Image and Vision Computing 29 (2011) 351–363

Contents lists available at ScienceDirect

Image and Vision Computing

journal homepage: www.elsevier.com/locate/imavis

Enhancement of historical printed document images by combining Total Variation regularization and Non-local Means filtering ☆

Laurence Likforman-Sulem [a,*], Jérôme Darbon [b,c], Elisa H. Barney Smith [d]

[a] Telecom ParisTech, Signal and Image Processing Department Paris, France
[b] Mathematics Department, UCLA, Los Angeles, California, USA
[c] CMLA, ENS Cachan, CNRS, PRES UniverSud, France
[d] Boise State University, Electrical and Computer Engineering Department, Boise, Idaho, USA

ARTICLE INFO

Article history:
Received 24 February 2010
Received in revised form 2 December 2010
Accepted 5 January 2011

Keywords:
Document image enhancement
Image processing
Variational approach
Non-local Means
Historical documents
Character recognition

ABSTRACT

This paper proposes a novel method for document enhancement which combines two recent powerful noise-reduction steps. The first step is based on the Total Variation framework. It flattens background grey-levels and produces an intermediate image where background noise is considerably reduced. This image is used as a mask to produce an image with a cleaner background while keeping character details. The second step is applied to the cleaner image and consists of a filter based on Non-local Means: character edges are smoothed by searching for similar patch images in pixel neighborhoods. The document images to be enhanced are real historical printed documents from several periods which include several defects in their background and on character edges. These defects result from scanning, paper aging and bleed-through. The proposed method enhances document images by combining the Total Variation and the Non-local Means techniques in order to improve OCR recognition. The method is shown to be more powerful than when these techniques are used alone and than other enhancement methods.

1. Introduction

A large number of document images are available for consulting, exchange and remote access purposes. These images have been scanned from collections of historical documents in libraries or archives thanks to digitization projects [3,15,20]. Accessing the content of document images is fully enhanced when textual transcriptions are attached to them: this allows users to index and search images through textual queries. For establishing such transcriptions, automatic tools such as OCR systems (Optical Character Recognition) are used to convert document images into text lines and words in ASCII format. However OCR systems are very sensitive: when facing noise, they perform poorly for both segmentation and recognition tasks. Historical documents include many defects due to aging and human manipulations. These defects include bleed-through ink, folding marks, ink fading, holes and spots. The grain of the paper can also produce texture in the background. The scanning process can produce uneven illumination in the image, modify character edges and create other types of defects such as streaks due to microfilms [8]. Thus, reducing or removing noise in document images is an important issue for improving OCR recognition. The main goal of this paper is to show that two recent powerful image restoration techniques can be combined to drastically improve the recognition performance.

Several approaches have been proposed for enhancing document images. Leung et al. [24] enhance contrast with the POSHE method based on sub-block histogram equalization: the readability of very low-contrast images is improved. Sattar and Tay [32] deblur noisy document images using fuzzy logic and multi-resolution approaches. Pan et al. [29] correct uneven illumination and remove wood grain and shading on images of text incised on wood tablets by filtering: this allows handwritten strokes to be more easily extracted. Removing shades in document margins produced while scanning thick documents with a region growing method has been studied by Fan et al. [17]. Shades in the background can also be detected by morphological operations and lightened for removal [28]. The water flow model of [21] can extract the different background layers of a document while binarizing it. Restoring character edges by PDE-based (Partial Differential Equations) approaches has been proposed in [14]. This approach regularizes a document image using anisotropic diffusion filtering through an iterative process. Removing bleed-through pixels has been proposed by [27,38]. These approaches require co-registration of the recto and verso images. Source separation is another framework to address document enhancement [37]. It assumes that each pixel results from the mixture of different sources (background, foreground and, in the case of palimpsests, another writing layer). With such an approach, smoothing, noise removal and thresholding are performed jointly. It contrasts with our filtering-based approach

☆ This paper has been recommended for acceptance by Peyman Milanfar.

* Corresponding author. Tel.: +33 1 45 81 73 28; fax: +33 1 45 81 71 44.
E-mail address: likforman@telecom-paristech.fr (L. Likforman-Sulem).

0262-8856/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.imavis.2011.01.001

which is fast thanks to efficient algorithms and does not assume any structure of the document.

Document image enhancement can be the first step of a binarization task. To handle the variations of foreground and background brightness in historical documents, local binarization methods may be preferred to global ones [33]. In [19], noise reduction by Wiener filtering is performed prior to the local (adaptive) binarization of the document image. In [6] a combination of a global and a local method is used to binarize handwritten characters whose edges are blend with the background. The local method classifies within a local window those pixels connected to the seed image of a noisy character obtained from the global thresholding. Markov Random Field (MRF) approaches [39,41,42] are also suitable for document enhancement. Such approaches include in a single model both the data (the spatial local context of a pixel) and a degradation model. In [39], a MRF-based blind deconvolution is used to segment touching characters. In [42], both character repair and binarization are performed through MRF modeling. Wolf in [41] uses a MRF with two hidden label fields to distinguish between recto pixels, verso pixels and those superimposed with the verso due to bleed-through. The labeling is obtained by energy minimization.

Most reported methods require an elaborate process to describe the document structure. For instance MRF approaches which build a statistical model for each layer of the document, have a parameter estimation process. Such methods have the ability to separate the different layers of a document (characters, noise, one or several backgrounds) but at the cost of increased time and computational complexity. In contrast, the method we propose is simpler and fast since it is based on filtering (TV and NLmeans). Compared with other filtering approaches (see Section 4.4), our approach is more efficient and needs only to tune a single parameter.

Our approach is based on regularization and filtering and aims at reducing the noise level in the background and on character edges of document images. The background noise comes from ink bleeding from the verso or from defects of the recto. The existence of such noise makes document segmentation and recognition difficult: additional pixels may fill the inter-line and inter-word spaces or create confusing character shapes. Character edges may not be smooth due to scanning. The proposed method combines two restoration steps based respectively on the Total Variation regularization approach (TV) and Non-local Means (NLmeans) filtering.

Two methods of combining TV and NLmeans are proposed and analysed. Fig. 1 shows the flowcharts of the proposed combinations. In Fig. 1a the input image is pre-processed by TV in order to reduce background noise. A mask is constructed from this TV-processed image through binarization and dilation operations. The resulting binary image is applied as a mask on the TV-processed image. This results in a grey level image with low background noise and smoothed edges. The NLmeans filter is then applied to enhance character details. This combination is called the *type-A* combination. In the second combination scheme, the *type-B* combination shown in Fig. 1b, the mask is applied to the image pre-processed by NLmeans rather than TV but the mask construction is identical. The decision whether to apply the first or second combination scheme is related to character dimensions and document contrast (see Section 4.3.3). Our approach contrasts with PDE-based approaches since it relies on both regularization and on the use of a mask for a better noise removal between text lines, while PDE relies on an evolution process that requires a stopping time. We compare both approaches in Section 4.5. Our approach is also different from the MRF approach of [41] since we do not attempt to detect noisy pixels but rather regularize the grey level values of both background and text.

The paper is organized as follows. In Section 2, the background noise reduction based on Total Variation is presented, as well as the mask construction. Section 3 presents the foreground noise reduction based on NLmeans. In Section 4.3.1, we describe the parameter setup for the main regularization β parameter. In Section 4, we validate the proposed method on printed documents from various periods showing its robustness to a range of degradations. Our evaluation is based on edge quality and on the recognition rate of an open source OCR, at the character level. Section 5 concludes the paper.

2. Background noise reduction by Total Variation regularization

The first step of our enhancement method is based on the Total Variation (TV) regularization approach. The image is first filtered by TV, then from the filtered image a mask is constructed (see Fig. 1). TV is a filtering method which preserves edges and sharp boundaries while reducing noise in the background. This property is desirable since background noise reduction increases the ability of an OCR system to extract textual data at different levels: zones, text lines and characters. The mask construction mainly uses the property of background noise reduction in regions distant from the characters but TV can also be used for enhancing character edges (see Section 2.2).

