**RESEARCH**  **Open Access**

CrossMark

# Textline detection in degraded historical document images

Byeongyong Ahn[1], Jewoong Ryu[2], Hyung Il Koo[3] and Nam Ik Cho[1*]  iD

## Abstract

This paper presents a textline detection method for degraded historical documents. Our method follows a conventional two-step procedure that the binarization is first performed and then the textlines are extracted from the binary image. In order to address the challenges in historical documents such as document degradation, structure noise, and skews, we develop new methods for the binarization and textline extraction. First, we improve the performance of binarization by detecting the non-text regions and processing only text regions. We also improve the textline detection method by extracting main textblock and compensating the skew angle and writing style. Experimental results show that the proposed method yields the state-of-the-art performance for several datasets.

**Keywords:** Historical documents, Handwritten documents, Document image binarization, Textline detection

## 1 Introduction

Historical documents are valuable cultural heritage and thus there are increasing demands to digitize them for archiving, indexing, and recognition purposes. However, historical documents suffer from various kinds of degradations and their understanding remains a challenging problem. In this paper, we present a textline detection algorithm for historical documents, which is a key step to document understanding.

### 1.1 Textline detection in historical documents

Textline detection is an essential step in many document processing tasks (e.g., layout analysis and optical character recognition), and numerous methods have been proposed for decades. However, most of conventional methods focused on machine-printed [8, 22] and/or clean (noise-free) documents [2, 9, 14–17, 28, 31–34], so they cannot be directly applied to historical documents. In historical document processing, binarization is a challenging task due to degradations (e.g., bleed-through and faint characters) and structure noises. In addition to difficulties in binarization, historical documents also suffer from a variety of challenges as they are mostly handwritten [25].

### 1.2 Our approach

In this paper, we propose a new textline detection method for degraded historical documents. The proposed method follows a conventional two-step procedure, i.e., binarization and then textline extraction from the binarized document. However, both of the steps have many challenges due to the degradation stated above. Hence, we develop new approaches for the two steps, i.e., new binarization and textline extraction methods that are robust to such degradations. To be precise, we develop a binarization algorithm that extracts text pixels in the presence of structure noises (e.g., page boundaries, blank regions, and scan errors as shown in Fig. 1). For the textline extraction in binary images, we adopt an algorithm for handwritten documents in [25]. However, the conventional work is based on the assumption that each input image has a single textblock and a global skew is already compensated [10, 25], while the historical documents contain metadata (scribbles or distracting text that are written later—see Fig. 1) and often skewed. Hence, we need to find main textblocks, remove metadata, and compensate the skew to improve the performance of the textline detection.

## 2 Methods

### 2.1 Methods on document image binarization

Many early methods used (adaptive) thresholding techniques, which can be classified into global and local methods. Global methods use a single threshold for the

*Correspondence: nicho@snu.ac.kr
[1]Department of Electrical and Computer Engineering and INMC, Seoul National University, Seoul, South Korea
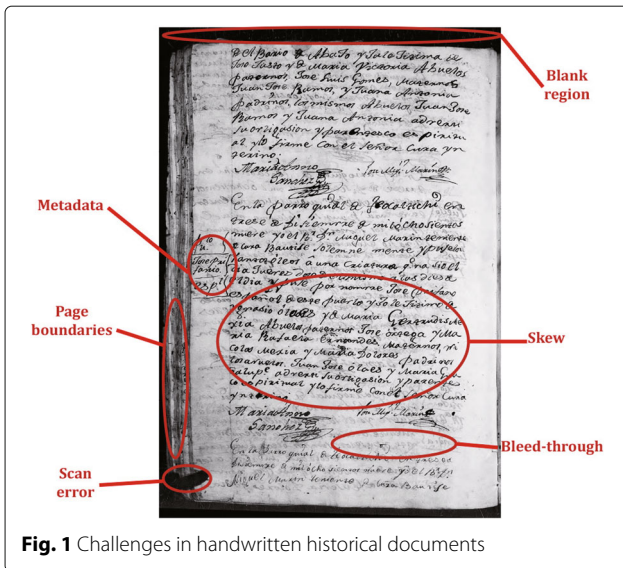Full list of author information is available at the end of the article

**Fig. 1** Challenges in handwritten historical documents

whole image. For example, Otsu's method [23] finds the threshold by minimizing the intra-class variance. On the other hand, the local methods analyze the statistics of local regions and set the locally different thresholds. For example, Niblack [19] and Sauvola et al. [27] used the linear combination of local mean and standard deviation for the thresholds. However, as shown in Fig. 2, they have difficulties in handling the historical documents. Specifically, Otsu's and Sauvola's methods sometimes miss characters and Niblack's method yields many noisy blobs.

More recently, Markov random field (MRF)-based or feature-based methods were proposed in [7, 11, 12] which obtain the binarized image as a result of minimizing an energy function. Although these optimizatio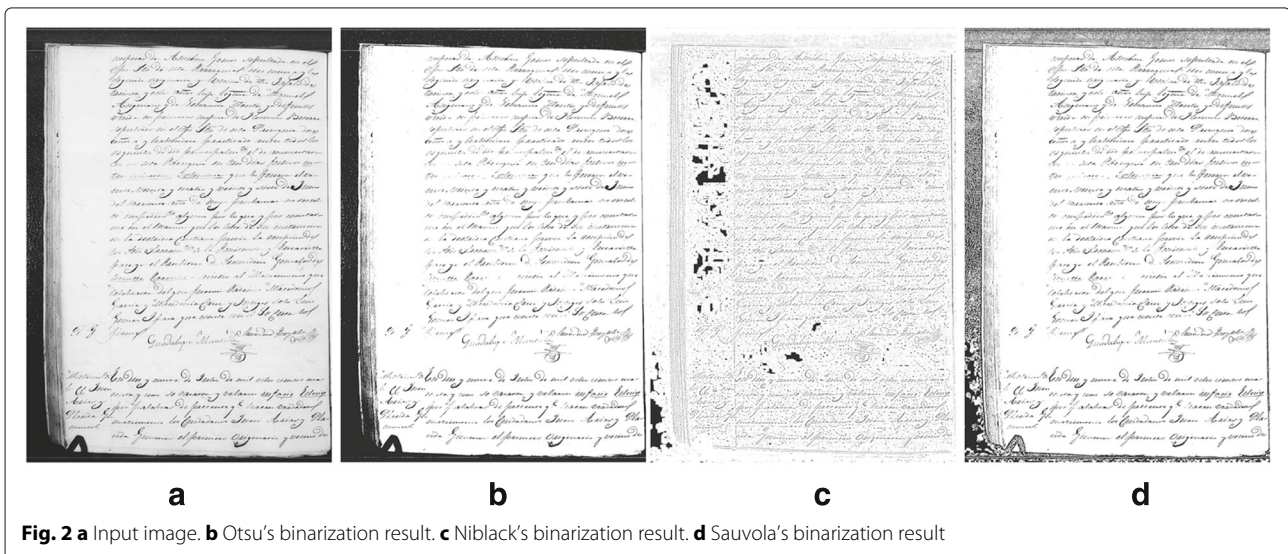n-based methods show promising results, they are computationally demanding compared with the thresholding approaches. Meanwhile, a combined approach has also been proposed to alleviate the problems of global and local methods. Gatos et al. [6] used the majority vote to combine the results of several methods. Ntirogiannis et al. [21] proposed a background reconstruction method and computed normalized images by using the reconstructed backgrounds. Then, local and global methods are applied to the normalized images and the final results are obtained by combining these two complementary results. This method shows the state-of-the-art performance in the 2009 and 2010 Document Image Binarization Contests (DIBCOs) [5, 24].

### 2.2 Methods on textline detection in handwritten documents

For the textline detection in handwritten documents, many researchers exploited the vertical quasi-periodicity of textlines and developed projection-profile-based methods. Bruzzone et al. [2] calculated horizontal projection profiles and partitioned the input image into horizontal strips. Stafylakis et al. [28] analyzed the patterns of projection profiles and estimated the transition probability between the two states (i.e., textlines and their gaps). Although their methods were simple, they could not handle skewed and/or curved textlines.

