language-independent text lines extraction using seam carving. Their method is applied to the binary image. In the preprocessing step, they calculated the average height of connected components and classify them (according to their height) into four categories: additional strokes, ordinary average components, large connected components, and vertically touching components. In this method, the seam carving concept is applied to find the medial axis of the text lines. The signed distance transform is used as energy function to compute the energy map. The dynamic programming is finally used to compute the minimal cost seam that passes from the left side of the image to the right side. The collection of connected components is then performed by applying some rules based on the position of the extracted medial axis of the text lines.

A new approach for text line segmentation based on the seam carving method that works directly on grayscale document images has been also proposed.[24,26] In this method, two types of seams are calculated: the medial seams and separating seams. Stoll et al.[27] used the simple gradient magnitude function as the energy function. Arvanitopoulos and Susstrunk[24] used the derivative image of the grayscale manuscript as the energy function, as follows

$$E_{i,j} = \left| \frac{I^{\sigma}_{i,j+1} - I^{\sigma}_{i,j-1}}{2} \right| + \left| \frac{I^{\sigma}_{i+1,j} - I^{\sigma}_{i-1,j}}{2} \right|,$$

where I^{σ} is the smoothed grayscale image with Gaussian filter of standard deviation σ. For both works, the seam map is generated by the following function:

$$M(i,1) = E(i,1)$$

$$M(i,j) = E(i,j) + \min \begin{cases} M(i-1, j-1) \\ M(i, j-1) \\ M(i+1, j-1) \end{cases}.$$

In the work of Asi et al.,[26] the distance transform is adopted as the energy function to generate the energy map. In their case, the medial seams are determined by local minimum points of the seam map, and the separating seams are determined by maximum points. To generate an accurate energy map and produce robust seams, they used different weights for the horizontal and diagonal distances. The seam map is generated by the following function:

$$M(i,j) = 2 * E(i,j) + \min \begin{cases} \frac{1}{\sqrt{2}} * M(i-1, j-1) \\ 1 * M(i, j-1) \\ \frac{1}{\sqrt{2}} * M(i+1, j-1) \end{cases}.$$

In our experimental studies, we first investigated the performance of the basic seam carving scheme by minimalizing the preprocessing and postprocessing steps. We tested the seam carving method on binary image and grayscale image. We followed the generic basic scheme for the seam carving method, which consists of three steps: calculate the energy map of the image by using an energy function, generate the seam map, and trace the minimal path by following the minimal cost provided by the seam map.

The distance transform is used as the energy function. The distance transform calculates the minimum distance of each pixel on the image from the nearest background pixel. In our experiments, to calculate the distance transform for binary images, we used MATLAB® function *bwdist* with the Euclidean distance metric, and for grayscale images, we used MATLAB® function *graydist* with quasi-Euclidean distance metric. The seam map is generated by using different weights for the horizontal and diagonal distances and by using two passes from left to right and from right to left as proposed by Asi et al.[26] We then calculated the final seam map from the average of two seam maps from two directions. We generated the separating seam paths by tracing the minimum seam cost from both directions. All separating seam paths directly define the text line areas (Fig. 10). We did not perform any further postprocessing task.

Second, as comparison with a more complete seam carving scheme, we also evaluated the seam carving method that is implemented by Arvanitopoulos and Susstrunk.[24] In their implementation, the medial seams are first computed based on the projection profile matching approach. The image is divided into some slices of column, and the Sobel operator is applied to compute the edge of image. For each slice of image, the smoothed horizontal projection profiles are

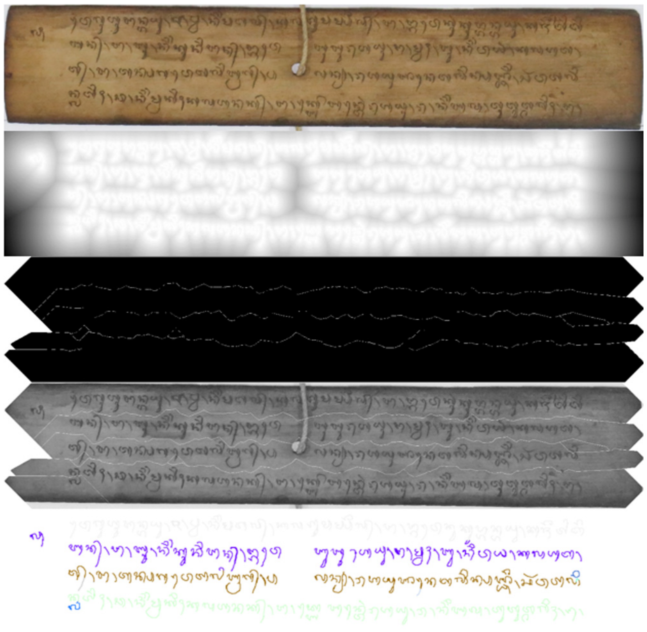

Fig. 10 From top to bottom: original image, energy map, minimum path, and text line segmentation result.

calculated independently. The local maxima locations of the profile in all slices are then detected and connected to create piecewise linear seams that approximate the medial axis of the text lines in the manuscript page. The separating seams are then computed with these medial axis as the constraint. The medial axis force the separating seam to pass between two consecutive text lines. The energy map is calculated from the derivative image of the grayscale image smoothed with Gaussian filter (Fig. 11).

3.5 *Energy Function for Shredding Method*

After analyzing the jumping and the joining path problem on the seam carving method, we have proposed an energy function that can still represent a high energy value in an empty area within one text line. For each pixel, we define the energy value as the number of text (foreground) pixels in an ellipse area that is centered in that pixel. The sizes of the minor axis and the major axis of the ellipse area are based on the approximated average character height and width that are calculated from the connected component analysis. The minor axis of the ellipse area is set with the character height to limit the transfer energy between two text lines, to avoid the jumping and the joining path. The major axis of ellipse area should be set large enough (n times of character width) to be able to transfer the energy from one side to the other side of a text line in an empty text line area.

