# Enhancement of Camera-captured Document Images with Watershed Segmentation

Jian Fan
*Hewlett-Packard Laboratories*
*jian.fan@hp.com*

## Abstract

*Document images acquired with a digital camera often exhibit various forms of degradation, with some commonly encountered forms being uneven illumination, color shift, and blurry text. To complicate matters, the content of modern documents has become increasingly complex. The potential combination of poor image quality and complex content is very challenging for an image processing algorithm to cope with. Towards this end, we present an image enhancement algorithm based on watershed segmentation. The problem of over-segmentation is alleviated by noise thresholding of gradient magnitude. The segmentation is then used for illumination and color correction, as well as for direct text sharpening. Our results show that the watershed segmentation is more robust to noise and accurate in object boundaries than a direct region growing. We also show that the proposed method works well with both text-only documents and with mixed text/graphical documents.*

## 1. Introduction

Digital cameras possess several unique advantages for document capture. Compact digital cameras, especially camera phones, are convenient to carry around. Professional digital cameras, with resolutions now commonly exceeding ten million pixels, have been widely used for various large-scale book digitization projects, showcasing the non-destructive nature of digital camera capture. However, document capture with digital cameras has many inherent limitations [1]. It is very difficult to project uniform lighting onto a document surface, and this often results in uneven illumination and color shift in the acquired images. For documents captured with handheld compact cameras, text blur is also commonplace. These degradations are of interest to this paper.

For the purpose of correcting non-uniform illumination, an illumination-reflectance model $f(x,y)=I(x,y)R(x,y)$, where $f(x,y)$ is the observed image, $I(x,y)$ is the illuminant and $R(x,y)$ is the reflectance, is commonly used [2,3]. In practice it is further assumed that the scale of the illumination's fluctuation is far larger than the affected objects and that the original document contains flat and white background areas that cover a significant portion of the total document area. A classic method for removing the illuminant is homomorphic filtering [2]. In [3], Pilu and Pollard presented a method with illuminant estimation by block averaging followed by a direct correction of $R(x,y)=f(x,y)/I(x,y)$. In an effort to reduce the possible influence exerted by the dark text pixels, Shih-Chang Hsia et al proposed a more sophisticated method in which averaging is done only on a number of maximum values within a line section [4]. However, the performance of these methods may be directly affected by the parameter of block size. In general, the block size should be larger than all text strokes such that no block will fall completely within a stroke of text. If this were to occur, the estimated illuminant surface may dip in this area, adversely affecting the quality of the text or object. On the other hand, increasing the block size generally reduces the scale of light fluctuation that the algorithm can cope with.

An alternative to fixed blocks is image segmentation. One such method, proposed by Yanowitz and Bruckstein, relies on edge detection [5]. However, edge detection has two major drawbacks: 1) it is sensitive to noise and threshold selection, and 2) it does not guarantee closed boundaries on all perceived regions. These drawbacks may significantly impair the robustness of an application's performance. Another commonly used tool is watershed transform [6, 7]. Watershed transform is a parameter-free segmentation method based on mathematical morphology. Instead of applying watershed transform directly to grayscale images, S. Beucher applied the watershed transform to image gradient [8]. This framework was adopted by Wang et al for scene text extraction [9]. The main drawback of the watershed-based segmentation method is over-segmentation. Fundamentally, the problem is largely due to noise. It may be alleviated by using various noise filtering techniques and appropriate application-specific heuristics. For scene text extraction, Wang et al used a weighted median filter based anisotropic diffusion (WMFAD) for noise filtering and heuristics of background regions. Their two-step region-clustering process utilizes heuristics of the size, location, border length, and mean color and intensity of the regions. In their text extraction step, they further incorporated heuristics of the height-to-width ratios of connected components. The authors caution that their method "is only suitable for square character extraction" [9].

Document image enhancement differs from text extraction and binarization in two major aspects. First, the output of document image enhancement should closely resemble the original hardcopy. In other words, the integrity and appearance of pictorial regions within a document should be preserved. Secondly, document image enhancement should be effective for documents of wide-ranging text alphabets, fonts, and orientations. These two requirements rule out the use of many heuristics derived from certain text properties for the application.

