2013 12th International Conference on Document Analysis and Recognition

Character Shape Restoration of Binarized Historical Documents by Smoothing via Geodesic Morphology

K. Shirai*, Y. Endo*, A. Kitadai†, S. Inoue‡, N. Kurushima‡,
H. Baba§, A. Watanabe§, M. Nakagawa¶
*Shinshu Univ., †J.F. Oberlin Univ., ‡Univ. Tokyo, §NRICP Nara, ¶Tokyo Univ. Agri. Tech.

***Abstract*—We propose a method which performs anisotropic morphological dilation via implicit smoothing for the purpose of restoring the degraded character shapes of binarized images. Exploiting the idea of geodesic morphology that the binary image and its distance transformed image are interconvertible, we apply a smoothing method not to the binary image but to the distance transformed image, and then reconvert it by binarization. This allows us to apply conventional smoothing methods for continuous intensity, i.e., gray scale, images to the discrete intensity, i.e., binary, image implicitly. For instance, by using anisotropic diffusion together with geodesic dilation, anisotropic dilation along the stroke direction is obtained and brings better results.**

I. Introduction

Japanese historical documents [1], [2] were often written with brush and ink called *Sumi* on wooden tablets *Mokkan* and Japanese paper *Washi*, as shown in Fig. 1. Due to the age deterioration of these media, character strokes become illegible. The degree of deterioration is sometimes too severe to extract the character shape by binarization (BZ). As a result the binarized black and white (*bw*) patterns are fractionated at wide intervals and also noisy background patterns are mixed in. Therefore, the kind of deterioration differs from what conventional OCR methods have been dealing with [3], [4]. Worse yet, sometimes only *bw*-images which were scanned a few decades ago are available because of further deterioration of the original materials. In a such case, existing restoration methods [5] for gray scale images cannot be applied directly. These are the problems for OCR of Japanese historical documents [6], and therefore, we deal with *bw*-character patterns and examine the shape restoration problem.

Basically, to improve the condition of *bw*-images, the original color images are smoothed before BZ, by using edge-aware denoising or smoothing filters such as the Wiener filter [7], bilateral filter [8] and anisotropic diffusion [9], since the original images have more color and gradient information. After that, a morphological operation is applied to the improved *bw*-images to remove the remaining noise. However, when the stroke structure is corrupted over a wide region, conventional filters for continuous intensity, i.e., gray scale images, do not work since they are designed to deal with the correlation within a compact region. Although texture smoothing [10] has the ability to smooth and inpaint the gap between structures, the structure intervals considered here are too wide to deal with.

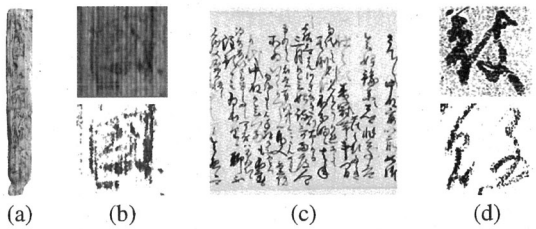


Fig. 1. Deteriorated character shapes written on a wooden tablet *Mokkan* and Japanese paper *Washi* provided by the databases of Tokyo Univ. [1] and Nara NRICP [2], (a) original *Mokkan*, (b) a character of *Mokkan* and its BZ, (c) original *Washi*, (d) *bw*-characters of *Washi*.

As for discrete intensity, i.e., binary images, morphological dilation and erosion are traditionally used to combine the fractionated *bw*-patterns. More recently, geodesic morphological filtering has been proposed in [11], [12], which extend traditional methods by utilizing the distance transform (DT). Because the dilation and erosion are isotropic, however, they tend to incorrectly combine stroke components and neighboring noisy background patterns. Therefore, [11], [12] are not directly suited to directional strokes. Binary Markov random fields with an Ising model [13] can also deal with the problem discussed here. Unfortunately, a long execution time is required to obtain stable results, and this method cannot restore strokes fractionated at wide intervals.