We introduce the following notation for our enhancement method. It is assumed that an image is defined on a 2D regular grid S. Let us denote by Ω the set of such images. The input image and its restored version are denoted by v and $\hat{u}$, respectively. The grid S is

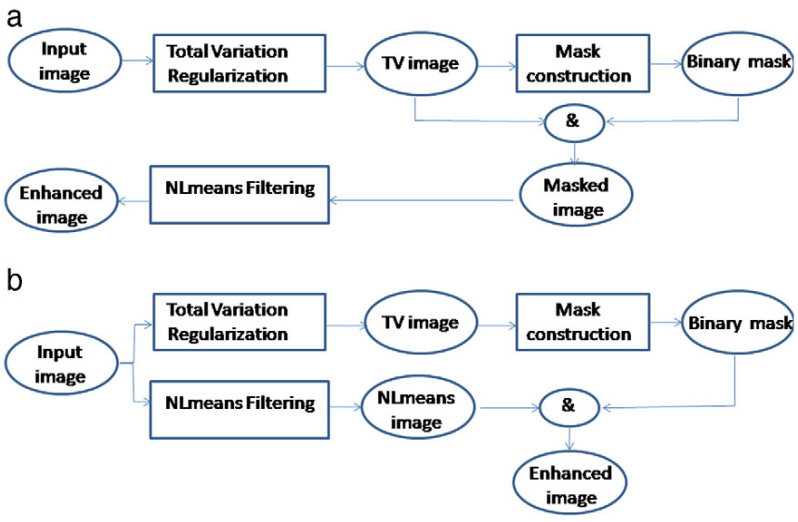

Fig. 1. Flowcharts of the two proposed enhancement methods combining TV regularization and NLmeans filtering. a) type-A combination b) type-B combination.

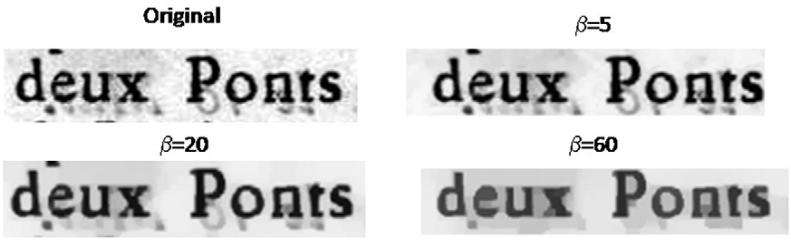

Fig. 2. Total Variation: input and regularized words for increasing values of regularization parameter β. Background noise is flattened as β increases.

endowed with a neighborhood system $\mathcal{N}$ and we denote by $\mathcal{N}(s)$ the set of sites that are neighbors of site s (relative to $\mathcal{N}$).

2.1. Total Variation

The TV approach relies on image modeling and a variational point of view. Variational approaches are popular methods for image restoration. Few works use a variational approach in the domain of document analysis and recognition. In [7], a variational approach is used at the character level for the restoration of degraded character contours. This approach needs a reference character for the contour in the degraded region of the character, to converge to the desired position. In contrast, our method aims to apply the TV approach directly to document images as a noise reduction step. Variational and Markovian formulations of the image restoration problem consist of minimizing an energy that is generally a weighted combination of two terms, namely the data fidelity and the regularization (also called prior). Since a discrete framework is considered in this paper, we consider a Markovian point of view as presented in [13]. The data fidelity D measures how far the current solution u is from the observed image v. It is defined from the nature of the noise that corrupts the image. For instance, a separable quadratic data fidelity term corresponds to the assumption that the noise is additive Gaussian noise. We assume here such noise and thus the data fidelity is defined as

$$D(u|v) = \frac{1}{2}\sum_{s\in S}(u(s)-v(s))^2 \quad \forall u\in\Omega, \tag{1}$$

where u is a candidate image defined on Ω. The value of the image u at site s is denoted by $u(s)$.

The prior should embed the knowledge we have of the nature of the problem, e.g., the statistical properties of the image. Among many regularization terms that have been proposed (see [40], for instance), Total Variation has been a popular one since the seminal works of [9,31]. The main characteristics of the TV prior are that the solution lives in the space of functions of Bounded Variation that allows for sharp edges and discontinuities. We follow [13] and define TV as the l^1-norm of a discrete gradient. More formally we have

$$TV(u) = \sum_{s}\sum_{t\in N(s)}|u(s)-u(t)|. \tag{2}$$

Recall that $N(s)$ denotes the set of pixels that are neighboring the site s. In this paper, we consider the 4-nearest neighbors.

The restored image $\hat{u}$ which corresponds to the candidate image u which minimizes the energy $E(.|v)$ that is a weighted combination of the two terms above

$$\hat{u} = \arg\min_{u} E(u|v) = \arg\min_{u}\,[D(u|v) + \beta\cdot TV(u)], \tag{3}$$

where the parameter β is non-negative. This coefficient is a parameter that governs the balance between the data fidelity and the regularization; a large value for β will produce an image with few details while a tiny one will yield an image that leaves v almost unchanged. It was shown in [26] that such an approach is prone to a loss of contrast. This prevents us from using a high regularization value since in addition to removing noise, it would both remove small features that can be text, and reduce too much the contrast so that the binarization process would fail to produce the desired result. Fig. 2 shows the effect of the regularization parameter β on sample words. The background has been regularized with values of $\beta=5$ and $\beta=20$. Larger values of β such as $\beta=60$, produce more regularization which can also remove smaller desired image features and fill intra-character spaces. If the image contains small characters from low resolution scanning (around 13–14 pixels in height), the value of β should be a smaller value, typically lower than 10.

For highly degraded documents (see Fig. 3), TV can considerably reduce the background noise on character contours. However small character components from fading characters may vanish.

2.2. Construction and application of the mask

The first step of both type-A and type-B enhancement is the construction of a mask from a TV regularized image to denote pixels that clearly belong to the background. The TV-regularized image is thresholded by the Otsu algorithm to locate the set of pixels which in majority belong to text. Second, the thresholded image is dilated by a square structuring element of side 9 pixels. This results in a binary mask where regions around characters are set to 0 and the remaining ones to 1.

The mask is superimposed on the initial image to construct the masked image. The masked image consists of saturating at value 255 the image values corresponding to mask value 1, and recovering the initial pixel values corresponding to mask values 0. The mask can be applied to any image and we choose to apply it either to the TV-regularized image which has been created during mask construction or to the NLmeans filtered image (see also Section 3). This corresponds to the type-A and type-B combination methods respectively. The choice of the image to be masked depends on the character size since spaces within small-sized characters processed by TV are more likely to be filled (see Section 4.3). There is no such effect with NLmeans. The choice of the combination type according to the characteristics of the document set is described in Section 4.3.3.

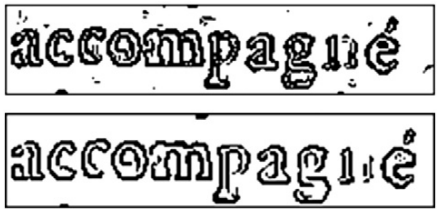

Fig. 3. Edges from a highly degraded word from set XVIII-b. From top to bottom: without regularization, after regularization by TV.

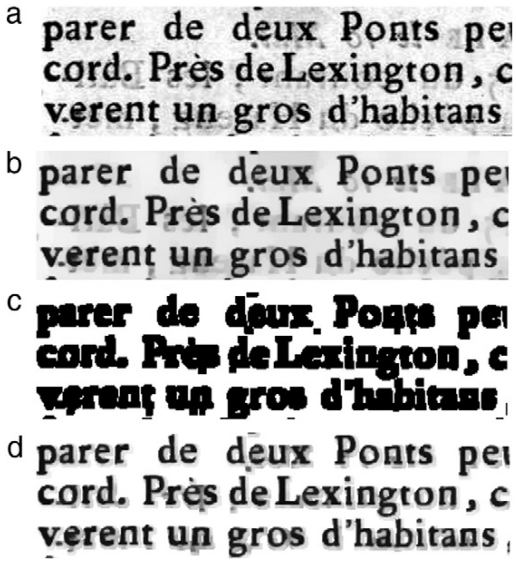

Fig. 4. a) Input image b) TV-regularized image c) mask and d) enhanced image resulting from the application of the mask on the TV image. Pixel values outside character regions are saturated to 255.

Fig. 4 shows the TV regularized image obtained from an input image. For this document, the TV image has been masked and the masked image (Fig. 4d) is different from the TV image (Fig. 4b) since background pixels distant from character pixels have been saturated.

3. Non-local Means filtering

The NLmeans filtering is used in our combined approach in two ways (see Section 1). The filtering can be applied either to the TV-regularized masked image as an additional filtering step (type-A), or to the input image which is then masked (type-B). The mask construction is described in Section 2.2. NLmeans is a non-local filter which can smooth character parts from neighboring data. NLmeans averages neighboring parts of the central pixel but the averaging weights depend on the similarities between a small patch around the pixel and the neighboring patches within a search window [10]. NLmeans capitalize on the redundancy present in most images [10]. Document images may contain even more redundancy than other types of images. As previously defined, the input image and its restored version are denoted by v and $\hat{u}$ respectively. The NLmeans filter considers the similarity of a block of neighboring pixels to the block centered on the pixel under evaluation. We denote such a block by Δ. For the sake of clarity we assume it is a square patch whose side is $(2P+1)$. The similarity measure $w(s,t)$ for the two sites s and t is defined as

$$w(s,t) = g\left(\sum_{\delta\in\Delta} (v(s+\delta) - v(t+\delta))^2\right), \tag{4}$$

where $g(x) = \dfrac{1}{1+(x/2)^2}$.