In this paper, we apply the watershed-based segmentation framework to the enhancement of camera-captured document images. The block diagram of the complete algorithm is shown in Figure 1. First, a linear Gaussian filtering is performed. Second, a color gradient magnitude is computed. This is followed by a hard thresholding, which has proved very effective in eliminating over-segmentation in the background regions. In the fourth step, the watershed transform is applied to the gradient magnitude and the background regions are determined. Finally, segmentation is used to estimate the illuminant surface $I(x, y)$ and color correction multipliers, and to guide a spatial selective sharpening. We shall point out that for the enhancement application it is not necessary to have a complete segmentation such as that for text extraction and binarization. In particular, we do not need to identify small background regions isolated within text characters. Under the smoothness assumption of the illuminant surface, a complete surface $I(x, y)$ may be reconstructed with linear interpolation from incomplete data points. The tolerance of incomplete segmentation simplifies the algorithm and improves its robustness.

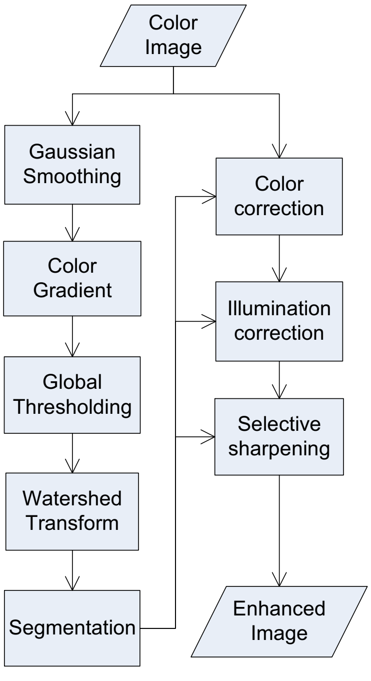


Figure 1. The block diagram of the proposed enhancement algorithm.

The remainder of this paper is organized as follows. Section 2 details the scheme for background segmentation. Section 3 describes image enhancements. Experimental results and comparisons are shown in Section 4. Section 5 summarizes and concludes the paper.

## 2. Background segmentation via watershed transform

We use the watershed-based framework for the background segmentation. Two major issues are over-segmentation and the identification of the background regions.

### 2.1. Noise filtering

Since noise is the main source of over-segmentation, noise filtering is critical to the performance of the overall algorithm. Most noise-filtering techniques can be classified as either linear or non-linear. Gaussian filters are commonly-used linear filters. Most edge-preserving and rank filters are non-linear. It has been shown that Gaussian bilateral filtering achieves very similar results as anisotropic diffusion without the complexity of multiple iterations [10, 11]. These filters operate in the image domain. There are others that operate in transform domains. Wavelet shrinkage is a simple and efficient denoising technique in the wavelet domain [11]. For our application, we applied a hard thresholding to gradient magnitude by setting gradient magnitude values below a threshold to zero. The threshold value is a critical parameter. Ideally, it should be determined by background noise level. In practice, the threshold may be set proportional to the standard deviation of the gradient magnitudes:

$$th_g = \begin{cases} k \cdot \sigma_g, & \text{if } (k \cdot \sigma_g) > th_{\min} \\ th_{\min}, & \textit{otherwise} \end{cases}$$

where $\sigma_g$ is the standard deviation of the gradient magnitudes, $k$ is a real number and $th_{min}$ is a pre-determined minimum threshold value.

Our experiments showed that thresholding of gradient magnitude is significantly more effective than bilateral filtering in reducing over-segmentation of background regions.

## 2.2. Image gradient

Experiments have shown that better and more complete region boundaries may be detected from color gradient than from grayscale gradient. There have been many color gradient operators proposed in the literature [12, 13]. For this application, we achieved very similar results with either Di Zenzo's color gradient or the simpler "max component gradient." The thresholding operation as described in the last section is then applied to the gradient magnitude to remove the noise component. The denoised gradient magnitudes are then linearly mapped onto an 8-bit grayscale image.

