IJDAR
DOI 10.1007/s10032-011-0157-5

ORIGINAL PAPER

A local linear level set method for the binarization of degraded historical document images

David Rivest-Hénault · Reza Farrahi Moghaddam · Mohamed Cheriet

Received: 1 June 2010 / Revised: 28 February 2011 / Accepted: 15 March 2011
© Springer-Verlag 2011

Abstract Document image binarization is a difficult task, especially for complex document images. Nonuniform background, stains, and variation in the intensity of the printed characters are some examples of challenging document features. In this work, binarization is accomplished by taking advantage of local probabilistic models and of a flexible active contour scheme. More specifically, local linear models are used to estimate both the expected stroke and the background pixel intensities. This information is then used as the main driving force in the propagation of an active contour. In addition, a curvature-based force is used to control the viscosity of the contour and leads to more natural-looking results. The proposed implementation benefits from the level set framework, which is highly successful in other contexts, such as medical image segmentation and road network extraction from satellite images. The validity of the proposed approach is demonstrated on both recent and historical document images of various types and languages. In addition, this method was submitted to the Document Image Binarization Contest (DIBCO'09), at which it placed 3rd.

Keywords Level set method · Local linear modeling · Multi-level classifiers · Binarization

D. Rivest-Hénault · R. Farrahi Moghaddam (✉) · M. Cheriet
Synchromedia Laboratory for Multimedia Communication in Telepresence, École de Technologie Supérieure, 1100 Notre-Dame West, Montréal, QC H3C 1K3, Canada
e-mail: rfarrahi@synchromedia.ca; imriss@yahoo.com

D. Rivest-Hénault
e-mail: dhenault@synchromedia.ca

M. Cheriet
e-mail: mohamed.cheriet@etsmtl.ca

1 Introduction

Document image binarization is an essential preprocessing step in many document-related applications, such as cleanup, storage, transmission, and offline analysis and recognition. It may be defined as the process of separating the text pixels from the background. Dramatically reducing the amount of information makes subsequent treatments simpler and more efficient. However, at the same time, any error at this step has a major influence on the performance of the following steps.

Although binarization of simple document images is straightforward, the difficulty quickly increases with document complexity. A large number of document defects may interfere with the binarization process. For handwritten documents, non uniformity of stroke pixel intensities, due to ink smearing and variations in pen pressure, are common problems. For printed documents, certain font shapes with very thin segments may be problematic. Also, many physical characteristics of the original media, along with digitalization artifacts, can complicate the binarization process: paper appearance and texture, stains, bleed-through effect, and various other degradations caused by exposure to humidity or light. As a result of these difficulties, no simple method could emerge as a general solution for the binarization problem. In this respect, the objective of this paper is to introduce a general purpose binarization method, which performs well in a variety of situations.

Global binarization methods [1,2] are interesting as they are parameterless. However, it is well accepted that, in many challenging cases, no global thresholding of pixel intensities could produce an acceptable binarization [3]. More complex schemes must then be considered in order to improve the quality of the output. In this respect, recent document image binarization methods tend to take advantage of more local cues. These approaches can be roughly classified in two main

Published online: 05 April 2011

categories: color/intensity-based [4–7] and edge-based [8,9]. The first category consists of methods which consider variations in color or intensity, while the second includes methods in which abrupt changes across edges are the main source of classification information. The intensity-based methods can be split into two subgroups: threshold-based methods and clustering methods [7]. Niblack's method [5] can be considered as the first local threshold method. However, its outputs tend to be noisy in background-only regions. Sauvola and Pietikinen [4] solved this problem by introducing a generalization of Niblack's method that tries to identify the background regions. Also, multi-scale approaches have been used in many works to differentiate between the text and the interfering patterns [6,10]. The main drawback of the local methods is their high computational cost and theirs tendency to fail to segment the inner parts of thick characters. Grid-based modeling, which was introduced in [6], reduces the computational time considerably. The integral image technique has also been used to reduce the computational cost [11]. Grid-based modeling also provides an opportunity to transform global thresholding methods into their local versions. For example, an adaptive method based on Otsu's global thresholding method has been introduced in [6]. Clustering methods, in contrast, use various features to classify pixels into two or more classes and to locate the text pixels. Several clustering methods, such as Markovian clustering [12], K-means [7], and PCA, can be used for this purpose. The main problem with these methods is the appearance of noisy pixels on the output. However, it seems that the recently introduced Markovian clustering method may have the ability to remove the noisy pixels [12]. The method that placed 2nd at DIBCO'09, proposed by Fabriozi and Marcotegui [13], will also be used for comparison purposes in this work. The method was developed based on the toggle mapping morphological operator toward text localization [14]. A few other successful approaches in binarization of document images are morphological operators [15], Markov Random Fields [16], local adaptive partitioning methods [17]. Despite huge diversity in the binarization methods, It is worth noting that the proposed method is the first level set-based approach in document binarization.

Edge-based methods usually use a measure of the changes across an edge, such as the gradient [18] or other partial derivatives, or use an edge detector, such as Canny edge detector, directly [8]. One of the most successful edge-based methods was presented in [19]. This method is of great interest, as a variation of it won first place in the first contest on document image binarization (DIBCO'09) [9,13]. In a most recent published work [9], the image contrast has been used to identify the text edges and to estimate the local threshold values in order to binarize highly degraded historical documents images. However, edge information alone is not always sufficient to perform a proper binarization of many degraded document images. Because of degradation and ink fading, the edges may not be very sharp, and therefore there is a high level of variation in the gradient across the edges over the image. This is one of the reasons why the gradient is used in this work in combination with other tools. Moreover, as a new contribution, and despite a high level of possible dissimilarity in color/intensity, we use the gradient not only for locating the edges, but also to follow the ink along the strokes. Its use, made possible using local linear modeling [20], is the key concept in the recovery of cuts and weak strokes and plays a major role in higher-level processes [6].

At a higher conceptual level, different methods can be combined in order to compensate for their individual drawbacks. Along the same line, an interesting voting scheme has been proposed in [21]. Those combination methods can be seen as high-level modeling based on the basic and fundamental methods. As our proposed method is a basic one, comparison with the combined method has not been performed. However, it could very well serve as an input to one of those high-level methods in order to achieve a higher performance.

The complexity of the binarization problem may be slightly reduced if some prior geometrical knowledge regarding the document is available or can be estimated. Typically, the size of the structures of interest is related to the pen width or to the font face. By knowing the stroke width, it is easier to reject small spurious objects or to fill unwanted gaps, as has been done in recent work on strongly degraded characters [22]. This information may also be considered in the selection of the scale used to compute local statistics.

In this paper, an innovative binarization method is introduced, based on the concept of *stroke map* extraction and optimization. The optimization is conducted using the level set framework, which offers an elegant way to integrate various priors. An implementation of the method was submitted to the *Document Image Binarization Contest—DIBCO'09*, held at the ICDAR 2009 international conference, where it placed 3rd [13].[1] However, this is the first time that a comprehensive description of the method is published. The local linear regional force used in the proposed model has been introduced by some of us in [20]. However, this work is based on widely different premises. Most significantly, the global process, the involved segmentation clues, and the initialization method are different.

To our knowledge this work is among the first ones to take advantage of the level set method to achieve good results in document image binarization. The quality of the results has been assessed during the DIBCO'09 contest. The level set framework is used in this work to integrate many document-related aspects. However, the level set method is not limited to the presented forces and could be used in conjunction with other functions. The local linear model, central to our

[1] Our method is identified by the number 24 in [13].

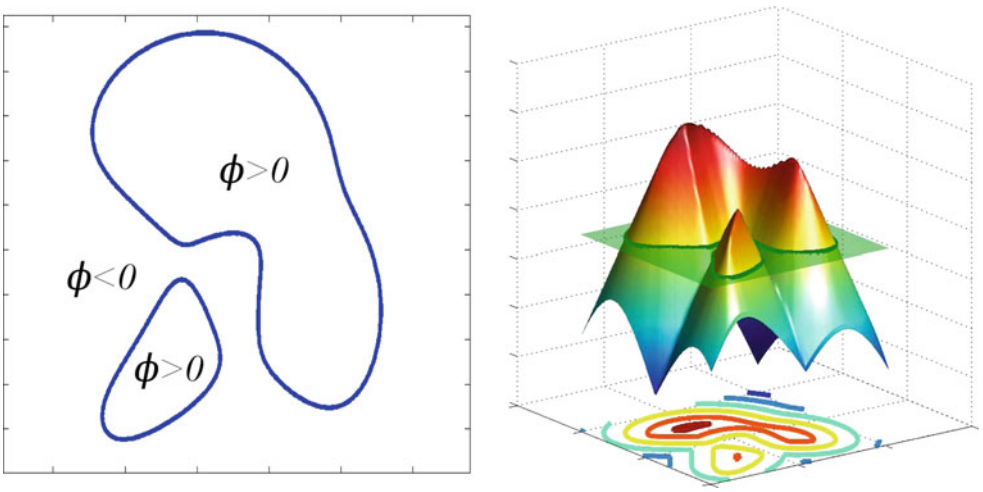


Fig. 1 An arbitrary contour (*left*) and the associated level set function (*right*)

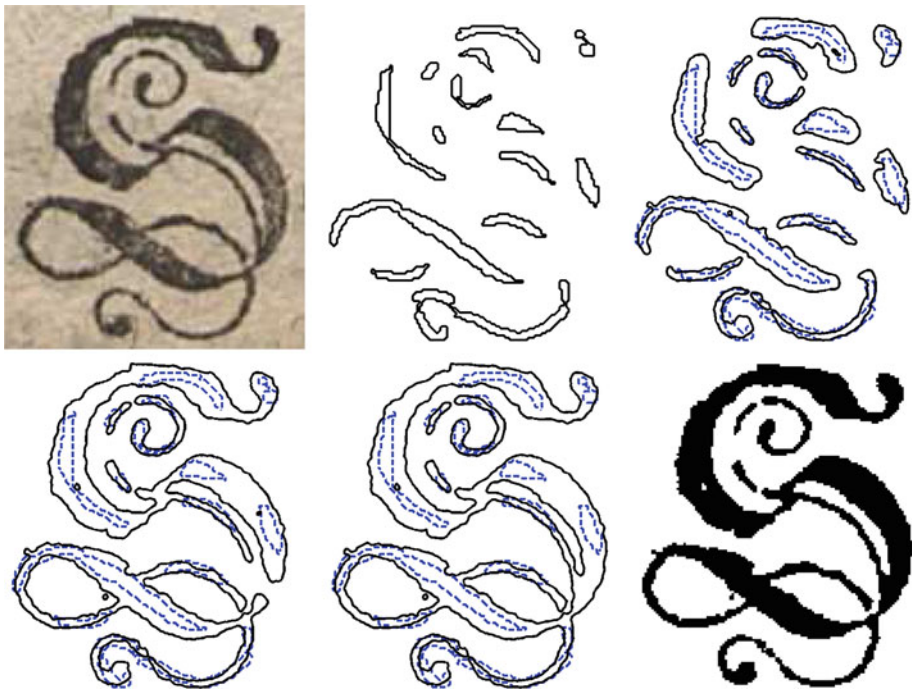


Fig. 2 Example of level set segmentation, using forces (12) and (15) in the level set Eq. (4) (described later). *From left to right. Top row* input image before grayscale conversion, contour from manual initialization, contour after 5 iterations. *Bottom row* contour after 20 and 30 iterations, final binarization

proposed method, used by the level set-based steps is able to capture configurations of smoothly varying stroke intensities. In addition, a stroke gray level (SGL) model is introduced to limit the propagation of the contour to areas that are unlikely to be part of the stroke. Finally, in our initialization step, we compute an estimate of the stroke width and use it as a prior knowledge to detect the potential stroke pixels.

The paper is organized as follows. Section 2 describes the proposed method. In Sect. 3, the experimental results, and both the qualitative and quantitative evaluations of the method, are presented. Finally, the conclusion is provided in Sect. 4.