In our method, we focus on the geodesic morphology [11], [12] and the interconvertibility of the *bw*-image and the distance transformed (*dt*) image. Because the latter has continuous intensity, we expect that well developed edge-aware smoothing filters can be applied to the *dt*-image. Our method is quite simple and consists of existing methods as described below (see also the graphical flowchart shown in Fig. 2).

1) **Distance transform (DT)**: a *bw*-image is converted into a *dt*-image.
2) **Anisotropic smoothing (AS)**: an AS filter is applied to the *dt*-image.
3) **Binarization (BZ)**: the smoothed *dt*-image is reconverted into a *bw*-image by BZ, where the effects of dilation and erosion are obtained by changing the threshold.

Note that conventional geodesic morphological filtering [11], [12] consists of 1) DT and 3) BZ, and they mainly discuss the DT method and the combination of *dt*-images. By embedding a smoothing filter into the framework of geodesic morphology, the characteristic of the AS filter appears in the resulting *bw*-images. Besides, it is known that character conversion via DT is robust to scale changes, and it is used for the real-time rendering of text strings in game graphics [14]. Therefore, in the DT domain, we can afford to change the image resolution according to the character fragment intervals. Furthermore,

1520-5363/13 $26.00 © 2013 IEEE
DOI 10.1109/ICDAR.2013.260

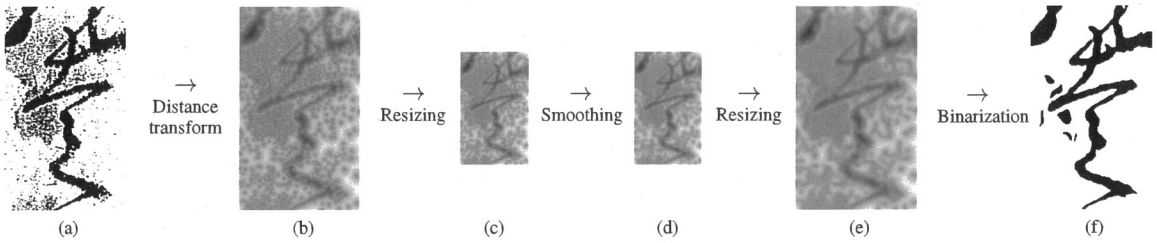


Fig. 2. Flowchart of our character shape smoothing method via the DT domain. First, a noisy input *bw*-image (a) is converted into a *dt*-image (b) (the luminance range of the *dt*-images is adjusted nonlinearly for display). Second, the *dt*-image is smoothed at an adequate resolution (c) and (d). Finally, the smoothed *dt*-image (e) is reconverted into a *bw*-image by BZ (f).

multi-resolution smoothing can be applied to the *dt*-image if needed.

The rest of this paper is organized as follows. Sec. II describes the details of the algorithm used in our method. Sec. III shows the experimental results and comparisons.

II. Smoothing via the Distance Transform Domain

We first describe the distance transform (DT) and binarization (BZ) as a part of conventional geodesic morphological filtering, and then describe anisotropic smoothing (AS).

A. Distance Transform and Binarization

The purpose of the DT is to estimate the distance from the nearest boundary of a character. Additionally, by using $\pm$ signs, we can distinguish the background $(+)$ and character fragments $(-)$. Fig. 3 (a) shows an example of the DT for a 1D signal. By thresholding the signed *dt*-image, a *bw*-image is obtained again. Additionally the change in threshold results in dilation and erosion effects.

Quasi Euclidean distance transform. Instead of the normal Euclidean DT, we employ the quasi Euclidean DT with a fast raster scan algorithm which is used in [12], [15], because the results obtained after AS are almost the same for all DT methods. The raster scan is performed twice, first in lexical order, then in inverse lexical order. At each pixel of the *dt*-image, the distance represented as a pixel value is propagated from adjacent pixels that have already been scanned .