## 2.3. Background segmentation

After applying Vincent and Soille's fast watershed transform to the 8-bit grayscale gradient image, every pixel is labeled either as a (catchment basin) region or as a watershed. For our application, the watershed pixels are simply merged with the neighboring region possessing the largest label number such that every pixel belongs to a region. The various steps and the effect of gradient thresholding are illustrated in Figure 2. It can be seen that both bilateral filtering (of size 11×11, $\sigma_d = 1.3$ and $\sigma_r = 35$ [10]) and Gaussian filtering (of size 9×9 and $\sigma = 1.3$) fail to completely remove background noise, resulting in severe over-segmentation in Figure 2 (d) and (e). The effect of gradient thresholding ($th_g = 4$, in this case) is apparent in Figure 2 (f) and (g) in that it essentially eliminated the over-segmentation in the main background region although the character regions are still fragmented.

To identify the background region, we compute the sum of normalized intensity for all pixels of each region $R_k$:

$S_k = \sum_{(i,i)\in R_k} \left( y_{i,j} / 255 \right)$, and select the region with the largest S sum as the background region, where $y_{i,j}$ is the pixel intensity at location $(i, j)$. For the example of Figure 2 (a), the segmentation results are shown in Figure 2 (h) and (i).

# 3. Image enhancements

Our enhancement components include illumination correction, color correction, and text sharpening.

## 3.1. Illumination correction

The illuminant surface may be estimated directly from the input image itself using the segmentation map. Assuming that the reflectivity of the document surface is uniform, illuminant values should be proportional to pixel luminance at background regions. Illuminant values at non-background regions may be interpolated from known background regions. One way to interpolate the two-dimensional surface is described in [5]. In practice, a separable 1-D (row and column) linear interpolation may be sufficient.

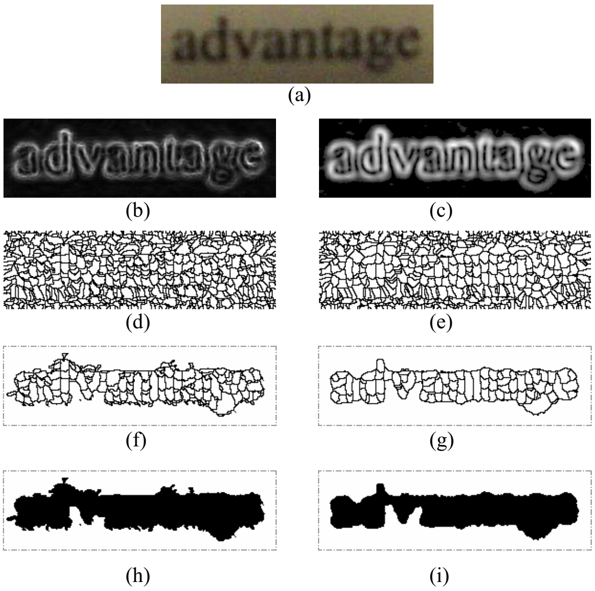

(a)
(b) (c)
(d) (e)
(f) (g)
(h) (i)

Figure 2. An example of background segmentation with the input image of (a). The left column shows the results of using a bilateral filter while the right column shows the results of a Gaussian filter. (b) and (c) are gradient images, (d) and (e) are watershed transforms, (f) and (g) are watershed transform using thresholded gradients, and (h) and (i) are segmentation results.

For an observed image $f(x, y)$ and the estimated illuminant surface $\hat{I}(x, y)$, the illumination-corrected image may be directly computed with

$$\hat{R}(x, y) = 255 * f(x, y) / \hat{I}(x, y).$$

## 3.2. Color correction

The segmentation is also used for color correction. Assuming that the true background color is a uniform neutral gray and that the observed average background color is $(\overline{R}_0, \overline{G}_0, \overline{B}_0)$, a set of three multipliers

$$(m_R, m_G, m_B) = \left( C_{\min} / \overline{R}_0, C_{\min} / \overline{G}_0, C_{\min} / \overline{B}_0 \right)$$

can be computed to convert the color $(\overline{R}_0, \overline{G}_0, \overline{B}_0)$ back into a neutral gray, where $C_{\min} = \min(\overline{R}_0, \overline{G}_0, \overline{B}_0)$. Notice that we assume the *R*,*G*,*B* color used here to be linear *R*,*G*,*B* values. The three multipliers are then applied to the *R*,*G*,*B* color planes of the whole image.

