

Article

Learning-Free Text Line Segmentation for Historical Handwritten Documents

Berat Kurar Barakat ^{1,*}, Rafi Cohen ^{1,†}, Ahmad Drobay ¹, Irina Rabaev ² and Jihad El-Sana ¹

¹ Department of Computer Science, Ben-Gurion University of the Negev, Be'er Sheva 84105, Israel; rafico@cs.bgu.ac.il (R.C.); drobya@post.bgu.ac.il (A.D.); el-sana@cs.bgu.ac.il (J.E.-S.)

² Department of Software Engineering, Shamoon College of Engineering, Be'er Sheva 8410802, Israel; irinar@ac.sce.ac.il

* Correspondence: berat@post.bgu.ac.il

† These authors contributed equally to this work.

Received: 18 October 2020; Accepted: 18 November 2020; Published: 22 November 2020



Abstract: We present a learning-free method for text line segmentation of historical handwritten document images. This method relies on automatic scale selection together with second derivative of anisotropic Gaussian filters to detect the blob lines that strike through the text lines. Detected blob lines guide an energy minimization procedure to extract the text lines. Historical handwritten documents contain noise, heterogeneous text line heights, skews and touching characters among text lines. Automatic scale selection allows for automatic adaption to the heterogeneous nature of handwritten text lines in case the character height range is correctly estimated. In the extraction phase, the method can accurately split the touching characters among the text lines. We provide results investigating various settings and compare the model with recent learning-free and learning-based methods on the cBAD competition dataset.

Keywords: text line segmentation; text line detection; text line extraction; learning-free; historical handwritten documents

1. Introduction

Digital handwritten documents are not easily explorable in their raw form but need to be transcribed further into machine readable text. Certainly, manual transcription of a large number of documents is not feasible in a reasonable time. Hence, there is a practical need for reliable handwritten document image processing algorithms. Text line segmentation is an essential operation and prerequisite for many document image analysis tasks. Advancement in text line segmentation performance will boost the performance of other tasks, such as word segmentation [1,2] and word recognition [3,4].

Text line segmentation consists of text line detection and text line extraction. Text line detection locates each text line by its baseline or x -height representation. Text line extraction in turn leads to polygonal or pixel level representation of text lines. Extraction level representation is more precise and useful for higher level document image analysis tasks. With the advances in deep learning, numerous learning-based methods have been proposed for text line segmentation of handwritten documents. Learning-based methods [5–8] can inherently handle the problems arising from complex layout of text lines and heterogeneity of documents. Henceforth, the recent competition datasets [9] (Figure 1) are more challenging than the prior ones [10–13].

In the last decade several learning-free methods for text line extraction of handwritten documents have been proposed [14–17]. However, we observe only limited attempts where learning-free algorithms have been used for text line detection of challenging historical documents [9]. This is particularly interesting as learning-based algorithms require a vast amount of labeling effort.

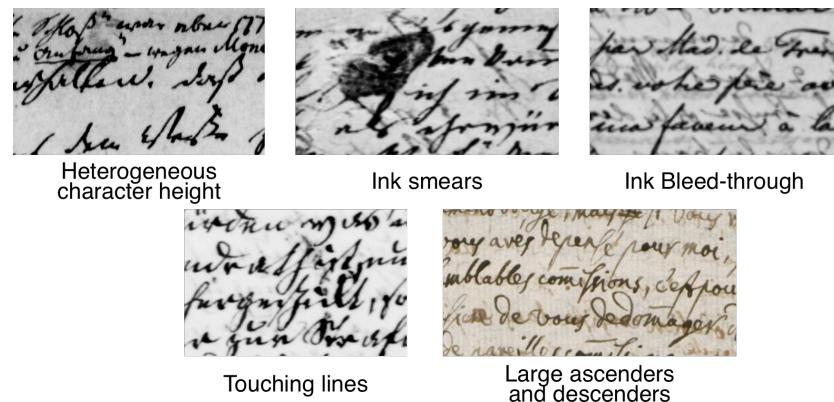


Figure 1. Examples of challenges in a recent historical handwritten text line detection dataset, cBAD.

A learning-free algorithm for text line extraction is proposed in [17] where text lines are detected using multiscale second derivative of Gaussian filters and extracted using an energy minimization (EM) function. The present article extends this approach for detecting text lines of challenging historical documents. The new method proposes a robust character height estimation even with the presence of noise and chains of touching characters among multiple text lines. In addition, the extraction phase includes a mechanism to split the consequently touching text lines. Apart from these, we mathematically formulate automatic scale selection together with second derivative of anisotropic Gaussian and thoroughly present and discuss the evaluation of parameters.

2. Related Work

Text line extraction approaches can be classified into three categories: top-down, bottom-up, and hybrid. Top-down approaches partition document images into text lines based on global features. Bottom-up approaches group pixels or components based on local features to form text lines. Hybrid approaches combine top-down and bottom-up techniques.

2.1. Top-Down Approaches

Top-down approaches are mainly based on projection profile, Hough transform, smearing, or seam carving. Projection profile sums pixel values along the horizontal axis for each y axis value of a binary image to determine locations of text lines.

Projection profile is commonly used for simple document images [18] but can also be adapted for gray scale images [19] and slightly skewed lines [20].

Hough transform calculates parameters of linear structures produced by the text components in the document image. Ref. [21] used Hough transform to find the global orientation of a document image and apply projection profile along this orientation. Hough transform can also be applied to the centroids of connected components and directly align them as text lines [22]. Hough transform is robust for dealing with skewed straight lines but entails high computational cost.

Smearing fills the white space between the consecutive black pixels along the same direction if their distance is within a predefined threshold [23]. Smearing is sensitive to overlapping text strokes, therefore, ref. [24] smeared background pixels running through the overlapping strokes to build line separators. Later on, ref. [25] adapted smearing to skewed lines by applying it in a strip-wise fashion. The main difficulty of smearing is determining the optimum threshold and is dealt in [26] by using steerable directional filters.

Seam carving computes the path of minimum energy cost from one end of the image to another. Ref. [27] employed medial seams for text line extraction of binary document images using signed distance transform as the energy map. Subsequently, they improved the method for gray scale documents using geodesic distance transform as energy map [16].

2.2. Bottom-Up Approaches

Top-down approaches process the document images at global level, which is problematic when the document does not have a Manhattan layout. Therefore, bottom-up approaches process document images at local level and do not assume straight lines. They group the elements into text lines. Elements can be pixels, super-pixels, or connected components. The counterpart of this is isolation of local elements, which is complicated for touching components across consecutive lines. Bottom-up approaches are mainly based on clustering or classification.

Clustering algorithms group elements according to their features in an unsupervised manner [28]. They can be applied both to the binarized document images [13,29] as well as to the gray scale document images [30,31]. The above mentioned works clustered super pixels into words and locally join them to form the text lines. Recently, ref. [6] clustered super pixels into text lines in a greedy manner. As a result, this approach is applicable to different datasets without parameter tuning. Clustering algorithms are suitable for heterogeneous document collections; however, the number of clusters has to be selected sensitively.

Classification algorithms classify the elements according to their features in a supervised manner. They are robust to noisy and transformed images but require a large amount of annotated data for training. Early methods used nonconvolutional classifiers with hand crafted features [12,32,33]. Recent methods are inspired by convolutional neural networks, which have proven to be efficient. Ref. [34] used recurrent neural network to segment isolated paragraphs. Ref. [35] used convolutional neural network, first to classify page pixels as paragraph, and then to classify paragraph pixels as text line or non text line. Text line extraction from full page is studied as a problem of predicting the bounding box around the text lines [36,37]. Sequential nature of these methods limit them to being used by sliding windows on horizontal text lines. Pixel classification in a sliding window fashion is not desirable due to redundant and expensive computation of overlapping areas in the sliding windows. As a remedy, dense prediction has been successfully used for text line segmentation of handwritten documents [5,8].