Let I be the input *bw*-image, and $I(p) \in \{0,1\}$ be the black (0) and white (1) discrete intensity at pixel $p \in \mathbb{Z}^2$. The neighboring 3×3 pixels are defined by $p+\Delta$, $\Delta \in \left\{\binom{-1}{-1}, \binom{-1}{0}, \ldots, \binom{+1}{+1}\right\}$. Black pixels adjacent to white pixels are defined as boundary pixels $\{p_b\}$. Then, let D be the *dt*-image, and initialize the pixel values, i.e., distances, by using the boundary pixels as follows:

$$D(p) = \begin{cases} 0 & \text{if } p = p_b \\ \infty & \text{otherwise} \end{cases}, \tag{1}$$

where ∞ is a large value at least larger than the total number of pixels.

The raster scanning in lexical order (left-to-right and top-to-bottom) is performed to propagate and update the minimum distance at each pixel as

$$D(p) = \min\{D(p+\Delta) + \|\Delta\|\}, \tag{2}$$

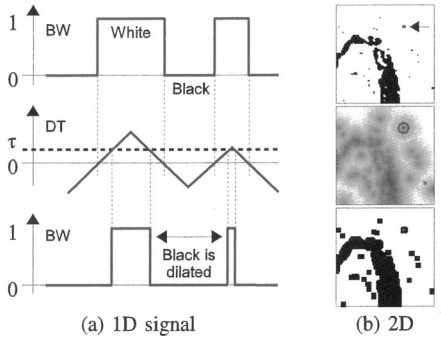


Fig. 3. An example of morphological dilation via DT and BZ. The original *bw*-signal (top) is transformed into a *dt*-signal (middle). By binarizing the *dt*-signal with a threshold $\tau > 0$, the result becomes a dilated *bw*-signal again ($\tau < 0$ results in erosion). In the case of using the 2D quasi Euclidean DT (b), the point indicated by the arrow (top), spreads into an octangle (middle).

where $\Delta \in \left\{\binom{-1}{-1}, \binom{0}{-1}, \binom{+1}{-1}, \binom{-1}{0}, \binom{0}{0}\right\}$ indicates the position relative to pixels that have already been scanned (the center pixel is also included as $(0,0)$), and the L_2 norm $\|\Delta\| \in \{0, 1, \sqrt{2}\}$ is the distance. On the other hand, scanning in inverse lexical order (right-to-left and bottom-to-top) uses $\Delta \in \left\{\binom{+1}{+1}, \binom{0}{+1}, \binom{-1}{+1}, \binom{+1}{0}, \binom{0}{0}\right\}$.

Finally, $\pm$ signs are assigned to each pixel $D(p)$ to distinguish character fragments $(-)$ from the background $(+)$:

$$D(p) = \begin{cases} -D(p) & \text{if } I(p) = 0 \\ +D(p) & \text{otherwise} \end{cases}. \tag{3}$$

Binarization by constant thresholding. We employ a simple non-adaptive thresholding filter with a single threshold value for all pixels, because it works well despite its simplicity. Let J be the reconverted *bw*-image. The pixel value $J(p)$ is given by using a threshold τ as follows:

$$J(p) = \begin{cases} 0 & \text{if } D(p) \le \tau \\ 1 & \text{otherwise} \end{cases}, \text{where } \begin{cases} \text{dilation} & \text{if } \tau > 0 \\ \text{none} & \text{if } \tau = 0 \\ \text{erosion} & \text{if } \tau < 0 \end{cases}. \tag{4}$$

Examples of *bw*-images yielded from different thresholds are shown in Fig. 4 of Sec. III. Incidentally, using pixel-wise adaptive thresholding, pixel-wise adaptive dilation and erosion are obtained. We consider automatic adaptive thresholding in future work.

B. Anisotropic Smoothing

As for the smoothing method, anisotropic smoothing methods tend to yield better results than isotropic smoothing methods. In particular, local gradient based methods such as anisotropic diffusion are suited to AS filtering. Examples of *bw*-images yielded from different smoothing filters are shown in Fig. 5 in Sec. III.