2 Proposed method

A novel binarization algorithm is introduced in this section, which benefits from the flexibility of the level set framework to integrate various regional information into one method. Since the level set method is a local optimization method with a tendency to fall into the nearest local minimum, a stroke map-based [23] initialization is used. This results in a three-step process: (1) A stroke map is extracted from the document image to serve as an initialization for the level set method; (2) Our level set method is used in *erosion* mode in order to remove less likely stroke pixels from the map; and,

Fig. 3 Example of local linear models corresponding to the image in Fig. 2, with the contour obtained after 30 iterations. *From left to right* model for the stroke $L_S(r)$, model for the background $L_B(r)$, and a composite image computed using $\mathrm{Img}(r) = L_S(r)*(\phi \geq 0)+L_B(r)*(\phi < 0)\forall r \in \Omega$. The last image corresponds to the output of the edge indicator function (7)

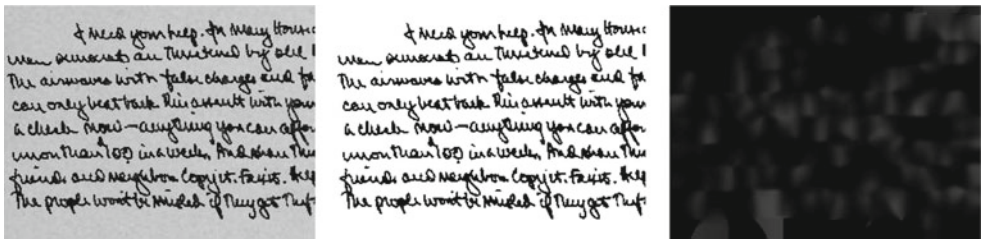


Fig. 4 Computation of the SGL map. *From left to right* input image, estimated stroke pixels, and the SGL map

finally, (3) The level set method is used in *free displacement* mode to produce the final delineation and binarization.

2.1 Problem statement

A gray-scale document image is given by $u(r) \in [0.0, 1.0]$, where $r = (r_x, r_y) \in \Omega \subset \mathbb{R}^2$. The image domain Ω is a rectangular lattice of height H and width W. If an RGB color image is given as input, only the luminance component is used. The conversion is then given by[2] $u = 0.288u_R + 0.587u_G + 0.114u_B$. The set of foreground pixels is denoted by $S \subset \Omega$ and the set of background pixels by $B = \Omega \backslash S$. It is assumed that S corresponds to the pixels of interest, which in this paper are limited to strokes and printed characters. The objective of the binarization is to produce a binary image $b(r)$, where

$$b(r) = \begin{cases} 0 & \text{if } r \in S \\ 1 & \text{otherwise} \end{cases} \tag{1}$$

Of course, the contents of the sets S and B are not known from the beginning, and therefore they must be estimated from the data. In the following, the subscripts S and B will be used to indicate relations with the set of stroke pixels or the set of background pixels.

[2] As defined by the ITU-R Recommendation BT.601.

2.2 Local linear level set framework

If a rough estimation of the stroke position is available, it is possible to set the binarization problem in a curve evolution framework.[3] The idea is to represent the stroke–background interface using closed curves. Then, the curves are evolved along an artificial time variable. At each time step, forces of various kinds attract the curve toward its best position, with respect to a certain local optimum (e.g. see Fig. 2). However, in the proposed method and the rest of the paper, the initial map is calculated automatically. In typical cases of document image binarization, there will be *many* curves to represent the boundaries of the multiple strokes. In such a complex case, explicit tracking of the contour evolution would be cumbersome, and possibly impractical, due to issues related to parameterization and the merging or splitting of the curves. It is worth noting that, for the sake of illustration, the initial map in this figure is selected manually.

The level set framework makes it possible to represent a set of curves implicitly. Its formulation is defined directly on the image grid and allows for topological change to be handled naturally. In addition, it permits the use of a large class of forces to drive boundary evolution [24]. This framework was introduced by Osher and Sethian in 1988 [25], and its potential for active contour segmentation began to be exploited during the second half of the '90s [26–29].

[3] In this section, this *rough estimation* of the stroke position, the initialization map, is taken as given. It is worth noting that in the rest of this paper an automatic initialization method is used. For details see Sect. 2.3.

While early segmentation methods mostly exploited edge information, Chan and Vese [30], and also Paragios and Deriche [31], among others, introduced *region-based* active contour segmentation methods which take advantage of global regional information. These methods make it possible to segment regions that are not delineated by strong gradients and to tackle a broader class of problems. They are also generally considered to be more robust than their edge-based counterparts.

Not withstanding their quality, many region-based segmentation models perform poorly on images where the statistics of the region vary significantly from one part to another. This is often the case with ancient documents, where the tone of the background may vary due, for example, to aging, water damage, or digitalization artifacts. Also, the variation in intensity within a single pen stroke, e.g. due to a change in the ink flow rate, may be problematic for some region-based algorithms. Local statistical models for region-based segmentation were recently introduced by several authors [32–35]. While this model adapts to local specificity, it also extracts some information from the structure of the gray level of the local pixels. Some of us introduced in [33] a local linear model for the segmentation of smoothly varying regions.

Introduction to the level set method The basic concept of the level set framework can be explained as follows. Within the framework, a set of closed, but possibly disjoint, contours C on Ω is represented by the intersection of a surface $\phi(r) : \Omega \rightarrow \mathbb{R}$ and of the zero level $z = 0$. By convention, $\phi(r)$, called the level set function (*lsf*), is approximated as a signed distance function of C. If $\phi(r_i) \geq 0$ the pixel at r_i is labeled as *stroke*, otherwise, the pixel is labeled as *background*. A sketch illustrating this mode of representation is shown in Fig 1.

Starting from an initial and arbitrary guess, C_0, the evolution of C subject to some driving force F acting in the contour's normal direction corresponds to the evolution of $\phi(r, t)$ along an artificial time variable t, as defined by the following governing equation [24]:

$$\frac{\partial \phi(r,t)}{\partial t} = -|\nabla \phi(r,t)|\, F(r), \tag{2}$$

with $\phi(r, t = 0)$ corresponding to C_0. The driving force F can be spatially variant and may correspond to a combination of both internal and external forces. Internal forces are computed from intrinsic geometrical properties of the contour, such as its volume and curvature, and external forces can come from various image features or from a priori knowledge. In the following, it is assumed that $|\nabla \phi(r, t)| \approx 1$ in (2), since ϕ is approximately a distance function. Also, all considered forces are restricted to only act on the boundaries by using a regularized Dirac function, δ_ϕ [30]:

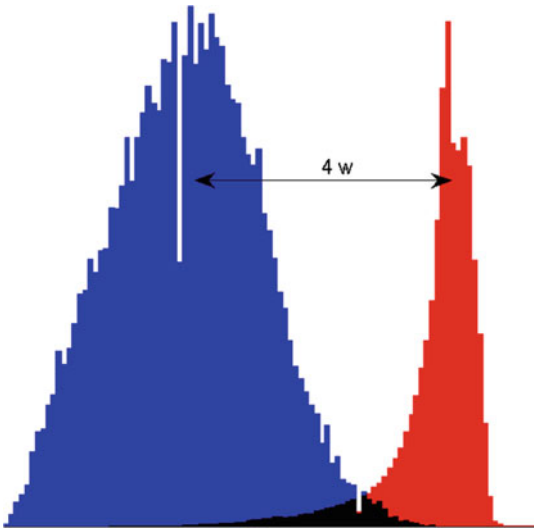


Fig. 5 Illustration of the parameter w for the SGL force. See text for description. Histogram values for the stroke (*blue*) and background (*red*) have been normalized

$$\delta_\phi = \frac{1}{\pi} \frac{1}{(1 + \phi^2(r,t))}. \tag{3}$$

Equation (2) can then be rewritten as:

$$\frac{\partial \phi(r,t)}{\partial t} = \delta_\phi F(r). \tag{4}$$

Equation (2) has been called the *level set equation* in [24] and is the main equation in the level set method. Equation (4) is a well-known approximation of (2) introduced in [30]. Basically, both equations represent a process by which a contour is updated iteratively subject to the net force $F(r)$ acting on it. As it will be seen in the following sections, all the forces considered in this work result from the minimization of certain quantities related to either the shape or the position of the contour, with respect to the input image intensities.

Before presenting the details of the various force acting on the binarizing contours, a high-level overview of a basic local linear level set segmentation method is presented below. Figures. 2 and 3 illustrate the basic concepts.

The method starts from a certain initialization, which is the contours of an arbitrary binarization of the document image. In Fig. 2, the initial binarization was manually drawn, but in the rest of this work, the initialization is estimated automatically (see Sect. 2.3). Also, as demonstrated in Sect. 3.1, a good result could be achieved starting from an almost random initialization in some cases.

After initialization, the iterative evolution of the contour begins. The displacement of the contour is driven by various forces that tend to minimize certain quantities, as detailed in Sects. 2.2.1–2.2.4. Central to this work are the forces induced by the local linear regional models. In order to drive the level

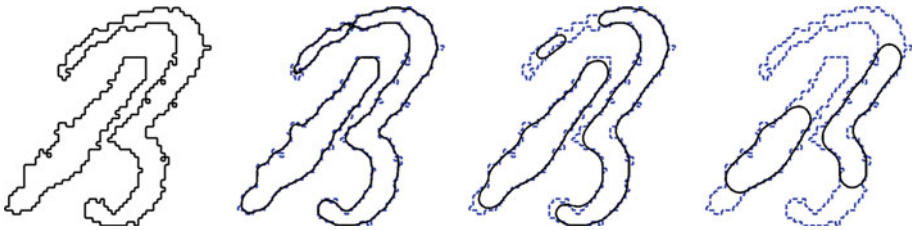


Fig. 6 Effect of the length regularization force, using (15) in (4). *From left to right* initialization, $t = 0.4$, $t = 4$, and $t = 40$

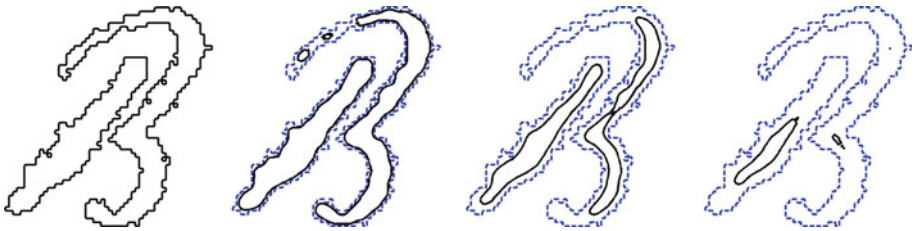


Fig. 7 Effect of the area contraction force, using (16) with $K = -1$ in (4). *From left to right* initialization, $t = 0.4$, $t = 1.4$, and $t = 2$

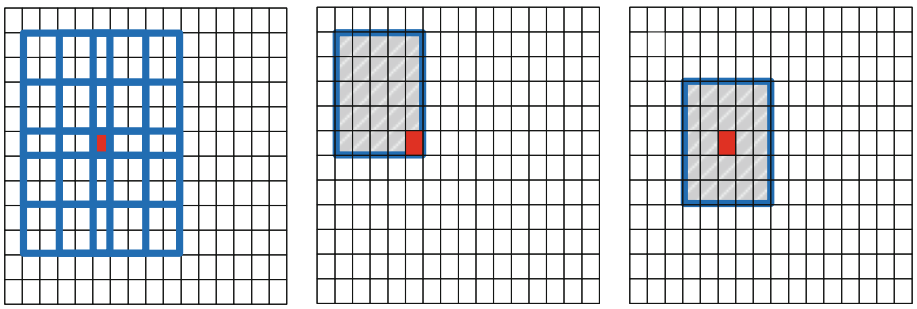


Fig. 8 (*Left*) the nine patches used to calculate the SM shown all together in *blue* (they are highly overlapping). (*Center* and *right*) Two of the patches are shown separately (the target pixel is shown in *red*)

set, the parameters of two distinct local linear models are computed: a model for the strokes and a model for the background. These model parameters are computed locally, at each pixel position, by fitting a bilinear model to the intensities of all the pixels that are inside the contour for the stroke model and outside the contour for the background model in a certain neighborhood. As an example, the average intensities of the local linear model for both the stroke and the background of the final contour in Fig. 2 are presented in Fig. 3. Once the local linear model parameters are computed, the contour is displaced in such a way that the error between each pixel's intensity and the corresponding model (the stroke model, if the pixel is inside the contour, and the background model otherwise) is minimized. Since the parameters of the local linear models depend on the position of the binarizing contour, their parameters must be updated when the position of the contour changes. The iterative evolution of the contour then consists of cycles of parameter updates and contour displacement. A few instances of the contour propagation and the final binarization result are also presented in Fig. 2.

The local linear-induced forces are the most important in this work. However, other requirements, such as the relative smoothness of the contour and the robustness to noise, call for the inclusion of other forces. Those forces are presented and discussed in the following sections.

2.2.1 Local linear regional force

The local linear regional force is the most important force in the binarization process. It is derived from the minimization of an energy function based on a probabilistic model, which tries to follow the local variations in the text intensities.

When considering a sufficiently small scale, smoothly varying stroke and background pixels can be modeled with interesting accuracy by using two local linear models: $L_S(r) = a_{S,r} + b_{S,r}r_x + c_{S,r}r_y$ for the stroke's pixels and $L_B(r) = a_{B,r} + b_{B,r}r_x + c_{B,r}r_y$ for the background pixels [20]—see Fig. 3. If the parameters of the local linear models are known or can be estimated, a segmentation can be achieved by iteratively moving the boundaries in order to maximize the model probabilities of the observed pixel. Below, the probabilistic model and the energy function are first presented. Then, the local linear force derived from the energy function is introduced.