### 3.3. Selective sharpening

Unsharp masking is a simple and effective method for text sharpening. For a linear unsharp masking, the enhanced image may be expressed as $q(x,y) = (\beta + 1)\cdot p(x,y) - \beta \cdot g(x,y) \otimes p(x,y)$, where $p(x,y)$ is an input image, $g(x,y)$ is a Gaussian lowpass kernel, $\beta$ is a real number controlling the amount of sharpening, and $\otimes$ denotes a 2D convolution.

The unsharp masking may be applied to the illumination-corrected image. However, it is not desirable to apply the unsharp masking uniformly to all pixels since it may also amplify background noise. Instead, we selectively apply the unsharp masking only to non-background pixels. To take into account the sharpness of the input image, we adaptively determine the size of the Gaussian kernel and the amount of sharpening from the maximum gradient magnitude $g_{\max}$:

$$x = \begin{cases} x_{\min}, & if(g_{\max} > g_H) \\ [x_{\min}(g_{\max} - g_L) + x_{\max}(g_H - g_{\max})]/(g_H - g_L), & g_L \le g_{\max} \le g_H \\ x_{\max}, & if(g_{\max} < g_L) \end{cases}$$

where $x$ is the parameter (window size or amount of sharpening) to be determined, $x_{\min}$ and $x_{\max}$ are the predetermined minimum and maximum values of the parameter, and $g_L$ and $g_H$ are the predetermined low and high reference gradient magnitude values.

## 4. Experimental results and discussion

In this section, we present experimental results of applying the proposed algorithm to several representative document images. The baseline parameters of the algorithm for the test images are the same: for pre-smoothing, 7×7 Gaussian kernel with $\sigma = 1.3$; for gradient thresholding, $th_{\min} = 4$ and $k = 0.5$; for selective unsharp masking, $g_H = 160$ and $g_L = g_H / 3$, $w_{\min} = 3$ and $w_{\max} = 7$ for the window size, and $\beta_{\min} = 0.5$ and $\beta_{\max} = 3$ for the amount of sharpening.

Figure 3 (a) shows a poor-quality image with text-only content captured with a camera phone. Figure 3 (b) is the result with the proposed method. Figure 3 (d) is the segmentation result. Figure 3 (c) shows the estimated illuminant surface. As a comparison, we also show the segmentation result using a direct region growing in Figure 3 (e). In this case, we simply selected the largest connected component of pixels with gradient magnitude equal to zero (after thresholding). It can be seen clearly that the segmentation with direct region growing is significantly noisier and less accurate in identifying boundaries.

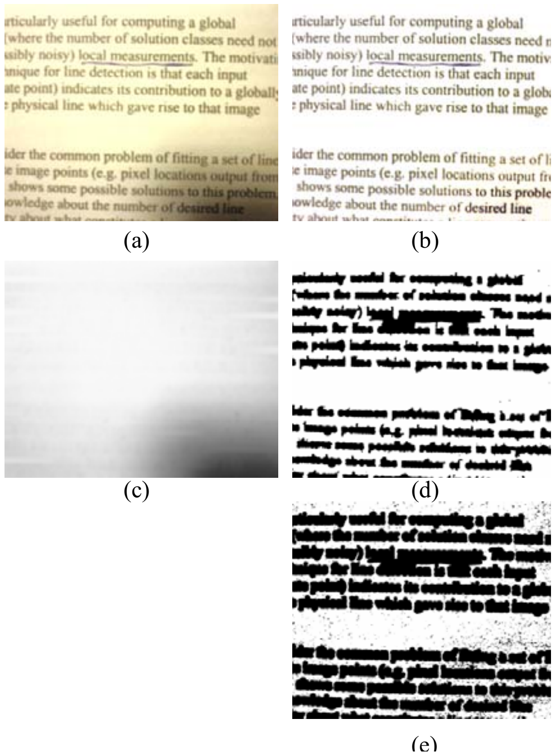
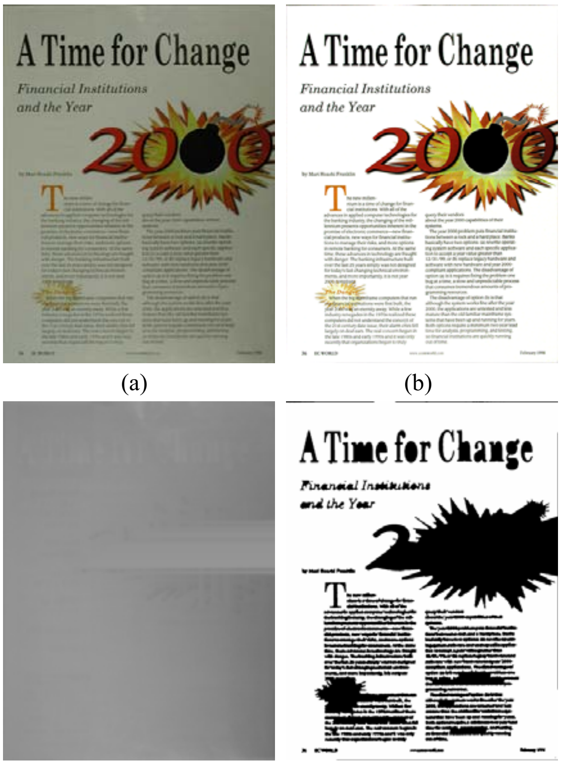