Anisotropic diffusion PDE in [9]. The gradients of *dt*-images are smooth in the range (intensity) direction, while variations are mainly observed in the spatial direction due to the curvature of the edges. Therefore, we expect that anisotropic diffusion, which has higher direction resolution than other filters, is suited to smoothing *dt*-images, and employ a multivalued regularization PDE (partial differential equation) [9] as our diffusion method.

First we compute the edge direction at each pixel p. Let $\boldsymbol{\theta}_+$ and $\boldsymbol{\theta}_- \in \mathbb{R}^2$ be the tangent and normal vectors of an edge, and let λ_+ and λ_- be gradients along the vectors (we omit $\boldsymbol{\theta}(p)$ and $\lambda(p)$ for simplicity). The vectors and gradients are obtained from SVD (singular value decomposition) of the structure tensor $\mathbf{G} \in \mathbb{R}^{2\times 2}$ which consists of gradient images of the *dt*-image $\{D_x, D_y\} = \nabla D$ as follows:

$$\mathbf{U}\boldsymbol{\Sigma}\mathbf{U}^T = \mathrm{svd}(\mathbf{G}), \text{where } \mathbf{G} = \begin{bmatrix} K * D_x^2 & K * D_x D_y \\ K * D_x D_y & K * D_y^2 \end{bmatrix}, \quad (5)$$

where $K*$ represents Gaussian filtering to smooth the gradient images. $\boldsymbol{\Sigma} = \mathrm{diag}(\lambda_+, \lambda_-)$ indicates the maximum and minimum eigenvalues of $\mathbf{G}$, and $[\boldsymbol{\theta}_+, \boldsymbol{\theta}_-] = \mathbf{U}$ indicates the corresponding eigenvectors.

Using the edge direction vectors $\{\boldsymbol{\theta}_+, \boldsymbol{\theta}_-\}$ and gradients $\{\lambda_+, \lambda_-\}$, the gradient tensor $\mathbf{T} \in \mathbb{R}^{2\times 2}$ is computed as follows:

$$\mathbf{T} = f_+(\sqrt{\lambda_+ + \lambda_-})\boldsymbol{\theta}_+\boldsymbol{\theta}_+^T + f_-(\sqrt{\lambda_+ + \lambda_-})\boldsymbol{\theta}_-\boldsymbol{\theta}_-^T, \quad (6)$$

where $f_\pm(\cdot)$ are functions which control the smoothness of the resulting image. In [9], they are given as

$$f_+(x) = \frac{1}{1+x^2}, \quad f_-(x) = \frac{1}{\sqrt{1+x^2}}. \quad (7)$$

In our implementation, we normalize the gradients $\{\lambda_+, \lambda_-\}$ for the *dt*-images by multiplying the standard deviation by $\varepsilon/\mathrm{std}(\{\sqrt{\lambda_-}\})$, because the original method seems to be designed for the intensity range $[0, 255]$. Here, ε controls edge enhancement.

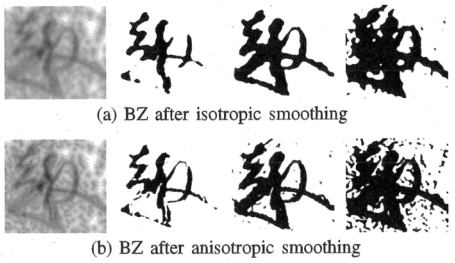
Finally, using the gradient tensor $\mathbf{T}$ and a Hessian matrix $\mathbf{H} = \nabla(\nabla D)^T = \begin{bmatrix} D_{xx} & D_{xy} \\ D_{xy} & D_{yy} \end{bmatrix}$, the iterative smoothing equation is given by

$$D^{(t+1)} = D^{(t)} + \gamma \mathrm{Tr}(\mathbf{TH}), \quad (8)$$

where t is the iteration number and γ controls the convergence. In the experiment in this paper, each parameter of AS is roughly set so as to yield reasonable results.

