Edge noise removal in bilevel graphical document images using sparse representation

Thai V. Hoang, Elisa H. Barney Smith, Salvatore Tabbone

EDGE NOISE REMOVAL IN BILEVEL GRAPHICAL DOCUMENT IMAGES USING SPARSE REPRESENTATION

Thai V. Hoang[1,2]*, Elisa H. Barney Smith*[3]*, Salvatore Tabbone*[1]

[1]LORIA, UMR 7503, Université Nancy 2, 54506 Vandœuvre-lès-Nancy, France
[2]MICA, UMI 2954, Hanoi University of Science and Technology, Hanoi, Vietnam
[3]Electrical and Computer Engineering, Boise State University, Boise, Idaho, 83725, USA

ABSTRACT

A new parametric method for edge noise removal in graphical document images is presented using geometrical regularities of the graphics contours that exists in the images. Denoising is understood as a recovery problem and is done by employing a sparse representation framework with a basis pursuit denoising algorithm for denoising and curvelet frames for encoding directional information of the graphics contours. The optimal precision parameter used in this framework is shown to have linear relationship with the level of the noise. Experimental results show the superiority of the proposed method over existing ones in terms of image recovery and contour raggedness.

***Index Terms*—** Edge noise removal, noise spread, sparse representation, basis pursuit denoising, curvelet transform

1. INTRODUCTION

The scanning and binarization processes that produce binary document images introduce noise concentrated on the edges of the image objects [1]. This edge noise has influence on later steps in the chain of automatic document processing. It could affect feature measurement in recognition, reduce image redundancy in compression, and distort skeletons in vectorization. For accurate analysis and recognition of document images, edge noise needs to be removed.

Many methods exist for noise removal in bilevel document images. The most famous and frequently used are median filtering [2], contour smoothing using chain codes [3] or morphological operators [2], and kFill filtering [4]. The median filter replaces the value of a pixel by the median of the values in the neighborhood of that pixel. It is known to perform well on binary images to remove salt-and-pepper noise, however, it unintentionally removes thread lines and corners. The kFill algorithm is designed to remove salt-and-pepper noise iteratively while maintaining readability of text by using a filter that retains text corners. The value of the parameter k of the filter can be chosen adaptively based on text size and image resolution. Its disadvantage is that the picky points of the contour like the end-points of thin lines are also removed. Contour smoothing based on chain codes replaces a sequence of a chain code by another simpler sequence, usually a shorter one corresponding to the shortest path. Even though this method can produce a smooth contour from a jagged one, it cannot be performed on a noisy contour such as that shown in Fig. 1(b). Morphological operators like erosion, dilation and their combinations opening and closing have their root in set theory. They can be used to remove salt-and-pepper noise but cannot guarantee smooth contours at the output. Thus, most of the existing denoising methods for binary document images are general-purpose, they are not designed specifically for edge noise and do not exploit the directional information in their operation. Contours denoised by these methods are jagged and sometimes shifted from their original positions. Another shortcoming of existing methods is that they do not take into account the information about the level of noise that exists in the binarized document images, denoising is performed in a "blind", non-adaptive way.

Recent research in image degradation modeling [5] shows that edge noise, which results from a combination of the optical system, the additive noise, and the thresholding, can be measured quantitatively and represented by a parameter called Noise Spread (NS), paving the way for the design of a new method to remove edge noise. This paper describes a new parametric method for edge noise removal in bilevel graphical document images exploiting the directional information of graphics contours. Information about the level of edge noise, represented by NS, is used as the input to optimize the method. Directional denoising is facilitated under a sparse representation framework by promoting sparse representation of the document image in an overcomplete dictionary using a basis pursuit denoising algorithm with curvelets as the dictionary. The denoised image reconstructed from this sparse representation is a grayscale one and it can be simply thresholded to obtain the final denoised bilevel image. Our contributions are twofold: directional denoising for edge noise in bilevel graphical document images using curvelet transform in a sparse representation framework and the use of the optimal linear relationship between the framework's parameter and NS, which leads to a superior performance over existing methods.