Let I be a binary image of size ($nb\text{row} \times nb\text{col}$) where the foreground pixel is 1 and the background pixel is 0, and r_{major} and r_{minor} be the size of the major axis and minor axis of an ellipse area. The energy value of a pixel in row x and column y is defined as

$$E(x,y) = \sum_{i=1}^{nb\text{row}} \sum_{j=1}^{nb\text{col}} I(i,j).\text{Ellipse}(i,j,r_{\text{min or}},r_{\text{major}},x,y),$$

where $\text{Ellipse}(i,j,r_{\text{minor}},r_{\text{major}},x,y)$ is a binary function to check whether the pixel $I(i,j)$ is inside the area of an ellipse with minor axis r_{minor}, major axis r_{major} and is centered on the coordinate (x,y), defined as

$$\text{Ellipse}(i,j,r_{\text{minor}},r_{\text{major}},x,y) = \begin{cases} 1, & \text{if}\left\{\frac{[\text{abs}(i-x)]^2}{r_{\text{minor}}^2} + \frac{[\text{abs}(j-y)]^2}{r_{\text{major}}^2}\right\} \le 1 \\ 0, & \text{if} - \text{not} \end{cases}.$$

After generating the image of the energy function, we directly apply the recursive tracing function of the shredding method (see Sec. 3.2). Figure 12 shows that the ellipse energy function can transfer the energy to fill the empty text line areas and then the recursive tracing function of the shredding method can easily separate the text line areas. We did not perform any further postprocessing task to investigate the performance of our proposed energy function for the shredding method.

In our experiments, we used $r_{\text{minor}} =$ average character height for the entire manuscript collection, $r_{\text{major}} = 5 \times$ average character width for the manuscript from Bali and Sunda, and $r_{\text{major}} = 8 \times$ average character width for the Khmer manuscript (Fig. 13). These values have been defined empirically based on the existence of the empty text line area width in the manuscript collection. Khmer manuscripts usually have a wider empty text line area than the manuscripts from Bali and Sunda.

3.6 *A* Path Planning Approach*

The A*PP line segmentation approach has been proposed by Surinta et al.[28] The objective of path planning is to compute the shortest path from a starting point to its destination avoiding obstacles along the way. A* (called A star) is one of the path planning algorithms that minimizes the travel costs between states from the starting state s_1 to the goal state s_n. To solve the line segmentation problem, paths separating text lines need to be traced from the left side (starting state) to right side (goal state) of the text, and the foreground (black) pixels are viewed as obstacles. Due to some handwritten text components from adjacent lines being touched or overlapped, the goal state can be unreachable. A modified A* path-planning technique is proposed to allow the path to pass through such components. The method now works as

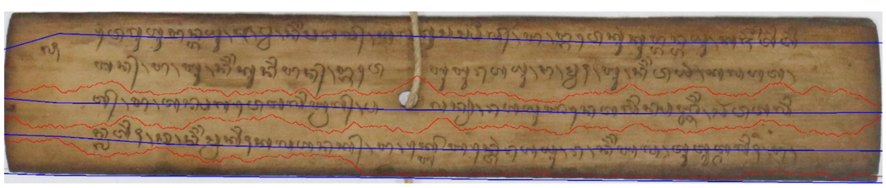

Fig. 11 The medial seams and the separating seams of the manuscript of Fig. 10

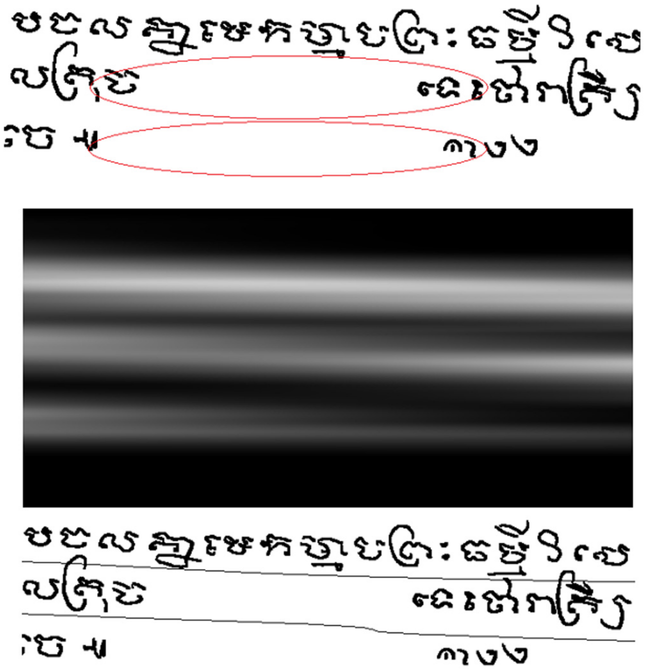

Fig. 12 From top to bottom: Ellipse area in the empty text line area, image of the energy function, and the text line segmentation path generated from the shredding method.

follows. The position of the starting state and the goal state of each path is computed from the valley points of the smoothed y-projection profile histogram. Five cost functions are then combined to determine the traveling cost between states:

- Foreground distance cost functions $D(n)$ and $D(n)^2$ control the path to stay along the gaps between foreground pixels. The two functions are defined as

$$D(n) = \frac{1}{1 + \min[d(n, n_{y_u}), d(n, n_{y_d})]},$$

$$D(n)^2 = \frac{1}{1 + \min\,[d(n, n_{y_u}), d(n, n_{y_d})]^2},$$

where $d(n, n_{y_u})$ and $d(n, n_{y_d})$ are the distances between the state n and the closest foreground pixel in the upward and downward directions, respectively.

- Map-obstacle cost function $M(n)$ gives a penalty if the path has to pass through foreground pixels. $M(n)$ returns 1 if the state n coincides with a foreground pixel, and it returns 0 otherwise.
- Vertical cost function $V(n)$: prevents the path from deviating from the y-position of the starting state and the goal state. The function is defined as

$$V(n) = |n_y - n_y^{\text{start}}|,$$

where n_y and n_y^{start} are the y-position of the current state n and the start state, respectively.