2.3. Hybrid Approaches

Hybrid approaches combine the strengths of top-down and bottom-up approaches while reducing the weaknesses of each one. Ref. [38] calculates the starting point and the skew of text lines with projection profile, and then uses them to search a piecewise linear separating path. Ref. [39] globally estimates coarse text lines and locally reassigns misclassified elements to split the touching text lines. Ref. [40] focuses on solving skewed and touching text lines using a combination of clustering connected components and projection profile analysis.

3. Scientific Background

In this section we present notations and definitions of scale-space representation with automatic scale selection, component tree, and energy minimization via graph cuts.

3.1. Scale-Space Representation with Automatic Scale Selection

The notion of scale is important when processing unknown image structures by automatic methods. This problem can be approached by representing image structures at different scales, so called scale-space representation [41]. However, scale-space representation does not address the problem of how to select local appropriate scales. Therefore, we use automatic scale selection [42] that adapts to the local scales of image structures.

Scale-space representation together with automatic scale selection apply to a large class of differential image descriptors. We adapt this for detecting blobs, regions brighter than their surroundings, and from text lines and term this procedure as blob line detection with automatic scale selection.

Given n scale values $(\sigma_1, \sigma_2, \dots, \sigma_n)$, scale-space representation is a stack of n layers where the i th layer is the convolution of image by Laplacian with the scale of σ_i (Figure 2). Spatially, the Laplacian response will be maximum at the center of blob when the scale is matched with the scale of the text line. However, Laplacian response decreases as the scale σ increases. To eliminate this decay, Laplacian response is multiplied by σ^2 and is called scale-normalized Laplacian. Then, automatic scale selection assigns the maximum scale-normalized convolution response to each pixel of the image.

Normally, the differential descriptor for forming a blob is the Laplacian of an isotropic Gaussian (LoG) [42]. However, we define a scale-space representation using Laplacian of anisotropic Gaussians, because a text line is elongated and has different scales along the two coordinate directions. In the following sections we define the mathematical formulas for scale-space of Laplacian of an anisotropic Gaussians and automatic scale selection using scale-normalized Laplacian.

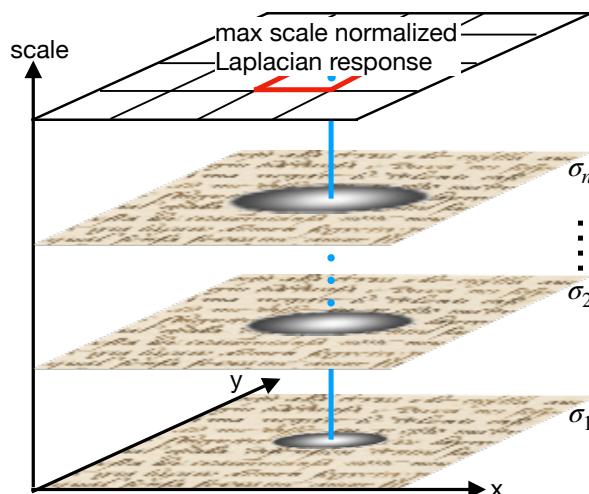


Figure 2. Scale-space representation with automatic scale selection is achieved by convolving the image with different scales $(\sigma_1, \sigma_2, \dots, \sigma_n)$ and assigning the maximum scale-normalized convolution response to each pixel of the image.

3.1.1. Scale-Space of Anisotropic Gaussians

Given a continuous image $I : \mathbb{R}^2 \rightarrow \mathbb{R}$, its scale-space of anisotropic Gaussians $L : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ is defined by following convolution:

$$L(x, y; \sigma_x, \sigma_y) = I(x, y) * g(x, y; \sigma_x, \sigma_y), \quad (1)$$

where $g : \mathbb{R}^2 \times \mathbb{R}^2 \rightarrow \mathbb{R}$ denotes the anisotropic Gaussian kernel

$$g(x, y; \sigma_x, \sigma_y) = \frac{1}{2\pi\sigma_x\sigma_y} e^{-\left(\frac{x^2}{2\sigma_x^2} + \frac{y^2}{2\sigma_y^2}\right)}. \quad (2)$$

In this representation $*$ is the convolution operator, σ_x is the scale parameter in horizontal direction, and σ_y is the scale parameter in vertical direction.

3.1.2. Scale-Space of Laplacian of Anisotropic Gaussians

Scale-space of Laplacian of anisotropic Gaussians is computed by differentiating the scale-space of anisotropic Gaussians with respect to x and y two times

$$L_{x^2y^2}(x, y; \sigma_x, \sigma_y) = \partial_{x^2y^2} L(x, y; \sigma_x, \sigma_y), \quad (3)$$

or, equivalently, by convolving the image with Laplacian of anisotropic Gaussians

$$L_{x^2y^2}(x, y; \sigma_x, \sigma_y) = I(x, y) * [g_{x^2y^2}(x, y; \sigma_x, \sigma_y)]. \quad (4)$$

Given the anisotropic Gaussian in Equation (2), its Laplacian is defined by

$$g_{x^2y^2} = g_{x^2} + g_{y^2}, \quad (5)$$

where

$$g_{x^2} = \left(\frac{x^2 - \sigma_x^2}{\sigma_x^4} \right) \cdot g(x, y; \sigma_x, \sigma_y), \quad (6)$$

and

$$g_{y^2} = \left(\frac{y^2 - \sigma_y^2}{\sigma_y^4} \right) \cdot g(x, y; \sigma_x, \sigma_y). \quad (7)$$

3.1.3. Automatic Scale Selection

In scale-space representation, the amplitude of the Laplacian in Equation (3) decreases with scale. Based on this phenomena, automatic scale selection states that local extreme over scales of scale-normalized Laplacian

$$L_{x^2y^2}^{\sigma-norm} = (\sigma_x \sigma_y) \cdot L_{x^2y^2}(x, y; \sigma_x, \sigma_y) \quad (8)$$

corresponds to the significant structures that are regions brighter than their surroundings.

3.2. Component Tree

Component tree [43] organizes the connected components of level sets in a tree structure. Let \mathcal{C}_t be the set of connected components obtained by thresholding with threshold t . The nodes in a component tree correspond to the components in \mathcal{C}_t for varying values of the threshold t . The root of the tree is the member of $\mathcal{C}_{t_{\min}}$, where t_{\min} is chosen such that $|\mathcal{C}_{t_{\min}}| = 1$. Level ℓ in the tree correspond to $\mathcal{C}_{t_{\min} + \ell d}$, where d is a parameter that determines the step size for the tree. We use $d = 1$ in all the experiments. There is an edge between $C_i \in \mathcal{C}_t$ and $C_j \in \mathcal{C}_{t+1}$ if and only if $C_j \subseteq C_i$. The maximal threshold t_{\max} used in the tree construction is simply the maximal value in the map. Figure 3 illustrates the above definitions.

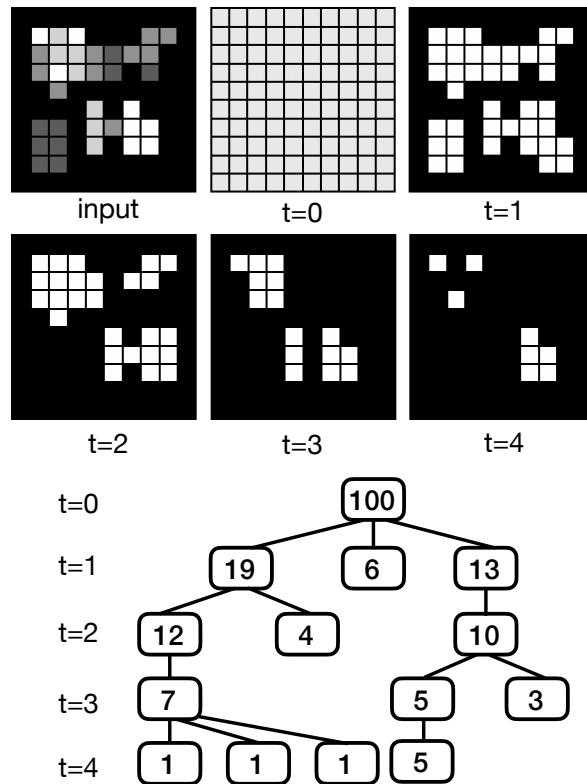


Figure 3. A gray-level input and its successive threshold sets from $t = 0$ to $t = 4$, where $d = 1$. Below is the component tree with each node showing the size of its connected component.

