A unified framework based on the level set approach for segmentation of unconstrained double-sided document images suffering from bleed-through

Reza Farrahi Moghaddam[†], David Rivest-Hénault[†], Itay Bar-Yosef[‡] and Mohamed Cheriet[†]
[†]Synchromedia Laboratory for Multimedia Communication in Telepresence
École de Technologie Supérieure, Montréal (QC), H3C 1K3, Canada
[‡]Department of Computer Science, Ben-Gurion university, Israel
{reza.farrahi-moghaddam, david.rivest-henault}.1@ens.etsmtl.ca,
itaybar@cs.bgu.ac.il, mohamed.cheriet@etsmtl.ca

Abstract

A novel method for the segmentation of double-sided ancient document images suffering from bleed-through effect is presented. It takes advantage of the level set framework to provide a completely integrated process for the segmentation of the text along with the removal of the bleed-through interfering patterns. This process is driven by three forces: 1) a binarization force based on an adaptive global threshold is used to identify region of low intensity, 2) a reverse diffusion force allows for the separation of interfering patterns from the true text, and 3) a small regularization force favors smooth boundaries. This integrated method achieves high quality results at reasonable computational cost, and can easily host other concepts to enhance its performance. The method is successfully applied to real and synthesized degraded document images. Also, the registration problem of the double-sided document images is addressed by introducing a level set method; the results are promising.

1. Introduction

Restoration and enhancement of historical documents is of expanding interest to the document processing community. The large number of un-processed old documents, the high degree of variations in writing style and the severe degradation are some of the most important challenges in this field.

One of the main causes of degradation encountered in historical documents is the bleed-through phenomenon. This process is characterized by the seepage of the ink from the verso side through the paper and appearing on the recto side. As a result of the seepage of the ink from each side of the page to the other, some interference patterns are visible on both image side. This effect results in a low level of readability and legibility of the document. It follows that this degradation interferes with subsequent understanding processes, which may be applied to the document image, and lowers their performance. For document that have been poorly conserved, e.g. because of a high humidity level, the effect of bleed-through degradation is usually important. Also, because the the verso side contains text information, the resulting interfering patterns on the recto side cannot be treated as random noise.

The bleed-through problem have been addressed from different points of view. For example, a global thresholding method based on entropy has been used. Statistical methods have also been applied to this problem. Multi-stage processing based on segmentation and inpainting, and edge extraction and stroke reconstructionhave be used. See [3], for a more complete discussion. A PDE-based direct restoration method, which uses the reverse diffusion of information, has also been developed and applied successfully [3]. In the latter, the interfering patterns are weakened using the reverse diffusion, and then the reminding parts are removed by the normal diffusion process. This approach provides a robust and powerful solution for the bleed-through problem. However, as the information is diffused during the process, the intensities of true stroke pixels are also changed, which can result in the creation of artifacts.

Each of the above mentioned methods uses characteristics of the phenomenon and tries to restore the degraded image based on those. Using a unified framework, all these characteristics can be integrated within one approach. Not only this has the potential to improve the performance of the restoration, but also it allows the user to select the best set of characteristics based on the complexity of his problem. A generic framework that is suitable for such an integration is the level set method. In addition, the evolutive nature of the level set method allows to refine the stroke and background models as the segmentation progresses. This "on-

978-0-7695-3725-2/09 $25.00 © 2009 IEEE
DOI 10.1109/ICDAR.2009.108

line" updating not only speeds up the process, but also it reduces the need for parameter estimation before the application of restoration method. Level set methods have been used extensively in many fields of image processing mainly in segmentation applications [1, 6]. This method provides inherently continuous boundaries and can support local and global forces. For example, if the concept of the reverse diffusion can be expressed into the level set framework, a segmentation method for degraded document images will be obtained which doesn't alter the true pixel intensities. The output of the new method is the contour of the true text, which is actually a segmentation of the input, while the input image has not been changed. Therefore, the weak strokes can be preserved. It worth noting that the extracted contours of the new method can then be supplied to other restoration methods (for example, PDE-based method [3]) as a priori information.

In this work, a novel restoration method for double-sided document images suffering from bleed-through effect is introduced. The method is based on the level set method and uses a global thresholding force and the reverse diffusion force to obtain the boundaries of the true text. The goal is to provide the continuous boundary of the original text on each side. This continuous boundary can be used for other tasks such as segmentation or enhancement and registration of the text in order to preserve weak and thin strokes. The method is applied to real and also synthesized [3] degraded document images. The results are promising.