Likforman-Sulem et al. [15] handled the skew problem by proposing a Hough-based approach. They applied the Hough transform to the centers of connected components (CCs) in order to find the textlines. Louloudis et al. [17] improved the Hough-transform-based method by introducing additional steps. They grouped CCs according to their sizes and each group was processed independently. Also, merging techniques and CC splitting were also employed to reduce the errors. These methods were



**Fig. 2 a** Input image. **b** Otsu's binarization result. **c** Niblack's binarization result. **d** Sauvola's binarization result

Ahn *et al. EURASIP Journal on Image and Video Processing*    (2017) 2017:82

Page 3 of 13

able to alleviate the skew problem; however, they have limitations in handling the curved textlines. Zahour et al. [32] proposed a partial projection procedure: they partitioned an input page into columns and drew horizontal strokes in each column. The horizontal strokes were connected to construct separating lines between the adjacent textlines. Arivazhagan et al. [1] horizontally divided the document into chunks. Then, an initial set of candidate lines were constructed by connecting the valleys in each chunk. These candidate lines were finally refined by using the Gaussian probability decision and distance metric decision.

Recently, some researchers formulated the textline detection as a path-finding problem. Saabni et al. [26] computed an energy map based on the sign distance transform. Since the result map had negative values in the textlines and positive values in the background, minimal energy paths that pass between the textlines were extracted by using the dynamic programming. Fernández-Mota et al. [3] also proposed a path-finding method based on the similar framework. Yin et al. [31] exploited the hierarchical structure in document images and presented a bottom-up grouping method: CCs of an input image were grouped into textlines by using a learned metric. Koo et al. [10, 25] developed textline extraction algorithms based on the energy minimization framework. They proposed a cost function that considers the distances between the textlines and the curvilinearity of each textline.

Finally, the overview of existing textline detection methods is summarized in in Table 1. Although the conventional methods were developed to handle a variety of textlines, most of them used binary images as the input or assumed that the characters are extracted from the background with a simple binarization algorithm. Therefore, applying these methods to degraded handwritten historical documents is not straightforward.

## 3    Overview of our method

In order to extract textlines in degraded documents like Fig. 1, we have to address the challenges in both binarization and textline detection steps. To this end, we propose a system consisting of two stages as shown in Fig. 3. In order to prevent the disturbance of structure noise (e.g., page boundaries, blank regions, and scan errors), we find the the region of interests (ROI) by detecting and excluding the non-page regions. Then we perform the binarization on the ROIs with the method in [21]. In the textline detection stage, we first detect main paragraphs by removing metadata, since lots of historical documents suffer from metadata as shown in Fig. 1. Then we adopt a CC grouping textline detection algorithm [25] and estimate the global skew of the document. Finally, we compensate the skew and perform CC re-grouping. The details will be discussed in Sections 4 and 5.

**Table 1** Textline detection methods for handwritten documents

| Category | Advantages | Disadvantages |
|---|---|---|
| Projection-profile [2, 28] | Simplicity and efficiency | Difficulties in handling skewed and/or curved textlines |
| Hough-transform [15, 17] | Availability to skewed textlines | Limitations in handling curved textlines |
| Partial projection [1, 32] | Availability to skewed and/or curved textlines | Limitations in handling overlapping components |
| Path-finding [3, 26] | Robustness to writing styles | Difficulties in handling touching lines |
| Bottom-up grouping [31] | Robustness to document geometry | Errors in splitting and merging lines |
| Energy minimization [10, 25] | High performance | High computational cost |

## 4    Document binarization

In historical document binarization, the combined approach introduced above [21] showed good performance. However, it was developed for page regions and hence has some difficulties in handling the images like Fig. 1. To be precise, the challenges are that (a) non-text regions have similar colors or intensities to the text pixels and hence the conventional methods fail to classify them correctly and (b) these misclassified pixels corrupt the statistics that will be used in text and non-text classification. In order to address these problems, we develop a non-page region detection algorithm and apply the binarization method in [21] to the ROIs as illustrated in the first stage of Fig. 3.

### 4.1    Non-page region detection

For the non-page detection, we first apply a global binarization method (Otsu's method) to input gray images. As shown in Fig. 1, the non-page regions consist of blank and page boundaries, and the binarization yields a few huge CCs on blank regions and vertically elongated CCs on page boundaries as shown in Fig. 2a. In order to effectively filter out these CCs, we merge them with a morphological operation (dilation). Figure 2a–c shows how the dilation works. Since the scale changes according to writing styles and/or scanning resolutions, we achieve the scale invariance by selecting filters adaptively. The size of dilation filter is given by $\alpha_{dil} \times \zeta$ where $\zeta$ is the median of character heights in a given image and $\alpha_{dil}$ is a constant. We use the median height because the heights of CCs do not change much even in cursive writings. With small filter size, some CCs on page boundaries can remain. On the other hand, some characters can be merged with the page boundaries
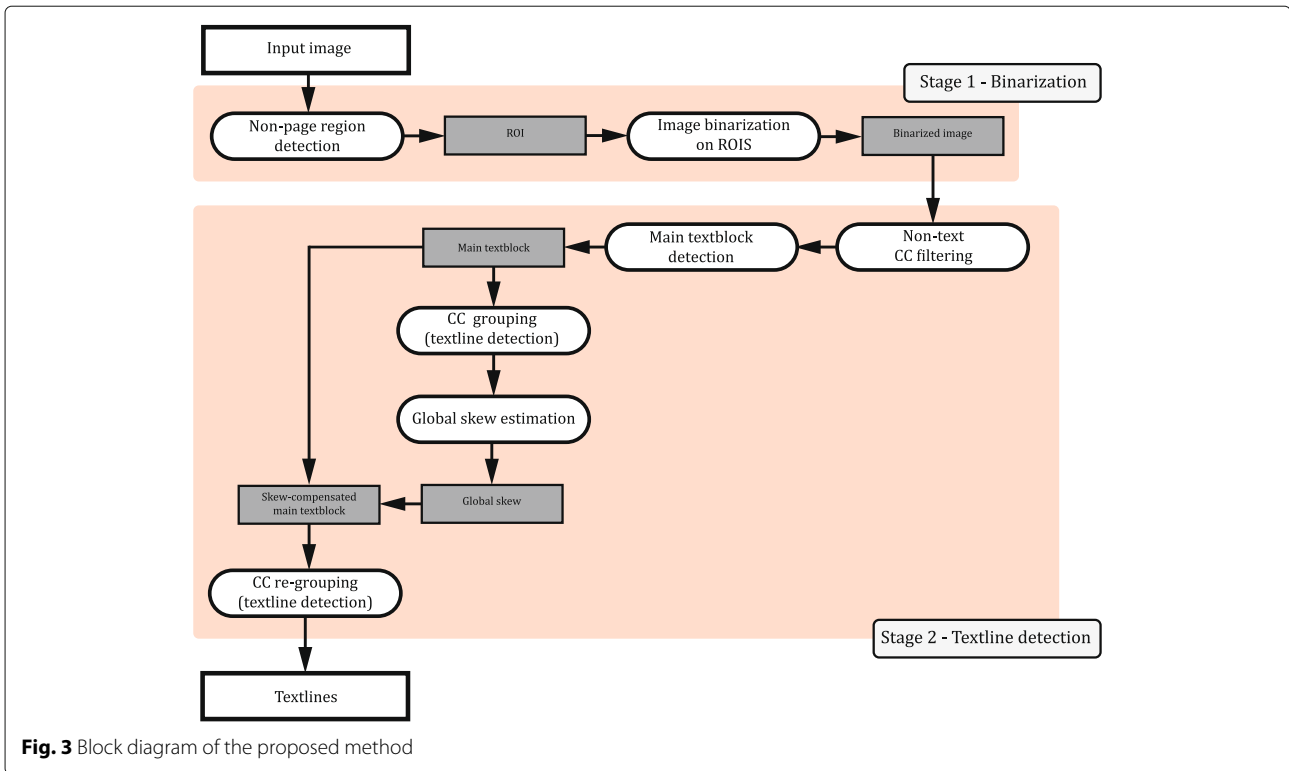
Ahn *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82