- Neighbor cost function $N(s_i, s_j)$ computes the shortest path between the starting state and the goal state. Like in the standard algorithm, this function returns 14 for diagonal directions, and it returns 10 for other directions in the eight-directional movements.

The combined cost function $C(s_i, s_j)$ is defined as follows:

$$C(s_i, s_j) = c_d D(s_i) + c_{d2} D(s_i)^2 + c_m M(s_i) + c_v V(s_i) + c_n N(s_i, s_j).$$

The parameters c_d, c_{d2}, c_m, c_v, and c_n are tuned empirically.

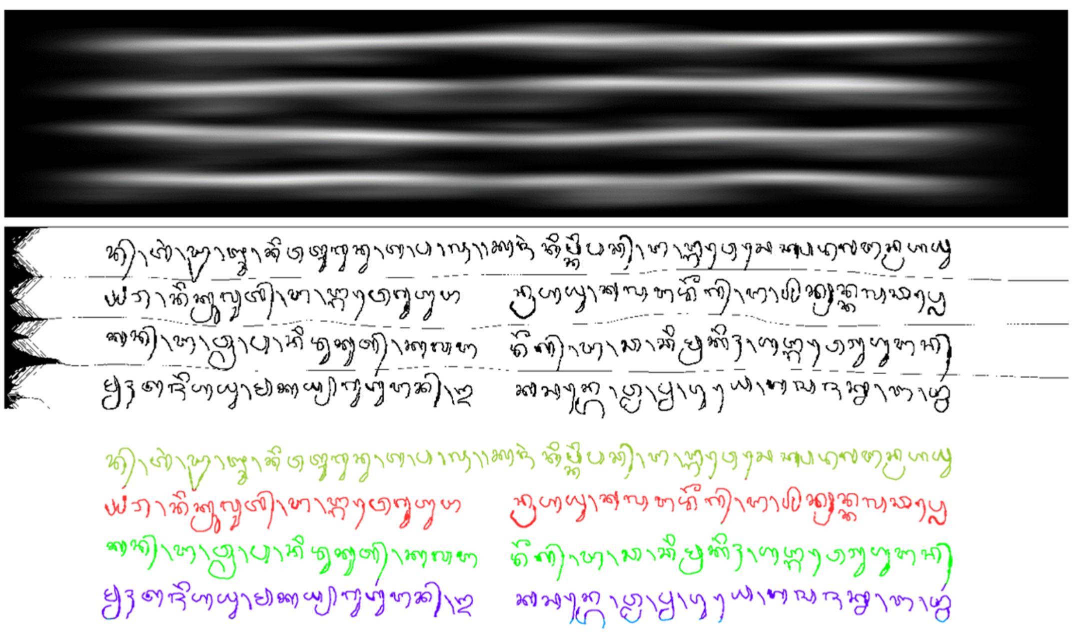

Fig. 13 From top to bottom: Example of results: image of the energy function, the text line segmentation path generated from shredding method, and the text line segmentation result.

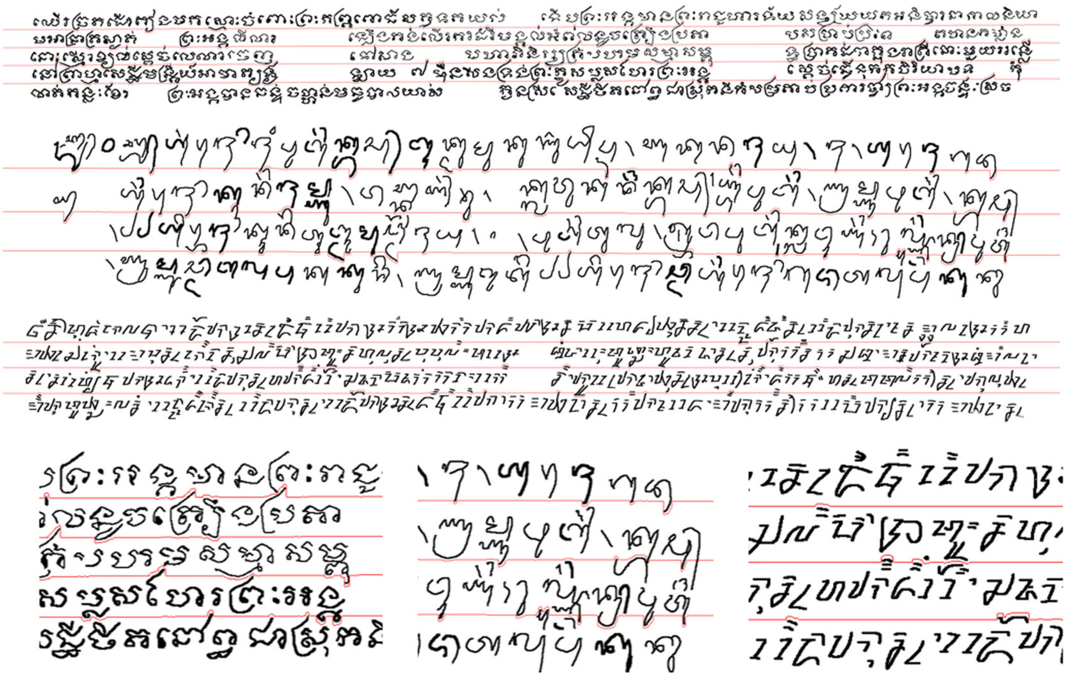

Fig. 14 Example of results of A* path planning approach.

There are two major drawbacks in this approach. The first is that the method assumes the y positions of the starting state and the goal state to be the same. Therefore, it does not work well with documents containing curved or slanted text lines. Another difficulty is caused by the tuning of the coefficient parameters to find the most efficient values to compute the final traveling cost $C(s_i, s_j)$.