The remainder of this paper is organized as follows. Section 2 gives some background on edge noise modeling using NS. The edge noise removal algorithm using a sparse representation is presented in Section 3. Experimental results are given in Section 4, and conclusions are drawn in Section 5.

2. NOISE SPREAD

In document and graphical images the image content contains large regions of white (0) background, with foreground image features displayed in black (1). When these are scanned in gray level, the edges change from a step function to ones sloped in the shape of the Edge Spread Function (ESF). Noise is added, which is generally assumed to be Additive White Gaussian Noise ($AWGN$). Document and graphical images are usually processed in bilevel, so this image is then thresholded, usually with a global threshold. While the $AWGN$ is evenly distributed over the whole gray level image, after the thresholding it is concentrated along the edges. This will turn a smooth edge into a rough one.

In [1] the noise was analyzed and the breadth of the region around the edge where the noise affects the image was defined as the Noise Spread (NS). The NS is not just dependent on the standard deviation of the $AWGN$, σ_{noise}, but also on the width w of the optics Point Spread Function (PSF), which determines the slope of the gray level edges, and the level of the binarization threshold, Θ, through the relation

$$NS = \frac{\sqrt{2\pi} \cdot \sigma_{noise} \cdot w}{LSF(ESF^{-1}(\Theta))}, \tag{1}$$

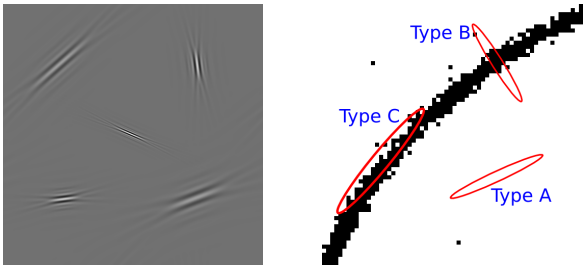
where LSF is the Line Spread Function, or one dimensional PSF. NS provides a measure that can numerically quantify the amount of edge noise, and this measure relates also to the noise visually observed on bilevel images. NS has been shown to be linearly related to the Hamming distance between the original and noisy images, as well as to be correlated with the ability to fit a line to a noisy edge. This gives the potential that the NS could be used as an input to a denoising algorithm that works on bilevel images, in a similar fashion to how the standard deviation of the additive noise is often used as an input to denoising algorithms that work on gray level images.

3. EDGE NOISE REMOVAL USING SPARSE REPRESENTATION

Results in computational harmonic analysis [6] show that curvelet frames which contain multi-scale, multi-directional, and elongated curvelets (Fig. 1(a)) can provide an optimal representation of objects with edges. Most of the energy of the objects is localized in just a few coefficients of curvelets which overlap and are nearly tangent to the object contours. In other words, the directional information of object contours is encoded in a small number of curvelet coefficients. Combining this code with the basis pursuit denoising algorithm [7] will constitute a sparse representation framework allowing directional denoising along object contours.

3.1. Curvelet transform

The radial and angular windows, $W(r)$ and $V(t)$, are two smooth, non-negative, and real-valued functions which are supported on $[1/2, 2]$ and $[-1, 1]$ respectively and satisfy the admissibility conditions $\forall\, r > 0$ and $t \in \mathbb{R}$:



(a) Spatial representation of some curvelets

(b) Alignment of curvelets with a noisy graphics contour

Fig. 1. Directional denoising using curvelets.

$$\sum_{j=-\infty}^{\infty} W^2\left(2^j r\right) = 1, \qquad \sum_{l=-\infty}^{\infty} V^2\left(t - l\right) = 1. \tag{2}$$

At each scale j, the mother curvelet φ_j is defined as:

$$\hat{\varphi}_j(r,\theta) = 2^{-3j/4} W\left(2^{-j} r\right) V\left(\frac{2^{\lfloor j/2 \rfloor}\theta}{2\pi}\right). \tag{3}$$