4. Method

The proposed approach utilizes scale-space representation with automatic scale selection to detect blob lines that strike-through the text lines. The detected blob lines are then binarized by component tree algorithm. Finally, energy minimization via graph cuts extracts the text lines with the help of the detected blob lines.

4.1. Blob Line Detection Using Automatic Scale Selection

We informally define a blob to be a connected region that is significantly brighter than its neighborhood. Text line detection aims to derive the blob lines that strike-through the text lines. These blob lines are derived by convolving text lines with the Laplacian of anisotropic Gaussians in a range of scales corresponding to the height range of the characters (Equation (5)).

Character height range σ_x is computed automatically using either way: (1) Component Evolution Map (CEM) [44] estimates character height range by analyzing the height distribution of connected components for each possible grayscale threshold; (2) Mean height of components estimates character height range as $[\mu, (\mu + \sigma)]$, where μ and σ are the average and standard deviation of components' heights in the document.

These Gaussian filters are elongated along the horizontal direction, as such for every value of σ_x , $\sigma_y = e \times \sigma_x$, where e is the elongation rate and its optimal value is experimentally determined in Section 6.2.6. Then, for each pixel we chose the strongest response along the scale-normalized Laplacian given by Equation (8). We investigate the effectiveness of horizontally elongated filters on text lines with varying inclination. The results suggest that this approach appears to be effective in detecting almost horizontal text lines (Figure 4).

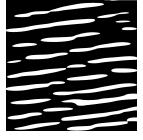
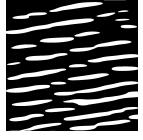
Rotation angle	0°	2°	4°	6°	8°	10°
Rotated image						
Grayscale blobs						
Binary blobs						

Figure 4. The effectiveness of horizontally elongated filters on text lines with varying inclination. The visual results show that as the inclination increases the effectiveness of the method decreases.

4.2. Blob Line Binarization

The blob lines detected by automatic scale selection are represented by a grayscale image as illustrated in Figure 5a. We investigate component tree algorithm to gather the binary blob lines that strike through the text lines.

4.2.1. Component Tree

Component tree [43] binarizes the grayscale blob lines image by fitting k knots linear splines of least squares to each of the candidate blob lines. A candidate blob line is labeled valid if it satisfies both fitting scores, Fitting Score 1 (FS1) and Fitting Score 2 (FS2). FS1 requires the average 1-norm of each blob pixel from the spline to be less than the maximum character height. This condition eliminates the fat blobs that include two consecutive lines. FS2 requires the ratio of the blob area and the sum of the distances of the contour pixels from the spline to be less than 0.9. This condition eliminates the nonconvex blobs with a partial merge (Figure 5c). We traverse component tree in a breadth first search manner with $d = 1$. At each node, if the connected component is valid, it is taken as a binary blob line, and the search along this branch is complete. Otherwise, component is refined by recursively processing the children of the node. Figure 5 illustrates component thresholding procedure which stops at valid blobs.

4.3. Text Line Extraction with Energy Minimization Using Graph Cuts

Binary blob lines that are detected in Section 4.2 correspond to the text line detection phase. Extraction level representation requires further assigning a text line label to each pixel in the document image. We use energy minimization [45] for assigning connected components to text line labels with the help of detected blob lines. It urges to assign components to the label of the closest blob line, while straining to assign closer components to the same label and not to assign any component to spurious blob lines. Let \mathcal{L} be the set of binary blob lines and \mathcal{C} be the set of connected components in the binary document image. Energy minimization finds a labeling f that assigns each component $c \in \mathcal{C}$ to a label $l_c \in \mathcal{L}$, where energy function $E(f)$ has the minimum.

$$E(f) = \sum_{c \in \mathcal{C}} D(c, \ell_c) + \sum_{\{c, c'\} \in \mathcal{N}} d(c, c') \cdot \delta(\ell_c \neq \ell_{c'}) + \sum_{\ell \in \mathcal{L}} h_\ell \quad (9)$$

Energy function has three terms:

1. Data cost is the cost of assigning component c to label ℓ_c . For every $c \in \mathcal{C}$, $D(c, \ell_c)$ is defined as the Euclidean distance between the centroid of c and the blob line ℓ_c . The closer the component to a blob line, the higher the chance the component will be assigned to the label of this blob line.
2. Smoothness cost is the cost of assigning closer components to different labels. Let N be the nearest component pairs. For every $\{c, c'\} \in N$, $d(c, c') = \exp(-\alpha \cdot d_e(c, c'))$ where $d_e(c, c')$ is the Euclidean distance between the centroids of the components c and c' . $\alpha = (2 \langle d_e(c, c') \rangle)^{-1}$ where $\langle \cdot \rangle$ denotes expectation over all pairs of adjacent elements [46]. $\delta(\ell_c \neq \ell_{c'})$ is 1 if the condition inside the parentheses holds, and 0 otherwise.
3. Label cost For every blob line $\ell \in \mathcal{L}$, h_ℓ is defined as $\exp(2 \cdot r_\ell)$ where r_ℓ is the normalized number of foreground pixels overlapping with blob line ℓ . The higher the label cost, the higher the probability of discarding the blob line as spurious line.

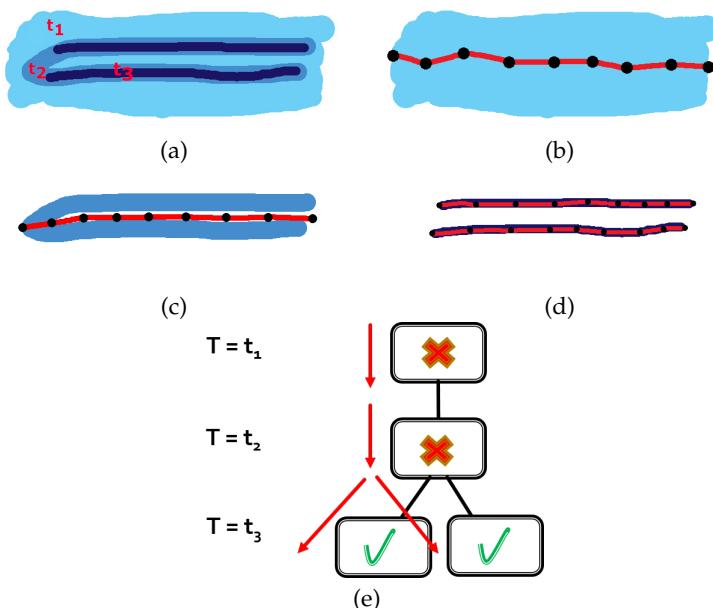


Figure 5. Illustration of thresholding on a synthetic component. (a) Component is composed of three gray levels with ascending values t_1, t_2, t_3 . (b–d) Child components from thresholding the parent component with $T \geq t_1, t_2, t_3$ respectively and their approximating splines. (e) Bread first traversal of the component tree.

4.4. Merging Broken Blob Lines

Once the EM removes the spurious blob lines, there may still exist the problem of broken blob lines. Given the binary image of blob lines, for each blob line we extract its left and right endpoints and define its direction as the vector connecting the left endpoint to the right endpoint. Two adjacent blob lines are merged if (1) the direction of the vector connecting the right of the first component to the left of the second one falls between the direction of the two blob lines, (2) their vertical distance is less than the maximum character height.

4.5. Splitting Touching Characters

Text line extraction results contain unsegmented touching characters from adjacent text lines. To check whether a connected component c overlaps more than one blob line, we relabel c by assigning each pixel in c to the label of nearest blob line.