Probabilistic modeling and energy function The binarization problem can be considered as a classification problem where each pixel of the document image is either labeled

as *stroke* or *background*. Given a certain binarizing contour described by a level set function ϕ, the pixels on the document image are labeled *stroke* if $\phi > 0$ and *background* otherwise. For computational purpose, a regularized Heaviside function is used to produce a soft assignment. The regularized Heaviside function used in this work is defined as follows: $H(\phi) = H_{1,0}(\phi)$ where

$$H_{w,p}(x) = \begin{cases} 0 & (x-p) < -w \\ \frac{1}{2}\left(1 + \frac{x-p}{w} + \frac{1}{\pi}\sin\left(\pi\frac{x-p}{w}\right)\right) & |x-p| \le w \\ 1 & (x-p) > w. \end{cases} \tag{5}$$

Given this membership, a Gaussian formulation is used to compute the stroke and the background probabilities of the local linear model for a particular pixel:

$$P\left(u(r) \mid R\right) = \exp\left(-\frac{(u-[a+bx+cy])^2}{\sigma_e^2}\right) \cdot \exp\left(-\mu \cdot g(u)\left[\frac{(u_x-b)^2}{\sigma_{eb}^2} + \frac{(u_y-c)^2}{\sigma_{ec}^2}\right]\right). \tag{6}$$

Here, $R \in \{L_S, L_B\}$, u_x and u_y denote the image gradients, $\mu \in \mathbb{R}^+$ is a balancing parameter, and g is a monotonically decreasing function of the document image gradient that decreases the weight of the gradient term close to the edges. Its definition is chosen from [27]:

$$g(u) = \frac{1}{1 + \nabla(G * u)}, \tag{7}$$

where G is a 3×3 Gaussian smoothing kernel with a standard deviation of 0.5. A sample output is presented in Fig. 3. The various σ_e terms in (6) represent the variance of the error. Because minimizing (6) when those terms are not fixed is difficult, a simplifying assumption that they are constant and equal everywhere was made.

By using (6) and (5), the probability of a certain labeling for the whole document image can be written as follows:

$$P(u \mid R) = \prod_{r\in\Omega} \left(H(\phi_r)P(u_r|S) + (1 - H(\phi_r))P(u_r|B)\right) \tag{8}$$

This corresponds to the probability of a given contour position with respect to the *stroke* and *background* local linear models. Related probabilistic models have also been presented by other authors [31,34,36,37]. It is worth noting that the model is intrinsically balanced. This is because boundary pixels, which almost consist of the same number of stroke and background pixels, are the most influencing ones in (8). This advantage of the proposed method prevents high-numbered background pixels from dominating over low-numbered stroke ones.

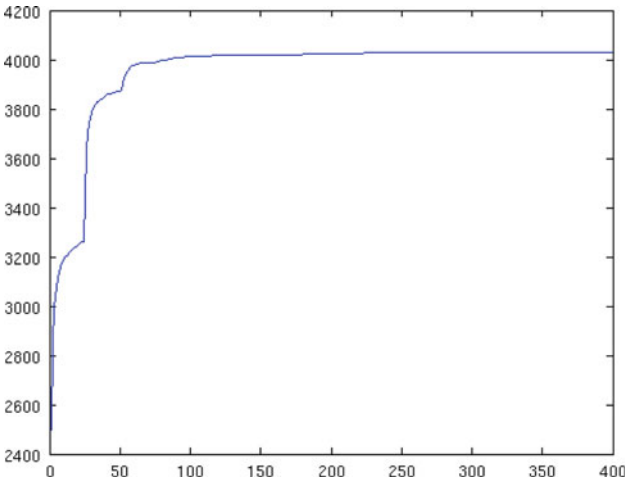


Fig. 9 Progression of the number of pixels that changed their label with respect to the number of level set iterations. This plot corresponds to the stroke map erosion step for the image presented in Fig. 17

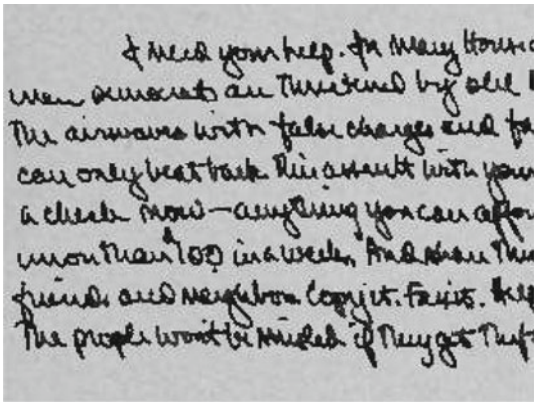


Fig. 10 Input image used in the sensitivity experiment

Since, from a numerical point of view, it is easier to minimize the negative log-likelihood of (8) than to perform its direct maximization, the following expression will be used:

$$P_1(u \mid R) = -\sum_{r\in\Omega} \left(H(\phi_r)\log P(u_r|S) + (1 - H(\phi_r))\log P(u_r|B)\right). \tag{9}$$

This equation defines the local linear energy. The energy level corresponding to a given contour is the minimum if the error between the document image pixel intensities and the local linear models is minimal. A hypothesis of this work is that this state will correspond to a good binarization of the document image.

Energy minimization and local linear force The minimization of (9) is performed by computing its Euler–Lagrange equations and by following an approximated gradient descent procedure. The derivation of the local linear force is briefly introduced below. For more details, we refer the interested reader to [20].

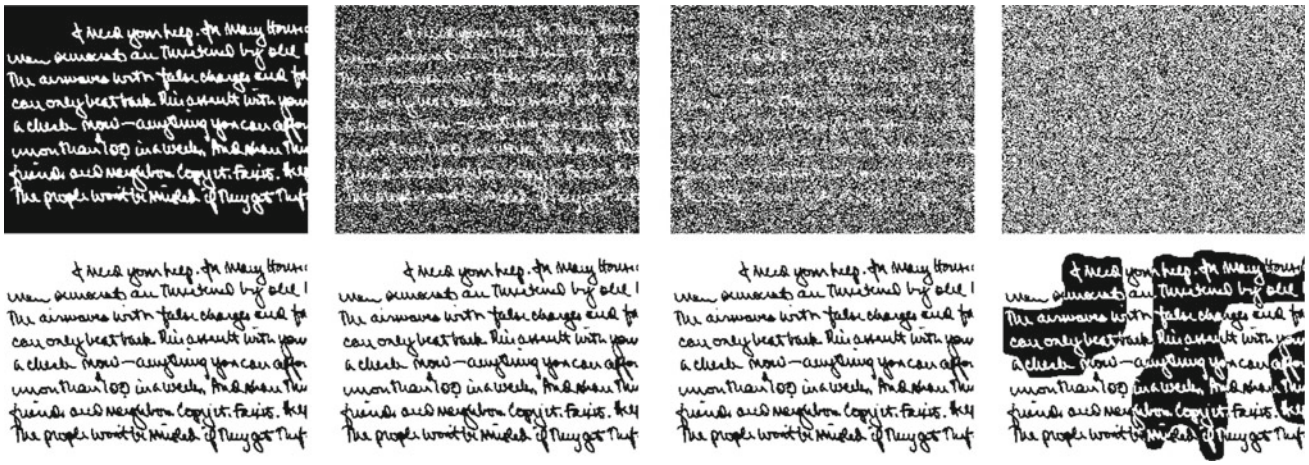


Fig. 11 *Top row*, *from left to right* the SM computed using the method described in Sect. 2.3, and the SM after adding respectively 50, 70, and 100% of the pixels by salt and pepper noise. *Bottom row* the final binarization corresponding to the SM initializations from the above row

Table 1 The details of the four scenarios used to evaluate the performance on synthesized degradation

Degradation type	Generation method
Bleed-through	Application of a physical, diffusion-based PDE model [42] considering various values for the parameter d which controls the ratio of the verso diffusion to the normal diffusion on the rector side (see the Eq. (5) in [42]. Except for d, default parameter values were used). The values of this parameter are 0.0390, 0.0552, 0.0782, and 0.2218 respectively
Gaussian blurring	Application of a numerical Gaussian filter with parameter $\sigma_B \in [0, 2]$ and kernel size of $k = 2\lceil 3\sigma_B \rceil + 1$
Gaussian white noise	Addition of white noise from a normal generating function with a standard deviation of $\sigma_N \in [0, 40]/255$
Loss of contrast	Application of a linear transformation $u_{\text{out}} = M * u_{\text{in}} + (1 - M)/2$, where the degradation parameter $M \in [0.01, 1]$ is the Michelson contrast

Following this scheme, the force that tends to minimize (9) is derived by keeping the local linear models L_S and L_B fixed and by computing the gradient descent equation of (9) with respect to ϕ [20,34]:

$$\begin{aligned} F_{L0}(r) = &- \left(a_{s,r} + b_{s,r} r_x + c_{s,r} r_y - u(r)\right)^2 \\ &- g \cdot \mu \left[\left(b_{s,r} - \frac{\partial u(r)}{\partial r_x} \right)^2 + \left(c_{s,r} - \frac{\partial u(r)}{\partial r_y} \right)^2 \right] \\ &+ \left(a_{b,r} + b_{b,r} r_x + c_{b,r} r_y - u(r)\right)^2 \\ &+ g \cdot \mu \left[\left(b_{b,r} - \frac{\partial u(r)}{\partial r_x} \right)^2 + \left(c_{b,r} - \frac{\partial u(r)}{\partial r_y} \right)^2 \right]. \end{aligned} \tag{10}$$

When integrated into the level set Eq. (2), this force can drive the level set contour toward a position that is a local minimum of (9).

Since, at initialization, the parameters of the two local linear models (stroke and background) are unknown, they must be estimated from the data. Also, any motion of the level set is likely to disturb the models' parameters, so the estimation will need to be refined at every few iterations. Keeping ϕ fixed, the parameters of the linear models are estimated at every spatial position r. Let us introduce two sets $\tilde{S}(r)$ and $\tilde{B}(r)$ that correspond to the current estimation of the local stroke and background pixels. Also, let $K_s(r) = \{s \in \Omega : \max(s_x - r_x, s_y - r_y) < W\}$ be the set of pixels in a square neighborhood of the pixel r and of the size W. Then, the set $\tilde{S}(r)$ membership is given by the regularized Heaviside function (5). With those sets defined, it is now possible to compute the parameters $a_s(r)$, $b_s(r)$ and $c_s(r)$ of the local linear model $L_S(r)$ by using a weighted 2D linear regression, which corresponds to the optimization of (6) with respect to the parameters of L_S by keeping ϕ fixed. Numerically, this is achieved by solving the following system of linear equations using any classical technique:

$$\begin{aligned} &\begin{bmatrix} \sum_{K_s} H & \sum_{K_s} xH & \sum_{K_s} yH \\ \sum_{K_s} xH & \left(x^2 + \sum_{K_s} \frac{\mu g}{\alpha}\right) H & \sum_{K_s} xyH \\ \sum_{K_s} yH & \sum_{K_s} xyH & \sum_{K_s} \left(y^2 + \frac{\mu g}{\alpha}\right) H \end{bmatrix} \cdot \begin{bmatrix} a_s \\ b_s \\ c_s \end{bmatrix} \\ &= \begin{bmatrix} \sum_{K_s} uH \\ \sum_{K_s} \left(xu + \frac{\mu g}{\alpha} \frac{\partial u}{\partial x}\right) H \\ \sum_{K_s} \left(yu + \frac{\mu g}{\alpha} \frac{\partial u}{\partial y}\right) H \end{bmatrix} \end{aligned} \tag{11}$$

The local linear parameters of the background are computed in a similar way, but with $\tilde{B}(r)$ membership given by

Fig. 12 Bleed-through degradation. *Left column* Input image. *Middle* and *right columns* Binarization and OCR results for various level of degradation: **a** $d = 0.0390$, **b** $d = 0.0552$, **c** $d = 0.0782$, and **d** $d = 0.2218$. See text in Sect. 3.2, Table 1, and [42] for details. Please note that the degradation is better visualized on screen

$1 - H(\phi(r))$. Also, it is worth noting that the sums over K_r can be computed using convolutions.

The system of linear equations (11) is solvable in most practical situations. However, three limiting cases where the system is underdetermined are of interest. Let N_{Ks} be the number of pixels in $K_s(r)$. (1) If $N_{Ks} = 0$, that it means that the position r is far from any boundary and, in this situation, the values of a_s, b_s and c_s are not of important and they can safely be set to 0. (2) If $N_{Ks} = 1$, it is not possible to compute the slope coefficients b_s and c_s. In this case, it is assumed that $a_s = u, b_s = 0$, and $c_s = 0$. (3) Finally, if $N_{Ks} > 1$, but all pixels are on the same one-pixel-wide line, it is not possible to compute the slope in the direction perpendicular to that line. In this case, it is assumed that $b_s = 0, c_s = 0$, or $b_s = kc_s,\ k \in \mathbb{R}$, depending on the specific configuration.

