IEEE Signal Processing Letters: SPL-00466-2002

1) *Paper Title*

Distance-Reciprocal Distortion Measure for Binary Document Images

2) *Authors*

Haiping Lu*, *Student Member, IEEE,*

Alex C. Kot, *Senior Member, IEEE,*

Yun Q. Shi, *Senior Member, IEEE*

3) *Affiliations*

Haiping Lu* : Nanyang Technological University, Singapore

Alex C. Kot : Nanyang Technological University, Singapore

Yun Q. Shi : New Jersey Institute of Technology, USA

4) *Contact Information of Corresponding Author*

Mr. Haiping Lu

Mailing address: WRL, S2-B3b-10, School of EEE, NTU, Nanyang Avenue, Singapore 639798

Telephone: (65) 6790-6462

Fax: (65) 6793-3318

E-Mail: ehplu@ntu.edu.sg

5) *Software Used*

Windows 2000/LaTeX2e

6) *EDICS Category*

SPL.IMD

Abstract

In this letter, we present a novel objective distortion measure for binary document images. This measure is based on the reciprocal of distance that is straightforward to calculate. Our results show that the proposed distortion measure matches well to subjective evaluation by human visual perception.

Index Terms

Objective distortion measure, document image, human visual perception.

I. INTRODUCTION

Digital document image processing is receiving more and more attention recently. Digital document images are essentially binary images. In some binary document image applications, such as watermarking and data hiding, visual distortion may be present and it is necessary to measure such distortion for performance comparison or evaluation [1]. There are two ways to measure visual distortion, as discussed in [2]. One is subjective measure and the other is objective measure. Subjective measure is costly, while it is important since human is the ultimate viewer. On the other hand, objective measure is repeatable and easier to implement, while such a measure does not always agree with the subjective one.

In this letter, we propose an objective distortion measure for binary document images that is based on human visual perception. Binary document images here refer to binary images that have sharp contrast of black and white and there are clear boundaries between black and white areas in the images. The distance between pixels is found to play an important role in human perception of distortion in these images. Hence, the reciprocal of distance is used to measure visual distortion in digital binary document images. Subjective testing results show a good correlation between the proposed objective measure and human visual perception.

II. PSNR AND DOCUMENT IMAGES IN HUMAN EYES

The PSNR (peak signal-to-noise ratio) is a popular distortion measure used in image and video processing. For an image processing system with $f(x, y)$ as the input image and $g(x, y)$ as the processed output image, the PSNR is defined as:

$$PSNR(dB) = 10 \log_{10} \frac{P^2 MN}{\sum_{x=0}^{M-1} \sum_{y=0}^{N-1} (g(x, y) - f(x, y))^2} \tag{1}$$

where M and N are the dimensions of the image, and P is the maximum peak-to-peak signal swing. E.g. P is 255 for 8-bit images.

For binary document images, the PSNR does not match well with subjective assessment since it is a point-based measurement and mutual relations between pixels are not taken into account. For instance, a simple document image is shown in Fig. 1(a), and four differently distorted images are shown in Fig. 1(b). Each distorted image has four pixels having opposite binary values compared with their counterpart in Fig. 1(a). According to (1), these four distorted images have the same PSNR, but then, the distortion perceived by our human visual system (HVS) is quite different.

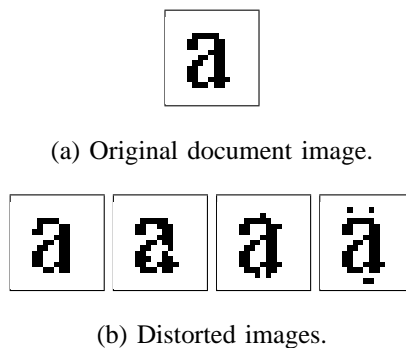(a) Original document image.

(b) Distorted images.

Fig. 1. Distortion in document image.

Human visual perception of document images is different from that of natural images, which are usually continuous multiple-tune images. Document images are essentially binary and there are only two levels, black and white. On the other hand, documents are mostly consisting of characters, which are more like "invented symbols/signals" rather than physical objects in natural images. Hence, the HVS models built for natural images, such as color and contrast, may not be well suited for document images.