Page 4 of 13

**Fig. 3** Block diagram of the proposed method

when the filter size is too large. Therefore, it would be beneficial to find an appropriate value of $\alpha_{dil}$. The parameter setting step would be explained in Section 6.1. After the dilation, the non-page region forms large CCs around the image boundaries. Let us denote a set of CCs in the dilated image as $\{C_i\}$. Then we can detect non-page regions by finding the CCs satisfying

$$|C_i| > \alpha_1 \times \max(H, W) \tag{1}$$

$$|C_i \cap \partial I(\delta)| > \alpha_2 \times \max(H, W) \tag{2}$$

where $|\cdot|$ is the number of elements in the set, $\partial I(\delta)$ is a set of pixels whose distance from the image border is smaller than $\delta$, and $\alpha_1, \alpha_2$ are parameters that control the performance of the non-page region detection algorithm. The first condition imposes a constraint on the scales of CCs and the second is a constraint on its location. In order to achieve the scale invariance, we use $\max(H, W)$, where $H$ and $W$ are the height and width of an image respectively. Finally, we consider the union of CCs satisfying (1) and (2) as the mask of non-page regions as shown in Fig. 4a.

For the CC merging step, the run length smoothing algorithm (RLSA) [30] that links foreground pixels can also be employed. However, experimental results show that the RLSA achieve similar results to dilation and hence we just adopt the dilation method.

### 4.2 Image binarization on ROIs
We remove detected non-page regions and then perform binarization on the remaining regions. The binarization method is the same as the combined approach in [21]. In detail, a background image is estimated from the pixels that are classified as backgrounds by the local method [19]. Then, an image is defined whose pixel value is the ratio of input image's pixel value to the background image's pixel value. The image is named "normalized image." Otsu's method is applied to the normalized image to obtain the global output. Some features of foreground and background pixel values (e.g., mean and standard deviation) and average stroke width of the character are estimated from the background image and global output. From these features, parameters for the local binarization method are selected and the local output is obtained. The global method usually results in clean output but the faint parts of the foreground are often disappeared. On the other hand, the local method preserves faint characters but yields noisy output. So these two methods are combined for recovering the faint parts of the global output with the local output. The detailed description can be found in [21].

Figure 4b, c shows the results of the combined method [21] without and with a non-page detection block. As shown in Fig. 4d, e, we can improve the binarization results of texts as well as removing non-page regions. The objective evaluation on the effects of the proposed
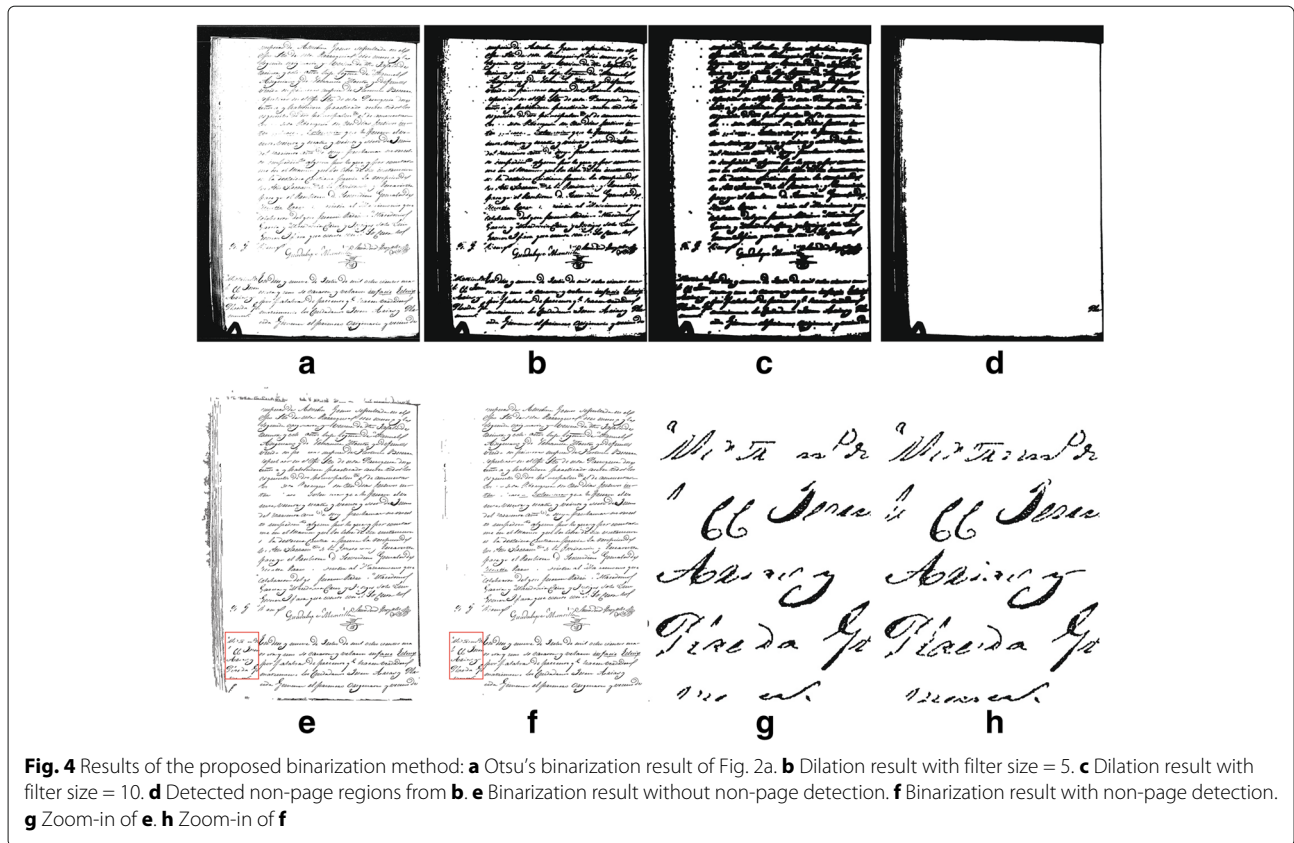
Ahn *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82

Page 5 of 13



**Fig. 4** Results of the proposed binarization method: **a** Otsu's binarization result of Fig. 2a. **b** Dilation result with filter size = 5. **c** Dilation result with filter size = 10. **d** Detected non-page regions from **b**. **e** Binarization result without non-page detection. **f** Binarization result with non-page detection. **g** Zoom-in of **e**. **h** Zoom-in of **f**

non-page detection block will be presented in the experimental section (Section 6.4).

For the combined method, it is desirable to choose a local method that achieves high precision on backgrounds and high recalls for faint characters. Although Sauvola's method [27] is better than Niblack's method [19] as an individual algorithm, Niblack's method is preferred as a part of a combined method since it achieves higher background precision and foreground recall.

## 5 Textline detection

From the binarization results, we extract the textlines. Although the handwritten textline detection method in [25] shows the state-of-the-art performance on several datasets, it assumed controlled inputs, i.e., (a) the input images do not contain non-text components, (b) each input image has a single textblock, and (c) there is no global skew. However, none of them hold in historical documents. In order to alleviate these problems and obtain improved results in complex document images, we develop a new textline extraction method as shown in stage 2 of Fig. 3.