A curvelet at scale j, orientation θ_l, and position $\mathbf{x}_{\mathbf{k}}^{j,l} = R_{\theta_l}^{-1}\left(2^{-j}k_1, 2^{-j/2}k_2\right)$ is defined as:

$$\varphi_{j,l,\mathbf{k}}(\mathbf{x}) = \varphi_j\left(R_{\theta_l}\left(\mathbf{x} - \mathbf{x}_{\mathbf{k}}^{j,l}\right)\right), \tag{4}$$

where R_{θ_l} is the rotation operator by $\theta_l = 2\pi 2^{\lfloor j/2 \rfloor} l$ radians with $l \in \mathbb{Z}^+$ such that $0 \le l < 2\pi$. The corresponding curvelet coefficients of $f \in L^2\left(\mathbb{R}^2\right)$ is defined as:

$$c_{j,l,\mathbf{k}} = \langle f, \varphi_{j,l,\mathbf{k}} \rangle = \int_{\mathbb{R}^2} f(\mathbf{x}) \overline{\varphi_{j,l,\mathbf{k}}(\mathbf{x})} d\mathbf{x}. \tag{5}$$

Curvelets constructed in this way define a tight frame in $L^2\left(\mathbb{R}^2\right)$, obey the parabolic scaling relation ($width = length^2$), and exhibit an oscillating behavior in the direction perpendicular to their orientation. They can be used as an overcomplete dictionary in a sparse promoting problem of graphical documents. The curvelet transform has a property that the coefficients of those curvelets whose essential support does not overlap with or overlap with but are not tangent to the edge are small and negligible. For example, in Fig. 1(b), coefficients of curvelets of type A and B are negligible, most of the energy of the graphics is localized in just a few coefficients of type C which overlap and are nearly tangent to the edge.

3.2. Basis pursuit denoising

The noisy image I is the result of scanning, followed by global thresholding, a noise-free input image I_0 of size $w \times h$, it will contain edge noise of a certain NS. Let $\mathbf{y}_0, \mathbf{y} \in \mathbb{R}^n (n = wh)$ be the vectors generated by stacking the columns of I_0, I respectively, $\mathbf{A} \in \mathbb{R}^{n \times K}$ with $n \ll K$ be an overcomplete dictionary allowing sparse representation of $\mathbf{y}_0$. Let $\mathbf{x}$ be the representation of $\mathbf{y}$ in $\mathbf{A}$ satisfying $\mathbf{y} = \mathbf{A}\mathbf{x} + \mathbf{z}$, where $\mathbf{z} \in \mathbb{R}^n$ stands for the unknown additive edge noise, then finding the sparse representation $\hat{\mathbf{x}}$ of $\mathbf{y}$ in $\mathbf{A}$ is equivalent to solving the following ℓ_0-optimization problem with precision parameter ϵ depending on $\mathbf{z}$:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_0 \ \text{ subject to } \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon. \qquad (6)$$

Finding the solution of an underdetermined system of linear equations like (6) is NP-hard [8]. However, recent results [9] show that if the solution of (6) is sufficiently sparse, it is equal to the solution of the following ℓ_1-optimization problem:

$$\hat{\mathbf{x}} = \underset{\mathbf{x}}{\operatorname{argmin}} \|\mathbf{x}\|_1 \ \text{ subject to } \ \|\mathbf{A}\mathbf{x} - \mathbf{y}\|_2 \leq \epsilon. \qquad (7)$$

This is a convex optimization problem and it can be solved using methods like the interior-point algorithm [7]. For the edge noise removal problem, the curvelet frame is selected as $\mathbf{A}$. The denoised grayscale image can be obtained from the sparse reconstruction $\hat{\mathbf{y}} = \mathbf{A}\hat{\mathbf{x}}$, which is then converted to a binary image by simple thresholding $\tilde{\mathbf{y}} = T(\hat{\mathbf{y}})$ where T is the thresholding operator with the threshold value of 0.5.