In turn, the perception of distortion in document images is also different from that in natural images. In a particular language, such as English, people know very well what a certain alphabetic character should look like. Hence, distortion in document images could be more obtrusive than distortion in natural images, and the distortion measures proposed for color/gray-level images [3]–[5] are not often applicable to binary document images.

Baird developed a document image defect model [6] for the image defects that occur during printing and scanning. This model has been used to construct "perfect metrics" in optical character recognition (OCR) applications [7]. In [8], Baddeley proposed an error metric for binary images to measure edge detection and localization performance in computer vision applications. The model and metrics are mainly for the estimation of classification errors rather than the measurement of visual distortion. Wu *et al.* in [9] measure the visual distortion in data hiding through the change in smoothness and connectivity caused

by flipping a pixel. However, the analysis involved is extensive for a larger neighborhood.

III. DISTANCE-RECIPROCAL DISTORTION MEASURE

We use a number of single-letter images to study distortion in binary document images. Each single-letter image is converted from a letter typed in MS Word with a font size of 10 or 12, including both uppercase and lowercase, using Adobe Acrobat 5.0 with a resolution of 150 dots per inch (dpi). One of them is shown in Fig. 1(a).

We observed that for a binary document image, the distance between two pixels plays a major role in their mutual interference perceived by human eyes. As discussed above, readers are so familiar with alphabetic characters that even a single-pixel distortion can be perceived easily. Therefore, the main factor in distortion perception is focusing, i.e. whether the distortion is in a viewer's focus. The distortion (flipping) of one pixel is more visible when it is in the field of view of the pixel in focus. The nearer the two pixels are, the more sensitive it is to change one pixel when focusing on the other. Further, from a magnified viewing, each pixel is essentially a black or white square. Therefore, a diagonal neighbor pixel is considered to be further away from a pixel in focus than a horizontal or vertical neighbor one. Hence, diagonal neighbors have less effect on a center pixel in focus than horizontal or vertical neighbors.

Based on these observations, we propose an objective distortion measure here for binary document images. This method measures the distortion of a processed image $g(x, y)$ compared with the original image $f(x, y)$ using a weighted matrix with each of its weights determined by the reciprocal of a distance measured from the center pixel. We name it the distance-reciprocal distortion measure (DRDM) method. Specifically, the weight matrix $\mathbf{W}_m$ is of size $m \times m, m = 2n+1, n = 1, 2, 3, 4, 5, ...$ The center element of this matrix is at (i_C, j_C), $i_C = j_C = (m + 1)/2$. $\mathbf{W}_m(i, j)$, $1 \leq i, j \leq m$, is defined as following:

$$\mathbf{W}_m(i, j) = \begin{cases} 0 & \text{for } i = i_C \text{ and } j = j_C \\ \frac{1}{\sqrt{(i-i_C)^2+(j-j_C)^2}} & \text{otherwise.} \end{cases} \quad (2)$$

This matrix is normalized to form the normalized weight matrix $\mathbf{W}_{Nm}$. The weight matrix after normalization is shown in Table I for $m = 5$.

$$\mathbf{W}_{Nm}(i, j) = \frac{\mathbf{W}_m(i, j)}{\sum_{i=1}^{m} \sum_{j=1}^{m} \mathbf{W}_m(i, j)} \quad (3)$$

Suppose that there are S flipped (from black to white or from white to black) pixels in $g(x, y)$, each pixel will have a distance-reciprocal distortion $DRD_k, k = 1, 2, 3, ..., S$. For the kth flipped pixel at $(x, y)_k$ in the output image $g(x, y)$, the resulted distortion is calculated from an $m \times m$ block $\mathbf{B}_k$ in

TABLE I

WEIGHT MATRIX AFTER NORMALIZATION

0.0256	0.0324	0.0362	0.0324	0.0256
0.0324	0.0512	0.0724	0.0512	0.0324
0.0362	0.0724	0	0.0724	0.0362
0.0324	0.0512	0.0724	0.0512	0.0324
0.0256	0.0324	0.0362	0.0324	0.0256