III. Experimental Results

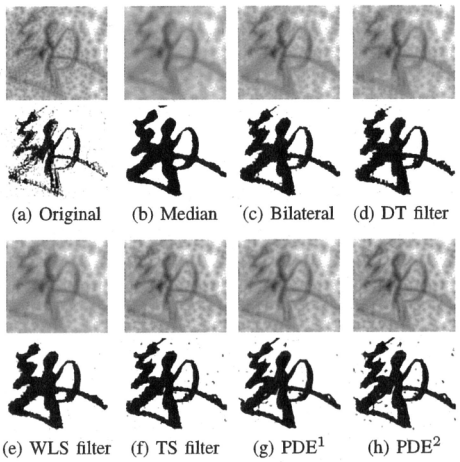
We show some comparison results in this section. The resulting images using our method are yielded by using the same parameter setting as for the AS in [9]: the resizing scale is $1/4$; the standard deviation of the Gaussian filter K is $\sigma = 1.5$; the edge enhancement is $\varepsilon = 20$; the number of iterations and its time step are $t = 30$ and $\gamma = \frac{1}{4}$. On the other hand, the parameters of conventional methods are tuned so as to yield similar results.



(a) BZ after isotropic smoothing

(b) BZ after anisotropic smoothing

Fig. 4. Comparison of BZ with different thresholds, (a) isotropic Gaussian filter, (b) anisotropic diffusion [9]. From the left, *dt*-image, *bw*-images using thresholds $\tau = \{0, 2, 4\}$.



(a) Original (b) Median (c) Bilateral (d) DT filter

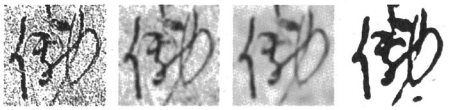
(e) WLS filter (f) TS filter (g) PDE^1 (h) PDE^2

Fig. 5. Comparison of smoothing filters, (a) original *dt*-image (top) and *bw*-image (bottom) for reference, (b) median filter, (c) bilateral filter [8], (d) domain transform (DT) filter (with recursive filtering) [16], (e) WLS filter [17], (f) texture smoothing (TS) [10], (g) anisotropic diffusion (PDE^1) in [18], (h) anisotropic diffusion (PDE^2) in [9]. The same threshold $\tau = 2$ is used for BZ. The parameters of each method are adjusted so as to yield similar results.

There seems to be many parameters required to be set, however, the only important parameter is the BZ threshold τ described later. The other parameters for the AS are not so sensitive for the resulting images. Although the parameters usually should be changed depending on image sizes, in our method, we can resize the distance image for the size normalization (because the distance images have robustness for scaling) and use the fixed parameters for that resolution. When applying our method to the whole image including many characters (as shown in Fig. 1), we need to adjust the BZ threshold for a few characters respectively, depending on the deterioration degree.

A Comparison of BZ thresholds is shown in Fig. 4. Reasonable thresholds are around the range $\tau \in [0.5, 2]$ in this experiment. Note that the morphological dilation via DT allows non-integer thresholds, which is different from traditional dilation, and it gives more latitude in the threshold selection. The use of anisotropic smoothing tends to preserve the stroke shapes.

A Comparison of AS methods is shown in Fig. 5. The



(a) (b) (c) (d)

Fig. 6. Imitative smoothing: the DT is substituted with a smoothing filter with a Laplacian distribution. The input *bw*-image (a) is smoothed by a smoothing filter with a Laplacian distribution (b). Then, the image is further smoothed by AS (c). Finally, the smoothed image is reconverted into a *bw*-image by BZ (d).