In our experiment, the position of the starting state and the goal state are calculated separately to adapt to the skewness of the text lines. The y-positions of the starting states and the goal states are extracted from the y-projection histogram of the first one-third of the document and the last one-third of the document, respectively. The vertical cost function now becomes

$$V(n) = \left| n_y - \left[\left(1 - \frac{n_x}{l} \right) n_y^{\text{start}} + \frac{n_x}{l} n_y^{\text{goal}} \right] \right|,$$

where $l = n_x^{\text{goal}} - n_x^{\text{start}}$ and $(n_x, n_y), (n_x^{\text{start}}, n_y^{\text{start}}), (n_x^{\text{goal}}, n_y^{\text{goal}})$ are the coordinates of the current state n, the starting state, and the goal state, respectively. Normally, n_x^{start} is 0 and n_x^{goal} correspond to the end of the document page, so l is equal to the width of the document. The cost $V(n)$ is now the vertical distance from the current state at position (n_x, n_y) to the slanted line constructed from the two points $(n_x^{\text{start}}, n_y^{\text{start}})$ and $(n_x^{\text{goal}}, n_y^{\text{goal}})$. To improve the execution time, only five directional steps (S, SE, E, NE, E) are computed for the

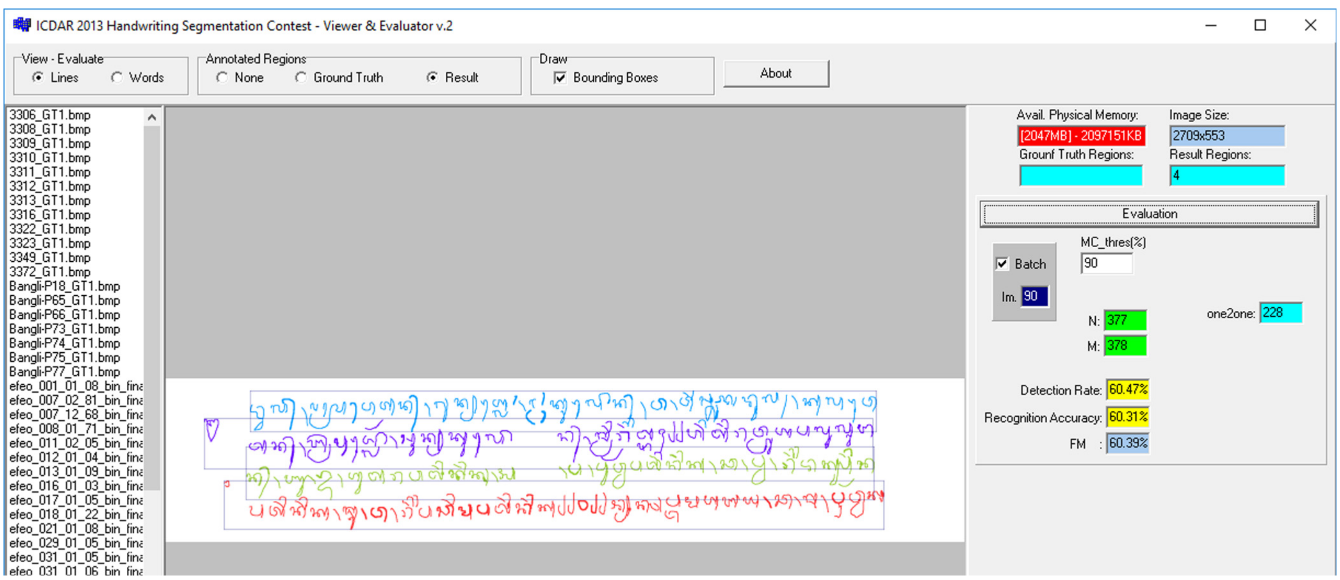

Fig. 15 ICDAR2013 handwriting segmentation contest—viewer and evaluator.

Table 2 Result of performance evaluation.

Method	Collection	*N*	*M*	o2o	DR (%)	RA (%)	FM (%)
APP (binary)	Bali	140	168	100	71.42	59.52	64.93
	Sunda	46	51	32	69.56	62.74	65.97
	Khmer	191	207	164	85.86	79.22	82.41
	All collection	377	426	296	78.51	69.48	73.72
Shredding (binary)	Bali	140	167	123	87.85	73.65	80.13
	Sunda	46	142	25	54.34	17.6	26.59
	Khmer	191	185	91	47.64	49.18	48.4
	All collection	377	494	239	63.39	48.38	54.87
Shredding with improved energy function (binary)	Bali	140	178	128	91.42	71.91	80.5
	Sunda	46	50	46	100	92	95.83
	Khmer	191	190	181	94.76	95.26	95.01
	All collection	377	418	355	94.16	84.92	89.3
Basic seam carving scheme (binary)	Bali	140	820	73	52.14	8.9	15.2
	Sunda	46	137	13	28.26	9.48	14.2
	Khmer	191	205	0	0	0	0
	All collection	377	1162	86	22.81	7.4	11.17
Basic seam carving scheme (grayscale)	Bali	140	1087	80	57.14	7.35	13.03
	Sunda	46	172	14	30.43	8.13	12.84
	Khmer	191	214	1	0.52	0.46	0.49
	All collection	377	1473	95	25.19	6.44	10.27
Complete seam carving scheme (binary)	Bali	140	143	131	93.57	91.60	92.57
	Sunda	46	46	46	100.0	100.0	100
	Khmer	191	189	51	26.70	26.98	26.84
	All collection	377	378	228	60.47	60.31	60.39
Complete seam carving scheme (grayscale)	Bali	140	167	128	91.42	76.64	83.38
	Sunda	46	43	36	78.26	83.72	80.89
	Khmer	191	145	57	29.84	39.31	33.92
	All collection	377	355	221	58.62	62.25	60.38
ALCM (grayscale)	Bali	140	322	20	14.28	6.21	8.65
	Sunda	46	66	4	8.69	6.06	7.14
	Khmer	191	392	59	30.89	15.05	20.24
	All collection	377	780	83	22.01	10.64	14.34
A* path planning (binary)	Bali	140	141	137	97.85	97.16	97.5
	Sunda	46	46	100	100	100	100
	Khmer	191	190	182	95.28	95.78	95.53
	All collection	377	377	365	96.81	96.81	96.81

neighbor cost function $N(s_i, s_j)$. The values of the parameters used on all data sets are: $c_d = 150$, $c_{d2} = 0$, $c_m = 50$, $c_v = 5$, and $c_n = 1$ (Fig. 14).