### 5.1 Non-text CC filtering

As shown in Fig. 5a, b, binarization results may contain non-text components (e.g., noises, stamps, and graphs in

documents) and we filter them out by analyzing their sizes. As discussed in the Section 4.1, we achieve the scale invariance using the median height of characters. Let us denote a set of CCs in the binarized image as $\{B_i\}$. Then, we remove the CCs violating

$$\beta_{low} \times \zeta^2 < |B_i| < \beta_{high} \times \zeta^2. \tag{3}$$

where $\zeta$ is the median of character heights in a given image and $\beta_{low}$, $\beta_{high}$ are thresholds for the size-based filtering. Also, we impose constraints on $w(B_i)$ and $h(B_i)$, the widths and heights of $B_i$: we remove CCs that violate either of

$$h(B_i) < \beta_h \times H, \tag{4}$$
$$w(B_i) < \beta_w \times W \tag{5}$$

where $\beta_h$ and $\beta_w$ are thresholds for CC height and width respectively. The $\beta_{low}$ affects the ability of this step to distinguish small characters from noise components. When using large $\beta_{low}$, some small characters may be classified as noise and filtered out. Also, $\beta_{high}$, $\beta_h$, and $\beta_w$ allow us to reject large non-text components such as stamps. Thus, large values for these parameters may introduce the removal of connected texts. The input and output of this block are shown in Fig. 5b, c.
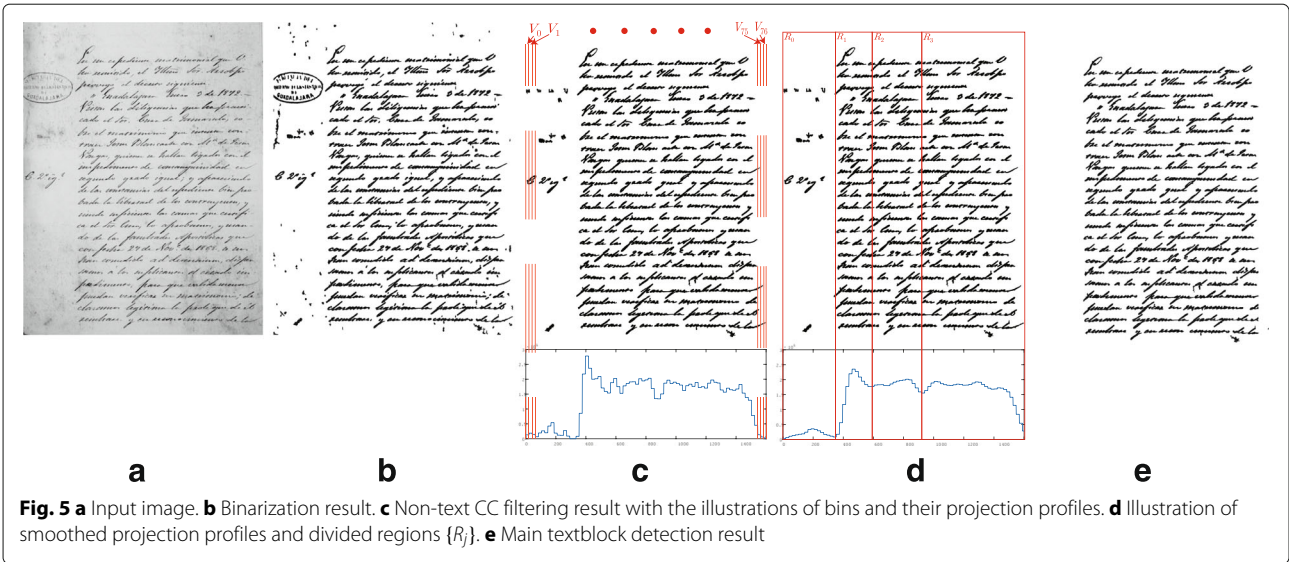
Ahn *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82

Page 6 of 13



**Fig. 5 a** Input image. **b** Binarization result. **c** Non-text CC filtering result with the illustrations of bins and their projection profiles. **d** Illustration of smoothed projection profiles and divided regions $\{R_j\}$. **e** Main textblock detection result

## 5.2 Main textblock detection via projection profile analysis

Document images sometimes contain text information (such as metadata and/or texts in other pages) that should not be considered in the textline extraction. In order to handle this problem, we use the projection profile analysis (PPA). First, an input $I$ is divided into vertical strips $\{V_i\}$ as shown in Fig. 5c and we compute a projection profile $|V_i|$ that represents the number of text pixels in the corresponding strip. As we can see in Fig. 5c, there are local minima between the main textblocks and other regions and we partition the strips $\{V_i\}$ into several regions $\{R_j\}$ according to the locations of local minima. In order to reduce the effect of outliers, a moving average filter that has three taps is applied to the projection profiles prior to the partitioning. Smoothed projection profiles and divided regions are shown in Fig. 5d. Among these candidates, we find main textblocks based on two observations. The first observation is that the strips in the main text regions are dense (i.e., more text pixels) and we filter out the regions $R_j$ satisfying

$$\text{average}_{V_i \in R_j} |V_i| < \beta_s \times \text{med}_{V_i \in I} |V_i| \qquad (6)$$

where $\beta_s$ is the threshold of the density of regions. In this equation, the left-hand side and the right-hand side indicate the text-pixel density in an individual region and whole image respectively. We used the median for the overall density of the image to achieve the robustness to outliers.

The second observation is that the main textblocks have some large width, and thus we remove the regions $R_j$ satisfying

$$W(R_j) < \beta_f \times \max_k(W(R_k)) \qquad (7)$$

where $W(R_k)$ is the width of a $k$th region $R_k$ and $\beta_f$ is the threshold handling the width of textblocks.

Thresholds $\beta_s$ and $\beta_f$ control the classification performance between the metadata and textblocks. When large values are selected for these parameters, some textblocks can be classified as metadata, and Fig. 5e shows the proposed result.

## 5.3 CC grouping (textline detection)

After extracting the CCs in the main textblocks, we find textlines in the textblocks by using the method in [25], which addressed the textline detection problem by partitioning extracted CCs into subsets corresponding to textlines. For each CC, its stroke width and size are computed. The normalized length of CC is defined as the
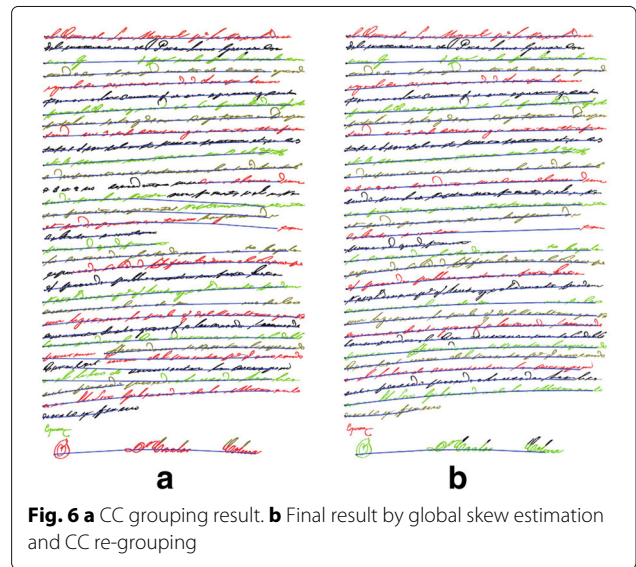


**Fig. 6 a** CC grouping result. **b** Final result by global skew estimation and CC re-grouping

Ahn *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82
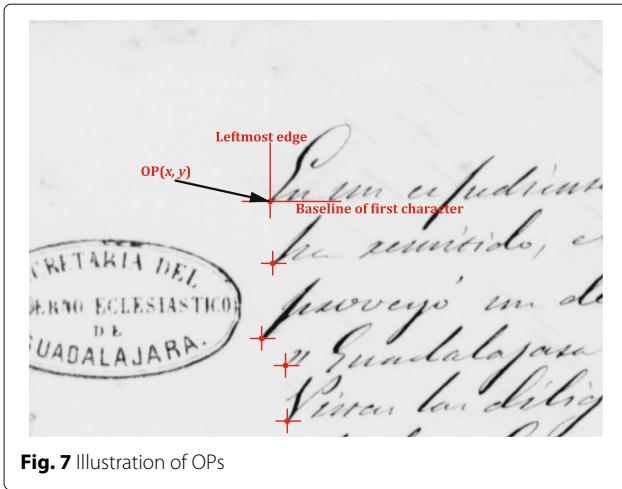
Page 7 of 13



**Fig. 7** Illustration of OPs

ratio of its size to its stroke length, and if the length is larger than a constant, then the CC is considered touching CCs and hence partitioned into smaller CCs. Then, CCs are grouped by optimizing a cost function that consists of two terms: a local term and a layout term. The local term considers the fitting errors of individual textlines and becomes small when the fitting function of each line yields small errors. On the other hand, the layout term becomes small when the distance between two textlines is large. The minimization of the cost function consists of two steps. At the first step, a coarse solution is obtained with a bottom-up grouping method. Then, the coarse solution is iteratively refined in a greedy manner. The refinement is based on four heuristic proposals: (1) merge, (2) split, (3) merge-split, and (4) merge-merge-split. The merge proposal merges two neighboring clusters into one cluster, the split proposal splits a cluster into two clusters, and other proposals are the combination of merge and split proposals. The refinement step is repeated until the energy function does not decrease further. The detailed

description of the method can be found in [10] and [25]. Although this approach shows superior performance on several datasets, it yields errors when global skews are present as illustrated in Fig. 6a.