The regional parameters a_i, b_i and c_i depend on the level set position and their value must be updated during the level set evolution, when the position of the contour defined by ϕ changes. However, since they are costly to compute and vary slowly from one iteration of (4) to another, they may be computed every few iterations without compromising the final result.

For practical purposes, it may not be desirable to rely on the local linear model if too few pixels are inside the window defined by W. Therefore, the following formulation is used in the proposed level set schemes:

$$F_L(r) = \begin{cases} 0 & N_{Ks} < W_{\min} \text{ or } N_{Kb} < W_{\min} \\ F_{L0}(r) & \text{otherwise} \end{cases}, \tag{12}$$

where N_{Kb} is the number of pixels in $K_b(r)$, and $W_{\min}$ is a small number related to the precision of the grid. In this work, we use $W_{\min} = 9$, which corresponds to a small block of 3×3 pixels. This force cancels itself out if the number of strokes or background pixels inside the window defined by W at a certain position r is less than a predefined number.

Fig. 13 Gaussian blurring. *Left column* Input image. *Middle* and *right columns* binarization and OCR results for various level of Gaussian blurring: **a** $\sigma_B = 0.5$ and $k = 5$, **b** $\sigma_B = 1.0$ and $k = 7$, **c** $\sigma_B = 1.5$ and $k = 11$, and **d** $\sigma_B = 2.0$ and $k = 13$. See text in Sect. 3.2 and Table 1 for details

2.2.2 Stroke gray level regional force

A stroke gray level (SGL) regional force is now introduced to guide the level set evolution from a larger scale than the local linear force, thereby ensuring greater consistency over the document image.

Key to this force is to define an SGL map [23] which represents the expected gray level of the stroke at the position of any pixel of the document image. In this work, a fast and effective technique to calculate the SGL map is introduced. Starting with a few stroke pixels identified by another mean, $\hat{S}$, the SGL map can be constructed by solving a heat diffusion problem on the image domain, where the pixels from $\hat{S}$ act as constant heat sources. A more computationally efficient technique, which gives similar results, has been used: 1. A binary stroke indicator image, $b_s(r)$, and a stroke only image, I_s, are created from $\hat{S}$, as follows:

$$b_s(r) = \begin{cases} 1 & \text{if } r \in \hat{S} \\ 0 & \text{otherwise} \end{cases}$$

$$I_s(r) = u(r) \cdot b_s(r).$$

2. The SGL map is created by smoothing I_s with a large Gaussian kernel K_σ and by normalizing this result with the smoothed binary stroke indicator image response:

$$\text{sgl}(r) = \frac{K_\sigma * I_s(r)}{K_\sigma * b_s(r) + \epsilon} = \frac{K_\sigma * (b_s(r) \cdot u(r))}{K_\sigma * b_s(r) + \epsilon}, \quad 0 \leq \text{sgl}(r) < 1, \tag{13}$$

where $*$ denotes the convolution operator, and ϵ is a small positive real number. (3) Finally, from (13), it can be noted that some region of $\text{sgl}(r)$ may have a value of 0, not because the average stroke value in this area is 0, but because at this location $K_\sigma * b_s(r) = 0$, which indicates that no stroke pixel has been involved in the computation. To circumvent

Fig. 14 Gaussian white noise. *Left column* Input image. *Middle* and *right columns* binarization and OCR results for various level of Gaussian white noise degradation: **a** $\sigma_N = 10/255$, **b** $\sigma_N = 20/255$, **c** $\sigma_N = 30/255$, and **d** $\sigma_N = 40/255$. See text in Sect. 3.2 and Table 1 for details

this problem, values where $\mathrm{sgl}(r) \leq \epsilon$ are canceled, and the value of the closest pixel of the SGL for which $\mathrm{sgl}(r) > \epsilon$ is assigned by using a distance transform algorithm [38]. In this work, the Matlab function `bwdist` has been used for this purpose. An illustration of the result is presented in Fig. 4.

Once the SGL map has been computed, the SGL regional force can be defined as follows:

$$F_G(r) = H_{w,p}(u(r) - \mathrm{sgl}(r)), \tag{14}$$

where $H_{w,p}$ is given by (5). This definition makes use of the fact that the background pixel intensities will be lighter than the stroke pixel intensities, on average. If this is not the case, then $F_G(r) = H_{w,p}(-u(r) + \mathrm{sgl}(r))$ should be used. The parameter w and p are automatically set based on the intensities of the stroke and of the background pixels. Let $\bar{s}$ and $\bar{b}$ be the average intensity of the pixel corresponding to the stroke $\hat{S}$ and the background not $(\hat{S})$, respectively. The parameters are calculated as follows: $p = \frac{\bar{b}-\bar{s}}{2}$ and $w = \frac{\bar{b}-\bar{s}}{4}$. Here, parameter w can be seen as a rough estimation of the standard deviation of the stroke pixel intensities. If the distance between some pixel intensity and the SGL map is less than w, the SGL force is zero. The value of the force smoothly raises to 1.0 as the intensity of the pixel gets closer to that of the background. Finally, the value of the force is 1.0 if the difference between the pixel intensity and the SGL map is more than $3w$. See Fig. 5 for an illustration.

2.2.3 Length regularization force

With the level set methodology, it is classical to add a curvature-driven force to the evolution equation. This force tends to minimize the contour length, thereby augmenting its regularity. It is defined as follows [25]:

$$F_R(r) = \nabla \cdot (\nabla \phi / |\nabla \phi|). \tag{15}$$

See Fig. 6 for an illustration of the effect of this force on a contour.

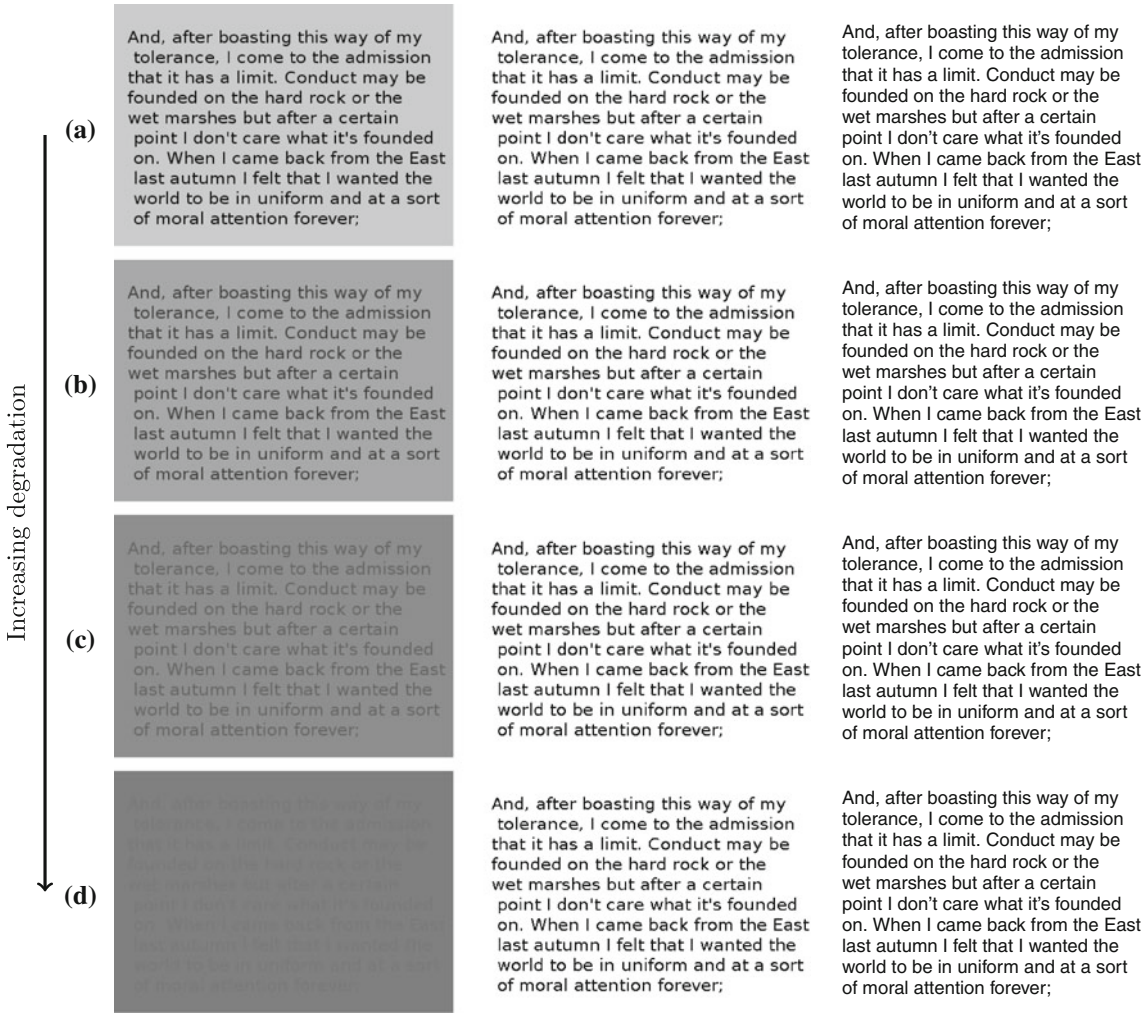


Fig. 15 Loss of contrast degradation. *Left column* Input image. *Middle* and *right columns* binarization and OCR results for various level of Michelson contrast: **a** $M = 0.50$, **b** $M = 0.25$, **c** $M = 0.10$, and **d** $M = 0.01$. See text in Sect. 3.2 and Table 1 for details

2.2.4 Area expansion or contraction force

A simple constant force may be added to (4) to cause the stroke area enclosed by the zero level set $(\{x, y\} \mid \phi(r) = 0)$ to tend to either expand or contract. Since crossing an image edge is usually not beneficial, this force can be modulated by using the edge indicator function (7). The resulting force can be written as follows:

$$F_A = Dg(r), \tag{16}$$

where $D \in \{-1, 1\}$ defines the direction. The estimated stroke area will tend to expand if $D = 1$ and to contract if $D = -1$. See Fig. 7 for an illustration of the effect of this force on a contour. In contrast to the length regularization force, the area contraction force shrinks the contour uniformly in all directions, regardless of its local properties.

As the above discussion suggests, the level set framework allows the simultaneous application of a large variety of forces at the stroke–background interface. Our binarization algorithm involves the combination of those forces in a specific way, as described in Sect. 2.4.

2.3 Stroke map and a priori information

An initialization map is required to identify high-probability text pixels. The other pixels, which may be degraded, will be recovered by the local linear evolution of the level set function. For this purpose, we use one of the multi-level classifiers [23], the stroke map (SM). Multi-level classifiers use different features to locate text pixels. Although the information at the pixel level is helpful, a major part of the image information is carried within the spatial relations. The classifiers at the content level, such as the SM, the stroke profile (SP) [23], and the stroke cavity map (SCM) [22], search for this information based on the stroke-based features. In other words, these classifiers try to use the document-related nature

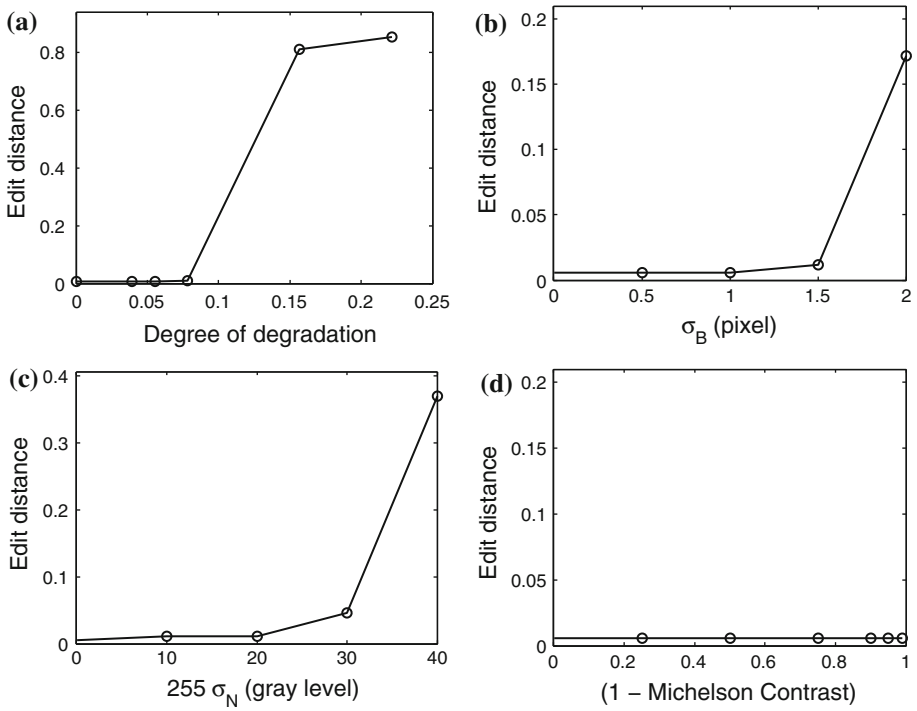


Fig. 16 Edit distance in function of the degradation level for 4 types of degradation: **a** bleed-through, **b** Gaussian blur, **c** Gaussian white noise, and **d** loss of contrast. The lower the edit distance is, the better

of the images. In the case of the SM, the likelihood of having a stroke around the pixel in question is examined based on the structure of the text pixels around it. In this analysis, the average stroke width, w_s [6], is used to determine the possibility of there being a stroke around the pixel. In a new kernel-based approach, on a neighborhood of size $2w_s + 1$, a score is calculated based on which an SM value is assigned to the pixel. The SM can operate on different operational regimes. For the purposes of this work, which is to avoid as many false positive pixels as possible, the SM is set to an internal high-confidence operating mode. As is obvious, the SM itself needs an initialization map in order to secure pre-estimation of the text pixels. We use the grid-based Sauvola method [6] to generate a fast and adequate initialization map for the SM.