$f(x, y)$ that is centered at $(x, y)_k$. The distortion DRD_k measured for this flipped pixel $g[(x, y)_k]$ is given by

$$DRD_k = \sum_{i,j} [\mathbf{D}_k(i, j) \times \mathbf{W}_{Nm}(i, j)] \tag{4}$$

where the $(i, j)^{th}$ element of the difference matrix $\mathbf{D}_k$ is given by

$$\mathbf{D}_k(i, j) = |\mathbf{B}_k(i, j) - g[(x, y)_k]| \tag{5}$$

Thus, DRD_k equals to the weighted sum of the pixels in the block $\mathbf{B}_k$ of the original image that differ from the flipped pixel $g[(x, y)_k]$ in the processed image. The pixel $f[(x, y)_k]$ does not contribute directly to DRD_k since its weight is always zero. For the possibly flipped pixels near the image edge or corner, where an $m \times m$ neighborhood may not exist, it is possible to expand the rest of the $m \times m$ neighborhood with the same value as $g[(x, y)_k]$, which is equivalent to just ignoring the rest of the neighbors.

After $\mathbf{W}_{Nm}$ walks over all the S flipped pixel positions, we sum the distortion as seen from each flipped pixel visited to get the distortion in $g(x, y)$ as:

$$DRD = \frac{\sum_{k=1}^{S} DRD_k}{NUBN} \tag{6}$$

where $NUBN$ is to estimate the valid (non-empty) area in the image and it is defined as the number of non-uniform (not all black or white pixels) 8×8 blocks in $f(x, y)$. The total pixel number $M \times N$ is not used in the denominator because uniform areas (e.g. all white pixel blocks) are common in binary document images and they may have a significant effect on the distortion value if used.

The proposed DRDM method provides an efficient way to measure distortion in binary document images. It is superior over PSNR in the sense that it takes human visual perception into account and hence correlates well to subjective assessment, which is the ultimate judge on distortion. This correlation is demonstrated by the experimental results presented below.

IV. EXPERIMENTAL RESULTS

We carried out experiments to test how well the distortion measure proposed is matched with human visual perception. We designed the image shown in Fig. 2 to be the original binary document image used in the subjective testing. It is converted from MS Word in the same way as described in Section III, and the characters in the image are with different fixed-size fonts. The image is of size 198×109 and there are 122 ($\approx 39\%$) non-uniform 8×8 blocks out of 312 blocks.

These *are* testing.
Thank   *You!*
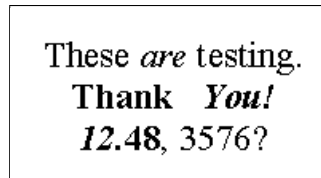***12.48****, 3576?*

Fig. 2.   Original binary document image.

It is important to design a distortion generator that can generate a number of independent test images with various amount of visual distortion. The design criteria is that under the constraint that the number of flipped pixels is the same in each test image, test images generated through this distortion generator should have a wide variety in terms of how noticeable the flipping is. It has been shown through experiments that the flipping of a number of pixels randomly selected will result in images with very similar amount of visual distortion, measured both in DRD values and by human eyes. This is not desirable since it will be very hard for the observers to give a reasonable ranking.

We have built a number of distortion generators and tested them against the above criteria. After careful testing, we choose a distortion generator that is designed to do random flipping in a restricted neighborhood of black pixels in the image. This generator is described below by showing its operation on the original binary document image in Fig. 2, which has 1763 black pixels. It flips 40 pixels in the original image with some randomness to generate the test images with various amount of visual distortion:

1) The positions of all 1763 black pixels are recorded in a 2×1763 matrix.

2) 40 black pixels out of 1763 are randomly chosen using a random number generator with uniform distribution.

3) For each black pixel chosen, one pixel is flipped in its neighboring area. As shown in Fig. 3, the pixel to be flipped is randomly selected from the Band1 pixel (the black pixel itself), or eight Band2 pixels, or sixteen Band3 pixels, with probability of $p1$, $p2$ and $p3$, respectively, and $p1+p2+p3 = 1$. For the Band2 and Band3 pixels, one neighbor is randomly chosen among the band pixels.