### 5.4 Global skew estimation and CC re-grouping (textline detection)

To improve the performance, we estimate the skew from the above CC grouping results and re-apply the grouping process after the skew compensation. We estimate the skew angle of each textline by fitting the centers of CCs (in the textline) with the random sample consensus algorithm [4]. For each line, we choose two CCs from the same textline and compute a straight line connecting their centers. CCs whose centers fit the straight line well (the distance from the line is less than $\eta/2$) are considered the elements of a consensus set, and when the consensus set contains more than 80% of the CCs in the line, the straight line is accepted. Finally, the straight line is refined using all elements in the consensus set. The slope of the refined line is considered the skew angle of the textline. Then, we consider the median of estimated skews of all textlines as the global skew. This idea can be justified because the majority of textlines are correctly estimated with the initial grouping method. Using the estimated value, we rotate the images and repeat the CC grouping. The result of the proposed method is shown in Fig. 6.

### 6 Results and discussion

In order to demonstrate the performance of the proposed method, we conducted experiments on publicly available datasets. Also, we submitted the proposed method to "ICDAR 2015 textline detection in historical documents competition" and won the first place [18]. The proposed algorithm is implemented with C++ and tested on a PC with a 3.4-GHz dual core processor. It takes an average of 12.31 s in handing one document image (its average size is $2168 \times 3787$). Steps that correspond

**Table 2** Experimental results on the ICDAR 2015 dataset [18]

|  | Average cost | Number of one-to-one matches | Number of detection misses | Number of false positives | Number of detection failure | DR | RA | FM |
|---|---|---|---|---|---|---|---|---|
| UNIFR | 19.00 | 2578 | 6456 | 267 | 3022 | 27.72% | 23.16% | 25.23% |
| IA-1 | 15.40 | 5626 | 5268 | 127 | 883 | 51.05% | 50.53% | 50.79% |
| IA-2 | 14.51 | 5655 | 6032 | 102 | **407** | 59.30% | 53.66% | 56.34% |
| A2iA-1 | 15.09 | 6590 | 2264 | 628 | 2393 | 69.50% | 59.19% | 63.93% |
| A2iA-2 | 13.35 | 5974 | 4020 | 79 | 1791 | 59.30% | 53.66% | 56.34% |
| A2iA-3 | 13.20 | 6523 | **2263** | 181 | 2490 | 72.74% | 58.59% | 64.91% |
| Proposed | **9.77** | **7741** | 2700 | **25** | 948 | **73.96**% | **69.53**% | **71.68**% |

The highlighted values show the best results

Ahn *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82
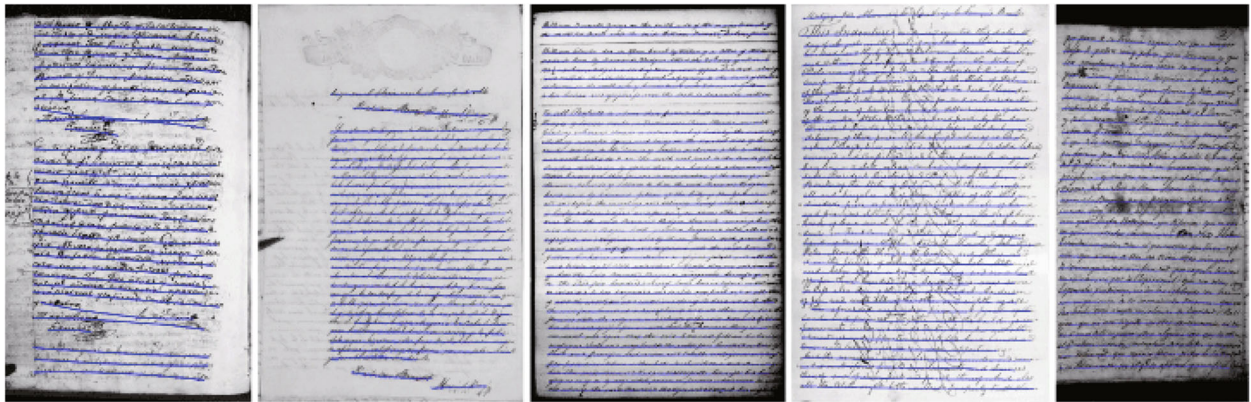
Page 8 of 13



**Fig. 8** Results of the proposed method on ICDAR 2015 dataset. As shown, our algorithm is able to find exact textlines in the presence of non-text objects, page border, metadata, and uneven background

to existing methods [21, 25] are based on our own implementation.

### 6.1 Parameter selection

Many parameters are involved in the proposed method. In order to set their values, we adopt a greedy method. Specifically, we tested five values {2, 5, 10, 20, 50} for $\delta$ while fixing other parameters, and selected a value showing the best performance on the ICDAR 2015 training dataset. After selecting $\delta$, we repeat a similar procedure for other parameters one by one. After repeating the whole procedure three times, we finally got the following values: $\delta = 5, \alpha_{dil} = 0.5, \alpha_1 = 10, \alpha_2 = 2, \beta_{low} = 0.5, \beta_{high} = 50, \beta_h = 0.3, \beta_w = 0.5, \beta_s = 0.5$, and $\beta_f = 0.3$. In all experiments, we used these values.
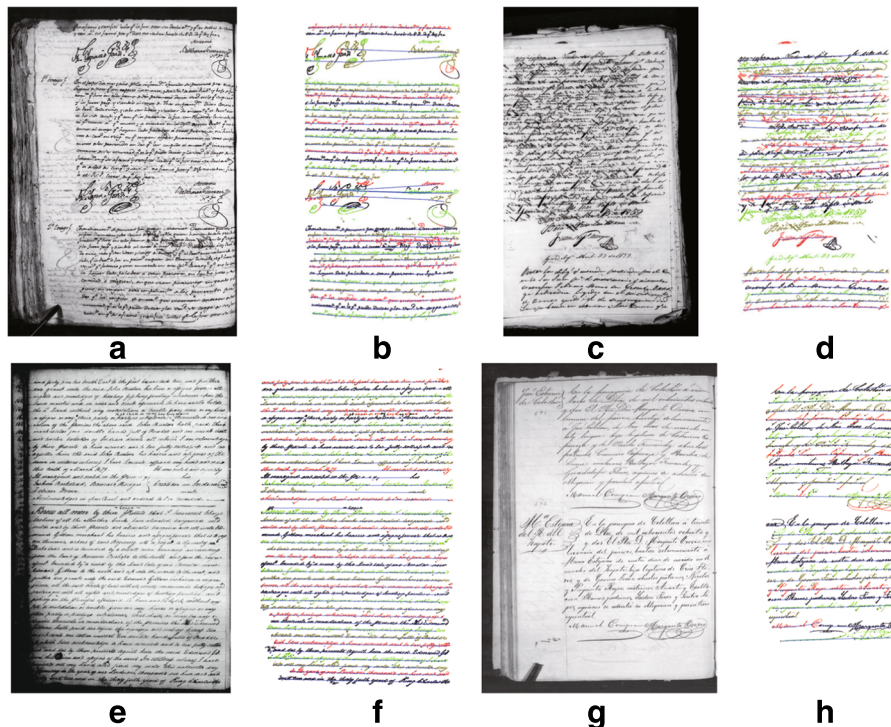