4 Experiments: Evaluation and Results

4.1 *Evaluation Metric and Tool*

We use the evaluation criteria and tool provided by ICDAR2013 Handwriting Segmentation Contest[29] (Fig. 15). First, the one-to-one (o2o) match score is computed for a region pair based on the evaluator's acceptance threshold. In our experiments, we used 90% as the acceptance threshold. Let N be the count of ground truth elements and M be the count of result elements. With the o2o score, three metrics are calculated: detection rate (DR), recognition accuracy (RA), and performance metric (FM).

4.2 *Results*

We performed the evaluation test for each collection of palm leaf manuscripts in our corpus. Table 2 shows the results of the performance evaluation of each method on each collection of manuscripts. We also calculated the performance of each method on the total collection.

In general, the methods that are applied on binary images achieved a good enough result. It is because we performed our experiments on the binary ground truth image of the manuscripts. These images are the ideal condition of the binary images that we can expect from the original manuscripts. The performance of binary image-based methods will be greatly influenced by the quality of the binary image.

In our experiments, without doing any postprocessing task, the A* path planning approach achieves the best results for all manuscript collection. Nevertheless, this approach greatly depends on the result of binarization and projection profile analysis in localizing the text line and the starting state of each line. This approach did not perform well when it failed to detect the starting points of lines because the manuscript contains some short lines (Fig. 16).

The APP method performs better on the collection of Khmer manuscripts because they often contain some spaces between words or between shorter subtext lines. This characteristic fits well with the APP method as it divided the manuscripts into smaller vertical zones. But the APP approach greatly relies on information extracted from the global horizontal projection profile at the beginning of the process to extract some important referenced values, such as the number of text lines, the average line position of each text line, and the average height of text lines. Most of the rules that are applied in the next step of the APP approach depend on the spatial information of text lines provided by those referenced values. For example, the average height of text lines is used to detect the baseline of the vowel, and the number of text lines is used to check (to insert or delete) the correct baseline to each column. If those values are incorrect, the accuracy of the approach will drop significantly. It, therefore, does not work well with document image with skewed and curved or fluctuated text lines. The APP method did not perform well on the collection of Balinese and Sundanese manuscripts because the size of ascenders and descenders of the character in Balinese and Sundanese scripts is almost the same as the size of the character itself and it occupies the space between two consecutive text lines. The APP approach considers it as a new baseline (Fig. 17). The APP technique also has some problems caused by the components that are too far away from their main lines.

The shredding method achieved the best performance on the Balinese manuscripts. The Balinese manuscripts have normally a wider space between two consecutive text lines

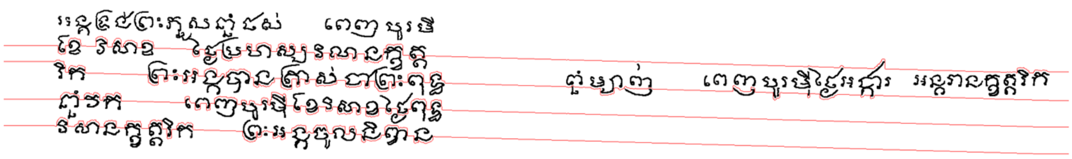

Fig. 16 The A* path planning approach failed to determine the starting state of each line in some manuscripts.

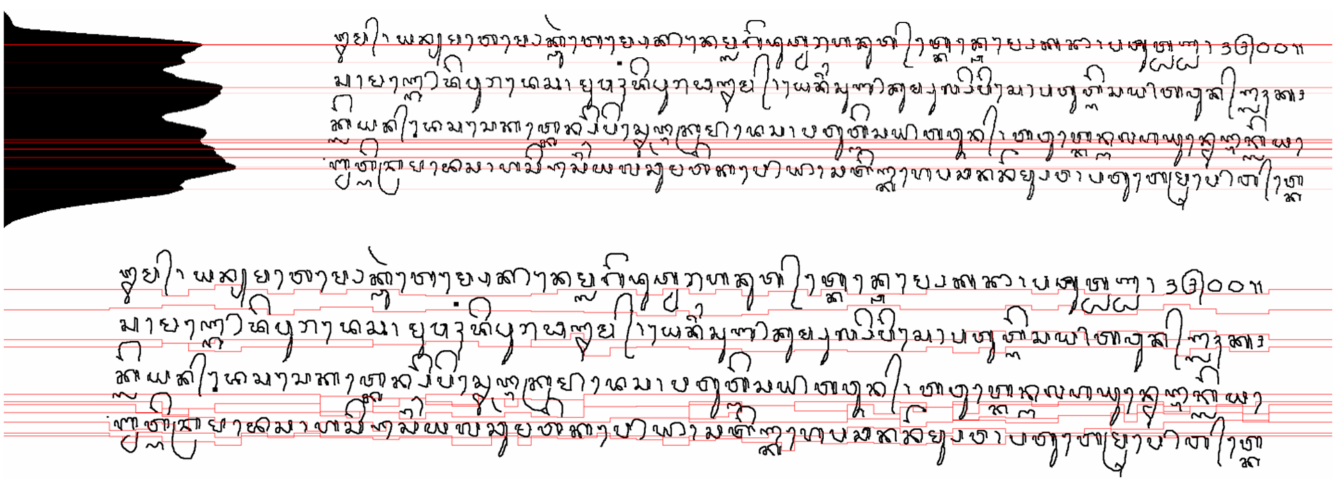

Fig. 17 The rules in the APP approach failed to determine the baseline of a Balinese manuscript.