(a) (b) (c) (d) (e)

Figure 3. A text-only partial document. (a) the original image (1280×1024); (b) enhanced image with the proposed method; (c) estimated illumination surface; (d) segmentation map; (e) segmentation with direct region growing.

Figure 4 shows the case of a full-color magazine page. The original image (2048×1536, not shown) was captured using a 3MB Olympus C3000 digital camera. Figure 4 (a) is the image after rectification and cropping, and is the input for the enhancement algorithm. Figure 4 (b) is the result with the proposed method. Figure 4 (c) shows the estimated illuminant surface. Notice that the illumination and color correction worked well in making the background uniformly white without damaging the graphical regions.

Figure 5 shows a book page captured with a professional digital camera. The image is dominated by a very large picture region. Even though the background area is smaller, it was identified correctly and the enhancement result is satisfactory.

Figure 6 shows a whiteboard image with irregular hand drawings. Even though the segmentation algorithm missed some background areas enclosed by hand drawings, the illuminant surface is still estimated quite well and a good enhancement result is achieved.

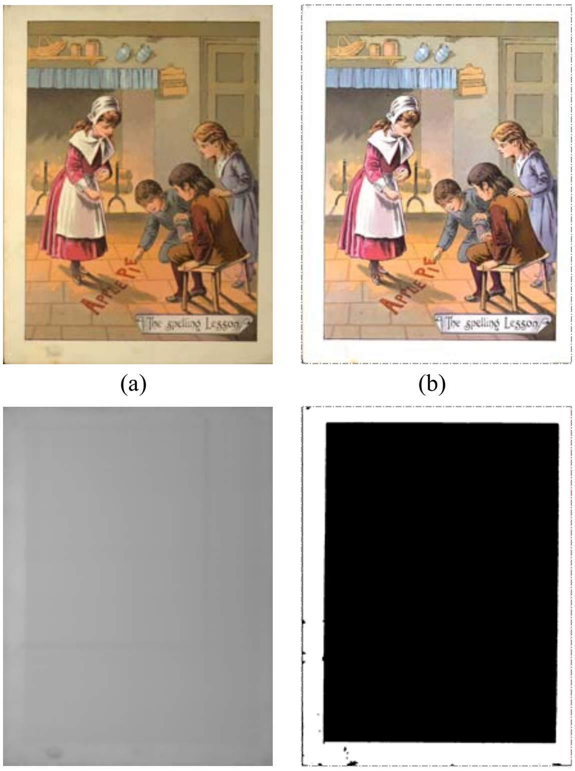

A Time for Change
Financial Institutions and the Year 2000
A Time for Change
Financial Institutions and the Year 2000
A Time for Change
Financial Institutions and the Year

(a) (b)

(c) (d)

Figure 4. A full magazine page. (a) rectified and cropped image (1419×1878) for enhancement; (b) enhanced image with the proposed method; (c) estimated illumination surface; (d) segmentation map.