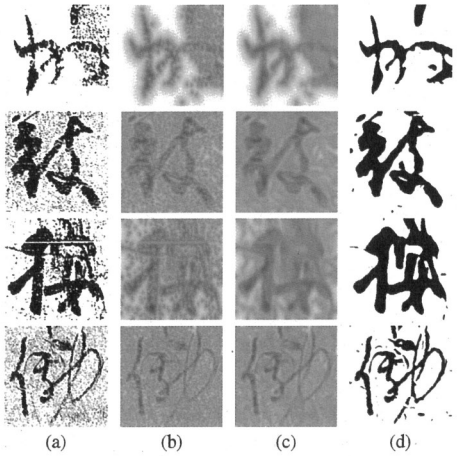
methods (b-d) have weak anisotropy, while methods (e-h) are gradient based methods and have strong anisotropy. The differences mainly appear in the smoothness of the stroke boundaries. In practice, the anisotropic diffusions (g) and (h) correctly capture the concave shape near the center of the image. The noisy background patterns remaining in the diffusion methods can be removed by applying morphological region properties such as area and aspect ratio.

Imitative smoothing without DT (IS). To show the effectiveness of using the DT domain, we substitute the smoothing filter with a Laplacian distribution for the DT. Since the shape of the Laplacian $\frac{1}{C}\exp(-\frac{|x|}{\sigma})$, where C is a coefficient for normalization, has a cuspate shape, the intensity of the resulting images has a similar shape to the *dt*-images. Fig. 6 shows a sample of IS where the standard deviation of the Laplacian is $\sigma = 15$. The bottleneck of this IS filter is it requires a large filter size for better results. Even if the Fourier transform is used, boundary expansion is required to reduce the artifacts caused by the circularity of the FFT.

Comparison with conventional methods. In Fig. 7 and Fig. 8, we show the original bw image (a) and an ideal image (j) for reference, and four kinds of denoising and smoothing methods which seem to be effective for *bw*-images: **morphological methods** (b)(c), **our DT based methods** (d)(e), **statistical methods** (f)(g), and **segmentation based methods** (h)(i). In addition, we show two samples: a sample with a noisy background (Fig. 7) and a sample with fractionated thin strokes (Fig. 8), because some methods tuned for the former sample are not good at dealing with the latter sample.

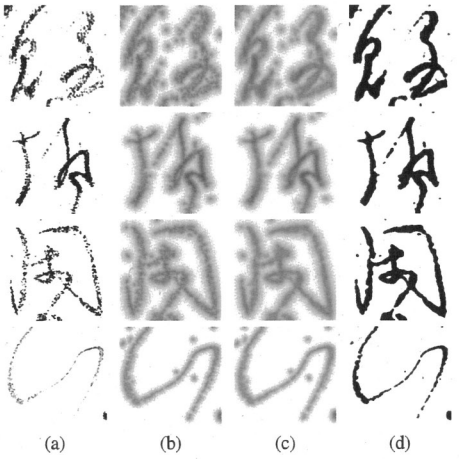
From the results, one can see the segmentation based methods (h)(i) and our method (d) extract the stroke shapes. Actually the median filter (g) has a basic property of the TCP (h), and it yields a similar result. The problem with the segmentation methods is their execution time (for random memory access) to search the thick density direction of the point cloud. In contrast, our method can be performed in a short time. The execution time of our method implemented in MATLAB is 2 (sec/Mpix) on a Core i7 2.67GHz PC.

The other results using our method are shown in Fig. 9 and Fig. 10. The upper images with minor deterioration are restored well. However, it is still difficult to restore fractionated patterns with wide intervals. In such cases, additional interpolation using manual annotation is required in the OCR stage [6]. Though, since one can roughly see the character shapes in the *dt*-images, especially in Fig. 10, it is possible to extract the correct stroke shapes with adaptive thresholding. Incidentally, different thresholds $\tau = 0.7$ and $\tau = 3$ are used in each figure. We think it is possible to decide the adequate threshold from the degree of background noise.



(a) (b) (c) (d)

Fig. 9. Our results for other samples with a noisy background. From left, (a) original *bw*-image, (b) *dt*-image, (c) smoothed *dt*-image, (d) resulting smoothed *bw*-image. The threshold used here is $\tau = 0.7$.



(a) (b) (c) (d)

Fig. 10. Our results for other samples with fractionated thin strokes. The order of the images is the same as in Fig. 9. The threshold used here is $\tau = 3$.