5. Evaluation

This paper proposes a learning-free text line segmentation method for challenging historical handwritten documents as such cBAD dataset (<https://zenodo.org/record/835441>). In addition, we also evaluated the method on another recent handwritten text line segmentation dataset, DIVA-HisDB (<http://diuf.unifr.ch/hisdoc/diva-hisdb>). For each dataset, we use only its test set because our method is not learning based. The proposed method's output is defined as pixel labels but it is manipulated according to the ground truth definition of each dataset.

5.1. Cbad Dataset

cBAD dataset [9] contains 539 document images from 7 different archives. This dataset's ground truth is defined as baselines whereas our method extracts pixel labels. To find baselines from pixel labels, first we get tight polygons around the pixel labels of text lines. Then, for each bounding polygon we get its lower contour points and iteratively fit a regression line among these points by excluding the outliers (Figure 6). Using the extracted baselines, the performance is measured by means of Precision (P), Recall (R), and F-measure (FM), as described in [9].

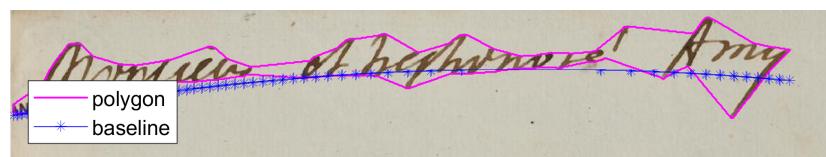


Figure 6. Baseline extraction from bounding polygon.

5.2. Diva-Hisdb Dataset

DIVA-HisDB dataset [47] contains 30 pages from 3 medieval manuscripts. Ground truth is defined as bounding polygons. Therefore, we get the tight polygons around the pixel labels of our output and measure the performance by means of the Intersection over Union (IU) as described in [47].

6. Experiments

6.1. Experimental Setting

In this section we do ablation studies for adapting the learning-free algorithm proposed in [17] to challenging historical handwritten documents. Default parameters of the learning-free algorithm proposed in [17] are like the following:

1. Estimating character height range using CEM as described in Section 4.1
2. Blob line binarization using component tree with both of the fitting scores given in Section 4.2.1 and using number of knots $k = 20$.
3. Merging broken blob lines as described in Section 4.4
4. Binarizing the document image using Otsu thresholding, to be used with the EM function.

cBAD dataset is a relatively large dataset and the proposed algorithm takes around one day to run on the whole dataset. For this reason we randomly sample a representative subset of cBAD dataset, one page from each collection, and do the ablation studies on this subset.

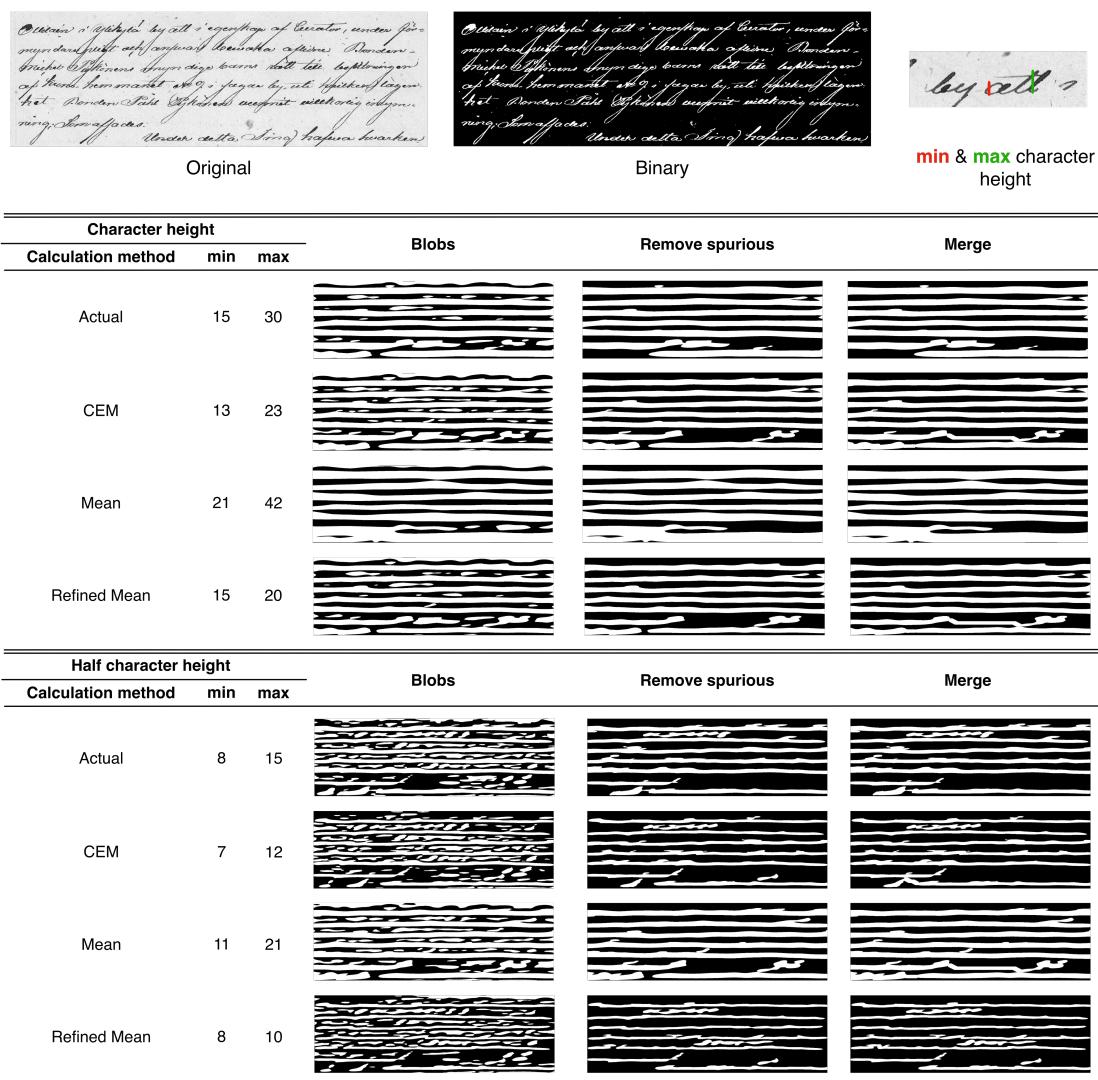
6.2. Text Line Detection Experiments

Text line detection stage is based on automatic scale selection by convolving the document image with second derivative of anisotropic Gaussian filters with a range of scales estimated as the character height range of the text in the document. The output of this filtering is a response map, a grayscale image of the detected blobs. This grayscale image is binarized to gather the binary blob lines that strike-through the text lines.

6.2.1. Effect of Character Height Range

Character height range that is used with the automatic scale selection inevitably effects the detected blob lines. Ref. [17] uses CEM for character height range estimation, but many touching text lines mislead the algorithm. To tackle this problem, we experiment with mean height of connected components (Section 4.1). Recognizing that the noise and long chain of touching text lines is also misleading, we also use refined mean estimation by excluding the components that are bigger than a maximum threshold (100 pixels) and smaller than a minimum threshold (10 pixels). Considering that the text is mostly located in the center region of a document, we also exclude the outer 2% of the document image. We found that the results are less than ideal and further experiment with half of the above range estimations is needed.

Figure 7 shows that there are no sharp changes between the results of different height estimations. This is valid for such sample document without severe noise. As can be seen, the half values of range values provide thinner blob lines that are more apart while provoking spurious blob lines. Qualitative analysis revealed a considerably higher performance and lower run time of the half-refined mean estimation for character height range (Figure 8).



* The minimum and maximum character height values are measured in pixels

Figure 7. Visualization of the effect of character height estimation method. Apparently the half values of range values provide thinner blob lines that are more apart while provoking spurious blob lines.