**Fig. 9** Failure cases of the proposed method in the ICDAR 2015 dataset [18], (**a**,**c**,**e**,**g**) Samples from ICDAR 2015 dataset, (**b**,**d**,**f**,**h**) corresponding results
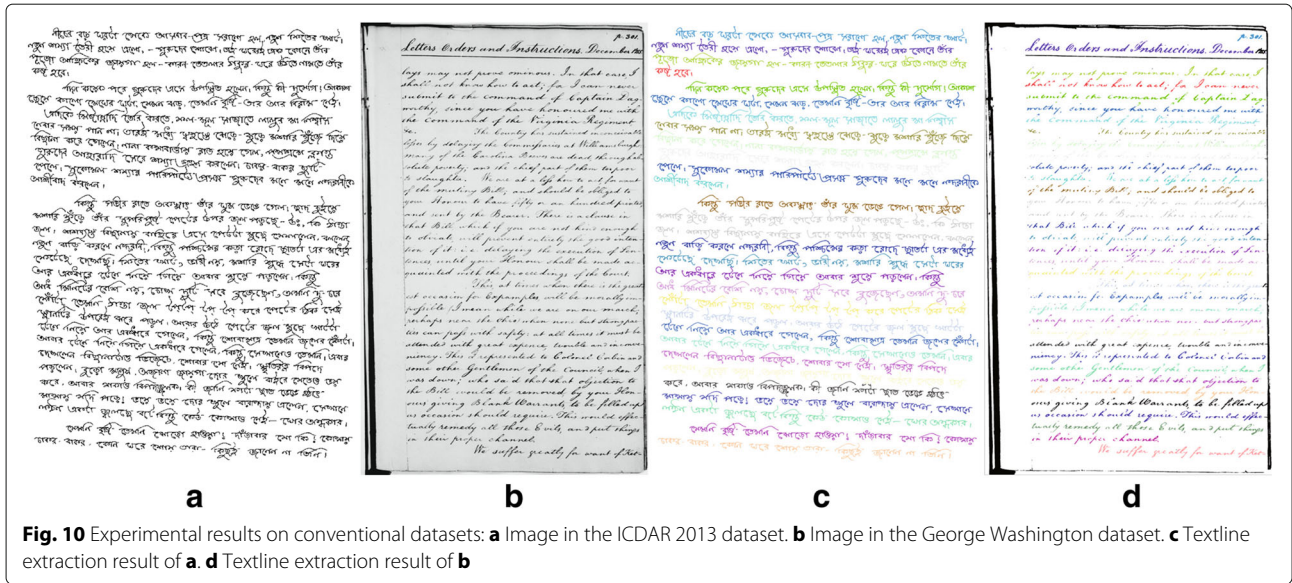
**Fig. 10** Experimental results on conventional datasets: **a** Image in the ICDAR 2013 dataset. **b** Image in the George Washington dataset. **c** Textline extraction result of **a**. **d** Textline extraction result of **b**

## 6.2 Evaluations on ICDAR 2015 competition

The ICDAR 2015 competition dataset consists of documents written in several languages (German, English, and Spanish) during the eighteenth and nineteenth centuries, and it presents a variety of challenges in historical document processing (e.g., touching lines, skewness, and structure noise).

### 6.2.1 Evaluation metrics

In the competition, an origin point (OP)-based scoring method was employed. As shown in Fig. 7, an OP coordinate is defined as an intersection of the baseline of each textline and the left-most edge of the first character. In particular, estimated origin points (EPs) are matched to the ground truth OPs and each pair is classified into five cases as follows:

- R1-Detection-Hit : an EP is very close to a corresponding OP. The cost for this case is 0.
- R2-Detection-Hit : an EP is somewhat close to a corresponding OP. The cost is proportional to the distance between the EP and OP.
- Detection-Miss : an EP is far from a corresponding OP. The cost is 20.
- False-Positive : Multiple EPs are matched to a single OP. The cost for each false-positive is 10.
- Detection-Failure : There are no EPs that match to the OP. The cost is 20.

In addition to the OP-based cost, we also computed detection rate (DR), recognition accuracy (RA), and F-measure (FM), which are defined as

$$DR = o2o/N \qquad (8)$$

$$RA = o2o/M \qquad (9)$$

$$FM = \frac{2 \times DR \times RA}{DR + RA} \qquad (10)$$

where $o2o$ is the number of one-to-one matches (R1-Detection-Hit+R2-Detection-Hit), $N$ and $M$ are the numbers of OPs and EPs respectively.

**Table 3** Experimental results on the ICDAR 2013 dataset [29]

|  | $DR_l$ | $RA_l$ | $FM_l$ |
|---|---|---|---|
| CUBS | 97.96% | 96.94% | 97.45% |
| GOLESTAN-a,b | 98.23% | 98.34% | 98.28% |
| LRDE | 96.94% | 97.57% | 97.25% |
| MSHK | 91.66% | 90.06% | 90.85% |
| NUS | 98.34% | 98.49% | 98.41% |
| QATAR-a | 90.75% | 91.55% | 91.15% |
| QATAR-b | 91.73% | 93.14% | 92.43% |
| CVC | 91.28% | 89.06% | 90.16% |
| IRISA | 97.85% | 96.93% | 97.39% |
| †ILSP [28] | 96.11% | 94.82% | 95.46% |
| †NCSR [17] | 92.37% | 92.48% | 92.43% |
| †TEI [20] | 97.77% | 96.82% | 97.30% |
| †Koo et al. [10] | 93.58% | 92.29% | 92.93% |
| †Fernández-Mota et al. [3] | 96.30% | 94.58% | 95.43% |
| †Ryu et al. [25] | 98.64% | 98.68% | 98.66% |
| Proposed | **98.67**% | **98.82**% | **98.75**% |

In addition to competition results, the performances of other conventional methods are also presented †
The highlighted values show the best results

Ahn *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82

Page 10 of 13

**Table 4** Experimental results on the George Washington dataset [13]

|  | $DR_l$ | $RA_l$ | $FM_l$ |
|---|---|---|---|
| Bruzzone et al. [2] | 47.20% | 46.40% | 46.70% |
| Fernández-Mota et al. [3] | 91.30% | 94.20% | 92.70% |
| Ryu et al. [25] | 95.18% | **96.19**% | 95.69% |
| Proposed | **95.74**% | 95.89% | **95.82**% |

The highlighted values show the best results

### 6.2.2 Evaluation results

The evaluation results on 363 images having 11,333 textlines are summarized in Table 2. As shown, the proposed method showed the best performance among 7 participating methods. Some results are shown in Fig. 8: the proposed method is able to detect the textlines in the presence of non-text object, uneven background, and non-paragraph metadata. Failure cases are shown in Fig. 9. As shown in the left column, some signatures are vertically long and they were considered multiple lines. Also, as shown in Fig. 9c, d, some documents suffer from severe degradations and even humans may have difficulties in detecting textlines without the understanding of contents. There are some failure points for the individual steps of the proposed algorithm. In Fig. 9e, f, some characters are mixed with the blank region and therefore excluded from the ROI. Some metadata invade the domain of main text and the proposed algorithm fails to filter them out completely as shown in Fig. 9g, h. Although these failures do not disturb detecting textlines, it makes the estimated EPs to be biased either to the left or to the right.

### 6.3 Evaluation on conventional datasets

In order to compare the performance of our algorithm with the conventional methods, we also evaluate our method on two conventional datasets: ICDAR 2013 handwritten segmentation contest dataset [29] and George Washington's manuscript dataset [13]. The ICDAR 2013 handwritten segmentation contest dataset consists of 150 binary images with 2649 textlines. In the set, textlines are relatively well-separated and images contain only text components (i.e., perfect binarization results). The George Washington's manuscript dataset consists of 20 grayscale images with 715 textlines. Although the set consists of gray images, its binarization is simple due to its high quality. Example images of these two datasets are shown in Fig. 10.