4) A total of $60,000$ test images are generated in the experiment by running the generator $10,000$ times for each of the six sets of $p1$, $p2$ and $p3$, with $p3 = 0, 0.2, 0.4, 0.6, 0.8$ and 1, $p1 = (1-p3)/10$ and $p2 = 9p1$.

5) The images generated with the number of flipped pixels less than 40 are ignored. That is, the cases where at least one pixel is flipped more than once are dropped.
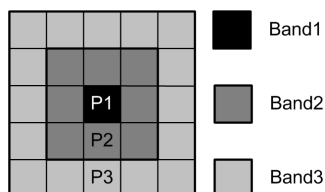

Fig. 3. Distortion generation for subjective testing.

Since all the test images are generated from the same original image and they have the same number of flipped pixels, they have the same PSNR of $27.32dB$ according to (1). One set of the test images generated is shown in Fig. 4.
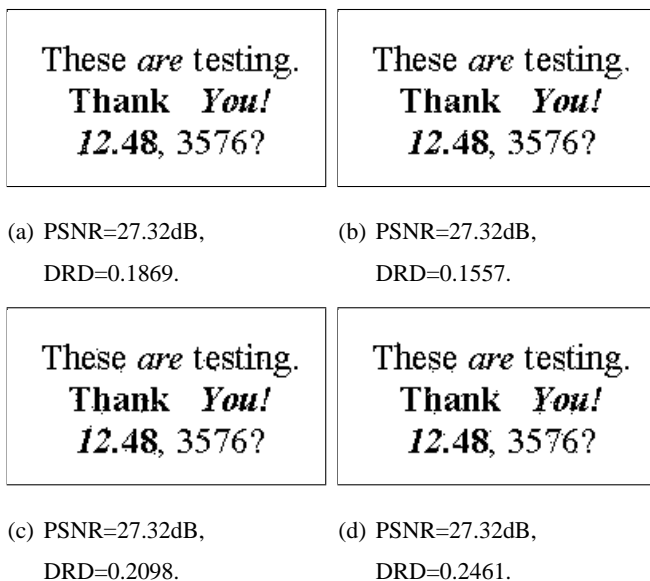

(a) PSNR=27.32dB, DRD=0.1869.

(b) PSNR=27.32dB, DRD=0.1557.

(c) PSNR=27.32dB, DRD=0.2098.

(d) PSNR=27.32dB, DRD=0.2461.

Fig. 4. One set of test images.

Next, we divided all the test images generated into four groups according to the DRD values computed, with group 1 having smallest values, group 4 having largest values and so on.

The subjective assessment is done by 60 observers. Each observer is given the original image and four sets of test images, which are printed on a piece of 80 GSM quality paper using a HP LaserJet 4100 printer. Each set of test images consists of four test images randomly chosen from the four groups. We asked the observers to rank the visual quality of the four images in each set according to the visual distortion that he or she perceives when he or she views the images at a comfortable distance under normal indoor lighting conditions in labs. A smaller ranking score indicates less distortion. There are four rankings (1, 2, 3 and 4) with score 1 for the least distortion and 4 for the most distortion perceived.

The ranking scores collected from the 60 observers are analyzed and compared with the rankings according to the average DRD values (with $m = 5$), as shown in Table II. Although the PSNR is the same for all the test images, the DRD values obtained are different for different distorted images and their average values for the four groups have a normalized correlation of 0.964 with the mean subjective rankings, indicating a very good match between our objective measure and the subjective evaluation.

We choose $m = 5$ here to demonstrate the high correlation between our measure and the subjective rankings. Based on our experimental data, for other values of m, it was found that the correlations calculated using various values (3, 5, 7, ..., 15) of m are quite close. The DRD values for a larger m have a slightly lower correlation with the subjective measure.

The distribution of the subjective ranking scores for each group is shown in a sub-figure in Fig. 5. In each sub-figure, the abscissa represents four ranking scores (1, 2, 3 and 4), and the ordinate shows the counts of the corresponding ranking scores given by the 60 human evaluators. Since each of the 60 observers is given four sets of test images, there are 240 scores in total for each group.