3.3. The precision parameter ϵ

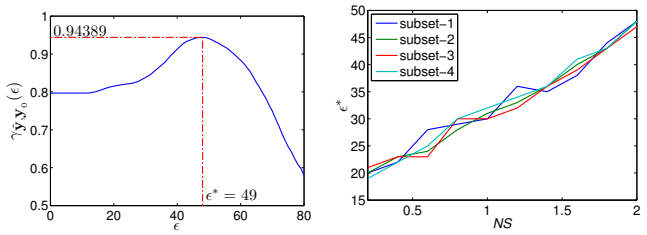
In (7), ϵ is a non-zero precision parameter that describes the desired fidelity of the reconstructed $\hat{\mathbf{y}}$ to the noisy $\mathbf{y}$. It is the only parameter that controls the performance of the denoising algorithm. When $\epsilon = 0$, $\mathbf{y} = \hat{\mathbf{y}} = \tilde{\mathbf{y}}$ and as the value of ϵ increases, the measure of sparsity of the solution of the problem in (7) must monotonically increase, since the feasible set is wider, meaning a better alignment of curvelets with graphics contour, which consequently increases denoising performance. However, due to the thresholding operation, deformation in the thresholded $\tilde{\mathbf{y}}$ will appear from a certain value of ϵ. The selected value of ϵ should thus depend on the level of noise that exists in the images. To our knowledge, there exists no work discussing the dependance of ϵ on the noise level in image denoising. For zero-mean white and homogeneous Gaussian noise with a known standard-deviation σ, the value of ϵ is usually chosen as $cn\sigma^2$, with $0.5 \leq c \leq 1.5$ [10, Chapter 14]. For edge noise, the relation $\epsilon(NS)$ will be derived to have the form kNS from experiment in the next section.

4. EXPERIMENTAL RESULTS

Two types of experiments have been carried out, one for the derivation of the relation $\epsilon(NS)$ and the other for the demonstration of the superiority of the proposed method over existing ones in removing edge noise.

4.1. The relation $\epsilon(NS)$

A dataset (SetA) containing noisy images has been generated from a ground-truth image "symbol017" which is a 256×256 pixel graphical symbol image taken from the GREC2005 database [11], its noisy version is given in Fig. 3(b). The testing database is divided into four subsets, each contains 10 noisy images corresponding to 10 values of NS ranging from 0.2 to 2.0 with increments of 0.2. In order to have best performance, at a specific value of NS, ϵ has the value ϵ^* that corresponds to the peak in a measure of fidelity between $\hat{\mathbf{y}}$ and $\mathbf{y}_0$, denoted by $\gamma_{\tilde{\mathbf{y}},\mathbf{y}_0}(\epsilon)$. The measure employed in this work is the normalized cross-correlation. Fig. 2(a) demonstrates



(a) A plot of $\gamma_{\tilde{\mathbf{y}},\mathbf{y}_0}(\epsilon)$ at $NS = 2.0$ (b) $\epsilon^*(NS)$ relation

Fig. 2. Determination of the precision parameter ϵ^* using noise level information represented by NS.

the determination of ϵ^* employing $\gamma_{\tilde{\mathbf{y}},\mathbf{y}_0}(\epsilon)$. The noisy image in Fig. 3(b) with $NS = 2.0$ is taken as the input image. At each possible value of ϵ, the ℓ_1-optimization problem in (7) is solved for $\hat{\mathbf{x}}$ and the value of $\gamma_{\tilde{\mathbf{y}},\mathbf{y}_0}(\epsilon)$ can be calculated thereafter. The plot of $\gamma_{\tilde{\mathbf{y}},\mathbf{y}_0}(\epsilon)$ in this case has its maximum value of 0.94389 at $\epsilon^* = 49$. This means that if the input noisy image I has $NS = 2.0$, ϵ in (7) should have the value 49.

The relation $\epsilon^*(NS)$ obtained from these four subsets is plotted in Fig. 2(b). It is realized that ϵ^* has a nearly linear relationship with NS for all four subsets, or $\epsilon^* \simeq kNS$. The deviation of ϵ^* from kNS is small and, moreover, due to the blunt maxima in $\gamma_{\tilde{\mathbf{y}},\mathbf{y}_0}(\epsilon)$ this deviation has almost no effect on the performance of the proposed method.