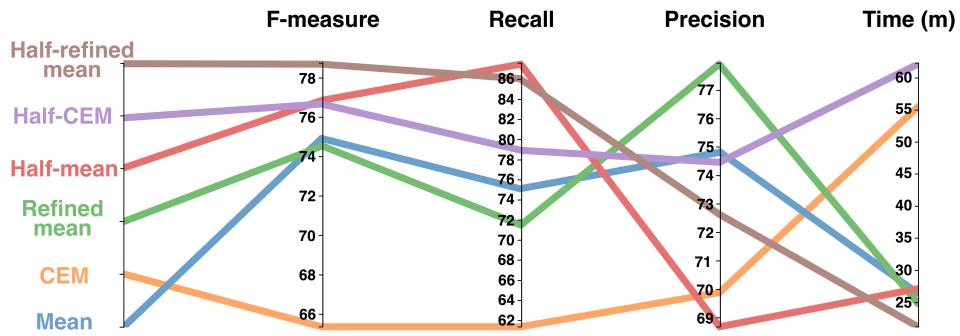


Figure 8. Effect of height estimation method on the performance.

6.2.2. Effect of Merging Blob Lines

In case of a big gap between words, a blob line can be disconnected. In this experiment we tested to which extent merging the blob line improves the results. We found that merging slightly increases the effectivity (Table 1).

Table 1. Effect of merging blob lines.

	Precision	Recall	F-measure	Time (m)
Merged	72.63	85.97	78.74	21.20
Nonmerged	70.48	85.22	77.16	20.90

6.2.3. Effect of Blob Line Fitting Scores

The quality of detected blob lines is a key factor that influences the final performance. A blob line is preferable as plain, continuous, and apart as possible. Ref. [17] defines this preferability by two fitting scores (Section 4.2.1). The results using only the FS1 provide compelling evidence that nonconvex blobs with a partial merge are negligible (Table 2).

Table 2. Effect of using only the FS1.

Fitting Score	Precision	Recall	F-measure	Time (m)
FS1 and FS2	72.63	85.97	78.74	21.20
FS1	74.66	85.50	79.71	19.96

6.2.4. Effect of Fitting Score Threshold

The results in Section 6.2.3 indicate that using only the FS1 is reasonable. FS1 eliminates the fat blobs that include two consecutive lines. [17] conditions the FS1 to be less than the upper character height. Relaxing this condition decreases the run time whereas stressing it increases the run time without an evident performance gain (Figure 9).

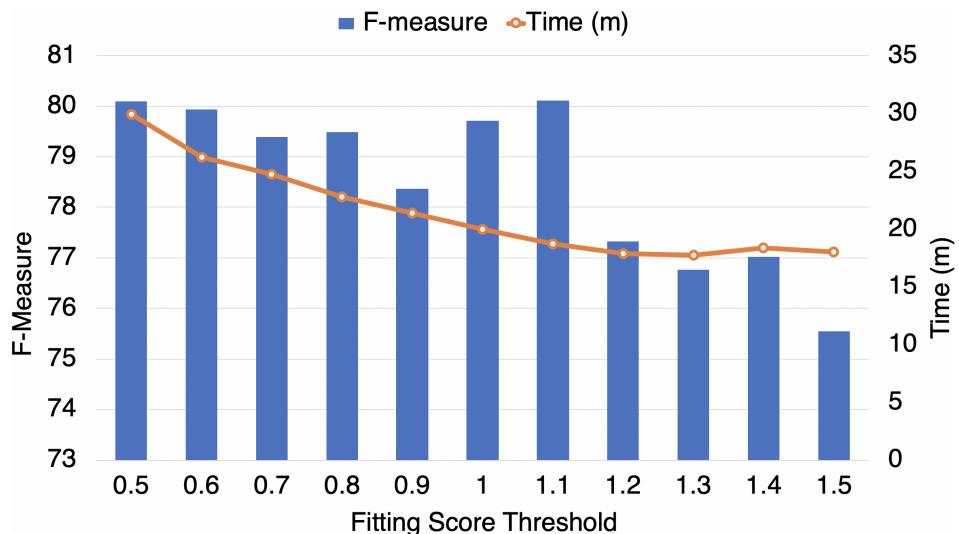


Figure 9. Effect of fitting score threshold (t) on performance and time.

6.2.5. Effect of Number Of Knots

The aim of using splines is their ability to fit on curved blob lines as well as straight blob lines. Ref. [17] uses 20 knots to fit the linear splines. We investigate different number of knots and show their effect on the method's performance. Results are presented in Figure 10. As expected, varying the number of knots does not impact the performance measure, because almost all text lines in the evaluation datasets are horizontal. However, the spline fitting makes the proposed method suitable for curved text lines [48]. It is apparent that recall values follow a higher trend than the precision values, possibly due to the spurious blob lines that could not be eliminated by the label cost because of crowded ascenders and descenders overlapping these spurious lines.

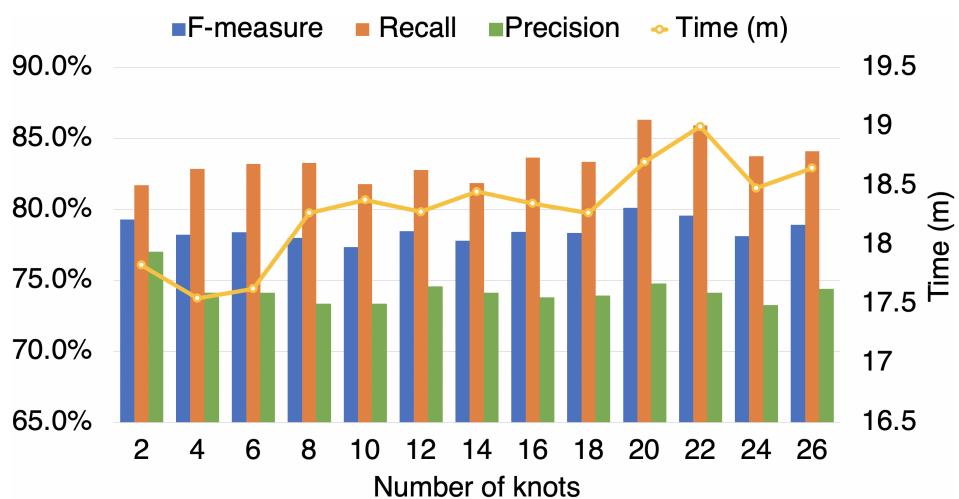


Figure 10. Effect of number of knots (k) on performance and time.

6.2.6. Effect of Elongation Rate

The proposed method assumes the text lines are almost horizontal. Therefore, convolution of a text line with the second derivative of an anisotropic Gaussian elongated along the horizontal direction generates a blob line that strike-through the text line. Figure 11 shows the blob lines resultant from two marginal cases: $e = 1$ and $e = 7$. There is a trade-off between the marginal cases. As e decreases as the blob lines are tight-fitting, e increases as the blob lines are loose-fitting. Tight-fitting misses the spatial context and may consider two touching text lines as one (Figure 11b). Loose-fitting grasps

the unrelated spatial context and may generate spurious blob lines (Figure 11f). From the performed experiments, the optimal point seems to be when $e = 2$ (Table 3).

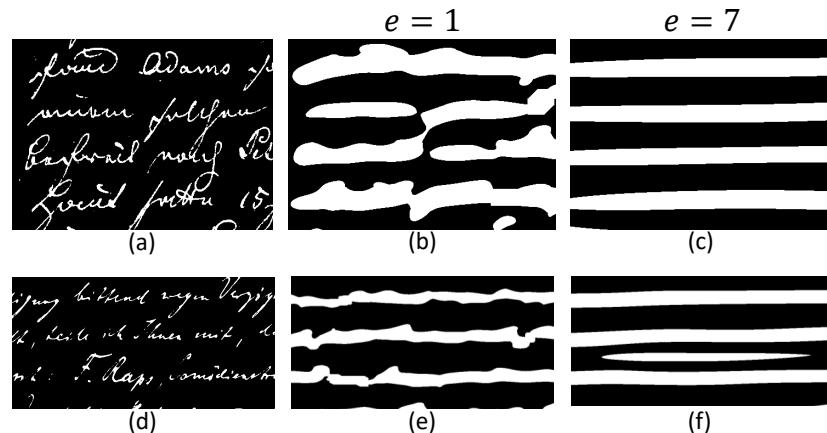


Figure 11. Effect of elongation rate (e) on blob lines. (a,d) are two different binary input images. (b,e) are the resultant blob lines when $e = 1$. (c,f) are the resultant blob lines when $e = 7$. As e decreases the spatial context is missed and two touching text lines are considered as one (b). As e increases the unrelated spatial context is grasped and a spurious blob line is generated (f).