The SM classifier makes use of a rough binarization of the input image in order to estimate the score of each pixel in terms of strokes. Let us assume that a rough binarization is u_{BW}. Based on the calculated score, each pixel is assigned to strokes (1) or to background (0). Therefore, the output of the SM is a binary image. The SM score of each pixel is calculated based on the SM score matrix, $a_{\mathrm{SM,k,l}}$ $(k, l = -1, \ldots, 1)$, which will be discussed later:

$$\mathrm{SM} = \begin{cases} 1 & \text{if } \ \mathrm{Thr}_{\mathrm{SM,max}} > \sum_{k,l} a_{\mathrm{SM},k,l} > \mathrm{Thr}_{\mathrm{SM,min}} \\ 0 & \text{otherwise} \end{cases} \tag{17}$$

$\mathrm{Thr}_{\mathrm{SM,min}}$ catches the isolated regions that are not strokes, and is between 1 and 9. We use $\mathrm{Thr}_{\mathrm{SM,min}} = 1$. $\mathrm{Thr}_{\mathrm{SM,max}}$ differentiates between strokes and big black regions, and is between 3 and 9. We use $\mathrm{Thr}_{\mathrm{SM,max}} = 9$. The score matrix a_{SM} provides a measure of the presence of black pixels around the target pixel based on the average stroke width, w_s. The matrix evaluates this relation on the area around the target pixel in different directions. For this purpose, an overlapping system of nine patches, $W_{k,l}$, is used (see Fig. 8):

$$W_{k,l}(\xi, \eta) = \begin{cases} 1 & ||(i + k, j + l) - (\xi, \eta)|| < w_s \\ 0 & \text{otherwise} \end{cases} \tag{18}$$

and $a_{\mathrm{SM},k,l}$ is defined as follows:

$$a_{\mathrm{SM},k,l} = \begin{cases} 1 & \frac{\sum_{\xi,\eta} W_{k,l}(\xi,\eta) u_{BW}(\xi,\eta)}{\sum_{\xi,\eta} W_{k,l}(\xi,\eta)} > \mathrm{Thr}_a \\ 0 & \text{otherwise} \end{cases} \tag{19}$$

where Thr_a is a value close to 1 (0.7) and is used to ensure that the corresponding patch is full of black pixels.

2.4 Combined-level set method

The complete method is now described using previously defined formulations. It is composed of three steps: (1) initialization, (2) level set evolution in *erosion* mode for stroke map consolidation, and (3) level set evolution in *free displacement* mode for contour optimization and final binarization.

Fig. 17 *Top row* samples of a document images with synthetic degradation added. *From left to right* The degree of degradation is 1, 7, 8, and 9. *Bottom row* The binarization results corresponding to the images in the *first row*

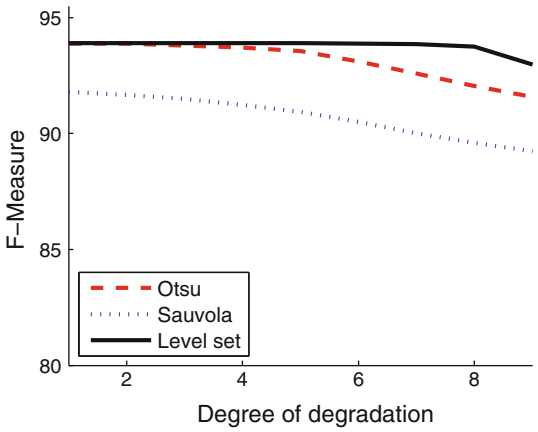


Fig. 18 Comparison of the results obtained with the images with synthetic degradations using: the proposed method, Otsu's method, and Sauvola's method

2.4.1 First step: stroke map extraction

In this step, an initial binarization is estimated using the method described in Sect. 2.3. At this point, the goal is to obtain a rough binarization that retains the main features of the input image.

2.4.2 Second step: stroke map erosion

The purpose of this step is to remove any miss-labeled pixels from the stroke map in order to compute, in the next step, the best possible gray level stroke map.

A level set function ϕ is first computed from the stroke map extracted in the previous step using a distance transform algorithm[4] [38]. The level set function is constructed such that region where $\phi > 0$ corresponds to the estimated stroke pixels. Then, the stroke map is eroded by updating the level set function ϕ with (4) subject to

$$F = F_L + \beta F_A + \nu F_R \tag{20}$$

to propagate the interface. Here, β and ν are positive real numbers that are used to balance the effects of the three forces. After propagation of the interface, a binarization can be produced by thresholding $\phi > 0$. Force F_L is the main driving force in (20) and is based on local region modeling of the underlying stroke and background properties. However, since the stroke map extracted during initialization may be imprecise, an area reduction force F_A is included to promote the removal of less likely stroke pixels. In order for this to work as expected, parameter D of (16) is set to -1. Finally, force F_R tends to smooth the boundary and to produce more natural-looking results.

Strategies to choose the value of the various parameters is discussed in Sect. 3.

2.4.3 Third step: stroke map optimization

In this step, the goal is to capture the weaker parts of the strokes and to produce the final binarization. After the second step has been completed, a map indicating the most likely stroke pixels is available. From this map, an SGL is computed using the method from Sect. 2.2.2, and the level set function ϕ is further refined by updating it with (4) subject to

$$F = F_L + \gamma F_G + \nu F_R \tag{21}$$

until the level set function stagnates. When comparing (20) and (21), it can be noted that in (21) the area contraction force F_A has been removed in favor of the SGL-based force F_G. The purpose of the latter is to provide segmentation cues based on the local average gray level of the stroke pixels and not to enforce any geometric a priori information. In this respect, F_G is complementary to F_L. However, two important distinctions apply: (1) the local linear model of F_L is different from the local average model of F_G, the local linear model being more suitable in order to follow strokes with increasing or decreasing pixel intensities and (2) the SGL map of F_G is computed once, during the *second step*, while the local linear model parameters of F_L are updated every few iterations of (4). Roughly speaking, F_L drives the binarization process based on the local structure, while F_G prevents the contour from leaking in regions with very different intensities from those of the SGL map. At the end of this process, the final binarization is obtained by a simple thresholding of the level set function, i.e. $\phi < 0$.

2.4.4 Level set evolution and stopping criterion

The level set-based steps described in Sects. 2.4.2 and 2.4.3 are both iterative. Indeed, after the level set function ϕ has been initialized to some position, it evolves based on (4)

[4] Numerically, this is done by using the function `bwdist` of Matlab.

Fig. 19 Sample image from the Juma Al Majid Center for Cultural Heritage dataset (Dubai): **a** input image, **b** binarization obtained using Otsu's method, and **c** binarization obtained using Sauvola's method. The *middle column* shows two intermediate results of the proposed method: **d** the stroke map initialization and **e** the stroke map erosion. The final result is in **f**. Images **g**, **h**, and **i** in the *right column* correspond to the framed areas in figures **d**, **e**, and **f**

subject to the net force (20) or (21), for the stroke map erosion step or the stroke map optimization step, respectively. The iterations are terminated when the contours *stagnate*. This condition is defined formally as follows: If the number of pixels that have changed their label did not increase after a period of N iterations, then the iterations are stopped. Figure 9 shows a typical example of the progression of the number of pixels that changed their label with respect to the number of iterations. As we can deduce from the shape of the curve, the final result is not very sensitive to the number of iterations N used to detect the stagnation of the contour. At the same time, it is important that $N > 1$, because the typical time step δt in (4) will be smaller than 1, so it may take several iterations before a pixel changes its label. In this work, we use an arbitrary large number and we fixed $N = 100$.

2.5 Parameter selection

The initialization method presented in Sect. 2.4.1 can be considered as parameterless. However, the level set formulations of Sects. 2.4.2 and 2.4.3 have various parameters that must be handpicked. More specifically, in (20) of the second step, F_L has parameters μ_1, W_1, β, and ν_1, which are used to balance the various forces. In the third step, the computation of the SGL map depends on the width σ of the Gaussian kernel and (21) has the following parameters: μ_2, W_2, γ, and ν_2. We note that the parameters W_1 and W_2 are related to the stroke width and that the parameters μ_i and ν_i are related to the document image contrast. Thus, we set $W = W_1 = W_2$, $\mu = \mu_1 = \mu_2$, and $\nu = \nu_1 = \nu_2$. The total number of free parameters is then only 6.

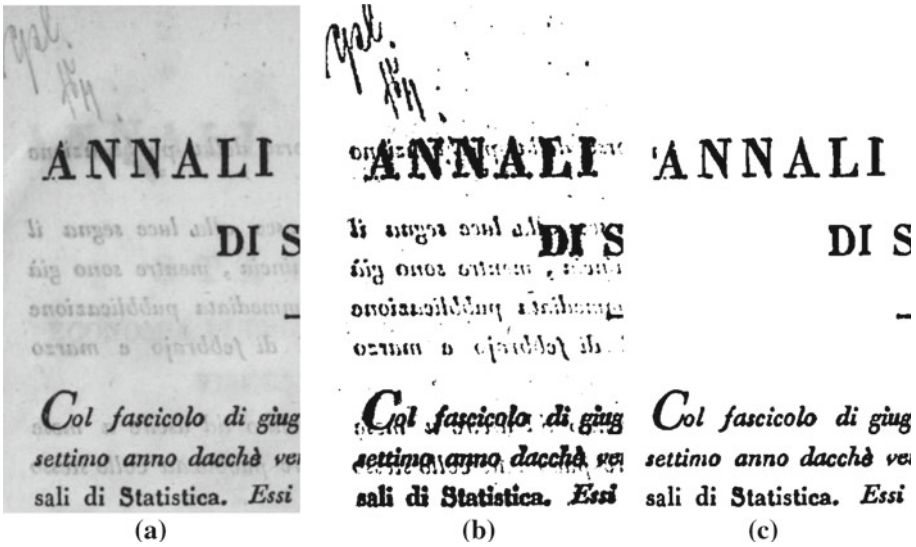


Fig. 20 Sample image from the Google Book Search dataset: **a** input image, **b** binarization results obtained using Sauvola's method and **c** the proposed method

The parameter W depends on the extend of the text blobs. Therefore, we find it useful to relate it to the estimate of the stroke width w_s computed during the initialization step (for details, see Sect. 2.4.1 and Appendix A). The following relations have been used to adapt the parameters to the input image:

$$W = \lambda w_s.$$

In addition, because parameters μ, ν, β, and γ account for the intensity-dependent quantities, they will give more consistent results across images if the pixel intensities are normalized. In this work, a simple linear rescaling, which maps the pixel intensities to values in [0.0, 1.0], has been used.

Since the process is divided into three steps, not all the parameters have to be adjusted at the same time, but only few of them. In this work, the parameters used during the stroke map erosion step (λ, μ, ν, and β) are first adjusted. Once this is done, the value for σ and γ is fixed. It is worth noting that the parameters have a rather independent behavior, which make their adjustment easier. For example, λ sets the sensitivity of the method, ν favors straight lines, and β determines the robustness to noise (maybe at the expense of recall performance). The method is much less sensitive to the specific value of μ, σ, and γ.

3 Experimental results and discussion

The proposed algorithm was tested on document images representing both handwritten and printed characters. The robustness of the proposed method with respect to the SM initialization is first demonstrated. Then, results obtained on images with synthetic degradations are presented. Finally, results obtained on real images are shown and analyzed.