4.2. Comparison with existing methods

The proposed method has been evaluated on two datasets, the first one is SetA as described in the previous subsection and the second one (SetB) is generated from four ground-truth graphical symbol images "symbol016", "symbol017", "symbol024", "symbol081" from the GREC2005 database (their noisy version is given in Fig. 3(a)-3(d)). These images are selected for experiment due to the existence of all possible graphics contour directions. For each ground-truth image, 5 noisy images have been generated corresponding to 5 values of $NS = 0.2, 0.6, 1.0, 1.5, 2.0$. Fig. 3 provides examples of denoising graphical symbols. The original noisy images with $NS = 2.0$ are given in the first row. Evidence of directional denoising along noisy contours exists in the denoised images in grayscale in the second row. Edge noise is smoothed out in the direction that is perpendicular to the noisy contour. The final denoised images in binary are given in the third row.

To demonstrate the effectiveness of the proposed method, it has been compared with four frequently used methods: median filtering using a 3×3 neighborhood, kFill filtering with the parameter $k = 3$, closing then opening using a 3×3 structuring element, and opening then closing using a 3×3 structuring element. The criteria used for comparison are the ability to recover the original image (measured by the normalized cross-correlation between the denoised and the ground-truth images), and the raggedness of the graphics contours (a moving average based on [12]). The proposed method and four other denoising methods are each applied to datasets SetA and SetB. Normalized cross-correlation and raggedness are measured for each

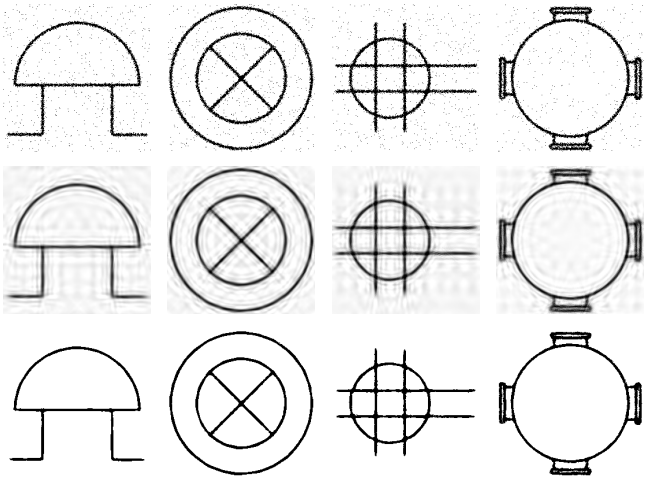


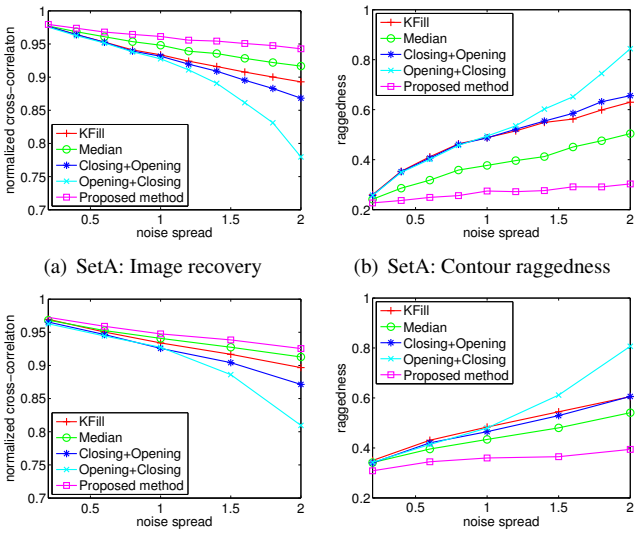
Fig. 3. Noisy and denoised images from the dataset SetB at $NS = 2.0$. First row: original noisy symbol images, $\mathbf{y}$ (from left to right: 016, 017, 024, 081). Second row: denoised images, $\hat{\mathbf{y}}$. Third row: binarized denoised images, $\tilde{\mathbf{y}}$.