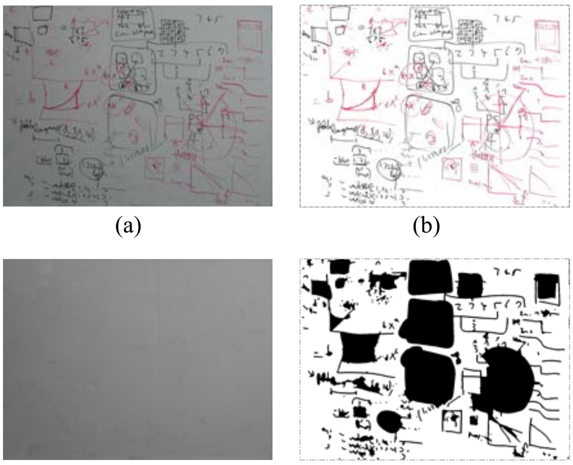
For the purpose of comparison, we implemented the block-based illumination correction algorithm proposed by *Hsia et al* [4]. The parameters for their algorithm include the size *C* of the sections of the raster scan-lines and the number *M* of maximum values for averaging. For the comparison results shown in Figures 7 and 8, two settings, with $M = 5$ and the number *N* of sections equal to 5 and 10, are included. The comparisons use the grayscale version of the two images, and only the illumination correction part of the proposed algorithm is applied. The results clearly verify theoretical analysis. For a large block size ($N = 5$), the prominent, solid “explosion” figure is left largely intact. However, processing with the block size did not make the background uniform, as is evident on the right side of Figure 8 (c) and in the bottom right corner of Figure 7 (c). On the other hand, with smaller block size ($N = 10$), background pixels are more uniformly white. However, damage to large objects (the “explosion” in Figure 8 (d) and handwritten underline in Figure 7 (d)) becomes evident. In contrast, illumination correction with the proposed method achieved satisfactory results in both cases.



(a) (b)

(c) (d)

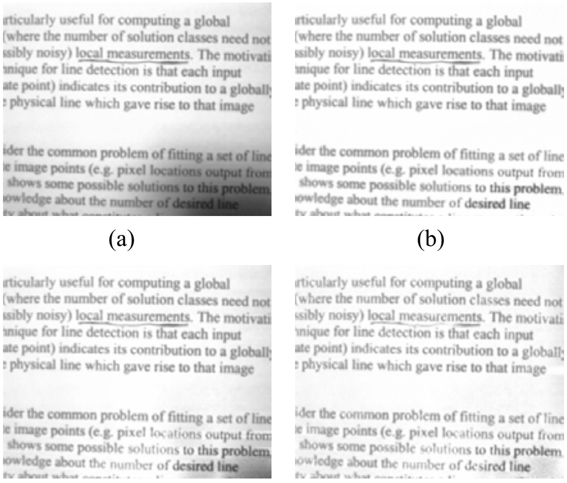
Figure 5. A figure-only book page. (a) cropped image (2409×3224) for enhancement; (b) enhanced image with the proposed method; (c) estimated illumination surface; (d) segmentation map.



(a) (b)

(c) (d)

Figure 6. A whiteboard image. (a) rectified and cropped image (1344×1002) for enhancement; (b) enhanced image with the proposed method; (c) estimated illumination surface; (d) segmentation map.



(a) (b)

(c) (d)

Figure 7. Comparison with Hsia *et al*'s method. (a) grayscale version of the image of Figure 3 (a). (b) illumination corrected using the proposed method; (c) correction with Hsia's method, with N = 5, (d) correction with Hsia's method, with N = 10.