The paper is organized as follows. In section 2, a novel level set formulation for bleed-through problem and its corresponding forces are discussed. Then, in section 3, the model selection and experimental results are presented. Finally, a level set method for registration of unregistered double-sided images is provided in section 4. The summary of the paper and future prospects are presented in the conclusion.

2. Level set framework

A double-sided document image, which consists of the recto and verso images (u_R and u_V respectively), is given. It is assumed that the verso side is registered on the recto side and has been flipped horizontally to correspond with the interfering patterns on the recto-side image. Because of symmetric nature of problem, only the processing of the recto side will be discussed in the following. However, it is implied that the same procedure is applied to the verso side simultaneously. Our method is inspired from the Chan-Vese approach [1]. Consider an active curve $C(t)$ which is expected to approaches the continuous boundaries of the text on the recto side as $t \to \infty$. $C(t)$ can be represented by the intersection of a surface, the level set function (lsf) $\phi_R(x, y, t)$, and the $z = 0$ plane. With our method, the evolution of ϕ_R is controlled by three forces, a thresholding force, a regularization force, and a reverse diffusion force, according to the following governing equation:

$$\partial \phi_R / \partial t = \delta_{\phi_R} \left(F_T + \nu F_{REG} + \mu F_V \right) \qquad (1)$$

where F_T is the thresholding force, F_{REG} is the regularization force and F_V is the reverse diffusion force from the verso side. The details of these forces is discussed in the following subsections. δ_{ϕ_R} is a stability term which limits the evolution of ϕ_R to regions which are close to the zero level. Many different forms can be used for δ_{ϕ_R} [1, 6]. In this work, we use the following form [1]:

$$\delta_{\phi_R} = (1/\pi) 1/(1 + \phi_R^2)$$

2.1. Threshold force

This force is a binarization force. It differentiates between the stroke pixels and the background pixels. F_T raises ϕ_R value of dark pixels and lowers ϕ_R values of white pixels. We propose to use a force that varies smoothly across the range of gray values. Therefore, there is enough room for other forces to take the control of the evolution. Given a threshold value $T(t)$, F_T is expressed as

$$F_T(t) = 2 \tanh \left(2(T(t) - u_R) \right) \qquad (2)$$

A plot of F_T is provided in Figure 1.

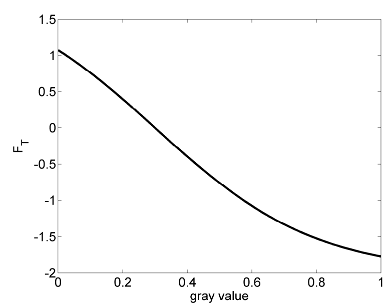

Figure 1. A typical behavior of the thresholding force with $T = 0.3$.

The thresholding parameter $T(t)$ is initialized with $T(0) = T_0$. Various thresholding technique can be used to select T_0, but we choose to use Otsu's method. During the evolution of the level set, T will be adjusted based on the *pdf*s of the current estimated text and background regions. Assume $hist_s$ and $hist_b$ are the histograms of $\phi_R > 0$ and $\phi_R < 0$ regions at a specific iteration, i.e., the regions are defined by $H(\phi_R) = 1$ and $H(\phi_R) = 0$, where H is the Heaviside function. Adjusting the threshold value T is accomplished by minimizing the error in the histograms after thresholding. This error is as follows:

$$e_{hist} = \frac{1}{\int hist_s^2 \,\mathrm{d}x} \int_0^T hist_s^2 \,\mathrm{d}x + \frac{1}{\int hist_b^2 \,\mathrm{d}x} \int_T^1 hist_b^2 \,\mathrm{d}x$$

The optimal solution for T will be

$$\begin{aligned} \tilde{T} &= \arg\min_T e_{hist}(T) \\ &= \arg\operatorname*{solve}_T \\ &\left(hist_s(T) \int hist_b^2 \,\mathrm{d}x = hist_b(T) \int hist_s^2 \,\mathrm{d}x \right) \end{aligned}$$

Because of the uniform and gradual behavior of T along time, it is not necessary to select a new optimal value of T at each iteration. In this work, the threshold value is updated after each 20 iterations.