3.1 Evaluation of the sensitivity of the level set method with respect to the SM initialization

The proposed method has been used to binarize the document image shown in Fig. 10. In order to assess the dependency of the proposed method on the stroke map initialization, a varying amount of salt and pepper noise have been added to the computed SM. The amount of noise used has been varied from 5 to 100% by 5% step, where $p\%$ of noise corresponds to the case where $p\%$ of the document image pixels has been randomly assigned to either 1 (*stroke*) of 0 (*background*).

As can be seen from Fig. 11, the final results are quite robust with respect to perturbation of the SM initialization. The binarization produced by the method remains unperturbed with level of salt and pepper noise of up to 95%. The F-measure, described in Sect. 3.4, has been used to quantify the result. A F-measure of 99.56% has been obtain at all levels of perturbation, except at 100% noise where a F-measure of 27.81 was obtained. When a totally randomized SM is used for initialization, the stroke and background region are inverted in some regions of the image. This phenomenon has been called *local twisting in* [39] and is a consequence of the local character of the method; one part of the image does not know about the other. This problem is easily avoided by using a SM initialization that contains at least some information.

Notwithstanding the good result presented in this experiment, it is important to note that a good initialization can be decisive in many cases. For example, when the image to be binarized contains large areas of stain, the local linear model

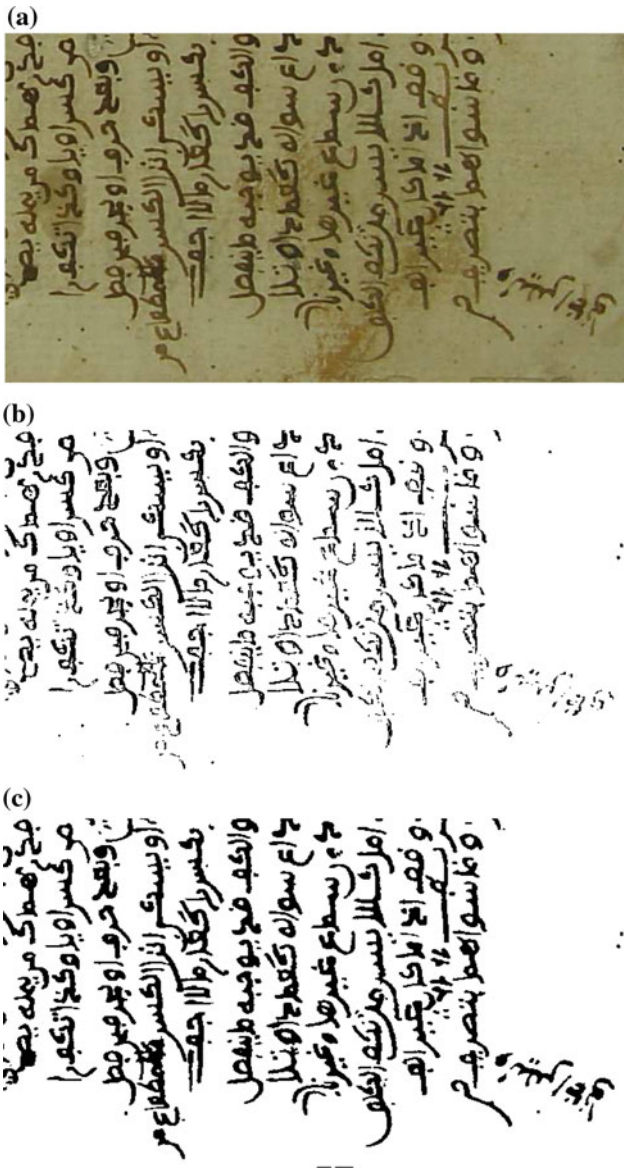


Fig. 21 Sample image from the Algeria dataset. **a** input image, **b** binarization results obtained using Sauvola's method, and **c** the proposed method

may not receive enough information in order to result in a good binarization of the document image.

3.2 Evaluation on images with synthetic degradation

3.2.1 Bleed-through, blurred, noisy, and low contrast documents

In this section, results obtained on images with synthetic degradations are presented. Four scenarios are investigated, as presented in Table 1. The proposed method has been applied to each degraded image, and the resulting binarization is fed to a commercial OCR software,

Table 2 Average F-measure for the ten images of the test dataset from the DIBCO'09 contest

Method	F-measure
Otsu's	78.60
Sauvola's	87.26
The multi-scale grid-based Sauvola [6]	89.26
Lu and Tan algorithm [13]	91.24
Fabriozi and Marcotegui algorithm [13]	90.06
The proposed, first step only	82.85
The proposed, first and second steps only	89.04
Proposed	89.34

The performances of Otsu's and Sauvola's methods are reported from [6]. Other results are available in [13]

Fine Reader 9.[5] As the proposed method has not been designed for such relatively high level of burring and noise, a preprocessing step has been included in scenario 2 and 3. In scenario 2, the blind deconvolution algorithm [40,41] with a Gaussian point spread function with size parameter $\sigma = 1/(\mathrm{D_{max}}\sqrt{2\pi})$ has been applied. $\mathrm{D_{max}}$ is the maximal image derivative, as computed with the Sobel's operator. In scenario 3, a median filter of size $w \times w, w = 2\lceil w_2/2 \rceil + 1$, where w_s is an estimation of the stroke width,[6] has been applied. Binarization and OCR results are shown in Figs. 12, 13, 14 and 15. The OCR error rate has been quantified using the *string edit distance*, as defined in [6]. For each scenario, the values obtained as a function of the degree of degradation are plotted in Fig. 16.

As can be seen from the results, the proposed method is fairly resilient to moderate level of bleed-through, blurring, and Gaussian white noise. In those three cases, the OCR performance was not linear: The computed string edit distance remained very low until the degradation level reaches a critical point, which is around $d = 0.12$ for bleed-through, $\sigma_B = 0.15$ for blurring and $\sigma_N = 30/255$ for the addition of white noise. Except for the bleed-through case, we believe that those values are much higher than what is commonly observed in practice. In the case of bleed-through, the achieved performance level seems sufficient for processing of most normal and historical documents. For documents that suffer from very high level of bleed-through degradation, other methods, which take into account the verso side of the document, exist and might be more appropriate [42]. Finally, the experiments demonstrate that the binarization method is independent of contrast.

[5] http://finereader.abbyy.com/. See also Sect. 3.5.

[6] See Appendix A for a description of the method used to automatically estimate the stroke width.

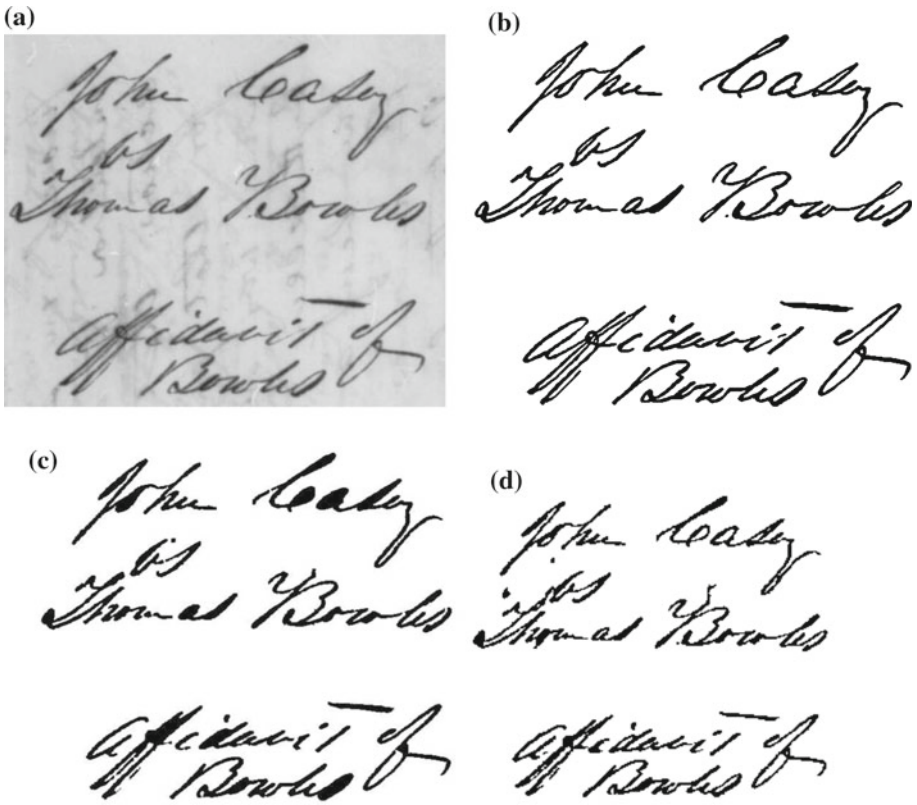


Fig. 22 Image H03 from the DIBCO'09 test dataset: **a** input image, **b** binarization results obtained using the ground truth, **c** the Lu and Tan algorithm [13], and **d** the proposed method

3.2.2 Simulated faded ink degradation

The proposed method was also applied on images with simulated faded ink degradation, which mimics the fading degradation of historical documents. A sample image is shown in Fig. 17. Starting from the uncorrupted image, progressive level of degradation has been added using the following procedure:

First, a random set of circles with a radius of a factor of stroke width are generated. The area within the circles is converted to gray scale using a Euclidean distance transform. This gray map, u_{gray}, is used to mix the input image and its estimated background. A parameter, called the degradation level, and denoted as L, controls the dominance of the background on the final degraded result as follows:

$$u_{\text{deg}} = (1 - L/10)u_{\text{gray}}(r)u(r) + (L/10)u_{\text{gray}}(r)\text{EB}(r)$$

where EB is the estimated background. Using a ground-truth binary map, this degradation is only restricted to the text regions. The total area of the circles is fixed to 40 percent of the image area in the following experiments.

The kind of synthetic degradation that has been added is very challenging for the proposed method because it results in faded spot on the stroke. Nevertheless, as can be seen from Fig. 17, the method is able to correctly binarize all the text up to a degradation level of 7. Also, based on the F-measure performance presented in Fig 18, the proposed method outperforms other methods such as Otsu's and Sauvola's. The degradations start to impact the final result at level 7 or 8, depending on the case, which we believe is quite satisfactory.

3.3 Subjective evaluation on real images

We applied our algorithm to difficult document images of Latin, Arabic, and Persian handwritten text taken from various databases: the Google Book Search dataset [43], which contains scanned, resized, and flattened images of books; a Latin and Arabic manuscript dataset [44], which contains a large number of ancient documents, including Arabic manuscripts from Tombouctou; a dataset from the national Library of Algeria; and two old manuscripts, courtesy of the Juma Al Majid Center for Cultural Heritage (Dubai),[7] which together contain about 500 pages. Images from the DIBCO'09 contest sample set were also used. As the ground truth for most of those images is not available, binarization results achieved by using our algorithm are presented, along with those obtained by using Otsu method [1], Sauvola's method [4], the PDE-based method [23], and the methods that placed first in the DIBCO'09. Sample results can be seen in Figs. 19, 20 and 21. The full images are available on the Internet.[8] The authors acknowledge that the DIBCO'09 participants did not have

[7] http://www.almajidcenter.org/English/Pages/default.aspx.

[8] http://www.synchromedia.ca/web/reza/expres/LSBIN10.

Fig. 23 Image P03 from the DIBCO'09 test dataset: **a** Input image, **b** binarization results obtained using Otsu's method, **c** binarization results obtained using Sauvola's method, and **d** binarization results obtained using the proposed method

access to the dataset, and therefore were not able to adapt the parameters of their methods.

A step-by-step illustration of our method is provided in Fig. 19d–f. As can be seen in Fig. 19d, the method described in [23] and used in this paper for the stroke map extraction step already performs better than both Otsu's and Sauvola's methods. The effect of the stroke map erosion step is illustrated in Fig. 19e and allows for the removal of many outlier regions. Finally, the final boundary optimization allows for both a finer delineation and the recovery of weaker strokes.

In addition, figures illustrating the behavior of the proposed method in presence of bleed-through (Fig. 20), stain (Fig. 21), and low-quality documents (21) are provided. In all cases, our methods clearly outperform the other methods in terms of the precision of the binarization. It is worth noting that the performance of the proposed method is not high in the case of a large amount of the bleed-through. The interfering patterns can pass through the rough binarization step onto the stroke map and then onto the final output. In this case, it is recommended to use the recursive binarized map introduced in [6] instead of the SM. The output of the recursive binarization methods consists of a scatter set of black pixels with a high degree of confidence that they belong to the text pixels. We will investigate this combination of recursive binarization methods as the initialization and level set framework in the future.

3.4 Objective evaluation against DIBCO'09

The binarization algorithm described in this paper was submitted to the DIBCO'09 binarization contest, the results of which were announced during the ICDAR'09 conference [13]. It placed 3rd among 43 algorithms submitted by 35 different international teams.

The DIBCO'09 database[9] is composed of a total of 14 images divided into two sets: 4 images were provided as samples before the contest, and 10 images were used for testing. The 10 test images were not available to participants before the contest. In the two sets, half the images were of handwritten text and the rest were of printed text. The document images present Latin texts, feature a variety of font faces and script styles, and suffer from various types of degradation.