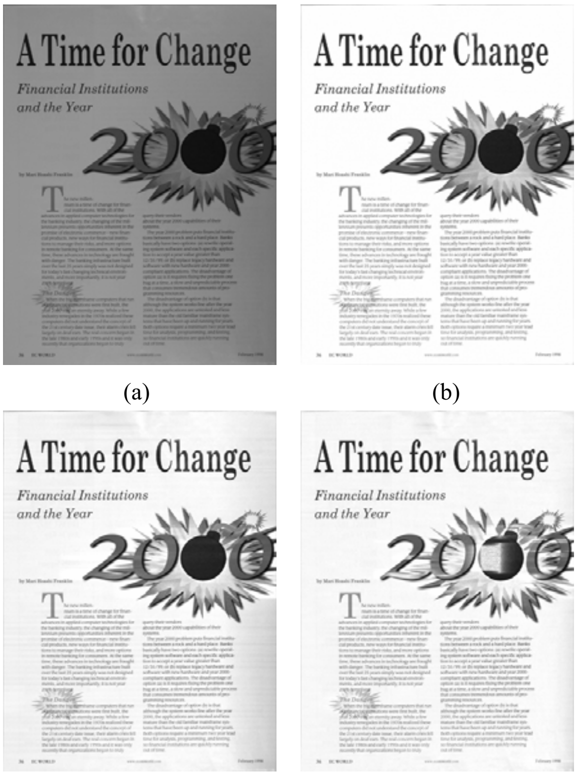
## 5. Conclusion

In this paper, we presented a segmentation-based image enhancement method and demonstrated its advantages over a fixed-block-based method. The segmentation is obtained using watershed transform on gradient magnitude. Both techniques lie on a solid mathematical foundation, and therefore their consistent behaviors can be expected. For the targeted enhancement task, we alleviate the over-segmentation problem by noise thresholding and by focusing on the background region. We demonstrated that satisfactory results can be achieved on images of various quality and content.

The main limitation of the current implementation lies in the assumption that the background region constitutes the largest catchment basin. Although this assumption is generally true for the large majority of documents, there are plenty of documents for which this assumption is invalid. In these cases, cross-region analysis is required to identify disconnected background regions. Other areas for future research include methods for better estimation of background noise and image sharpness.

## 6. References

[1] Jian Liang, David Doermann, Huiping Li, "Camera-based analysis of text and documents: a survey", IJDAR (2005) 7, p. 84–104



(a) (b)

(c) (d)

Figure 8. Comparison with Hsia *et al*'s method. (a) grayscale version of the image of Figure 4 (a). (b) illumination corrected using the proposed method; (c) correction with Hsia's method, with N = 5, (d) correction with Hsia's method, with N = 10.

[2] Rafael C. Gonzalez and Paul Wintz, *Digital image processing*, 2nd edition, Addison-Wesley, Reading, Massachusetts, 1987

[3] Pilu M., Pollard S., "A light-weight text image processing method for handheld embedded cameras", British Machine Vision Conference, Sept. 2002

[4] Shih-Chang Hsia, Ming-Huei Chen, and Yu-Min Chen, "A cost-effective line-based light-balance technique using adaptive processing", IEEE Trans. Image Proc., Vol. 15, No. 9, p. 2719-2729, Sept. 2006

[5] SD Yanowitz, AM Bruckstein, "A new method for image segmentation", CVGIP, v.46 n.1, p.82-95, April 1989

[6] Jos B.T.M. Roerdink and Arnold Meijster, "The watershed transform: definitions, algorithms and parallelization strategies", Fundamenta Informaticae 41 (2001) p. 187-228

[7] Vincent, L., and Soille, P. Watersheds in digital spaces: an e_cient algorithm based on immersion simulations. IEEE Trans. Patt. Anal. Mach. Intell. 13, 6 (1991), p. 583-598

[8] S. Beucher. The watershed transformation applied to image segmentation. Conference on Signal and Image Processing in Microscopy and Microanalysis, p. 299--314, September 1991

[9] Kongqiao Wang, Jari A. Kangas, Wenwen Li, “Character Segmentation of Color Images from Digital Camera”, ICDAR'01, p. 210-214

[10] C. Tomasi and R. Manduchi, “Bilateral filtering for gray and color Images”, ICCV, Bombay, India, 1998

[11] Danny Barash, “A fundamental relationship between bilateral filtering, adaptive smoothing, and the nonlinear diffusion equation”, IEEE Trans. PAMI, VOL. 24, NO. 6, p. 844-847, June 2002,

[11] D. Donoho. “De-noising by soft thresholding”, IEEE Trans. Information Theory, vol. 38(2), p. 613--627, 1995.

[12] Silvano Di Zenzo, “A note on the gradient of a multi-image”, CVGIP, Vol. 33, p. 116-125, 1986

[13] Jian Fan, “A local orientation coherency weighted color gradient for edge detection”, ICIP 05, p. 1132-1135, Sept. 2005