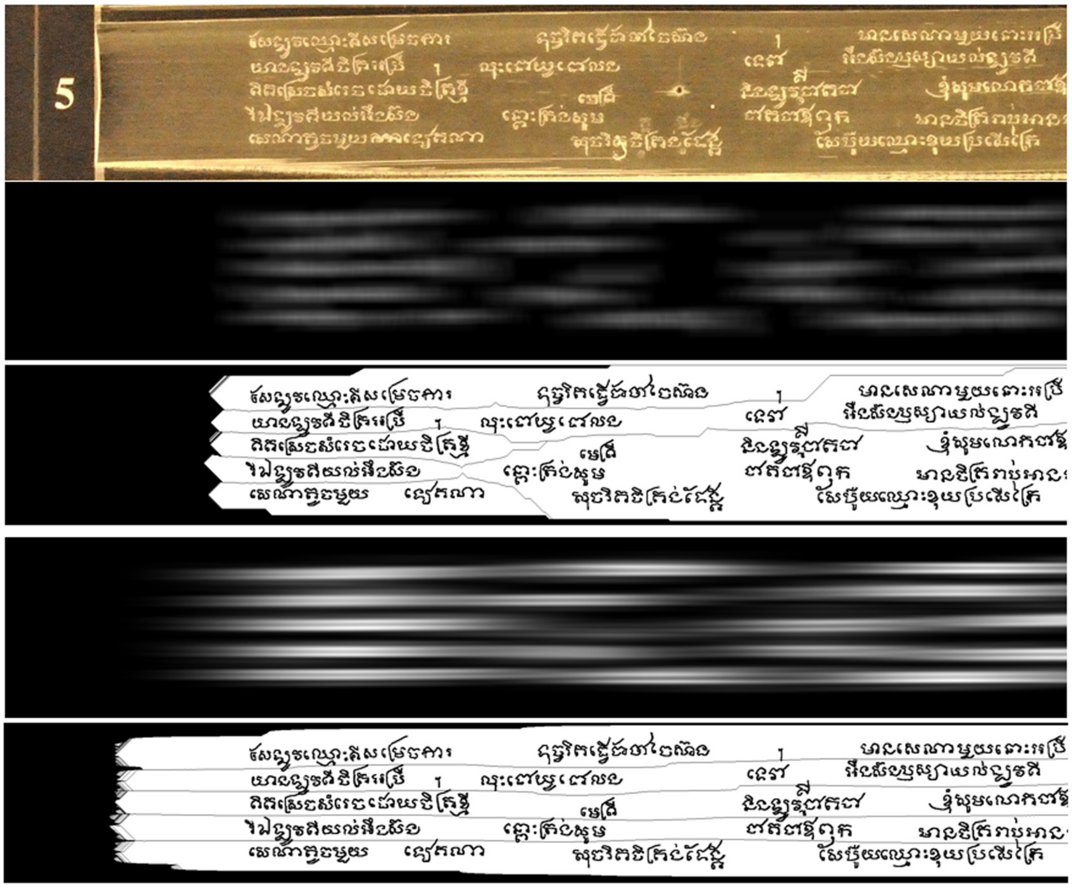

Fig. 18 The improved energy function for the shredding method. From top to bottom: original image of manuscript, original blurring function, detected line areas, improved energy function, and improved detected line areas.

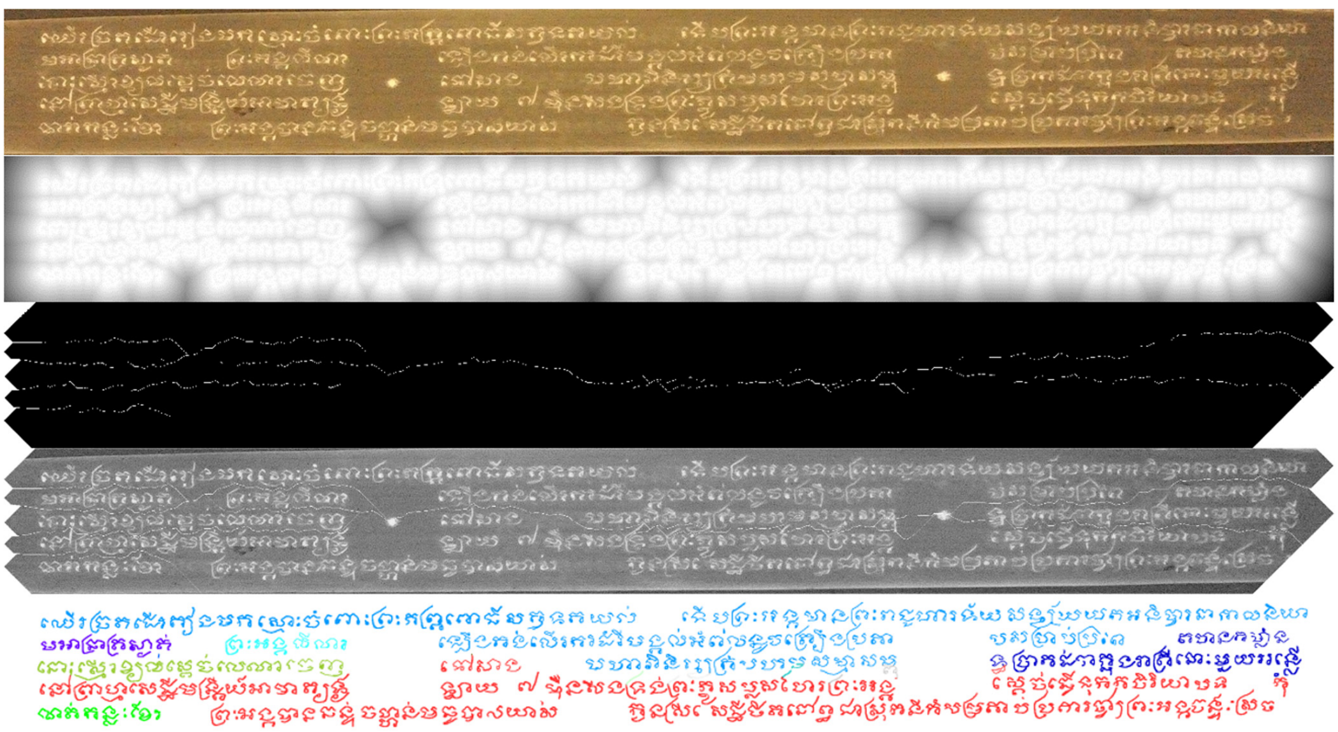

Fig. 19 An example of the jumping and joining separating seam paths on Khmer manuscript.