Once the similarity measures are available they are convexly combined to produce the filtered image (Fig. 5). The value of the filtered image at site s is

$$\hat{u}(s) = \frac{1}{Z(s)} \sum_{t\in\mathcal{N}(s)} w(s,t)v(t), \tag{5}$$

where $Z(s)$ is a normalization constant defined as $Z(s) = \sum_{t\in\mathcal{N}(s)} w(s,t)$ for all sites, and $\mathcal{N}(s)$ the set of sites that are neighbors of s. In the original NLmeans version [10], the similarity is assumed to be computed for every pair of pixels, and thus $N(s) = S\setminus\{s\}$ (all image sites not including s). This is extremely time consuming. Instead, the similar sub images for the pixel are searched for in a region around the considered pixel: the search window. Let us assume that this search window is a square whose side is $2K+1$. The number of patches within the search window is $(2K+1)(2K+1)-1$. In our implementation, when the patch reaches pixels outside the image, we replicate the image pixels in a mirror-based way. Otherwise, we consider the actual image pixels. The size of this search window affects the pixels that could be found and the level of regularization on the image. A larger search window possibly allows more similar patches to be found and thus would yield a filter that better preserves the features. This also tends to produce more regularized images since the restored pixel will be a weighted average of more values. This behavior gives a background that is more homogeneous. But since NLmeans never completely discards patches, this tends to smooth too much and we have shown in [25] that fading characters lose more pixels when the search window is too wide. Increasing K also produces a small loss of contrast. Let us emphasize that this loss is more modest than for the Total Variation based filter. Nevertheless, there will be less contrast between background and characters due to the averaging with more data patches. Moreover increasing K yields a higher computational load since complexity is $K^2 \cdot N$ (N is the image size).

The complexity of our implementation is independent of P. However, when increasing the size of patches, it is more unlikely to find similar patches. This is why P is generally kept small. If images

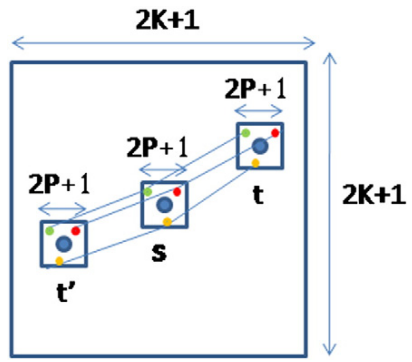

Fig. 5. NLmeans approach: two patches Δ of size $2P+1$ around t and t' within the size $2K+1$ search window centered on s. The similarity measure between patches is computed from all pairs of points at the same location within patches.

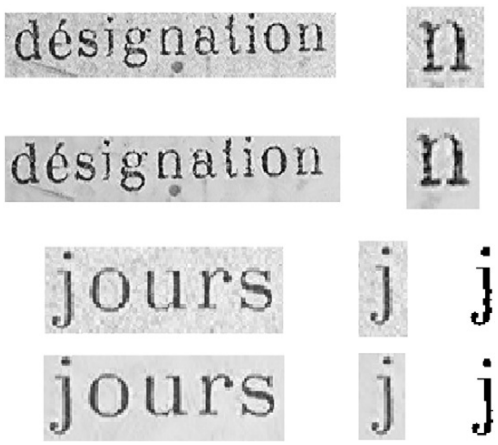

Fig. 6. From top to bottom. An input sample word and enlarged character. The same word and character enhanced by NLmeans. Input word, enlarged and binarized character. Same word and character enhanced by NLmeans. NLmeans filtering reduces background noise and preserves character details.

include characters of greater size, one could increase the size of K and P. In practice, this is not performed because of the computational load due to the higher value of K. Default parameters for the NLmeans approach, $K=4$ and $P=3$, are widely used for a number of applications. We also use the default parameters in our proposed enhancement method. Fig. 6 shows the effect of the NLmeans filtering on a sample word and characters of set XXa. We observe that the noisy background has been strongly smoothed since similar patches can be found for background pixels. The fading pixels of the character n have been preserved; they have been smoothed less since fewer similar patches have been found. We also observe a small improvement on the edges of the binarized character j since similar patches can be found along the long vertical stroke of this character. It can be noted that edges are also regularized by TV but small objects such as fading pixels are more likely to vanish with TV.

4. Experiments

Experiments are conducted to evaluate the combination of TV and NLmeans on printed documents of various periods. The effect that this combination has on sample images is illustrated in Section 4.2. Section 4.3 then shows how this combination fares when used for OCR both when TV and NLmeans are used as stand alone algorithms, and when they are combined. Sections 4.4, 4.5 and 4.6 compare our method with other preprocessing methods.

4.1. Data sets

Three sets of real degraded documents are used in these experiments. The sets are built according to the period in which the documents were created: the XVII, XVIII or XX century. Each set currently includes text images from two document collections, so that each set can be separated into set-a and set-b (see Fig. 7). Set XVII includes 1463 characters from the electronic collection of the British library [2]. Set XVII-a comes from a Hamlet theater piece, while set XVII-b is a festival book in French. Set XVIII includes 4560 characters of French Gazettes, newspapers from the 18th century [4]. Set XVIII-a (Gazette d'Avignon) is less degraded, while set XVIII-b (Gazette de Leyde) includes more degraded characters. Set XX includes 496,836 characters of twentieth century documents. Sample images from a French journal whose publishing period is around 1930 are used to form set XX-a [1] and the whole set News.3G provided by ISRI forms set XX-b [36]. Table 1 shows how the size of x-characters varies according to the sets. The smallest characters are found in set XVIIa.

4.2. Edge analysis on sample documents

An enhanced image resulting from the TV and NLmeans combination is presented in Fig. 8b for a noisy document of the XVIII century. Background noise comes from ink bleeding from the recto and the verso (see Fig. 8a). The TV regularized image has been

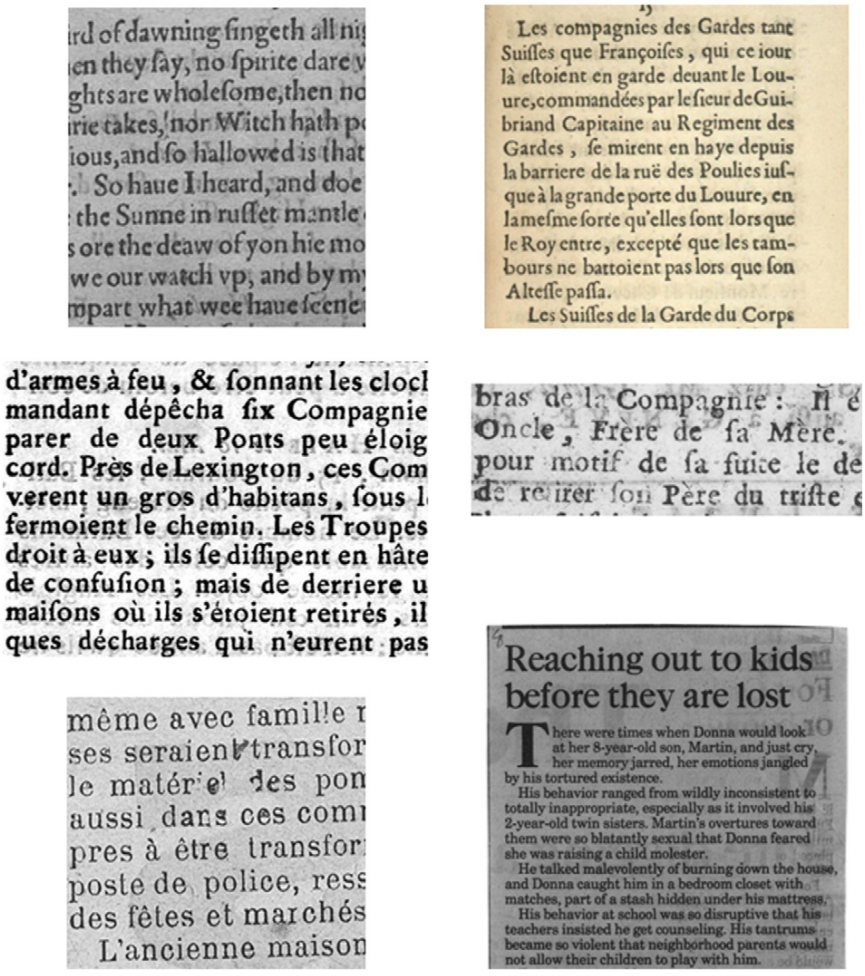

Fig. 7. Sample documents. From top to bottom, left to right: XVII-a, XVII-b, XVIII-a, XVIII-b, XX-a, and XX-b.

Table 1
Dataset image specifications. Size of x characters and qualitative description of the primary noise effects.