### 6.3.1 Evaluation metrics

For the comparison with the existing methods, we adopted the conventional pixel-based measure used in [29]. To be precise, in the conventional metric, the *MatchScore* is defined as

$$MatchScore(i, j) = \frac{|G_j \cap R_i|}{|G_j \cup R_i|} \tag{11}$$

where $R_i$ is a set of pixels in the $i$th detected textline and $G_j$ is a set of pixels in the $j$th ground truth textline. When the *MatchScore* is greater than 0.95, the pair is considered a one-to-one match. Then, detection rate ($DR_l$), recognition accuracy ($RA_l$), and F-measure ($FM_l$) are similarly defined to (8), (9), and (10) respectively. However, we use a subscript "$l$" to indicate that these values are obtained from textline-level correspondences.
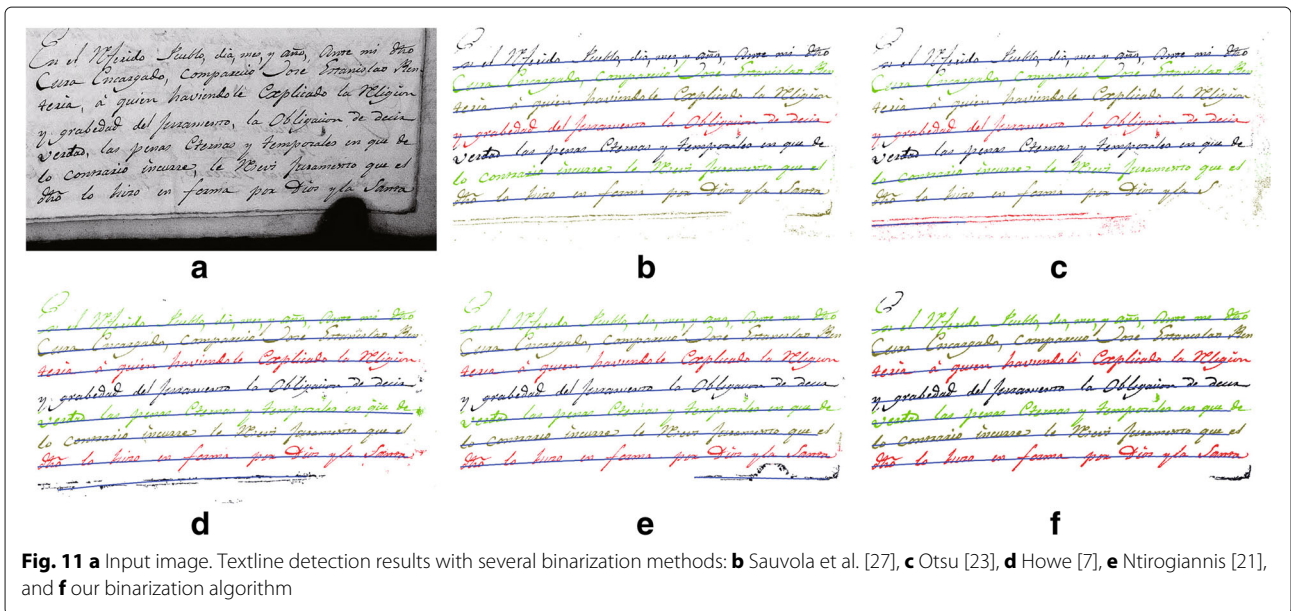


**Fig. 11 a** Input image. Textline detection results with several binarization methods: **b** Sauvola et al. [27], **c** Otsu [23], **d** Howe [7], **e** Ntirogiannis [21], and **f** our binarization algorithm

*Ahn et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82

Page 11 of 13

### 6.3.2 Evaluation results

Evaluation results on the ICDAR 2013 dataset and comparison to existing methods [3, 10, 17, 20, 25, 28] are summarized in Table 3. As shown, the proposed method compares favorably with existing methods. Although the improvement is not significant, these results show that the proposed method is able to handle noise-free inputs as existing methods, even though the proposed method is developed to handle the severely degraded documents. Also, the evaluation results on George Washington dataset are in Table 4. The proposed algorithm also shows the best performance. Figure 10 shows the textline detection results of the proposed method.

### 6.4 Evaluation of the proposed blocks

In order to evaluate the contribution of each step, we conducted additional experiments that replace the proposed binarization and textline detection methods with the corresponding conventional methods. To be precise, we evaluate the following methods on the ICDAR 2015 dataset: (a) the proposed textline detection algorithm in Section 5 with the conventional binarization methods [7, 21, 23, 27] and (b) the proposed binarization method in Section 4 with the state-of-the-art textline detection algorithms [10, 25]. However, unfortunately, we use the training set in the ICDAR 2015 dataset in the evaluation, since the test set is not available.

Some experimental results obtained by replacing the binarization methods are shown in Fig. 11. As shown, the proposed binarization method successfully filters out most of page borders and it is able to preserve the faint characters well compared with conventional methods as shown in Fig. 11e, f. The comparison results are summarized in Table 5, where it can be seen that the proposed binarization method shows the best result by improving FM by 3.19% compared with [21].

Table 6 shows comparison results obtained by replacing the proposed textline detection method with conventional methods. As shown, the proposed textline detection method improves FM by 10.76% compared with [25]. This is mainly because our method considers the structure of document images (e.g., the presence of metadata). As

can be seen in Tables 5 and 6, both blocks are essential for the overall performance. Thus, we believe that the main contribution of this paper is the development of a robust textline detection system for degraded historical document images, not just the improvement of each block.

Also, Table 7 shows the performance of the proposed algorithm with varying parameter values. It is difficult to find the general correlation between the parameter values and the performance. However, the performance definitely depends on almost every parameter and therefore the parameter setting step described in subsection 6.1 is an essential step for the algorithm optimization. In the case of the vertical strips $V_i$, its size does not affect the performance and thus not analyzed in Table 7.

**Table 6** Experimental results on the ICDAR 2015 dataset [18] with conventional textline detection methods

|  | DR | RA | FM |
|---|---|---|---|
| Koo et al. [10] | 61.21% | 59.93% | 60.57% |
| Ryu et al. [25] | 60.57% | 62.49% | 61.51% |
| Proposed | **71.49**% | **73.06**% | **72.27**% |

The highlighted values show the best results

**Table 7** Experimental results on the ICDAR 2015 dataset [18] with different parameter values

|  | DR | RA | FM |
|---|---|---|---|
| $\delta = 2$ | 70.41% | 67.75% | 69.06% |
| $\delta = 5$ | **71.49%** | **73.06%** | **72.27%** |
| $\delta = 10$ | 72.86% | 71.39% | 72.12% |
| $\delta = 20$ | 72.02% | 70.80% | 71.40% |
| $\delta = 50$ | 71.57% | 70.39% | 70.98% |
| $\alpha_{dil} = 0.1$ | 71.08% | 73.26% | 72.15% |
| $\alpha_{dil} = 0.5$ | **71.49%** | **73.06%** | **72.27%** |
| $\alpha_{dil} = 1.0$ | 71.61% | 72.77% | 72.18% |
| $\alpha_1 = 5$ | 72.35% | 72.15% | 72.25% |
| $\alpha_1 = 10$ | **71.49%** | **73.06%** | **72.27%** |
| $\alpha_1 = 20$ | 72.14% | 71.91% | 72.03% |
| $\alpha_2 = 1$ | 70.32% | 73.24% | 71.75% |
| $\alpha_2 = 2$ | **71.49%** | **73.06%** | **72.27%** |
| $\alpha_2 = 4$ | 71.52% | 72.69% | 72.10% |
| $\beta_s = 0.4$ | 71.15% | 71.14% | 71.15% |
| $\beta_s = 0.5$ | **71.49%** | **73.06%** | **72.27%** |
| $\beta_s = 0.6$ | 71.08% | 73.46% | 72.25% |
| $\beta_f = 0.2$ | 71.32% | 72.22% | 71.77% |
| $\beta_f = 0.3$ | **71.49%** | **73.06%** | **72.27%** |
| $\beta_f = 0.4$ | 71.49% | 73.00% | 72.24% |