Fig. 20 (a) The medial seams and (b) the separating seams of the manuscript of Fig. 19 are correctly detected.

Fig. 21 (a) The medial seams and (b) the separating seams of the manuscript are not correctly detected.

than the Sundanese and Khmer manuscripts, so the shredding function more easily separates the intertext line areas. The use of ellipse energy function significantly improves the performance of the recursive tracing function from the shredding method. For Khmer manuscripts, our proposed energy function is optimal for forcing the energy transfer in one text line while preventing the energy transfer between two text lines (Fig. 18).

For the seam carving-based method, without any preprocessing and postprocessing tasks, the performances of this method are not optimal, especially on the collection of Khmer manuscripts. The spaces between some shorter subtext lines greatly influence the minimum separating seam path. By passing these spaces, the separating seam paths jumped to other intratext lines area or joined together into a single separating seam path (Fig. 19). This behavior makes the seam carving method fail to separate the text lines. The complete scheme of seam carving significantly improves separating seam detection (Fig. 20). But it greatly depends on the previous step of medial seam detection as a constraint that is based on the projection profile matching approach. If this first step failed to detect the correct medial axis of textline, the separating seams will not be detected correctly (Fig. 21).

5 Conclusions and Future Works

We investigated the performance of six text line segmentation methods by conducting the comparative experimental studies for the collection of palm leaf manuscript images. Four methods work on the binary image: the APP line segmentation approach, the A* path planning approach, the shredding method, and our proposed energy function for shredding method. Two other methods can be directly applied on grayscale images: the ALCM method and the seam carving-based method. The results show that each method performed optimal on some specific characteristic of the manuscript collection. The behavior of some methods is greatly influenced by some challenges that are clearly present on each collection of the Southeast Asian manuscripts. For future works, a scheme to adopt and take into account all advantages from each method should be proposed.

Acknowledgments

The authors would like to thank the Museum Gedong Kertya, Museum Bali, Mr. Undang Ahmad Darsa from University of Padjadjaran, and all families in Bali, Indonesia, for providing us the samples of palm leaf manuscripts, and the students from the Department of Informatics Education and the Department of Balinese Literature, University of Pendidikan Ganesha, for helping us in the ground truthing process for this research project. This work is also supported by the DIKTI BPPLN Indonesian Scholarship Program, the STIC Asia Program implemented by the French Ministry of Foreign Affairs and International Development (MAEDI), and ARES-CCD (program AI 2014-2019) under the funding of Belgian university cooperation.

References

1. A. Nicolaou and B. Gatos, "Handwritten text line segmentation by shredding text into its lines," in *10th Int. Conf. on Document Analysis and Recognition (ICDAR '09)*, pp. 626–630, IEEE (2009).
2. N. Arica and F. T. Yarman-Vural, "Optical character recognition for cursive handwriting," *IEEE Trans. Pattern Anal. Mach. Intell.* **24**, 801–813 (2002).
3. N. Arica and F. T. Yarman-Vural, "A new scheme for off-line handwritten connected digit recognition," in *Proc. Fourteenth Int. Conf. on Pattern Recognition*, Vol. **2**, pp. 1127–1129, IEEE Computer Society (1998).
4. S.-W. Lee, D.-J. Lee, and H.-S. Park, "A new methodology for grayscale character segmentation and recognition," *IEEE Trans. Pattern Anal. Mach. Intell.* **18**, 1045–1050 (1996).