Test set	x-dimensions	Image degradation	#Characters
XVII-a	10×10	Low contrast	730
XVII-b	23×23	Bleed through	733
XVIII-a	16×16	Bleed through, textured background	1689
XVIII-b	37×37	Bleed through, textured background, scanning streams	2871
XX-a	21×21	Low contrast, folding marks	4756
XX-b	17×20 to 100×100	Bleed through, low contrast	492,080

obtained with $\beta = 20$. Background noise has been reduced by the type-A combination of TV and NLmeans (see Fig. 8b).

The method can be compared with two well-known restoration methods: the median filter and the Wiener filter which are popular methods for noise reduction (see also Section 4.4). For this experiment, the window size of the median filter is 3×3. Results in Fig. 8b, c and d show that our method reduces the number of extraneous edges compared to the median and Wiener filters and produces more regular edges. This can be seen for instance on the first character u of the third text line (Fig. 8f).

Since the proposed method is mask-based, we also show a result when the mask is constructed using the Wiener filter instead of TV. Results in Fig. 8e and f show that TV is more efficient than Wiener for removing background noise and regularizing edges, when combined with NLmeans.

Another result in Fig. 9b, c, d and e shows that our method reduces the effect of the streak due to microfilm. This streak is located in the area between the second and third lines. As mentioned earlier (see Section 2.1) small character components may vanish when regularized by TV as seen in Fig. 9b. Detecting such fading character areas would be useful in order to adapt locally the β parameter.

In summary, we have shown in this Section that:

- the proposed method has the ability to regularize background noise such as microfilm streaks and character edges,
- It is thus easier for an OCR to locate characters and recognize their shapes.

4.3. Evaluation through recognition

The proposed methods are evaluated through recognition performance on printed documents of various periods described in Section 4.1. To evaluate the performance of the proposed approach quantitatively, we pass the enhanced images through the OCR Tesseract [34]. This OCR was originally developed by HP and obtained good results at the UNLV accuracy test in 1995 [36]. It is now available open source through Google. The set XX-b is one of the sets tested in the UNLV evaluation as set News.3G. The OCR engine is a means for evaluating the improvements brought by the enhancement method we are proposing.

We consider two tunings for the OCR. One tuning consists of reducing the influence of dictionaries in order to test the improvement brought by the proposed enhancement approaches at the pattern recognition level, character by character. This is done by setting the Tesseract configuration variables *ok_word*,*good_word*, *non_word*, and *garbage* to one. This setting is suggested by [35] and allows us to run the OCR without dictionary-based corrections. The other tuning consists of using the Tesseract dictionary of the document language (here English or French) since dictionaries are

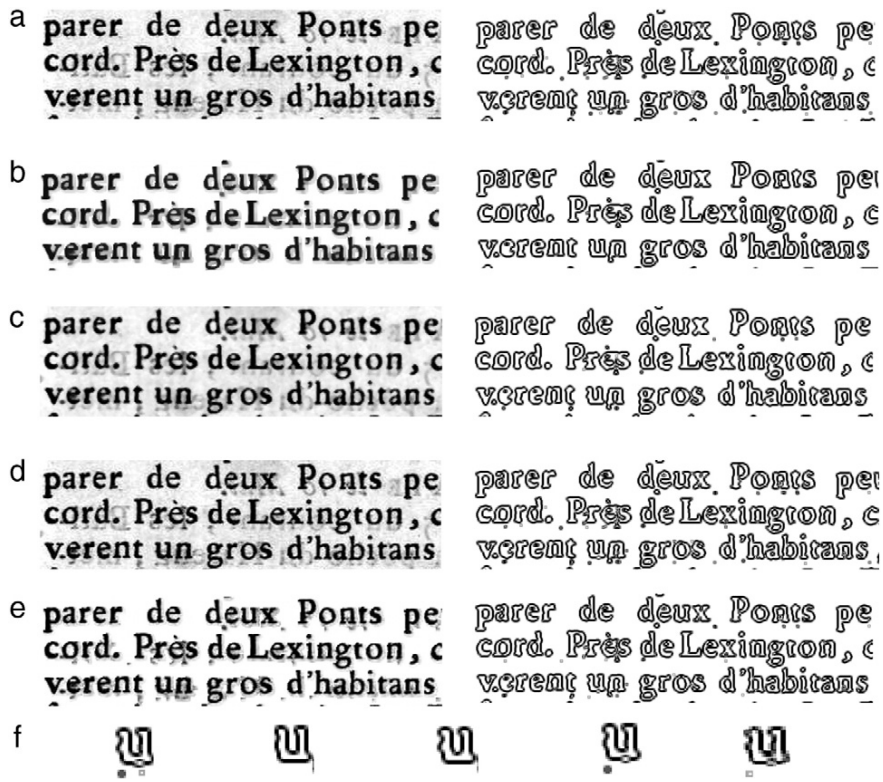

Fig. 8. Edges obtained with a) input image b) proposed method (type-A) c) Median filter d) Wiener filter e) combination method with TV replaced by Wiener f) sample character u (on third line) obtained with methods a)–e).

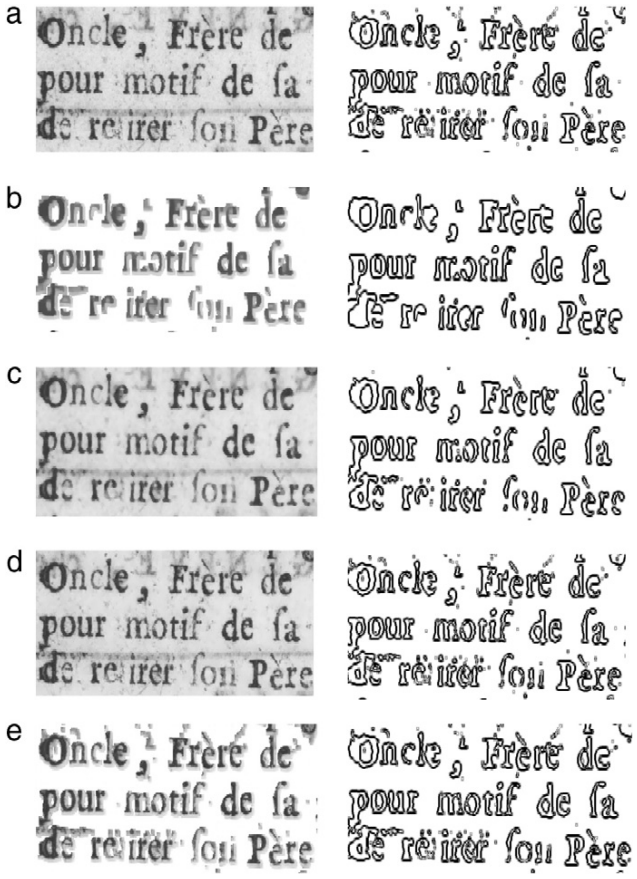

Fig. 9. Edges obtained with a) input image b) proposed method (type-A) c) Median filter d) Wiener filter and e) combination method with TV replaced by Wiener.

often used to compensate for OCR errors. The improvement brought by both preprocessing and dictionary-corrections is then observed.

We provide to Tesseract the grey level document images, unprocessed or enhanced. The OCR uses the Otsu thresholding algorithm to binarize the images. However in our combined approach, there are two kinds of background pixels: saturated and not saturated (see Section 2.2). Consequently, we have modified the Otsu algorithm within the OCR system to take into account the specificity of these images which include three modes. The modification consists of removing the saturated mode from the histogram. Similarly, the modification of the Otsu thresholding algorithm is necessary for set XX-b (News.3G) since those images have large white background zones around the clipped news articles whose paper intensities are moderately dark grey. The histograms of these images also have three modes due to the two backgrounds and the foreground. It can be noted that preliminary experiments on set XX-b [25] used a common global threshold for XX-b images, the value (75) being chosen as suggested in [36] for this data set. However for the sake of comparison, all sets are binarized in our experiments according to the same Otsu-based framework.

The main parameter of our proposed approach is the regularization parameter β. We first set the range of its values in Section 4.3.1. In Section 4.3.2, we evaluate separately the two enhancement approaches, TV and NLmeans, on the different datasets. Then we evaluate in Section 4.3.3 their combined performance.

4.3.1. Parameter setup

The most important parameter is the TV β parameter. In the following experiments we set the TV parameter β to increasing values with a step of 2, ranging from $\beta=2$ to $\beta=24$ for most sets, and up to $\beta=30$ for sets XVIII-a and b since high values of β still yield high performance. The results shown in Fig. 10 indicate that:

- the β parameter can be set to lower values when the resolution of the characters is low,
- and set to higher values when characters are of higher resolution.

Thus, for sets XVIIa and XXb which include small sized characters, β values should be lower (lower than 12 for set XVIIa and lower than 6 for set XXb). The β values for sets XVIIIa and XVIIIb including characters of higher resolution should be chosen greater than 20. However a too fine tuning of this parameter is not necessary since ranges of values yield comparable performance of the TV regularization.