3.4.1 Performance evaluation

Various objective measures can be used to evaluate the performance of binarization algorithms. During the DIBCO'09 binarization contest, four measures were used to evaluate the performance of the methods: the F-measure, the peak signal-to-noise ratio, the negative-rate metric, and the misclassification penalty metric. [13]. However, from what can be seen in [13], those values are highly correlated. Therefore, we chose to concentrate on the F-measure, because it is well-widely accepted and simple, and so is easy to interpret. The F-measure is defined as follows:

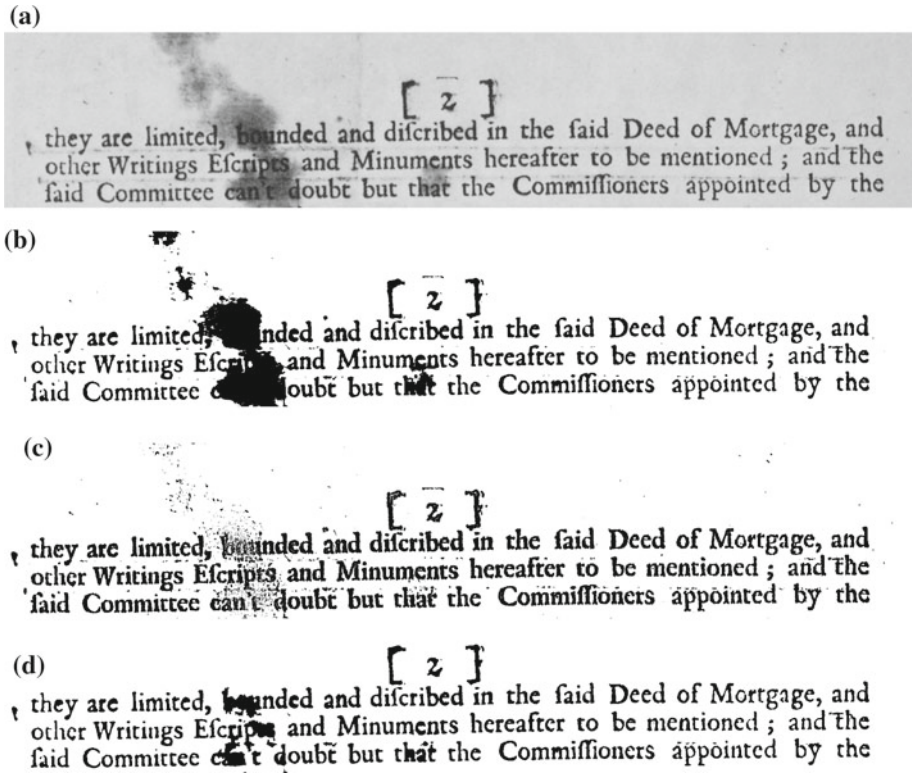
$$\text{F-measure} = \frac{2 \times \text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} \tag{22}$$

where $\text{Recall} = \frac{TP}{TP+FN}$, $\text{Precision} = \frac{TP}{TP+FP}$ and TP, FP, and FN representing the number of true positive, false positive and false negative values respectively.

3.4.2 Model selection

The four sample images provided for the DIBCO'09 contest were used for model selection. Initially, a set of parameters was handpicked by trial and error. Then, this set of

[9] Available online: http://users.iit.demokritos.gr/~bgat/DIBCO2009/benchmark/.



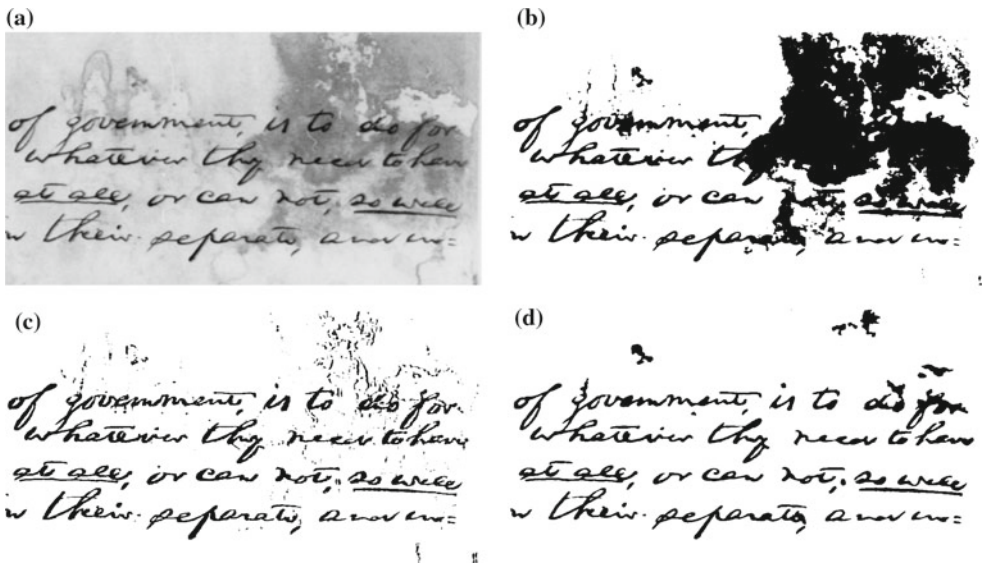


Fig. 24 Image P04 from the DIBCO'09 test dataset: **a** Input image, **b** binarization results obtained using Otsu's method, **c** binarization results obtained using Sauvola's method, and **d** binarization results obtained using the proposed method

Fig. 25 Image H04 from the DIBCO'09 test dataset: **a** Input image, **b** binarization results obtained using Otsu's method, **c** binarization results obtained using Sauvola's method, and **d** binarization results obtained using the proposed method

parameters was refined by a heuristic optimization procedure. Various values for the parameters related to the second step (μ_1, W_1, W_{min1}, β, and ν_1) were tested on the four sample images. For each test, only one parameter was different from those in the original parameter set. The parameter values resulting in the best average F-measure for all four images were then selected and fixed. The same procedure was repeated to find the best parameters related to the third process step (μ_2, W_2, W_{min2}, γ, w, p, and ν_2).

3.4.3 Results

The ten images of the DIBCO'09 test set were processed using the model found in Sect. 3.4.2. The values of the

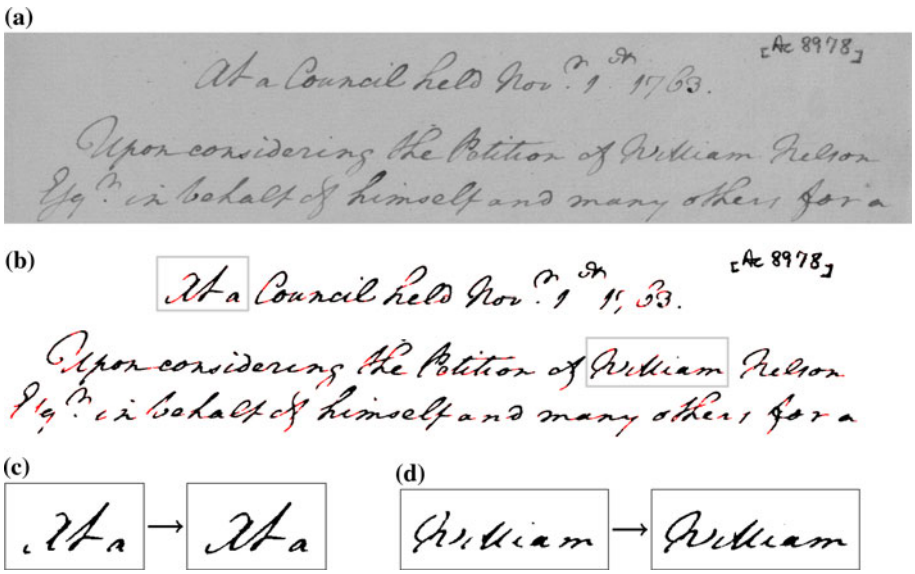


Fig. 26 Image H01 from the DIBCO'09 dataset: **a** Input image, **b** binarization results obtained using the proposed method, **c** and **d** details of the binarization after the stroke map erosion step (*left*) and after the stroke map optimization step (*right*). Note that (**b**) is best seen in color; the *red strokes* indicate the improvements due to the stroke map optimization

performance measures for the individual images are presented in Table 2. Also, sample results are presented in Figs. 22, 23, 24, 25 and 26 along with results obtained using the Otsu and Sauvola methods. It can be to noted that the proposed method leads to more consistent result with more continuous strokes.

3.5 OCR evaluation with printed DIBCO'09 images

A typical application of a binarization algorithm is to prepare a document image before performing OCR. In this respect, it is interesting to measure how the proposed algorithm might improve the recognition rate of a specific OCR engine. Binarized versions of the five images representing printed documents from the DIBCO'09 test set were analyzed using an OCR engine for analysis. For this purpose, we considered the performance of the commercial OCR software Fine Reader 9,[10] and opted to use it. The ground-truth text was generated manually in the form of a string of characters. All line breaks were replaced with space characters. The OCR error, defined in [6], of various methods is provided in Table 3. Binarized images were produced using the proposed method, and also Otsu's and Sauvola's methods. The results obtained are presented in Table 3.

As an example of the limit of the proposed method, an image from DIBCO'09 dataset is used in Fig. 24. The outputs of different methods including Sauvola's method and the proposed method are presented in the figure. The Sauvola's method works better in this case and discovers the degraded letters. The lower performance of the proposed method is because of the initialization limit which can be improved in future. Also, the lower performance on this image is the reason for slightly lower score of the proposed method in Table 3.

Table 3 OCR recognition error of the binarization methods

Method	Recognition error
Otsu's	0.392
Sauvola's	0.575
The multi-scale grid-based Sauvola [6]	0.337
The proposed method	0.357

3.6 Computational cost and complexity of the method

With our current implementation, the proposed method is computationally intensive. For example, the processing of the 512×512 pixel image from Fig. 19 with a window size of $W_1 = W_2 = 51$ takes 25, 253, and 348 s for steps 1, 2 m and 3 of the proposed method, respectively, on a single core of a 2.6 GHz AMD Opteron 885 processor.

More generally, the computational cost of the initialization procedure described in Sect. 2.3 is $O(n \cdot m^2)$, where n and m are the number of pixels and the size of the window, respectively. However, the complexity reduces to $O(n)$ using integral image representation [45].

For the two level set steps, the computation of (2) is linear in terms of the number of pixels, n, and is then performed in time $O(n)$. The sums in (11) computed at each pixel location are evaluated in time $O(n \cdot m^2)$. Other operations of (11) are linear in n. Overall, the computational time for those steps is then $O(n) + O(n \cdot m^2) + O(k) = O(n \cdot m^2)$ per iteration. It is also worth noting that the distance transform algorithm used to construct the level set function is linear in the number of pixels [38].

[10] http://finereader.abbyy.com/.

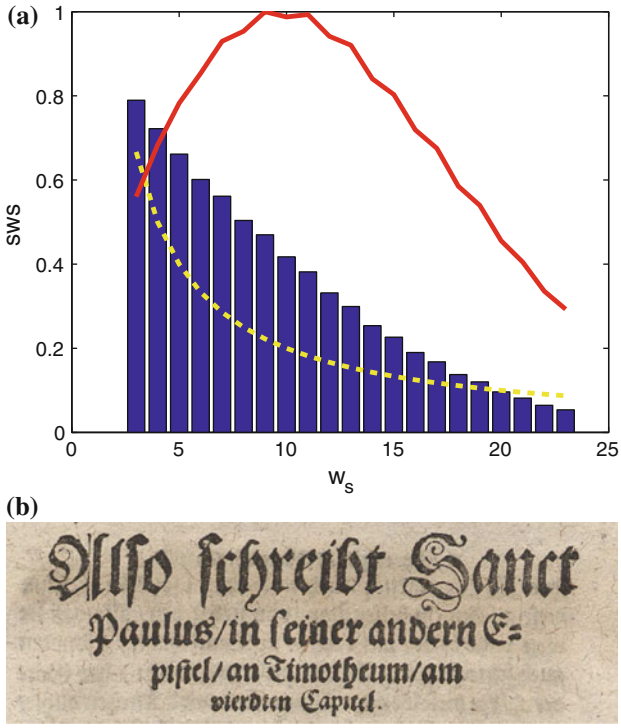


Fig. 27 **a** An example of the stroke width spectrum (SWS). The *model* is shown in *yellow*. As seen from the figure, the maximum is at $\bar{w}_s = 11$, and therefore $w_s = 11$ is obtained as the average stroke width. **b** The input image used for computing the SWS. The grid-based Sauvola method is used to obtain a rough binarization

A few strategies can be put in place to reduce the computational burden. Since the parameters of the local linear model vary slowly in time, it is not necessary to compute them at every iteration, thereby reducing the actual computational cost. In addition, as there is no global image operation, a good acceleration is expected from a narrow band level set implementation, which would reduce the complexity to approximately $O(\sqrt{n} \cdot m^2)$. Also, the formulation of the level set method allows for GPU implementation that could further reduce the execution time [46,47].