5. R. J. Ramteke, "Invariant moments based feature extraction for handwritten devanagari vowels recognition," *Int. J. Comput. Appl.* **1** (18), 1–5 (2010).
6. M. Blumenstein, B. Verma, and H. Basli, "A novel feature extraction technique for the recognition of segmented handwritten characters," in *Proc. Seventh Int. Conf. on Document Analysis and Recognition*, Vol. **1**, pp. 137–141, IEEE Computer Society (2003).
7. Z. Shi, S. Setlur, and V. Govindaraju, "Text extraction from gray scale historical document images using adaptive local connectivity map," in *Proc. Eighth Int. Conf. on Document Analysis and Recognition*, pp. 794–798, Vol. **2**, IEEE (2005).
8. G. Louloudis et al., "Text line and word segmentation of handwritten documents," *Pattern Recognit.* **42**, 3169–3183 (2009).
9. R. P. dos Santos et al., "Text line segmentation based on morphology and histogram projection," in *10th Int. Conf. on Document Analysis and Recognition (ICDAR '09)*, pp. 651–655, IEEE (2009).
10. X. Zhang and C. L. Tan, "Text line segmentation for handwritten documents using constrained seam carving," in *14th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR '14)*, pp. 98–103, IEEE (2014).
11. F. Yin and C.-L. Liu, "Handwritten Chinese text line segmentation by clustering with distance metric learning," *Pattern Recognit.* **42**, 3146–3157 (2009).
12. J. Kumar et al., "Handwritten Arabic text line segmentation using affinity propagation," in *Proc. of the 9th IAPR Int. Workshop on Document Analysis Systems (DAS '10)*, pp. 135–142, ACM Press (2010).
13. C. C. Fung and R. Chamchong, "A review of evaluation of optimal binarization technique for character segmentation in historical manuscripts," in *Third Int. Conf. on Knowledge Discovery and Data Mining (WKDD '10)*, pp. 236–240, IEEE (2010).
14. M. W. A. Kesiman et al., "An analysis of ground truth binarized image variability of palm leaf manuscripts," in *5th Int. Conf. on Image Processing, Theory, Tools and Applications (IPTA '15)*, pp. 229–233 (2015).
15. M. W. A. Kesiman et al., "An initial study on the construction of ground truth binarized images of ancient palm leaf manuscripts," in *13th Int. Conf. on Document Analysis and Recognition (ICDAR '15)* (2015).
16. I. Bar-Yosef et al., "Line segmentation for degraded handwritten historical documents," in *10th Int. Conf. on Document Analysis and Recognition (ICDAR '09)*, pp. 1161–1165, IEEE (2009).
17. A. Garz et al., "Binarization-free text line segmentation for historical documents based on interest point clustering," in *10th IAPR Int. Workshop on Document Analysis Systems (DAS '12)*, pp. 95–99, IEEE (2012).
18. L. Likforman-Sulem, A. Zahour, and B. Taconet, "Text line segmentation of historical documents: a survey," *Int. J. Doc. Anal. Recognit.* **9**, 123–138 (2007).
19. E. Saund, J. Lin, and P. Sarkar, "PixLabeler: user interface for pixel-level labeling of elements in document images," in *10th Int. Conf. on Document Analysis and Recognition, (ICDAR '09)*, pp. 646–650, IEEE (2009).
20. K. Ntirogiannis, B. Gatos, and I. Pratikakis, "Performance evaluation methodology for historical document image binarization," *IEEE Trans. Image Process.* **22**, 595–609 (2013).
21. R. Chamchong and C. C. Fung, "Text line extraction using adaptive partial projection for palm leaf manuscripts from Thailand," in *Int. Conf. on Frontiers in Handwriting Recognition (ICFHR '12)*, pp. 588–593, IEEE (2012).
22. R. Chamchong and C. C. Fung, "Character segmentation from ancient palm leaf manuscripts in Thailand," in *Proc. of the 2011 Workshop on Historical Document Imaging and Processing (HIP '11)*, pp. 140–145, ACM Press (2011).
23. J. Sauvola and M. Pietikäinen, "Adaptive document image binarization," *Pattern Recognit.* **33**, 225–236 (2000).
24. N. Arvanitopoulos and S. Susstrunk, "Seam carving for text line extraction on color and grayscale historical manuscripts," in *14th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR '14)*, pp. 726–731, IEEE (2014).
25. R. Saabni and J. El-Sana, "Language-independent text lines extraction using seam carving," in *Int. Conf. on Document Analysis and Recognition (ICDAR '11)*, pp. 563–568, IEEE (2011).
26. A. Asi, R. Saabni, and J. El-Sana, "Text line segmentation for gray scale historical document images," in *Proc. of the 2011 Workshop on Historical Document Imaging and Processing (HIP '11)*, pp. 120–126, ACM Press (2011).
27. C. Stoll, Y. Xiao, and Z.-H. Duan, "Text line extraction using seam carving," in *Int. Conf. on Image Processing, Computer Vision, and Pattern Recognition* (2015).
28. O. Surinta et al., "A path planning for line segmentation of handwritten documents," in *14th Int. Conf. on Frontiers in Handwriting Recognition (ICFHR '14)*, pp. 175–180, IEEE (2014).
29. N. Stamatopoulos et al., "ICDAR 2013 handwriting segmentation contest," in *12th Int. Conf. on Document Analysis and Recognition (ICDAR '13)*, pp. 1402–1406, IEEE (2013).

Made Windu Antara Kesiman is a PhD candidate at the Laboratory of Information, Image and Interaction (L3i), University of La Rochelle, and is also a lecturer at the University of Pendidikan Ganesha, Indonesia. He received his engineering degree in informatics in 2004 at Bandung Institute of Technology, Indonesia, and his MSc degree in informatics majoring in image processing at the University of La Rochelle, France, in 2006. His work focuses on document analysis of ancient palm leaf manuscript.

Dona Valy is a PhD candidate at the Université Catholique de Louvain, Belgium, and is also a researcher at the Institute of Technology of Cambodia (ITC). He received his engineering degree in computer science and communication in 2010 and his MSc degree in image synthesis and graphics at the University of Limoges, France, in 2011. His current PhD research work focuses on ancient palm leaf manuscript analysis and Khmer handwritten character recognition.

Jean-Christophe Burie received his PhD in automatic control engineering from University of Lille, France, in 1995. He was a research fellow in the Department of Mechanical Engineering for Computer-Controlled Machinery, Osaka University, Japan, from 1995 to 1997 in the framework of the Lavoisier Program of the French Foreign Office. He is currently a full professor at La Rochelle University. His research interests include computer vision, color image processing, and pattern recognition. His research topics concern color document analysis and character recognition.

Erick Paulus received his BSc degree in mathematics from the University of Padjadjaran, Indonesia, in 2004 and his MSc degree in computer science from the University of Gadjah Mada, Indonesia, in 2006. He is now working at the Robotics, Artificial Intelligence, and Digital Image (RAID) Laboratory at the University of Padjadjaran, Indonesia, as a researcher on the augmented reality and ancient manuscript of Sundanese culture and language. His major research interests include social informatics, document image analysis, and historical document processing.

I. Made Gede Sunarya is a lecturer at the University of Pendidikan Ganesha, Indonesia. He received his engineering degree in computer science in 2006 and his master's degree in computer science in 2012 at the University of Gadjah Mada, Indonesia. His current research work focuses on printed and handwritten character recognition for Balinese script.

Setiawan Hadi works as associate professor in Department of Computer Science, University of Padjadjaran, Indonesia. His research interest is in object detection and recognition especially for face-related objects. He has been working as a reviewer of several international and national journals, a member of national accreditation board, and a member of Robotics, Artificial Intelligence, and Digital Image (RAID) Laboratory, University of Padjadjaran.

Kim Heng Sok is deputy head of Computer Science Department at Institute of Technology of Cambodia (ITC), master degree from INSA de Rennes France. He has led mobile research activity at ITC. He is also the outstanding instructional designer for E-learning course in ASEAN Cyber University Project and also the best trainer for USAID-COMET mekong skill to work project for the CLMVT countries and joins STIC-ASIE Program for being the coadvisor of a PhD student.

Jean-Marc Ogier received his PhD in computer science from the University of Rouen, France, in 1994. He is a full professor at the University of La Rochelle, he is the author of more than 160 publications. He has been the chair of the TC10 (Graphic Recognition) of the IAPR, and is now the representative member of France at the governing board of the IAPR. He has been the chair of many international scientific events dealing with document analysis.