4.3.2. TV and NLmeans as single enhancement methods

TV and NLmeans can be used as single enhancement approaches and applied directly to document images [25]. We applied these methods in isolation at the ICDAR DIBCO 2009 document image binarization contest [18] and our results were reported in the top 10 (of over 43 participants) according to the four measures used in the contest, and in the top 6 according to the F-measure only. We use the fast implementation of TV proposed by Chambolle [11]. We also use the fast implementation of NLmeans proposed by Darbon et al. [12].

Our first experiment consists of evaluating each preprocessing algorithm in isolation and comparing their respective performance. Results in Fig. 11 show the performance of TV and NLmeans for each set. Results are reported for the value of β yielding the best performance in the range of values determined previously in Section 4.3.1. Results show that in all but one set (XX-b), TV and NLmeans as single enhancement methods, improve recognition. The improvement is small for less degraded sets such as XVII-b and XX-a, and higher for more degraded sets such as XVII-a, XVIII-a and b. The low performance observed when using TV for set XX-b can be explained by the loss of contrast resulting from the TV regularization. The loss of contrast is lower for NLmeans: this explains the better results obtained with NLmeans for this set. However, as will be shown in Section 4.3.3, the combination of both methods will increase performance further.

In Fig. 12, results are given when dictionary corrections are turned on in the OCR. TV regularization is performed with the same β value for each set, obtained when corrections are turned off (see Fig. 11). TV and NLmeans improve recognition for all but one set as described previously. Results in Figs. 11 and 12 show that the improvement brought by NLmeans or TV is more important when dictionary-based corrections are turned off. While disabling of the dictionary is generally not desirable for practical applications, it shows that isolated recognition is enhanced with the proposed preprocessing methods. In particular for the most degraded sets (XVIII-a and b), the relative improvement brought by the preprocessing is the highest as shown in Fig. 12. Even if the number of correctly recognized characters improves a little, the dictionary-based correction yields a much higher number of corrected characters. Similar tests were also performed using the ABBYY OCR system [5]. Results were comparable to those achieved with Tesseract, indicating the results are OCR engine independent.

The Otsu algorithm is the thresholding method provided with the OCR engine. However binarized images can be directly fed to the OCR. Thresholding can also be seen as an enhancement technique since it aims at removing unwanted background pixels and at detecting all foreground ones. To observe the effect of the thresholding on performance, we replace Otsu by the Kittler and Illingworth method [22]. We choose this method since it is also a clustering-based method. Results show that this thresholding globally performs worse on our document sets than Otsu. Performance is still acceptable for sets XVIIb and XVIIIa but for other sets there is a drastical decrease in performance. For set XVIIIb, the binarization of the unprocessed

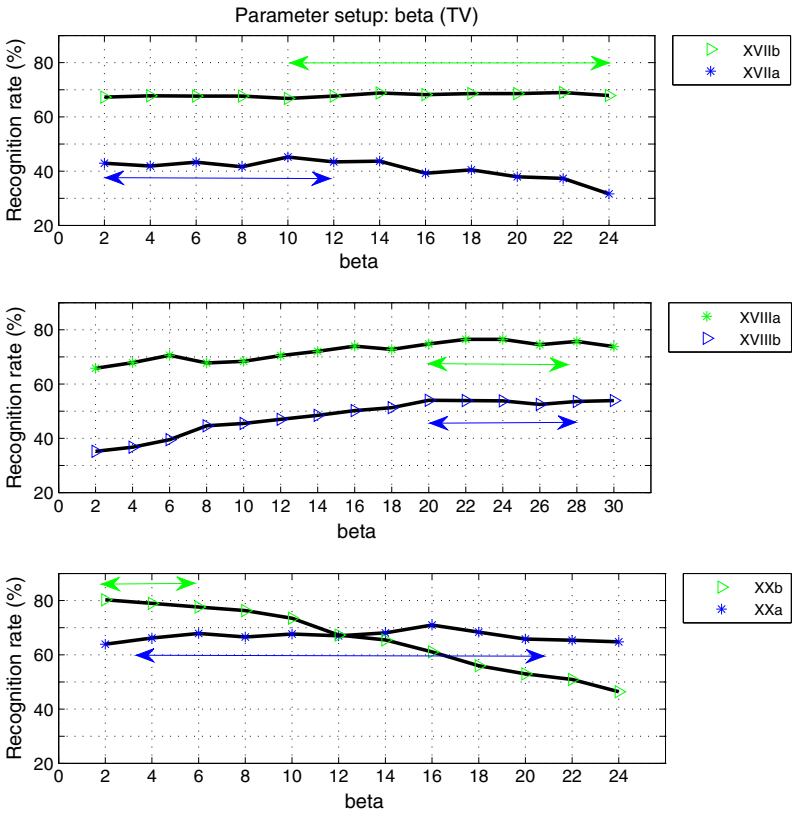

Fig. 10. Performance of TV regularization as a function of the β parameter. Arrows indicate for each set the range of operating values for β.

images could not even be exploited by the OCR. This may be due to the fact that image histograms do not follow the underlying assumption of a two Gaussian mixture. However we still observe a global improvement with TV and NLmeans over no preprocessing when changing the thresholding method (Fig. 13).

In summary in this Section, it has been shown that:

- TV and NLmeans as single enhancement methods globally improve OCR results,
- TV yields a higher loss of contrast compared to NLmeans,

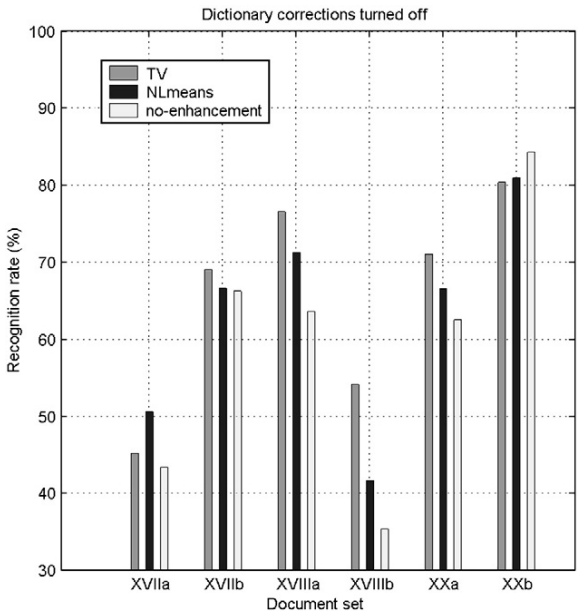

Fig. 11. Performance of TV regularization and NLmeans filtering as single enhancement methods (without combination). Dictionary corrections are turned off.

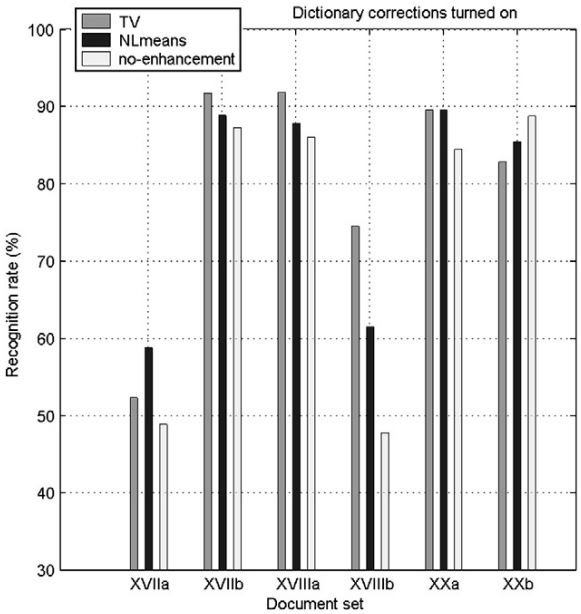

Fig. 12. Performance of TV regularization and NLmeans filtering when dictionary corrections are turned on.

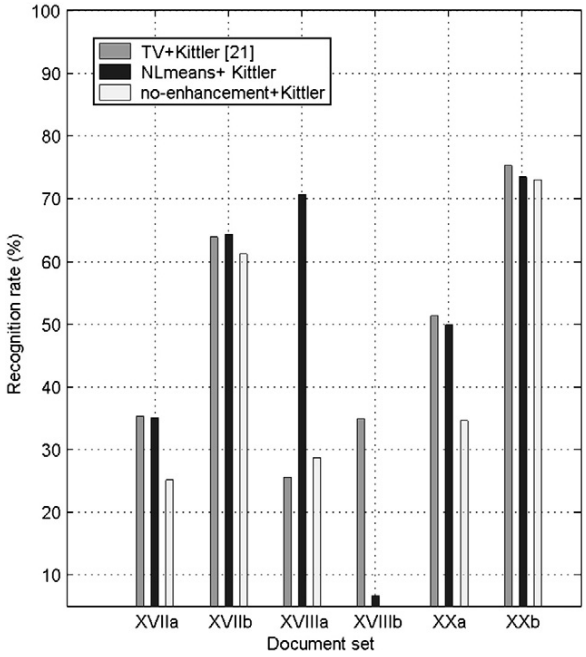

Fig. 13. Performance of TV regularization and NLmeans filtering followed by another thresholding technique [22].