resulting image. The performance of each method per noise level on one dataset is the average performance over all the noisy images in that dataset. The results are shown in Fig. 4 for these two criteria over a range of noise levels. It is observed that as NS increases, the ability to recover the original image decreases and the contour raggedness of the denoised images increases for all methods. However, the decrease in the image recovery and increase in the contour raggedness of the proposed method are nearly constant and are the smallest among all methods for both datasets. The proposed method thus results in a denoised image of best recovery with smoothest contours at all values of NS. It should also be noted from the comparison results that for noisier images, in terms of image recovery and contour raggedness the proposed method produces the most significant improvements.

5. CONCLUSION

A new parametric method for edge noise removal from graphical document images has been presented in this paper using sparse representation. Curvelet transform and basis pursuit denoising are employed for directional denoising along graphics contours. Noise Spread is used as the input parameter to select the precision parameter ϵ^* of the method, which turns out to have a nearly linear relationship with NS. Experimental results show the superiority of the proposed method over existing ones in terms of image restoration and contour raggedness.

Comparison using the recovery criterion is only feasible when the ground-truth images are available. Thus in this work, as a proof of concept and for the ease of evaluation, the experimental noisy images are generated from ground-truth images. To make the method applicable, in real applications, the value of NS can be estimated directly from the input images [1]. Performance evaluation of the proposed method using the estimated NS is left as future work.



(a) SetA: Image recovery (b) SetA: Contour raggedness

(c) SetB: Image recovery (d) SetB: Contour raggedness

Fig. 4. Comparison results with four other denoising methods in terms of image recovery and contour raggedness.

6. REFERENCES

[1] C. McGillivary, C. Hale, and E. H. Barney Smith, "Edge noise in document images," in *Proceedings of the 3rd Workshop on Analytics for Noisy Unstructured Text Data*, 2009, pp. 17–24.

[2] J. D. Gibson and A. C. Bovik, Eds., *Handbook of Image and Video Processing*, Academic Press, 2nd edition, 2005.

[3] D. Yu and H. Yan, "An efficient algorithm for smoothing, linearization and detection of structural feature points of binary image contours," *Pattern Recog.*, vol.30, no.1, pp. 57–69,1997.

[4] L. O'Gorman, "Image and document processing techniques for the RightPages electronic library system," in *Proc. of the 11th Int. Conf. on Pattern Recognition*, 1992, vol. 2, pp. 260–263.

[5] E. H. Barney Smith, "Modeling image degradations for improving OCR," in *Proceedings of the 16th European Signal Processing Conference*, 2008.

[6] E. J. Candès and D. L. Donoho, "New tight frames of curvelets and optimal representations of objects with piecewise C^2 singularities," *Comm. Pure Appl. Math*, vol. 57, no. 2, pp. 219–266, 2002.

[7] S. S. Chen, D. L. Donoho, and M. A. Saunders, "Atomic decomposition by basis pursuit," *SIAM Journal on Scientific Computing*, vol. 20, no. 1, pp. 33–61, 1998.

[8] G. Davis, S. Mallat, and M. Avellaneda, "Adaptive greedy approximations,"*Constr. Approx.*, vol.13, no.1, pp. 57–98,1997.

[9] D. L. Donoho, "For most large underdetermined systems of linear equations the minimal ℓ_1-norm solution is also the sparsest solution," *Comm. Pure Appl. Math*, vol. 59, pp. 797–829, 2004.

[10] M. Elad, *Sparse and Redundant Representations: From Theory to Applications in Signal and Image Processing*, Springer, 2010.

[11] P. Dosch and E. Valveny, "Report on the second symbol recognition contest," in *GREC*, Wenyin Liu and Josep Lladós, Eds. 2005, vol. 3926 of *LNCS*, pp. 381–397, Springer. *Available at*: http://symbcontestgrec05.loria.fr/.

[12] *ISO/IEC 13660:2001 Information technology - Office equipment - Measurement of image quality attributes for hardcopy output - Binary monochrome text and graphic images*, ISO, Geneva, Switzerland, 2001.