Table 3. Effect of elongation rate on performance and time.

e	Precision	Recall	F-measure	Time (m)
1	75.38	82.09	78.59	22.96
2	75.71	84.64	79.93	21.11
3	74.19	85.62	79.49	22.80
4	71.92	87.31	78.87	23.30
5	72.31	86.97	78.96	23.05
6	71.09	86.30	77.96	25.83
7	71.44	86.18	78.12	29.76

6.2.7. Conclusion about Text Line Detection Experiments

We have demonstrated that alternative values of parameters for text line detection produce slightly different results, with the exception of the character height estimation method. Half values of the refined-mean of the connected components achieve a superior performance over the default character estimation method, CEM.

6.3. Text Line Extraction Experiments

In all the text line extraction experiments, we use half values of refined-mean for character height estimation. FS1 with a threshold of $t = 1.1$ merge the binarized blob lines. We set the number of knots to $k = 20$ and elongation rate to $e = 2$.

The energy function is formulated in the binary image domain, thereby, any noise pixel remained from the binarization process breaks down the accuracy. Therefore, we examined the effect of two binarization methods, Otsu [49] and Bar-Yosef et al. [50]. Both binarization results suffer either from losing parts of text lines or from noise parts of text lines. However, the performance is not sensitive to the binarization method (Table 4). On the other hand, splitting the touching characters among two text lines lead to a significant improvement (Table 4).

Table 4. Effect of using different binarization methods, Otsu and Bar-Yosef et al., together with the effect of splitting the touching characters among text lines.

	Precision	Recall	F-measure	Time (m)
Nonsplit	75.71	84.64	79.93	21.11
Split and Otsu	81.74	92.25	86.68	23.95
Split and Bar-Yosef et al.	84.17	88.68	86.45	24.31

7. Results

We present results on two recent historical handwritten datasets, DIVA-HisDB and cBAD. For text line detection, we use half of refined-mean for character height estimation, FS1 with a threshold of $t = 1.1$, merge the binarized blob lines, and set the number of knots to $k = 20$ and elongation rate to $e = 2$. For text line extraction, we use Bar-Yosef et al. binarization and split the touching characters among text lines.

DIVA-HisDB is a historical document dataset with a high number of consequently touching text lines and heterogeneous in terms of text line height. The input to our algorithm is the ground truth of layout analysis task of the competition. Table 5 shows our results compared to two learning-free algorithms. One is the best performing algorithm in the ICDAR2017 Competition on Layout Analysis for Challenging Medieval Manuscripts [47]. The other [51] is an improved version of the best performing algorithm in the competition. The proposed algorithm can reach a line IU of 100%. The degradation in pixel IU mostly comes from the coarse splitting of touching characters among text lines. [51] performs perfect in terms of the overall Line IU at 100%. The advantage of the proposed algorithm is that it can detect text lines irrespective of touching components across the text lines and is robust to complex layouts. However, its performance is highly dependent on the correct estimation of average character height (Figure 8). Whereas the desired property of the work proposed by [51] is that their results are fairly stable in respect to the varying values of parameters. However, this method is not appropriate for complex layouts with heterogeneous text line lengths or separate text line groups because it labels the text lines with the intuition that all components belonging to the same text line have the same amount of seams below them.

Table 5. Comparison of our method with state-of-the-art results on DIVA-HisDB dataset.

Learning-Free	CB55		CSG18		CSG863		Overall	
	Pixel IU	Line IU						
[51]	-	-	-	-	-	-	97.22	100.0
[47]	96.67	98.36	96.93	96.91	97.54	99.27	97.05	97.86
Our method	97.04	100	98.14	98.38	97.31	100	97.50	99.46

cBAD is a historical document dataset with a larger evaluation set and contains documents with varying layouts, originating from different time periods and locations. It is heterogeneous in terms of text line height and length. Input documents are gray-scale images. Some examples of pages with extracted bounding polygons can be seen in Figure 12. Common error cases arise from a group of touching characters that mislead the algorithm to merge the text lines (Figure 12b). Another type of error occurs due to the artifacts in the frame of the page images. We compare our results with the results obtained in cBAD competition. Table 6 shows precision, recall, and F-measure values of the participants. The methods are grouped as learning-based and learning-free. The best performance with an F-measure value of 97.70 is for a learning-based method, DMRZ [9]. Two learning-free methods, our method and IRISA [9], achieve close F-measure values. The precision and recall values indicate that the IRISA method splits baselines more precisely but also misses more baselines compared to our method.

Table 6. Comparison of our results with the cBAD competition results and another recent work.

	Method	Precision	Recall	F-measure
Learning-based	DMRZ [9]	97.30	97.00	97.10
	UPVLC [9]	93.70	85.50	89.40
	BYU [9]	87.80	90.70	89.20
	LITIS [9]	78.00	83.60	80.70
Learning-free	IRISA [9]	88.30	87.70	88.00
	[52]	74.70	92.60	82.70
	Our method	82.09	91.13	86.38

Finally, in Figure 13 we present the results obtained during the RASM2018 competition [53] for historical scientific manuscripts in Arabic. We submitted output text lines and results were evaluated by the competition committee using a success rate defined in [53]. The results and the comparison with other participator methods are presented in Table 7.

Table 7. RASM2018 competition results in terms of success rate defined by the competition.

Tesseract 3	Tesseract 4	FRE11	Our Method	RDI
28.80	44.20	43.20	67.70	81.60

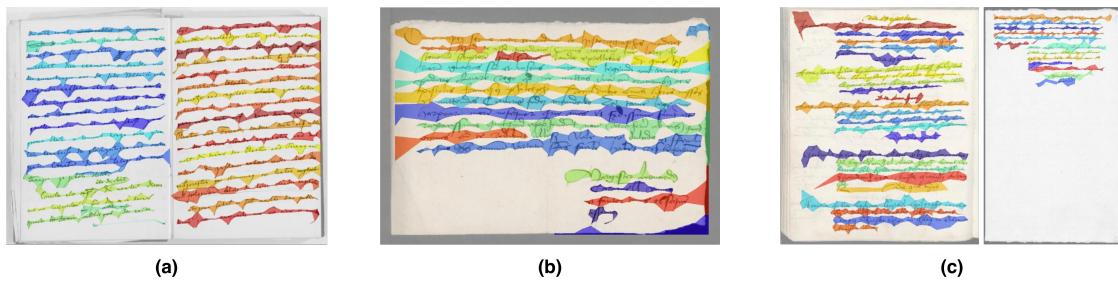


Figure 12. Example results from cBAD dataset. Slightly skewed text lines can be extracted (a). Common error cases arise from group of touching characters among two different text lines but that have relatively big gap with their neighbours in the same text line (b). Text lines with heterogeneous lengths can be extracted (c).