2.2. Regularization force

A characteristic shared among many level set methods is a regularization, curvature dependent, force that leads toward smooth boundaries. We use the following form [6, 7]:

$$F_{REG}(\phi_R) = \nabla \cdot (\nabla \phi_R / |\nabla \phi_R|)$$

2.3. Reverse diffusion force

We use a reverse diffusion force, introduced earlier in the context of PDE filtering [3], to counterbalance the effect of bleed-through interfering patterns. The reverse diffusion force for each pixel is calculated based on the relation between the recto and verso information:

$$F_V = \frac{-1}{2\tanh 2} \left(\tanh 2 + \tanh \left(\frac{u_R - \overline{u_V} - 2\delta_{rev}}{\delta_{rev}} \right) \right)$$

where $\overline{u_V}$ is a median filtered image of u_V. The profile of F_V versus $\Delta u = u_R - \overline{u_V}$ is plotted in Figure 2 with $\delta_{rev} = 0.1$. As it can be seem from the figure, F_V is effective in a range of $2\delta_{rev}$. In other words, this force takes advantage of any small difference, of the order of δ_{rev}, between the grayscale information originating from the recto and that originating from the verso. Therefore, this force is robust with respect to the level of degradation and complexity of input images.

3. Validation

3.1. Model selection

The most important parameter in our method is μ, which controls the importance of the diffusion force. This parameter depends directly on the degree of degradation in

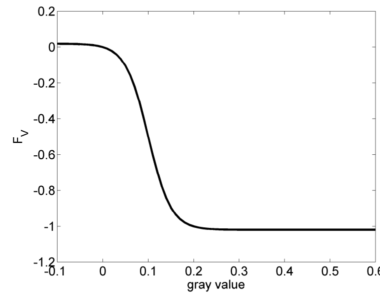

Figure 2. A typical behavior of the reverse diffusion force with $\delta_{rev} = 0.1$.

the input images. To estimate the degradation level, we first binarize both the recto and verso image and we define the set of recto verso stroke pixels $\Omega_{RV} = \{x : [u_R(x) > T_{0,R}]$ or $[u_V(x) > T_{0,V}]\}$, where, for example, $T_{0,R}$ is the same as in sec. 2.1 for u_R. Then, we estimate how close the bleed-through and stroke pixel intensities are using the following equation:

$$D = \frac{1}{\#\Omega_S} \sum_{\Omega_S} \left(\exp\left(-(u_V(x) - u_R(x))/\delta_{rev}\right)^2 \right),$$

where $\#\Omega_S$ is the cardinality of Ω_S. If the degradation is more severe, D will have a higher value. Figure 3 shows the relation between minimum required value of μ and D. As it is expected, μ has an increasing and uniform relation with respect to D.

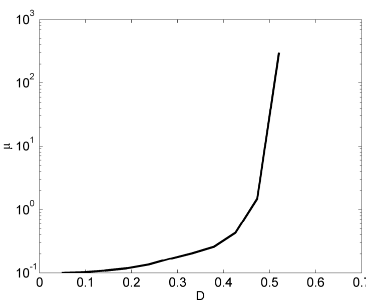

Figure 3. Relation between μ and D. See text.

Parameter ν controls the smoothness of the segmentation boundaries and as to be hand picked or learned for a specific database. We fixed $\nu = 2.0$ for all our experiment.

3.2. Numerical results

The performance of the method is evaluated on a set of real and synthesized double-sided images. In Figure 4, the

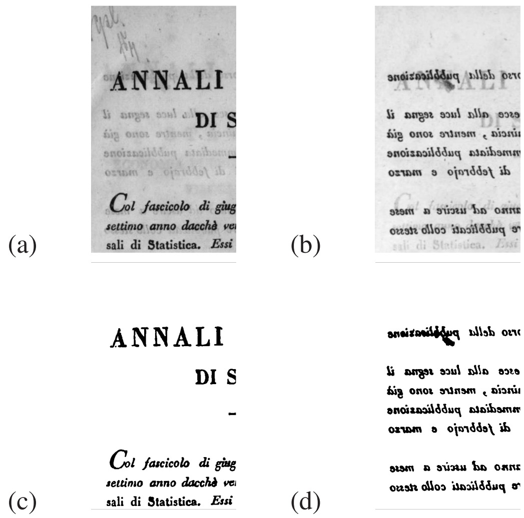

Figure 4. (a) and (b): recto and verso sample images from the Google book search dataset [4]. (c) and (d): corresponding outputs of the proposed method.

input and output images of an unconstrained real example from Google book search dataset are presented.

The method is also applied to a set of synthesized examples with different degree of degradation. In Figure 5, only two extreme cases have been shown. The input images are generated using a degradation model presented in [2,3]. For the sake of comparison, the results of a direct PDE-based restoration method [3] are also applied to the degraded input images. As one can see, in most cases, the results are almost the same. Only for very highly degraded cases, e.g. the last row in Figure 5, the PDE-based method resulted in the lost of important strokes. This is a limitation of the PDE-based method: overlapping regions, where the true strokes and interfering patterns intensities are almost the same, cannot be distinguished properly and the diffusion process results in the lost of valid strokes. From the computational point of view, both methods achieve the steady state after about 100 iterations. This good performance is the result of using an adaptive threshold in the level set method.