TABLE II

EXPERIMENTAL RESULTS

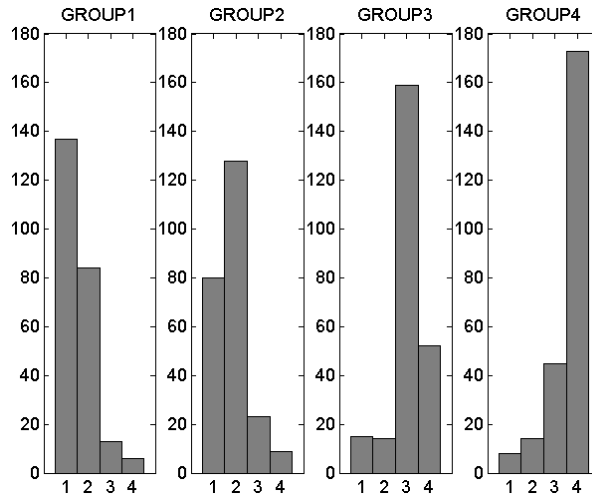| Test Images	Mean Subjective Rank	PSNR (dB)	Average DRD (m=5)
Group 1	1.5333	27.32	0.1566
Group 2	1.8375	27.32	0.1869
Group 3	3.0333	27.32	0.2098
Group 4	3.5958	27.32	0.2413

Fig. 5.   Distribution of subjective ranking scores.

V. CONCLUSIONS

In this letter, we propose an objective distortion measure for binary document images. This measure is derived from our observation that for binary document images, the distance between pixels plays a major role in their visual interference, and it is called the distance-reciprocal distortion measure. Experimental results have shown its high correlation with the subjective assessment. This measure is useful in a wide range of applications involving visual distortion in digital binary document images, such as watermarking, data hiding and lossy compression.

However, this distortion measure is not suitable for halftone (dithered) images, in which black and white pixels are well interlaced and graininess is desired. In this case, the weight matrix elements should be proportional to the distance rather than its reciprocal.

ACKNOWLEDGMENT

R EFERENCES

[1] M. Chen, E. K. Wong, N. Memon, and S. Adams, "Recent developments in document image watermarking and data hiding," in *Proc. SPIE Conf. 4518: Multimedia Systems and Applications IV*, Aug. 2001, pp. 166–176.

[2] Y. Q. Shi and H. Sun, *Image and Video Compression for Multimedia Engineering: Fundamental, Algorithm, and Standards*. Boca Raton, FL: CRC Press LLC, 1999.

[3] N. Damera-Venkata, T. D. Kite, W. S. Geisler, B. L. Evans, and A. C. Bovik, "Image quality assessment based on a degradation model," *IEEE Trans. Image Processing*, vol. 9, no. 4, pp. 636–650, Nov. 2000.

[4] S. A. Karunasekera and N. G. Kingsbury, "A distortion measure for blocking artifacts in images based on human visual sensitivity," *IEEE Trans. Image Processing*, vol. 4, no. 6, pp. 713–724, June 1995.

[5] P. C. Teo and D. J. Heeger, "Perceptual image distortion," in *Proc. IEEE International Conference on Image Processing*, vol. 2, Nov. 1994, pp. 982–986.

[6] H. S. Baird, "Document image defect models and their uses," in *Proc. IAPR Second International Conference on Document Analysis and Recognition*, Tsukuba Science City, Japan, Oct. 1993, pp. 62–67.

[7] T. Ho and H. S. Baird, "Perfect metrics," in *Proc. IAPR Second International Conference on Document Analysis and Recognition*, Tsukuba Science City, Japan, Oct. 1993, pp. 593–597.

[8] A. J. Baddeley, "An error metric for binary images," in *Robust Computer Vision: Quality of Vision Algorithms*, W. Forstner and S. Ruwiedel, Eds. Karlsruhe, 1992, pp. 59–78.

[9] M. Wu, E. Tang, and B. Liu, "Data hiding in digital binary image," in *Proc. IEEE International Conference on Multimedia and Expo*, vol. 1, New York City, July 31 to August 2, 2000, pp. 393–396.