IV. Conclusions

In historical documents written by *Sumi*, characters need to be considered as not only a shape but the stream of lines. Because some characters can be recognized only by their stroke orders. We have described a method for restoring the stream of strokes, using implicit smoothing for the morphological operation. This method is able to remove noisy patterns in the background and connect the fractionated character strokes in a short execution time. In future work, the final thresholding process has to be improved by employing adaptive thresholding. Also, anisotropic smoothing must be improved to suit *dt*-images.

Acknowledgment

This work was supported by the Grant-in-Aid for Scientific Research (S)-20222002, (A)-23240031, (C)-24520771, and (S)-25220401.

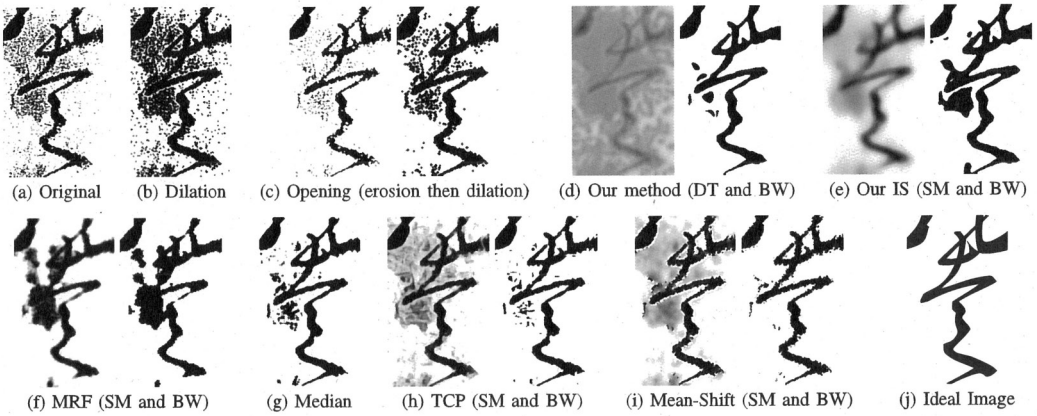


Fig. 7. Comparison with conventional methods with a noisy background, where SM means the smoothed image. In lexical order, (a) original BW, (b) dilation, (c) opening, (d) our method, (e) our imitative smoothing (IS), (f) binary Markov random fields (MRF) [13], (g) median filter, (h) tow-colored pixels (TCP) [19], (i) multiresolution mean-shift [20], and (j) ideal image. The parameters of each method are set so as to give good results.

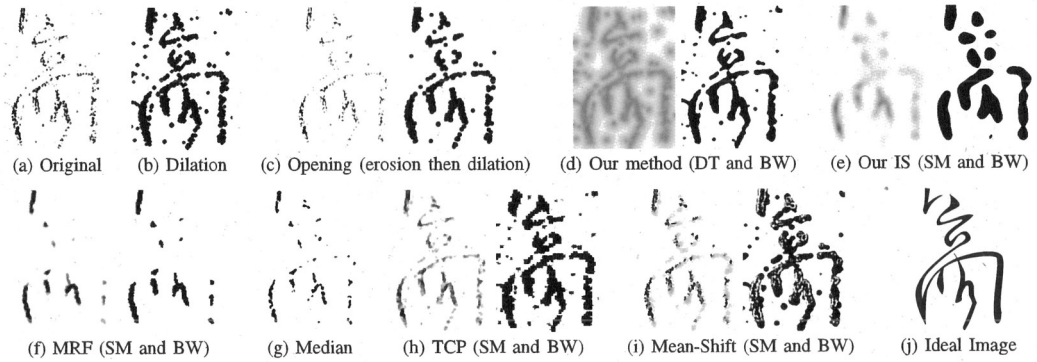


Fig. 8. Comparison with conventional methods for fractionated thin strokes. The order of the images is the same as in Fig. 7.