4. Registration of double-sided images

In this section, a model for registration of the unregistered images is presented. The registration process can be defined as finding a mapping between the original text in u_V to its corresponding bleed-through text on u_R (or Vice versa). For that purpose, we apply a joint registration and segmentation process to the bleed-through text, based on information extracted from the evolving lsf ϕ_R, ϕ_V.

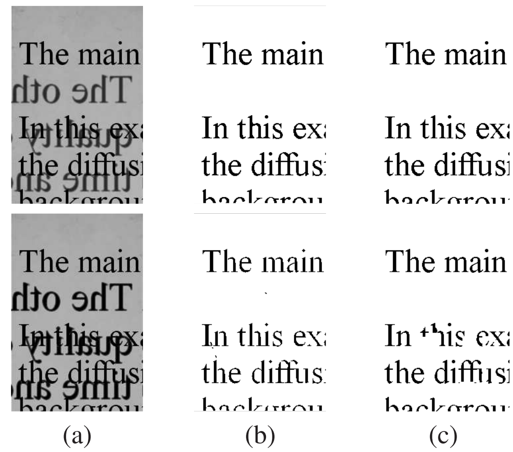

Figure 5. Two severe cases of bleed-through problem. (a): input images, (b): outputs of the proposed method, and (c): output of a diffusion based method [3]. Only the recto sides are shown.

To represent the misalignment of u_R and u_V, we incorporate another lsf, namely ϕ. If two level set functions ϕ_V and ϕ are aligned, the registration is done. Two forces control the evolution of ϕ, i.e., a data term extracted from u_R which represents the interfering patterns from verso side, and a shape prior term defined by the evolving level-set ϕ_V which represents the source text of those interfering patterns. The data term for our segmentation model is based on a modified version of the well known Chan-Vese (CV) model [1]. We propose the following modification:

$$\begin{aligned} E_{data} \;&=\; \gamma \int_\Omega (u_R - \hat{c}_1)^2 H_1(\phi, \phi_R)\, \mathrm{d}r \\ + \int_\Omega |\nabla H(\phi)|\, \mathrm{d}r \;&+\; \gamma \int_\Omega (u_R - \hat{c}_2)^2 H_2(\phi, \phi_R)\, \mathrm{d}r \end{aligned} \tag{3}$$

where $H_1(\phi, \phi_R) = [H(\phi) - H(\phi)H(\phi_R))]$ and $H_2(\phi, \phi_R) = [(1 - H(\phi))(1 - H(\phi_R))]$. $\hat{c}_1$ and $\hat{c}_2$ are the mean intensity values in the region defined by $H_1 = 1$, and $H_2 = 1$ respectively. The meaning of the above modification, is that regions defined as the recto text (according to $H(\phi_R)$) are ignored both in the calculations of $\hat{c}_1$ and $\hat{c}_2$, and during the level set evolution, thus enabling more accurate segmentation of the bleed-through text. Applying the CV model on the image u_R will result in a partial segmentation of the original text of u_R. Since our purpose is to segment the bleed-through text (originated from u_V), we exploit the fact that as the segmentation process progresses, the lsf ϕ_R more accurately describes the text in u_R.

The shape prior term is defined according to the lsf ϕ_V. Notice that ϕ_V is the lsf that segments the text on u_V, therefore can be used as a shape prior for its corresponding bleed-through text on u_R. Since the pose parameters of ϕ_V with respect to ϕ are unknown, we use the following shape dissimilarity measure [5]:

$E_{shape} = \int_\Omega [H(\phi) - H(\tilde{\phi}_V)]^2 \, \mathrm{d}r$,

where $\tilde{\phi}_V = A(\phi_V) = B((x, y, \phi_V)^T) \cdot (0, 0, 1)^T$, and B is a rigid transformation consisting of a rotation θ and a translation r_T in the xy-plane.