- Thus, for low-contrast documents NLmeans should be preferred to TV as a single enhancement method.

4.3.3. Combination of TV and NLmeans

The two enhancement steps, TV and NLmeans, are now combined as described in Section 1. There are two types of combination:

- the first type (type-A) applies the mask to the TV-regularized image,
- the second type (type-B) applies it to the NLmeans filtered image.

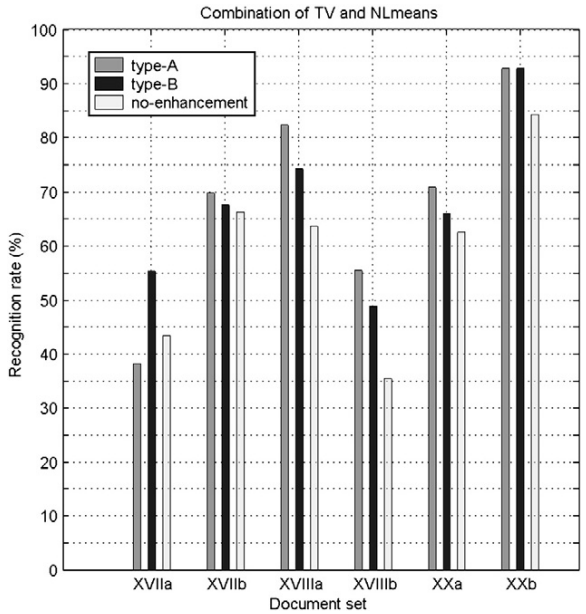

Fig. 14. Combination of TV and NLmeans. Performance of the combined methods (types A and B) through character recognition accuracy (in%).

Results are provided for the two types of combination in Fig. 14. The value of β yielding the best performance is chosen for each set and each type of combination. We observe that the type-A combination is more effective for sets XVII-b, XVIII-a, XVIII-b and XX-a. The type-B combination is more effective for set XVII-a which includes the smallest sized characters. On set XX-b, since TV and NLmeans perform similarly as single enhancement techniques, both types of combination perform similarly.

The improvement brought by the type-A combination is very high for sets XVIII-a and b: the increase reaches 20% in absolute value for set XVIII-b. This is due to the efficiency of the background noise reduction step on these highly degraded sets. Additional results are provided in Fig. 20.

The choice between the type-A and type-B combination depends on how TV performs as a single enhancement method. If the TV regularization brings improvement, the type-A combination should be chosen. When TV regularizes too much, this degrades performance and the type-B combination should be chosen. It can be noted that for the type-B combination, the final enhanced image is not regularized by TV but TV is however very useful for removing the background noise. All combination types construct the same mask obtained by thresholding the TV regularized image. Noise pixels which have their grey levels flattened by TV are thus not included in the mask. When the type of the combination is correctly chosen, the combination globally brings improvement over each enhancement method in isolation. Actually, the combination always yields an improvement except for the set XX-a for which TV alone performs slightly better. More precisely, the type-A combination has a 70.8% recognition rate (see Fig. 14) while the best TV ($\beta=16$) result reaches a 71% rate (see Fig. 11).

It can be noted that the optimal β value for TV regularization as a single enhancement method may be different than the one used for the type-B combination. During mask construction, a high β value will remove small objects, which is desirable in the background, and this property is fully exploited in the type-B combination. Thus, we found that the $\beta=20$ value is the most effective for sets XVIIa and XXb and for the type-B combination. Practically, setting the value of β for the combination may be performed on a training set when a collection of documents with similar characteristics (character size, contrast) is

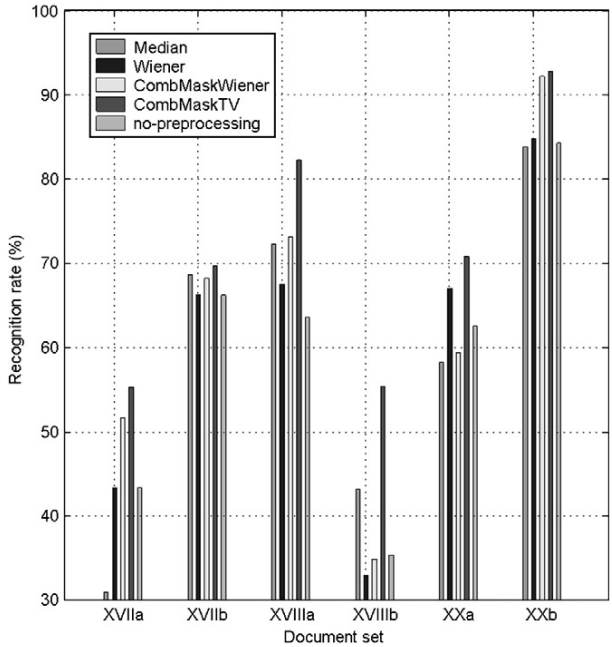

Fig. 15. Comparison of enhancement methods through character recognition accuracy (in%). Dictionary corrections are turned off.

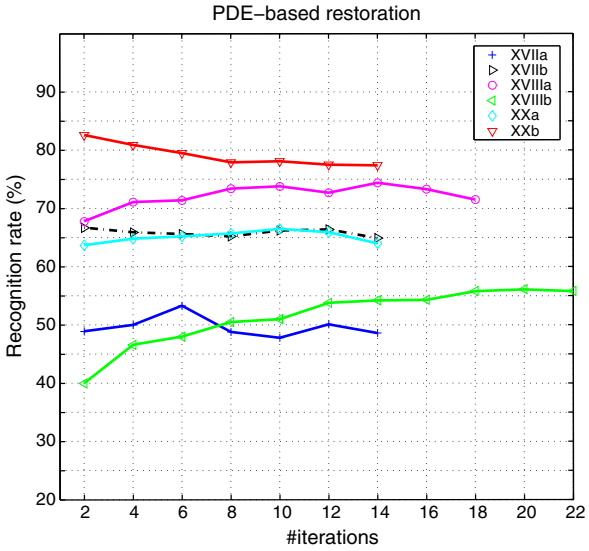

Fig. 16. Recogition accuracy as a function of the number of iterations when preprocessing images with the Perona–Malik PDE-based approach.

available. The peak of the graph representing the OCR accuracy as a function of β will indicate the best β value. Alternatively, the β value can be set to a default value. We found that $\beta = 20$ was an effective default value (see Section 4.6) and we provide additional results for this setting in Fig. 18.

In summary, in this Section it has been shown that:

- there are two ways for combining TV and NLmeans, namely the type-A and type-B combinations,
- the type-B combination is preferred for low-contrast documents,
- the β parameter used for constructing the mask can be set to a default value $\beta = 20$.

4.4. Comparison with standard enhancement methods

We have compared in Section 4.2 the edges obtained by the proposed approach with those obtained by other enhancement methods such as the Median and the Wiener filters. In the following experiment, we compare all enhancement approaches through the character recognition rate. Fig. 15 compares the accuracy at the character level without dictionary-based corrections for all data sets. The combined methods are of type-A for sets XVII-b, XVIII-a, XVIII-b and XX-a, and of type-B for sets XVII-a and XX-b. The combined approach always performs better than the other enhancement methods. Median filtering is a suitable approach for document restoration with low background noise (such as set XVIIb), but when background noise is high, it cannot remove all remaining edges as shown in Fig. 8 and those edges interfere with the recognition.

We also conduct a combination experiment by replacing the TV-filtered image by the Wiener-filtered image. As done previously we use the same combination type for each set: the mask is either applied to the Wiener-filtered image or to the NLmeans-filtered image. In both cases the mask is constructed using the Wiener filter. Results in Fig. 15 show that the Wiener filter globally performs better in combination than in isolation. The mask construction and image regularization with Wiener filtering is also effective when combined with the NLmeans filtering. However TV performs better.