4 Conclusion and future prospects

A binarization method for old and degraded document images has been developed based on local linear probabilistic models embedded within a level set implementation of an active contour scheme. The local linear models of the stroke and background intensities, introduced here in the context of document image binarization, are used to derive the main driving force for the displacement of the level set function. A complementary SGL model was used to restrain the progression of the contour outside regions of stroke-like intensity. Other forces, traditional in the level set framework, such as the curvature and area contraction forces, have also been used. Also, a multi-level classifier, the stroke map, has been used to provide a rough initialization of the level set function.

During our experiments, the proposed three-step binarization process proved very convenient. Each step has its own goal and produces intermediate results that facilitate understanding of the algorithm. The three goals are initialization, outlier removal, and final binarization, respectively. In many cases, the initial binarization will be incrementally improved during the process, and this effect is reflected in the results presented in Table 2. It is clear that the improvement in binarization quality due to each step of the proposed method will vary depending on the setting. For example, although the improvement due to the last binarization step in Fig. 19 might not stand out visually, inspection of the enlarged areas (Fig. 19g–i) reveals that this step did improve the connectivity, which might be an important cue for some recognizers. In contrast, the benefit of the last step is more perceptible in Fig. 26. Whether this benefit worth the computational investment depends on the intention of the end user.

The level set methodology is an appealing integration framework because of its elegant formulation, power, and simplicity. The method has been tested on different datasets including DIBCO'09 dataset, with promising results. It placed 3rd among 43 methods at the DIBCO'09 contest.

As a prospect for the future, application of the recursive binarization methods to initialize the level set framework will be considered in order to enable the method to remove the bleed-through signatures on the highly degraded document images. Also, improvement to the driving forces to make them more adaptable to the variations on the degraded document images will be considered.

Acknowledgments The authors would like to thank the NSERC of Canada for their financial support.

A Estimation of the average stroke width

In order to estimate the average stroke width, we first compute the *stroke width spectrum* (SWS), which is a measure of the frequency of various stroke width values on the document image. In order to determine this spectrum, a kernel-based approach on a binarized version of the image is used for each given stroke width candidate, $\bar{w}_s$. For each pixel on the image domain, if the ratio of text pixels in a patch of size $\bar{w}_s$ around that pixel is higher than a threshold (say 0.9), that pixel is counted as $\bar{w}_s$. In order to reduce the computational cost, integral image representation [11,45] is used to obtain the stroke width spectrum, a sample of which is shown in Fig. 27 as blue bars. Usually, $\bar{w}_s$ is tested from 3 to 23 pixels. As can be seen from the spectrum, it consists of many jumps, the highest of which can be associated with the average stroke

width on the input image. This behavior of the stroke width spectrum is very similar to the first ionization potential curve in inorganic chemistry [48]. However, to increase the accuracy of this estimation, we use a model-based approach. In this model, we assume that the ideal input images contain strips of a constant width w_s. For these images, it is easy to show that SWS is equal to $1/w_s$ for $\bar{w}_s = w_s$. This model is shown in Fig. 27 as a yellow-dashed line. By dividing the SWS by this model, a characteristic curve for the document is obtained which reaches its highest value at the average stroke width, w_s. The curve is shown in the same figure as a continuous line and is normalized to 1. An average stroke width of 11 can be estimated from the image in the figure.

References

1. Otsu, N.: A threshold selection method from gray-level histograms. IEEE Trans. Syst. Man Cybern. **9**, 62–66 (1979)
2. da Silva, J.M.M., Lins, R.D., Martins, F.M.J., Wachenchauzer, R.: A new and efficient algorithm to binarize document images removing back-to-front interference. J. Univers. Comput. Sci. **14**(2), 299–313 (2008)
3. Robert Milewski, V.G.: Binarization and cleanup of handwritten text from carbon copy medical form images. Pattern Recognit. **41**, 1308–1315 (2008)
4. Sauvola, J., Pietikäinen, M.: Adaptive document image binarization. Pattern Recognit. **33**(2), 225–236 (2000)
5. Trier, O., Jain, A.: Goal-directed evaluation of binarization methods. IEEE Trans. Pattern Anal. Mach. Intell. **17**(12), 1191–1201 (1995). doi:10.1109/34.476511
6. Farrahi Moghaddam, R., Cheriet, M.: A multi-scale framework for adaptive binarization of degraded document images. Pattern Recognit. **43**, 2186–2198 (2010). doi:10.1016/j.patcog.2009.12.024
7. Drira, F., Le Bourgeois, F., Emptoz, H.: Restoring ink bleed-through degraded document images using a recursive unsupervised classification technique. Doc. Anal. Syst. **VII**, 38–49 (2006). doi:10.1007/11669487_4
8. Tan, C.L., Cao, R., Shen, P., Wang, Q., Chee, J., Chang, J.: Removal of interfering strokes in double-sided document images. In: Fifth IEEE Workshop on Applications of Computer Vision, 2000, pp. 16–21. Palm Springs, CA (2000). doi:10.1109/WACV.2000.895397
9. Su, B., Lu, S., Tan, C.L.: Binarization of historical document images using the local maximum and minimum. In: DAS'10, pp. 159–166. ACM, Boston, Massachusetts (2010) doi:10.1145/1815330.1815351
10. Cheriet, M.: Extraction of handwritten data from noisy gray-level images using a multiscale approach. Int. J. Pattern Recognit. Artif. Intell. **13**(5), 665–684 (1999)
11. Shafait, F., Keysers, D., Breuel, T.M.: Efficient implementation of local adaptive thresholding techniques using integral images. In: Document Recognition and Retrieval XV. San Jose, USA (2008)
12. Hedjam, R., Farrahi Moghadam, R., Cheriet, M.: Text extraction from degraded document images. In: EUVIP'10, pp. 248–253. Paris, France (2010)
13. Gatos, B., Ntirogiannis, K., Pratikakis, I.: Icdar 2009 document image binarization contest (dibco 2009). In: ICDAR (2009)
14. Serra, J.: From Pixels to Features, Ch Toggle Mappings. pp. 61–72 . Elsevier, Amsterdam (1989)
15. Naegel, B., Wendling, L.: A document binarization method based on connected operators. Pattern Recognit. Lett. **31**(11), 1251–1259 (2010)
16. Cao, H., Govindaraju, V.: Preprocessing of low-quality handwritten documents using markov random fields. IEEE Trans. Pattern Anal. Mach. Intell. **31**(7), 1184–1194 (2009). doi:10.1109/TPAMI.2008.126
17. Dimov, D., Dimov, A.: Data driven approach to binarization of astronomical images. In: CompSysTech'10, pp. 478–484. ACM, Sofia, Bulgaria (2010). doi:10.1145/1839379.1839465
18. Bar-Yosef, I., Mokeichev, A., Kedem, K., Dinstein, I., Ehrlich, U.: Adaptive shape prior for recognition and variational segmentation of degraded historical characters. Pattern Recognit. **42**(12), 3348–3354 (2009). doi:10.1016/j.patcog.2008.10.005
19. Tan, C.L., Cao, R., Shen, P.: Restoration of archival documents using a wavelet technique. IEEE Trans. Pattern Anal. Mach. Intell. **24**(10), 1399–1404 (2002). doi:10.1109/TPAMI.2002.1039211
20. Rivest-Hénault, D., Cheriet, M.: Unsupervised mri segmentation of brain tissues using a local linear model and level set, Elsevier MRI (online) (2010). doi:10.1016/j.mri.2010.08.007
21. Gatos, B., Pratikakis, I., Perantonis, S.: Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information. In: ICPR'08, pp. 1–4. (2008)
22. Farrahi Moghaddam, R., Rivest-Hénault, D., Cheriet, M.: Restoration and segmentation of highly degraded characters using a shape-independent level set approach and multi-level classifiers. In: ICDAR'09, pp. 828–832. Barcelona, Spain (2009). doi:10.1109/ICDAR.2009.107
23. Farrahi Moghaddam, R., Cheriet, M.: RSLDI: restoration of single-sided low-quality document images. Pattern Recognit. **42**, 3355–3364 (2009)
24. Sethian, J.: Level Set Methods and Fast Marching Methods. Cambridge University Press, Cambridge (1999)
25. Osher, S., Sethian, J.A.: Fronts propagating with curvature dependent speed: algorithms based on Hamilton–Jacobi formulations. J. Comput. Phys. **79**(2), 12–49 (1988)
26. Malladi, R., Sethian, J.: A unified approach to noise removal, image enhancement, and shape recovery. IEEE IP **5**, 1554–1568 (1996)
27. Caselles, V., Kimmel, R., Sapiro, G.: Geodesic active contours. Int. J. Comput. Vis. **22**, 61–79 (1997)
28. Tsai, A., Yezzi, A., Willsky, A.S.: Curve evolution implementation of the Mumford–Shah functional for image segmentation, denoising, interpolation, and magnification. IEEE Trans. Image Process. **10**(8), 1169–1186 (2001)
29. Vese, L.A., Chan, T.F.: A multiphase level set framework for image segmentation using the Mumford and Shah model. Int. J. Comput. Vis. Vol. **50**(2), 271–293 (2002)
30. Chan, T.F., Vese, L.: Active contours without edge. IEEE Trans. Image Process. **10**(2), 266–277 (2001)
31. Paragios, N., Deriche, R.: Geodesic active regions: a new framework to deal with frame partition problems in computer vision. J. Visual Commun. Image Rep. **13**(2), 249–268 (2002)
32. Liu, J.: Robust image segmentation using local median. In: Proceedings of the of the 3rd Canadian Conference on Computer Robot Vision, p. 31. (2006)
33. Rivest-Hénault, D., Cheriet, M.: Image segmentation using level set and local linear approximations. In: Kamel, M., Campilho, A. (eds.) ICIAR 2007, LNCS 4633, pp. 234–245. Springer, Berlin (2007)
34. Rosenhahn, B., Brox, T., Weickert, J.: Three-dimensional shape knowledge for joint image segmentation and pose tracking. Int. J. Comput. Vis. **73**(3), 243–262 (2007)
35. Li, C., Kao, C.-Y., Gore, J.C., Ding, Z.: Minimization of region-scalable fitting energy for image segmentation. IEEE Trans. Image Process. **17**(10), 1940–1949 (2008)
36. Ben Ayed, I., Mitiche, A., Belhadj, Z.: Polarimetric image segmentation via maximum-likelihood approximation and efficient

multiphase level-sets. IEEE Trans. Pattern Anal. Mach. Intell. **28**, 1493–1500 (2006)
37. Brox, T., Cremers, D.: On local region models and a statistical interpretation of the piecewise smooth mumford-shah functional. Int. J. Comput. Vis. **84**, 184–193 (2009)
38. Breu, H., Gil, J., Kirkpatrick, D., Werman, M.: Linear time euclidean distance transform algorithms. IEEE PAMI **17**, 529–533 (1995)
39. Liu, J., Chelberg, D., Smith, C., Chebrolu, H.: Distribution-based level set segmentation for brain mr images. In: BMVC07. (2007)
40. Farrahi Moghaddam, R., Cheriet, M.: Beyond pixels and regions: a non-local patch means (NLPM) method for content-level restoration, enhancement, and reconstruction of degraded document images. Pattern Recognit. **44**(2), 363–374 (2011). doi:10.1016/j.patcog.2010.07.027
41. Haykin, S.S.: Blind Deconvolution. PTR Prentice Hall, Englewood Cliffs (1994)
42. Farrahi Moghaddam, R., Cheriet, M.: Low quality document image modeling and enhancement. IJDAR **11**(4), 183–201 (2009). doi:10.1007/s10032-008-0076-2
43. Google, Book Search Dataset, version v Edition (2007)
44. Deriche, R., Faugeras, O.: Les EDP en traitement des images et vision par ordinateur. Technical Report 2697, INRIA, 1–63 (1996)
45. Bradley, D., Roth, G.: Adaptive thresholding using the integral image. J. Graph. GPU Game Tools **12**(2), 13–21 (2007)
46. Cremers, D., Fluck, O., Rousson, M., Aharon, S.: A probabilistic level set formulation for interactive organ segmentation. In: SPIE Medical Imaging. (2007)
47. Mei, X., Decaudin, P., Hu, B., Zhang, X.: Real-time marker level set on gpu. In: International Conference on Cyberworlds, CW '08. (2008)
48. Mackay, K., Henderson, W., Mackay, R.: Introduction to Modern Inorganic Chemistry. CRC Press, Boca Raton (2002)