Minimization of the energy functional $E_{data} + \eta E_{shape}$ results in a lsf that both minimize the data term (3) while fitting the transformed shape prior ($\tilde{\phi}_V$) to the evolving level-set, and recovering the transformation parameters (θ, r_T). The corresponding coupled gradient descent equations are as follows (see [1], [5] for more details):

$$\begin{aligned}
\frac{\partial \phi}{\partial t} &= \delta_\phi [(-(u_R - \hat{c}_1)^2 + (u_R - \hat{c}_2)^2)(1 - H(\phi_R))\gamma \\
&+ F_{REG}(\phi)] + 2\eta (H(\phi) - H(\tilde{\phi}_V)) \qquad (4) \\
\frac{\partial \theta}{\partial t} &= 2\eta \int_\Omega \delta_{\tilde{\phi}_V} [H(\phi) - H(\tilde{\phi}_V)](\nabla \phi_V \cdot \frac{\partial}{\partial \theta} A) \, \mathrm{d}r \\
\frac{\partial r_T}{\partial t} &= 2\eta \int_\Omega \delta_{\tilde{\phi}_V} [H(\phi) - H(\tilde{\phi}_V)](\nabla \phi_V \cdot \nabla_{r_T} A) \, \mathrm{d}r
\end{aligned}$$

The process continues until the zero level set functions of $\tilde{\phi}_V$ and ϕ are aligned.

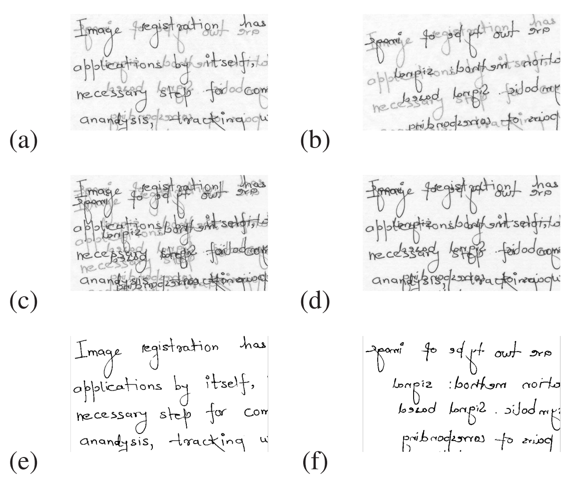

Figure 6. An example of registration of double-sided document images. (a) and (b): The close up of the recto and verso sides. (c): The superimposition before registration, (d): the superimposition after registration of the verso side using (4). (e) and (f): The final segmentation obtained using equation (1).

The registration process were evaluated on a set of artificially unregistered double-sided images. For each pair of registered images, we artificially rotated and translated the verso image by 20 different angles in the range $[-5\,°, 5\,°]$ with $0.5\,°$ difference, and translated by 30 different values in the range [-15,15]. The registration process was evaluated independently from the segmentation process. The parameters are set to $\gamma = 1$ and $\eta = 1$. Therefore, given a rough binarization of the recto and verso images (u_R and u_V), we set the level set functions ϕ_R and ϕ_V and computed the registration transformation using equations (4). The average error of the translation parameters is 0.26 pixels with standard deviation of 0.2 pixels. For the rotation angle θ, the average error is $0.24\,°$ with standard deviation of $0.2\,°$. An example of the registration process is provided in Figure 6. The unregistered verso side is obtained using a affine transformation with $\theta_0 = 5\,°$ and $r_{T0} = (10, 7)^T$ pixels. After applying (4), the following values for the parameters of the registration transformation are obtained: $\theta = -4.98\,°$ and $r_T = (-10.04, -7.02)^T$. The final segmentation results are shown in Figures 6(e) and 6(f).

5. Conclusion

Using adaptive global thresholding and a reverse diffusion force, a fast and powerful method for the restoration of double-sided document images has been developed. The adaptive threshold is selected based on the current state of the level set function and corresponding histograms. The integration is done using the level set framework. Application of the new method to both real and synthesized problems have resulted in promising segmentation outputs. In the second part, a registration method based on the level set framework and affine transformation has been introduced. Encouraging results have been obtained by application of the registration method to artificially unregistered images.

As the accuracy of the registration process depends on both ϕ_R, and ϕ_V, we plan to integrate the segmentation and registration processes in an alternate optimization scheme. Also, experiments on a large database will be conducted.

References

[1] T. Chan and L. Vese. *Image Processing, IEEE Transactions on*, 10:266–277, 2001.

[2] M. Cheriet and R. Farrahi Moghaddam. In *WOSPA'2008*, Sharjah, UAE, Invited paper, 2008.

[3] R. Farrahi Moghaddam and M. Cheriet. *IJDAR*, to be published, 2009.

[4] Google. *Book Search Dataset*. Version v edition, 2007.

[5] T. Riklin-Raviv, N. Kiryati, and N. Sochen. *International Journal of Computer Vision*, 72(3):309–328, May 2007.

[6] D. Rivest-Hénault and M. Cheriet. *Image Analysis and Recognition*, 4633/2007:234–245, 2007.

[7] L. A. Vese and T. F. Chan. *International Journal of Computer Vision*, 50(2):271–293, 2002.