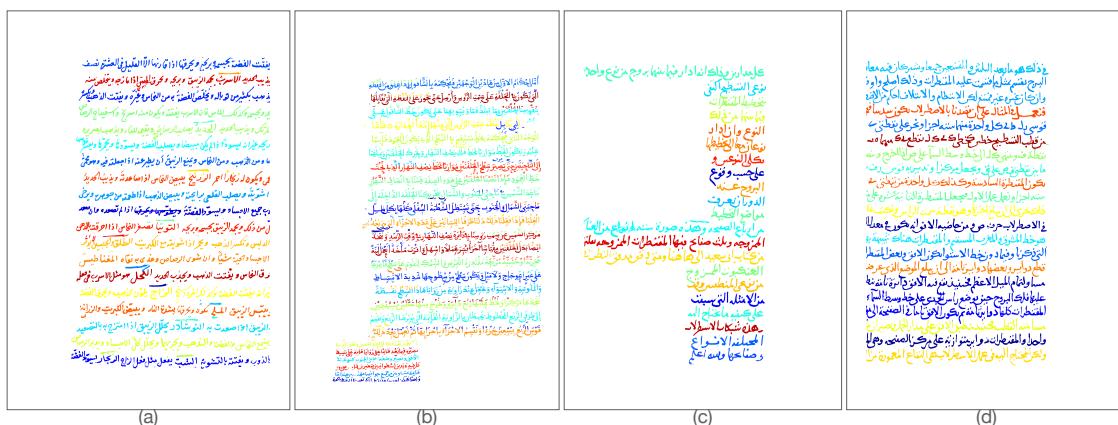


Figure 13. Example results from RASM2018 dataset. Some errors occur due to non-textual elements such as the underlines in (a). Some errors occur because of satellite textual elements are grouped separately (b). Text lines with heterogeneous lengths (c). Touching characters are split but not very adequately (d).

8. Conclusions

We presented a learning-free text line detection and extraction method for historical handwritten document images. Historical handwritten documents contain complex layouts with varying text line heights and lengths, touching characters, crowd of ascenders and descenders. Learning-based algorithms are currently an increased trend in text line segmentation of historical handwritten documents; however, they require labeling effort for training. The proposed method can fairly detect the blob lines that strike-through text lines with arbitrary heights and lengths by convolving the input image with second derivative of anisotropic Gaussian using automatic scale selection. On the other hand, EM formulation removes spurious blob lines and assigns the connected components to the closest detected blob line or to the label of the closest component. Ablation study shows that the method is not sensitive to the parameters except character height estimation. Another limitation of the method is that it can deal with severely skewed text lines. This limitation can be overcome using multiorientated and multiscale anisotropic Gaussians. The results on three different datasets are not always the state of the art but achieved using the same experimental setting. In the future, we plan to develop an unsupervised machine learning method for text line segmentation, using the blob lines detected by the proposed method as an annotation of the text lines.

Author Contributions: Conceptualization, J.E.-S.; methodology, J.E.-S.; software, R.C. and B.K.B.; validation, B.K.B.; formal analysis, B.K.B.; investigation, B.K.B.; data curation, B.K.B.; writing—original draft preparation, B.K.B.; writing—review and editing, A.D. and I.R.; visualization, A.D. and B.K.B.; supervision, J.E.-S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Acknowledgments: This research was supported by the Frankel Center for Computer Science in Ben-Gurion University of the Negev.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Manmatha, R.; Srimal, N. Scale space technique for word segmentation in handwritten documents. In *International Conference on Scale-Space Theories in Computer Vision*; Springer: Berlin/Heidelberg, Germany, 1999; pp. 22–33.
2. Varga, T.; Bunke, H. Tree structure for word extraction from handwritten text lines. In Proceedings of the Eighth International Conference on Document Analysis and Recognition (ICDAR'05), Seoul, Korea, 31 August–1 September 2005; pp. 352–356.
3. Graves, A.; Liwicki, M.; Fernández, S.; Bertolami, R.; Bunke, H.; Schmidhuber, J. A novel connectionist system for unconstrained handwriting recognition. *IEEE Trans. Pattern Anal. Mach. Intell.* **2008**, *31*, 855–868. [[CrossRef](#)] [[PubMed](#)]
4. Liwicki, M.; Graves, A.; Bunke, H. Neural networks for handwriting recognition. In *Computational Intelligence Paradigms in Advanced Pattern Classification*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 5–24.
5. Renton, G.; Soullard, Y.; Chatelain, C.; Adam, S.; Kermorvant, C.; Paquet, T. Fully convolutional network with dilated convolutions for handwritten text line segmentation. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2018**, *21*, 177–186. [[CrossRef](#)]
6. Gruenning, T.; Leifert, G.; Strauss, T.; Labahn, R. A robust and binarization-free approach for text line detection in historical documents. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 1, pp. 236–241.
7. Oliveira, S.A.; Seguin, B.; Kaplan, F. dhSegment: A generic deep-learning approach for document segmentation. In Proceedings of the 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), Niagara Falls, NY, USA, 5–8 August 2018; pp. 7–12.
8. Kurar Barakat, B.; Droby, A.; Kassis, M.; El-Sana, J. Text Line Segmentation for Challenging Handwritten Document Images using Fully Convolutional Network. In Proceedings of the 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), Niagara Falls, NY, USA, 5–8 August 2018; pp. 374–379.

9. Diem, M.; Kleber, F.; Fiel, S.; Grüning, T.; Gatos, B. cbad: Icdar2017 competition on baseline detection. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 1, pp. 1355–1360.
10. Gatos, B.; Stamatopoulos, N.; Louloudis, G. ICDAR2009 handwriting segmentation contest. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2011**, *14*, 25–33. [[CrossRef](#)]
11. Stamatopoulos, N.; Gatos, B.; Louloudis, G.; Pal, U.; Alaei, A. ICDAR 2013 handwriting segmentation contest. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 25–28 August 2013; pp. 1402–1406.
12. Baechler, M.; Liwicki, M.; Ingold, R. Text line extraction using DMLP classifiers for historical manuscripts. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition (ICDAR), Washington, DC, USA, 25–28 August 2013; pp. 1029–1033.
13. Diem, M.; Kleber, F.; Sablatnig, R. Text line detection for heterogeneous documents. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition (ICDAR), Washington, DC, USA, 25–28 August 2013; pp. 743–747.
14. Bukhari, S.S.; Shafait, F.; Breuel, T.M. Script-independent handwritten textlines segmentation using active contours. In Proceedings of the 2009 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 26–29 July 2009; pp. 446–450.
15. Cohen, R.; Asi, A.; Kedem, K.; El-Sana, J.; Dinstein, I. Robust text and drawing segmentation algorithm for historical documents. In Proceedings of the 2nd International Workshop on Historical Document Imaging and Processing, Washington, DC, USA, 24 August 2013; pp. 110–117.
16. Saabni, R.; Asi, A.; El-Sana, J. Text line extraction for historical document images. *Pattern Recognit. Lett.* **2014**, *35*, 23–33. [[CrossRef](#)]
17. Cohen, R.; Dinstein, I.; El-Sana, J.; Kedem, K. Using scale-space anisotropic smoothing for text line extraction in historical documents. In Proceedings of the International Conference Image Analysis and Recognition, Vilamoura, Portugal, 22–24 October 2014; pp. 349–358.
18. Antonacopoulos, A.; Karatzas, D. Document image analysis for World War II personal records. In Proceedings of the First International Workshop on Document Image Analysis for Libraries, Palo Alto, CA, USA, 23–24 January 2004; pp. 336–341.
19. Kesiman, M.W.A.; Burie, J.C.; Ogier, J.M. A new scheme for text line and character segmentation from gray scale images of palm leaf manuscript. In Proceedings of the 15th International Conference on Frontiers in Handwriting Recognition 2016, Shenzhen, China, 23–26 October 2016; pp. 325–330.
20. Ouwayed, N.; Belaïd, A. A general approach for multi-oriented text line extraction of handwritten documents. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2012**, *15*, 297–314. [[CrossRef](#)]
21. Shapiro, V.; Gluhchev, G.; Sgurev, V. Handwritten document image segmentation and analysis. *Pattern Recognit. Lett.* **1993**, *14*, 71–78. [[CrossRef](#)]
22. Gatos, B.; Louloudis, G.; Stamatopoulos, N. Segmentation of historical handwritten documents into text zones and text lines. In Proceedings of the 2014 14th International Conference on Frontiers in Handwriting Recognition (ICFHR), Heraklion, Greece, 1–4 September 2014; pp. 464–469.
23. Wong, K.Y.; Casey, R.G.; Wahl, F.M. Document analysis system. *IBM J. Res. Dev.* **1982**, *26*, 647–656. [[CrossRef](#)]
24. Shi, Z.; Govindaraju, V. Line separation for complex document images using fuzzy runlength. In Proceedings of the First International Workshop on Document Image Analysis for Libraries, Palo Alto, CA, USA, 23–24 January 2004; p. 306.
25. Alaei, A.; Pal, U.; Nagabhushan, P. A new scheme for unconstrained handwritten text-line segmentation. *Pattern Recognit.* **2011**, *44*, 917–928. [[CrossRef](#)]
26. Swaileh, W.; Mohand, K.A.; Paquet, T. Multi-script iterative steerable directional filtering for handwritten text line extraction. In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, 23–26 August 2015; pp. 1241–1245.
27. Saabni, R.; El-Sana, J. Language-independent text lines extraction using seam carving. In Proceedings of the 2011 International Conference on Document Analysis and Recognition (ICDAR), Beijing, China, 18–21 September 2011; pp. 563–568.
28. Yin, F.; Liu, C.L. Handwritten Chinese text line segmentation by clustering with distance metric learning. *Pattern Recognit.* **2009**, *42*, 3146–3157. [[CrossRef](#)]