4.5. Comparison with a PDE-based enhancement method

PDE-based approaches, originally developed for describing physical laws, rely on a diffusion process. As mentioned in Section 1, they can also be used for restoring images, including document images [14]. A popular PDE approach is the Perona–Malik diffusion approach [30]. This approach relies on an iteration process which smoothes

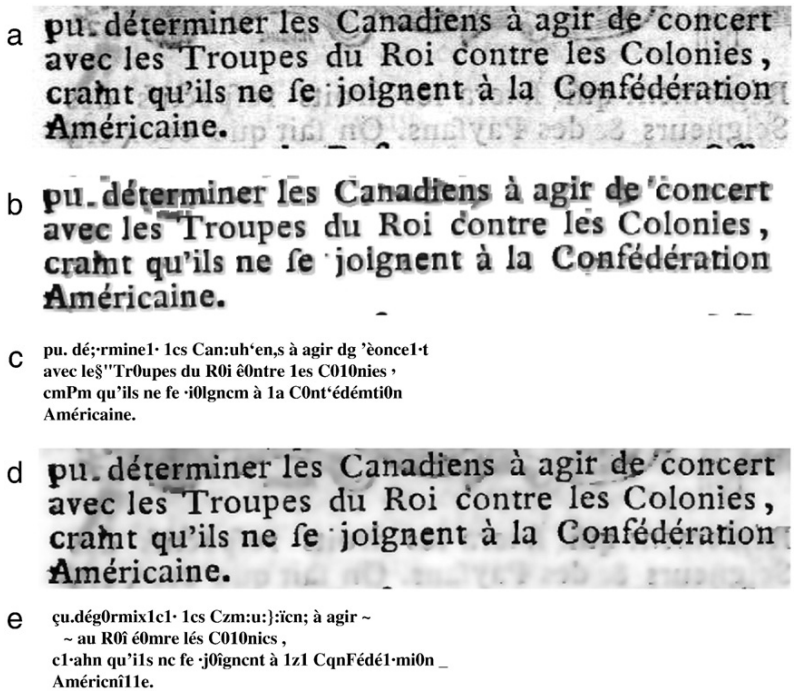

Fig. 17. a) Input image from set XVIIIa. b) type-A combination c) OCR result for the type-A combined image (dictionary corrections turned off) d) PDE processed image with six iterations and e) OCR result for the PDE-enhanced image.

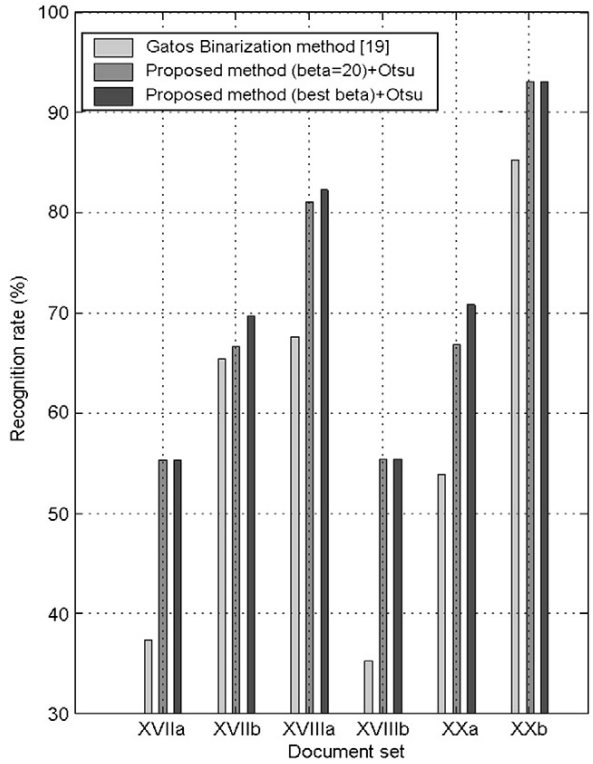

Fig. 18. Comparison of proposed enhancement method followed by Otsu binarization with the binarization method presented in [19]. Results for the proposed enhancement method are given for the default $\beta = 20$ and the best β values.

homogeneous regions while preserving edge discontinuities. Three main parameters govern the diffusion process. In Kovesi's implementation [23], they are: the number of iterations, the integration constant δt and the gradient modulus threshold κ which appears in the conduction coefficient function.

We conduct experiments on our datasets varying the number of iterations. Choosing a correct number of iterations can be critical since characters may vanish if the diffusion is not stopped early enough. We avoid this by slowly increasing the number of iterations until a maximum is reached for the recognition rate. It was found that 14 iterations are generally necessary, except for sets XVIIIa and b which need more iterations. We set the remaining parameters to default values: $\delta t = 1/7$ and $\kappa = 30$. The diffusion coefficient is equal to $\exp(-(\Delta I/\kappa)^2)$. Results in Fig. 16 show that there are sets (XVIIb and XXa) for which performance is stable within a range of iterations. For the other sets, performance shows more variations.

As a pre-processing method, PDE-diffusion improves performance except for set XXb. The images in set XXb include characters of different sizes and the contrast is low. The diffusion process results in a number of images where characters of the text body are merged, even for a small number of iterations.

Comparing Figs. 14 and 16 the PDE-approach with our combined approach, PDE performs slightly better for sets XVIIa and XVIIIb, and performs worse for the other sets (XVIIb, XVIIIa, XXa, and XXb). For sets XVIIa and XVIIIb, the absolute difference in accuracy is 0.5% and 0.7% respectively, when choosing the optimum number of PDE iterations for each set. We have observed that the PDE diffusion yields thinner characters while our approach results in thicker characters. Thick characters have lost less detail but when characters are small, intra-character spaces may be filled. However our approach better removes the background noise which results in an easier line and word segmentation for the OCR. This is shown in Fig. 17 where the last words of the first text line and the first words at the beginning of the second line were correctly found by the OCR with our approach, but not with the PDE approach.

4.6. Comparison with a mask-based binarization method

We have compared in the last sections the proposed enhancement method with several filters. All methods were followed by the same

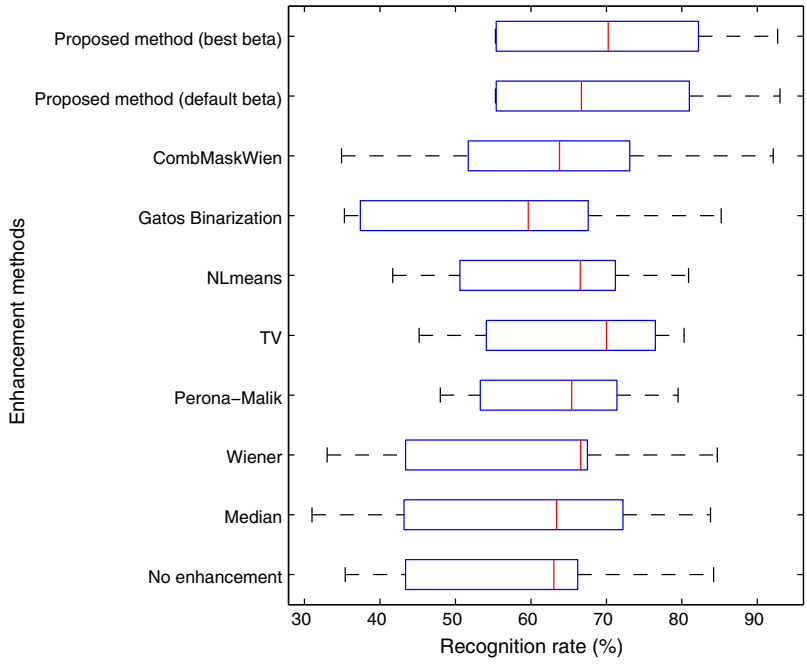

Fig. 19. Performance of enhancement methods through box-and-whisker plots.

Otsu-based binarization performed by the OCR. We now compare the binarization obtained with the proposed enhancement method followed by Otsu thresholding with the binarization method presented in [19]. This method has been chosen since it is also mask-based and dedicated to degraded documents. We implemented the following steps: enhancement (Wiener filtering), estimation of the foreground and background images and final thresholding. The estimation of the foreground image is performed through Sauvola adaptive binarization and it is used as a mask applied to the input image. Pixels belonging to the foreground region are interpolated in order to construct the background image.

We propose two ways of comparison. We first consider the proposed combination method as described in Section 4.3.3 using the best β parameter for each set. Then we set the value of parameter β to 20, which is the best β value when a single value has to be selected for all document sets. Setting β to a single value also corresponds to the case where the user does not know or does not want to set the β parameter and sets it to the default value. Moreover, since we compare the Gatos binarization method using its default parameter values, a fairer comparison with the proposed method is obtained when using a default β value. Results are shown in Fig. 18 for all sets. The default parameters of [19] are used in these experiments except for set XVIIa where the weighting parameter q is set to 0.2 in order to avoid resulting blank images. Using the implementation of the Gatos binarization method provided thanks to the Gamera project [16] gave us similar results. Compared to this latter method, we observe better results for our method, both using the default $\beta = 20$ value and when β is tuned for each set.

For the sake of comparison, performance for all methods presented here are provided in Fig. 19 through box-and-whisker plots. Each plot shows the median value of the recognition rate for each method as well as the spread of variation (25th and 75th percentiles) of this rate through document sets. There are no outliers for our rates, hence the whiskers extend to the most extreme rates. The improvement of the proposed combination method can be observed through its high median value. TV and NLmeans as single

Fig. 20. Enhancement results. Left: input image, right: enhanced image using a combination of TV and NLmeans.

enhancement technique also have high median values. The smallest spread of variation is observed for the Perona–Malik enhancement method. However the spread of enhancement methods based on TV and NLmeans (combined or not) is globally smaller than for the other methods.