The highlighted values show the best FM and are adopted in the proposed algorithm

**Table 5** Experimental results on the ICDAR 2015 dataset [18] with conventional binarization methods

|  | DR | RA | FM |
|---|---|---|---|
| Sauvola et al. [27] | 49.59% | 37.51% | 42.71% |
| Otsu. [23] | 45.20% | 57.25% | 50.52% |
| Howe [7] | 63.26% | 65.68% | 64.45% |
| Ntirogiannis et al. [21] | 70.41% | 67.75% | 69.06% |
| Proposed | **71.49**% | **73.06**% | **72.27**% |

The highlighted values show the best results

Ahn *et al. EURASIP Journal on Image and Video Processing* (2017) 2017:82

Page 12 of 13

## 7 Conclusions

In this paper, we have proposed a textline detection algorithm for degraded historical documents. Our method is based on the conventional procedure that extracts textlines after binarization. However, in order to address the challenges in historical handwritten documents, we have developed new methods for binarization and textline extraction. First, we proposed an ROI setting method to deal with the non-page region problem. Also, we developed a textline detection method that can handle the metadata and page skews. Experimental results on a variety of datasets showed that our method outperforms conventional methods and both steps in our system are essential for the overall performance.

### Authors' contributions
All authors equally contributed the original research in the paper. All authors read and approved the final manuscript.

### Authors' information
Authors are with the Dept. of Electrical and Computer Eng., Seoul National University, and also affiliated with the INMC (Institute of New Media and Communications).

### Competing interests
The authors declare that they have no competing interests.

## Publisher's Note
Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

### Author details
[1]Department of Electrical and Computer Engineering and INMC, Seoul National University, Seoul, South Korea. [2]DMC R&D Center, Samsung Electronics Co. Ltd., Suwon, South Korea. [3]Department of Electrical and Computer Engineering, Ajou University, Suwon, South Korea.

### References
1. M Arivazhagan, H Srinivasan, S Srihari, A statistical approach to line segmentation in handwritten documents. Electron. Imaging (2007)
2. E Bruzzone, MC Coffetti, in *International Conference on Document Analysis and Recognition (ICDAR)*. An algorithm for extracting cursive text lines, (1999), pp. 749–752
3. D Fernández-Mota, J Lladós, A Fornés, A graph-based approach for segmenting touching lines in historical handwritten documents. Int. J. Doc. Anal. Recogn. **17**(3), 293–312 (2014)
4. MA Fischler, RC Bolles, Random sample consensus: a paradigm for model fitting with applications to image analysis and automated cartography. Commun. ACM. **24**(6), 381–395 (1981)
5. B Gatos, K Ntirogiannis, I Pratikakis, DIBCO 2009: document image binarization contest. Int. J. Doc. Anal. Recogn. (IJDAR). **14**(3), 35–44 (2011)
6. B Gatos, I Pratikakis, SJ Perantonis, in *International Conference on Pattern Recognition (ICPR)*. Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information, (2008), pp. 1–4
7. NR Howe, Document binarization with automatic parameter tuning. Int. J. Doc. Anal. Recogn. (IJDAR). **16**(3), 247–258 (2013)
8. J Hull, Document image skew detection: survey and annotated bibliography. Ser. Mach. Percept. Artif. Intell. **29**, 40–66 (1998)
9. HI Koo, NI Cho, in *European Conference on Computer Vision (ECCV)*. State estimation in a document image and its application in text block identification and text line extraction, (2010), pp. 421–434
10. HI Koo, NI Cho, Text-line extraction in handwritten Chinese documents based on an energy minimization framework. IEEE Trans. Image Process. **21**(3), 1169–1175 (2012)
11. JG Kuk, NI Cho, KM Lee, in *IEEE International Conference on Image Processing (ICIP)*. MAP-MRF approach for binarization of degraded document image, (2008), pp. 2612–2615
12. JG Kuk, NI Cho, in *International Conference on Document Analysis and Recognition (ICDAR)*. Feature based binarization of document images degraded by uneven light condition, (2009)
13. V Lavrenko, TM Rath, R Manmatha, in *International Workshop on Document Image Analysis for Libraries*. Holistic word recognition for handwritten historical documents, (2004), pp. 278–287
14. Y Li, Y Zheng, D Doermann, S Jaeger, Script-independent text line segmentation in freestyle handwritten documents. IEEE Trans. Pattern. Anal. Mach. Intell. **30**(8), 1313–1329 (2008)
15. L Likforman-Sulem, A Hanimyan, C Faure, in *International Conference on Document Analysis and Recognition (ICDAR)*. A Hough based algorithm for extracting text lines in handwritten documents, (1995), pp. 774–777
16. L Likforman-Sulem, A Zahour, B Taconet, Text line segmentation of historical documents: a survey. Int. J. Doc. Anal. and Recogn. (IJDAR). **9**(2-4), 123–138 (2007)
17. L Louloudis, B Gatos, I Pratikakis, C Halatsis, Text line and word segmentation of handwritten documents. Pattern Recog. **42**(12), 3169–3183 (2009)
18. M Murdock, S Reid, B Hamilton, J Reese, in *International Conference on Document Analysis and Recognition (ICDAR)*. ICDAR 2015 competition on text line detection in historical documents, (2015), pp. 1–5
19. W Niblack, *An introduction to digital image processing*. (Prentice-Hall, 1986)
20. A Nicolaou, B Gatos, in *International Conference on Document Analysis and Recognition (ICDAR)*. Handwritten text line segmentation by shredding text into its lines, (2009), pp. 626–630
21. K Ntirogiannis, B Gatos, I Pratikakis, A combined approach for the binarization of handwritten document images. Pattern Recogn. Lett. **35**(1), 3–15 (2014)
22. L O'Gorman, The document spectrum for page layout analysis. IEEE Trans. Pattern. Anal. Mach. Intell. **15**(11), 1162–1173 (1993)
23. N Otsu, A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**(1), 62–66 (1979)
24. I Pratikakis, B Gatos, K Ntirogiannis, in *International Conference on Frontiers in Handwriting Recognition*. H-DIBCO 2010-handwritten document image binarization competition, (2010), pp. 727–732
25. J Ryu, HI Koo, NI Cho, Language-independent text-line extraction algorithm for handwritten documents. IEEE Signal Process. Lett. **21**(9), 1115–1119 (2014)
26. R Saabni, A Asi, J El-Sana, Text line extraction for historical document images. Pattern Recogn. Lett. **35**(7), 23–33 (2014)
27. J Sauvola, M Pietikäinen, Adaptive document image binarization. Pattern Recogn. **33**(2), 225–236 (2000)
28. T Stafylakis, V Papavassiliou, V Katsouros, G Carayannis, in *IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*. Robust text-line and word segmentation for handwritten documents images, (2008), pp. 3393–3396
29. N Stamatopoulos, B Gatos, G Louloudis, U Pal, A Alaei, in *International Conference on Document Analysis and Recognition (ICDAR)*. ICDAR 2013 handwriting segmentation contest, (2013), pp. 1402–1406
30. KY Wong, RG Casey, FM Wahl, Document analysis system. IBM J. Res. Devel. **26**(6), 647–656 (1982)
31. F Yin, C Liu, Handwritten Chinese text line segmentation by clustering with distance metric learning. Pattern Recogn. **42**(12), 3146–3157 (2009)
32. A Zahour, B Taconet, P Mercy, Ramdane S, in *International Conference on Document Analysis and Recognition (ICDAR)*. Arabic hand-written text-line extraction, (2001), pp. 281–285

Ahn *et al. EURASIP Journal on Image and Video Processing*   (2017) 2017:82

Page 13 of 13

33. SJ Ha, B Jin, NI Cho, in *IEEE International Conference on Image Processing (ICIP)*. Fast text line extraction in document images, (2012), pp. 797-800
34. C Yan, H Xie, H Liu, J Yin, Y Zhang, Q Dai, Effective Uyghur language text detection in complex background images for traffic prompt identification. IEEE Trans. Intell. Transp. Syst. (2017)