29. Roy, P.P.; Pal, U.; Lladós, J. Text line extraction in graphical documents using background and foreground information. *Int. J. Doc. Anal. Recognit. (IJDAR)* **2012**, *15*, 227–241. [[CrossRef](#)]
30. Garz, A.; Fischer, A.; Sablatnig, R.; Bunke, H. Binarization-free text line segmentation for historical documents based on interest point clustering. In Proceedings of the 2012 10th IAPR International Workshop on Document Analysis Systems (DAS), Gold Coast, Australia, 27–29 March 2012; pp. 95–99.
31. Garz, A.; Fischer, A.; Bunke, H.; Ingold, R. A binarization-free clustering approach to segment curved text lines in historical manuscripts. In Proceedings of the 2013 12th International Conference on Document Analysis and Recognition (ICDAR), Washington, DC, USA, 25–28 August 2013; pp. 1290–1294.
32. Mehri, M.; Héroux, P.; Gomez-Kramer, P.; Boucher, A.; Mullot, R. A pixel labeling approach for historical digitized books. In Proceedings of the 12th International Conference on Document Analysis and Recognition, Washington, DC, USA, 25–28 August 2013; pp. 817–821.
33. Chen, K.; Wei, H.; Liwicki, M.; Hennebert, J.; Ingold, R. Robust text line segmentation for historical manuscript images using color and texture. In Proceedings of the 2014 22nd International Conference on Pattern Recognition (ICPR), Stockholm, Sweden, 24–28 August 2014; pp. 2978–2983.
34. Moysset, B.; Kermorvant, C.; Wolf, C.; Louradour, J. Paragraph text segmentation into lines with recurrent neural networks. In Proceedings of the 2015 13th International Conference on Document Analysis and Recognition (ICDAR), Tunis, Tunisia, 23–26 August 2015; pp. 456–460.
35. Pastor-Pellicer, J.; Afzal, M.Z.; Liwicki, M.; Castro-Bleda, M.J. Complete system for text line extraction using convolutional neural networks and watershed transform. In Proceedings of the 2016 12th IAPR Workshop on Document Analysis Systems (DAS), Santorini, Greece, 11–14 April 2016; pp. 30–35.
36. Moysset, B.; Louradour, J.; Kermorvant, C.; Wolf, C. Learning text-line localization with shared and local regression neural networks. In Proceedings of the 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR), Shenzhen, China, 23–26 October 2016; pp. 1–6.
37. Moysset, B.; Kermorvant, C.; Wolf, C. Full-page text recognition: Learning where to start and when to stop. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 1, pp. 871–876.
38. Fischer, A.; Indermühle, E.; Bunke, H.; Viehhauser, G.; Stolz, M. Ground truth creation for handwriting recognition in historical documents. In Proceedings of the 9th IAPR International Workshop on Document Analysis Systems, Boston, MA, USA, 9–11 June 2010; pp. 3–10.
39. Kumar, J.; Kang, L.; Doermann, D.; Abd-Almageed, W. Segmentation of handwritten textlines in presence of touching components. In Proceedings of the 2011 International Conference on Document Analysis and Recognition (ICDAR), Beijing, China, 18–21 September 2011; pp. 109–113.
40. Clausner, C.; Antonacopoulos, A.; Pletschacher, S. A robust hybrid approach for text line segmentation in historical documents. In Proceedings of the 2012 21st International Conference on Pattern Recognition (ICPR), Tsukuba, Japan, 11–15 November 2012; pp. 335–338.
41. Witkin, A.P. Scale-space filtering. In *Readings in Computer Vision*; Elsevier: Amsterdam, The Netherlands, 1987; pp. 329–332.
42. Lindeberg, T. Feature detection with automatic scale selection. *Int. J. Comput. Vis.* **1998**, *30*, 79–116. [[CrossRef](#)]
43. Naegel, B.; Wendling, L. A document binarization method based on connected operators. *Pattern Recognit. Lett.* **2010**, *31*, 1251–1259. [[CrossRef](#)]
44. Biller, O.; Rabaev, I.; Kedem, K.; El-Sana, J.J. Evolution maps and applications. *PeerJ Comput. Sci.* **2016**, *2*, e39. [[CrossRef](#)]
45. Delong, A.; Osokin, A.; Isack, H.N.; Boykov, Y. Fast approximate energy minimization with label costs. *Int. J. Comput. Vis.* **2012**, *96*, 1–27. [[CrossRef](#)]
46. Rother, C.; Kolmogorov, V.; Blake, A. “GrabCut” interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph. (TOG)* **2004**, *23*, 309–314. [[CrossRef](#)]
47. Simistira, F.; Bouillon, M.; Seuret, M.; Würsch, M.; Alberti, M.; Ingold, R.; Liwicki, M. Icdar2017 competition on layout analysis for challenging medieval manuscripts. In Proceedings of the 2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR), Kyoto, Japan, 9–15 November 2017; Volume 1, pp. 1361–1370.

48. Kurar Barakat, B.; Cohen, R.; El-Sana, J.; Rabaev, I. VML-MOC: Segmenting a multiply oriented and curved handwritten text line dataset. In Proceedings of the 3rd International Workshop on Arabic and Derived Script Analysis and Recognition (ASAR), Sydney, Australia, 22–25 September 2019; pp. 151–155.
49. Otsu, N. A threshold selection method from gray-level histograms. *IEEE Trans. Syst. Man, Cybern.* **1979**, *9*, 62–66. [CrossRef]
50. Bar-Yosef, I.; Hagbi, N.; Kedem, K.; Dinstein, I. Line segmentation for degraded handwritten historical documents. In Proceedings of the 10th International Conference on Document Analysis and Recognition, Barcelona, Spain, 26–29 July 2009; pp. 1161–1165.
51. Alberti, M.; Vöglin, L.; Pondenkandath, V.; Seuret, M.; Ingold, R.; Liwicki, M. Labeling, cutting, grouping: An efficient text line segmentation method for medieval manuscripts. In Proceedings of the 2019 International Conference on Document Analysis and Recognition (ICDAR), Sydney, Australia, 20–25 September 2019; pp. 1200–1206.
52. Aldavert, D.; Rusiñol, M. Manuscript text line detection and segmentation using second-order derivatives. In Proceedings of the 2018 13th IAPR International Workshop on Document Analysis Systems (DAS), Vienna, Austria, 24–27 April 2018; pp. 293–298.
53. Clausner, C.; Antonacopoulos, A.; McGregor, N.; Wilson-Nunn, D. ICFHR 2018 Competition on Recognition of Historical Arabic Scientific Manuscripts—RASM2018. In Proceedings of the 2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR), Niagara Falls, NY, USA, 5–8 August 2018; pp. 471–476.

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



© 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).