REFERENCES

[1] "Databases of *Kuzushi ji*," Univ. Tokyo, http://wwwap.hi.u-tokyo.ac.jp/ships/db-e.html, cross search: http://r-jiten.nabunken.go.jp/kensaku.php.

[2] "Databases of *Mokkan*," NRICP Nara, http://jiten.nabunken.go.jp/index.html, cross search: http://r-jiten.nabunken.go.jp/kensaku.php.

[3] B. Gatos, I. Pratikakis, and S. J. Perantonis, "Adaptive degraded document image binarization," *Elsevier Trans. Pattern Recogn.*, vol. 39, no. 3, pp. 317–327, 2006.

[4] M. R. Gupta, N. P. Jacobson, and E. K. Garcia, "Ocr binarization and image pre-processing for searching historical documents," *Elsevier Trans. Pattern Recogn.*, vol. 40, no. 2, pp. 389–397, 2007.

[5] R. F. Moghaddam, D. R.-Hénault, and M. Cheriet, "Restoration and segmentation of highly degraded characters using a shape-independent level set approach and multi-level classifiers," in *Proc. IAPR ICDAR*, 2009, pp. 828–832.

[6] A. Kitadai, M. Nakagawa, H. Baba, and A. Watanabe, "Similarity evaluation and shape feature extraction for character pattern retrieval to support reading historical documents," in *Proc. IAPR Intl. WS. DAS*, 2012, pp. 359–363.

[7] *Extrapolation, interpolation, and smoothing stationary time series*. New York: Wiley, 1949.

[8] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color imagegs," in *Proc. IEEE ICCV*, 1998, pp. 839–846.

[9] D. Tschumperlé and R. Deriche, "Vector-valued image regularization with PDE's: A common framework for different applications," *IEEE Trans. PAMI*, vol. 27, no. 4, pp. 506–517, 2005.

[10] L. Xu, Q. Yan, Y. Xia, and J. Jia, "Structure extraction from texture via relative total variation," *ACM TOG (Proc. SIGGRAPH Asia)*, vol. 31, no. 6, pp. 139:1–139:10, 2012.

[11] P. Soille, *Morphological image analysis: principles and applications*, 2nd ed. Springer-Verlag New York, Inc., 2003.

[12] A. Criminisi, T. Sharp, C. Rother, and P. Perez, "Geodesic image and video editing," *ACM TOG (Invited talk at SIGGRAPH)*, 2010.

[13] B. Vidakovic, "Markov random fields," IAPR Researcher and Student Resources.

[14] C. Green, "Improved alpha-tested magnification for vector textures and special effects," in *ACM SIGGRAPH courses*, 2007, pp. 9–18.

[15] R. Azriel and J. Pfaltz, "Sequential operations in digital picture processing," *J. Association Computing Machinery*, vol. 13, no. 4, pp. 471–494, 1966.

[16] E. S. L. Gastal and M. M. Oliveira, "Domain transform for edge-aware image and video processing," *ACM TOG (Proc. SIGGRAPH)*, vol. 30, no. 4, pp. 69:1–69:12, 2011.

[17] Z. Farbman, R. Fattal, D. Lischinski, and R. Szeliski, "Edge-preserving decompositions for multi-scale tone and detail manipulation," *ACM TOG (Proc. SIGGRAPH)*, vol. 27, no. 3, pp. 67:1–67:10, 2008.

[18] R. Kimmel, R. Malladi, and N. Sochen, "Images as embedded maps and minimal surfaces: movies, color, texture, and volumetric medical images," *Intl. J. of Comput. Vision*, vol. 39, no. 2, pp. 111–129, 2000.

[19] D. Pavić and L. Kobbelt, "Two-colored pixels," *Computer Graphics Forum (Proc. Eurographics)*, vol. 29, pp. 743–752, 2010.

[20] D. Comaniciu and P. Meer, "Mean Shift: a robust approach toward feature space analysis," *IEEE Trans. PAMI*, vol. 24, pp. 603–619, 2002.