5. Conclusion

We have proposed a new enhancement method based on the combination of two powerful preprocessing methods, namely the Total Variation regularization and the NLmeans filtering which have the ability to reduce background noise and improve character detail, respectively. We have proposed two types of combinations, depending on the size of characters in a document set. Both types use TV regularization for eliminating noise in the background through the construction of a mask. We have compared our approach to other noise reduction methods, namely the Median and the Wiener filters, and the Perona–Malik diffusion filter. It has been shown that our approach performs better than these filters for most degraded documents on an OCR recognition task. An advantage of the proposed method is that it includes only one main parameter. Only the TV β parameter has to be set within a range of values according to character size and noise level. We have shown that when ones does not know or want to tune this parameter, a convenient default value is $\beta = 20$.

However, automatically tuning the β parameter is a major perspective beyond the scope of this study. Another perspective is to locally adapt parameter β instead of globally. This would be useful for instance for processing fading character areas which need a lower β. Finally, evaluating our approach on other types of document noise such as speckle and blur as well as studying the impact of several local and global binarization techniques are also included in our future work.

Acknowledgements

The authors wish to thank Marc Sigelle from Telecom ParisTech for fruitful discussions. Research of Jérôme Darbon has been supported by the Office of Naval Research through grant N000140710810.

References

[1] Archives departementales des Yvelines, http://www.yvelines.fr/archives/home.html.
[2] British Library: Treasures in Full, http://www.bl.uk/treasures/treasuresinfull.html.
[3] Googlebooks project, http://books.google.com/googlebooks/library.html.
[4] Les Gazettes europeennes du 18eme siecle, http://gazettes18e.ish-lyon.cnrs.fr/.
[5] ABBYY, ABBYY Fine Reader, http://www.abbyy.com/.
[6] I. Bar Yosef, Input sensitive thresholding for ancient hebrew manuscript, Pattern Recognition Letters 26 (8) (2005) 1168–1173.
[7] I. Bar-Yosef, A. Mokeichev, K. Kedem, U. Erlich, I. Dinstein, Global and local shape prior for variational segmentation of degraded characters, ICFHR, Montreal, 2008.
[8] E. Barney Smith, Characterization of image degradation caused by scanning, Pattern Recognition Letters 19 (1998) 1191–1197.
[9] C. Bouman, K. Sauer, A generalized gaussian image model for edge-preserving map estimation, IEEE Transactions on Signal Processing 2 (3) (july 1993) 296–310.
[10] A. Buades, B. Coll, and J.M Morel. A review of image denoising algorithms, with a new one, SIAM-Multiscale Modeling and Simulation 4 (2005) 490–530.
[11] A. Chambolle, An algorithm for total variation minimization and applications, J. Math. Imaging Vis. 20 (1–2) (2004) 89–97.
[12] J. Darbon, A. Cunha, T.F. Chan, S. Osher, G.J. Jensen, Fast nonlocal filtering applied to electron cryomicroscopy, Proceedings of the IEEE International Symposium on Biomedical Imaging: From Nano to Macro, (ISBI'08), 2008, pp. 1331–1334.
[13] J. Darbon, M. Sigelle, Image restoration with discrete constrained total variation, Part I: fast and exact optimization, J. Math. Imaging Vis. (JMIV) 26 (2006) 261–276.
[14] F. Drira, F. Lebourgeois, H. Emptoz, OCR accuracy improvement through a PDE-based approach, Proc. of ICDAR'07, 2007, pp. 1068–1072, Brasil.
[15] M. Droettboom, Correcting broken characters in the recognition of historical printed documents, Proc. of Joint Conference on Digital Libraries, JCDL'03, 2003.
[16] M. Droettboom, C. Dalitz, The Gamera Project, http://gamera.informatik.hsnr.de/download/index.html.
[17] K.-C. Fan, T.-R. Lay, Y.-K. Wang, Marginal noise removal of document images, Pattern Recognition 35 (2002).
[18] B. Gatos, K. Ntirogiannis, I. Pratikakis, ICDAR 2009 document image binarization contest (DIBCO 2009), Proceedings of the IEEE 10th International Conference on Document Analysis and Recognition (ICDAR'09) Barcelona, 2009.
[19] B. Gatos, I. Pratikakis, S.J. Perantonis, Adaptive degraded document image binarization, Pattern Recognition 39 (2006) 317–327.
[20] IMPACT, Impact: Improving access to text, description of work, http://www.impact-project.eu.
[21] K. Kim, D.W. Jung, R.H. Park, Document image binarization based on topographic analysis using a water flow model, Pattern Recognition 35 (2002) 265–277.
[22] J. Kittler, J. Illingworth, Minimum error thresholding, Pattern Recognition 19 (1) (1986) 41–47.
[23] P.D. Kovesi, Matlab and octave functions for computer vision and image processing, http://www.csse.uwa.edu.au/pk/research/matlabfns/.
[24] C.-C. Leung, K.-S. Chan, H.-M. Chan, W.-K. Tsui, A new approach for image enhancement applied to low-contrast–low-illumination IC and document images, Pattern Recognition Letters 26 (6) (2005) 769–778.
[25] L. Likforman-Sulem, J. Darbon, E. Barney Smith, Pre-processing of degraded printed documents by non-local means and total variation, Proceedings of the IEEE 10th International Conference on Document Analysis and Recognition, 2009, pp. 758–762,, Barcelona.
[26] Y. Meyer, Oscillating Patterns in Image Processing and Nonlinear Evolution Equations, volume 22 of University Lecture Series, American Mathematical Society, 2001.
[27] R. Farrahi Moghaddam, D. Rivest-Henault, I. Bar-Yosef, and M. Cheriet. A unified framework based on the level set approach for segmentation of unconstrained double-sided document images suffering from bleed-through. Proceedings of the IEEE 10th International Conference on Document Analysis and Recognition (ICDAR'09), Barcelona, 2009.
[28] S. Nomura, K. Yamanaka, T. Shiose, H. Kawakami, O. Katai, Morphological preprocessing method to thresholding degraded word images, Pattern Recognition Letters 30 (8) (2009) 729–744.
[29] X. Pan, M. Brady, A.K. Bowman, C. Crowther, R.S.O. Tomlin, Enhancement and feature extraction for images of incised and ink texts, Image Vision Comput. 22 (6) (2004) 443–451.
[30] P. Perona, J. Malik, Scale-space and edge detection using anisotropic diffusion, IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI) 12 (7) (1990) 629–639.
[31] L. Rudin, S. Osher, E. Fatemi, Nonlinear total variation based noise removal algorithms, Physica D 60 (1992) 259–268.
[32] F. Sattar, D.B.H. Tay, Enhancement of document images using multiresolution and fuzzy logic techniques, Signal Processing Letters 6 (1999) 249–252.
[33] M. Sezgin, B. Sankur, Survey over image thresholding techniques and quantitative performance evaluation, Journal of Electronic Imaging. 13 (1) (2004) 146–168.
[34] R. Smith, An overview of the Tesseract OCR engine, Proceedings of the IEEE 9th International Conference on Document Analysis and Recognition (ICDAR'07), 2007, pp. 629–633, Brasil.
[35] M. Sturgill, S. Simske, An optical character recognition approach to quantifying thresholding algorithms, Document Engineering 08, 2008, pp. 263–266.
[36] K. Taghva, T. Nartker, J. Borsack, A. Condit, UNLV-ISRI document collection for research in OCR and information retrieval, Document recognition and retrieval VII, 2000, pp. 157–164, San Jose CA.
[37] A. Tonazzini, L. Bedini, E. Salerno, Independent component analysis for document restoration, International Journal on Document Analysis and Recognition (IJDAR) 7 (1) (2004) 17–27.
[38] A. Tonazzini, G. Bianco, E. Salerno, Registration and enhancement of double-sided degraded manuscripts acquired in multispectral modality, Proceedings of the IEEE 10th International Conference on Document Analysis and Recognition, 2009, pp. 546–550, Barcelona.
[39] A. Tonazzini, S. Vezzosi, L. Bedini, Analysis and recognition of highly degraded printed characters, International Journal on Document Analysis and Recognition (IJDAR) 6 (2004) 236–247.
[40] G. Winkler, Image Analysis, Random Fields and Dynamic Monte Carlo Methods. A Mathematical Introduction, Springer, 2006.
[41] C. Wolf, Improving recto document side restoration with an estimation of the verso side from a single scanned page, Proceedings of the International Conference on Pattern Recognition (ICPR), Tampa, 2008, pp. 1–4.
[42] C. Wolf, D. Doermann, Binarization of low quality text using a Markov random field, Proceedings of the International Conference on Pattern Recognition (ICPR), Quebec, 2002, pp. 160–163.