

Distance-Reciprocal Distortion Measure for Binary Document Images

IEEE Signal Processing Letters,
vol.11, No.2, Feb. 2004

CMLAB

National Taiwan University
CMLAB, since 1991

I. Introduction

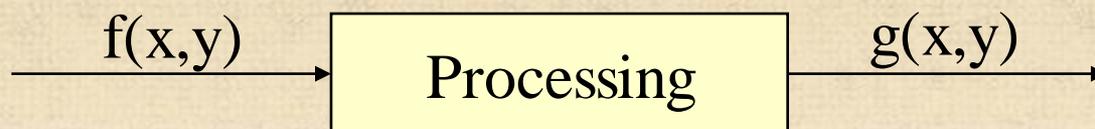
- It is inevitable to introduce visible artifacts when a Binary Image has been watermark embedded or data hidden. Therefore, an effective visual distortion measure is a must for performance comparison or evaluation of such an application.



- Subjective measure:
 - costly but important since a human is the ultimate viewer!
- Objective measure:
 - repeatable and easier to implement.
- Objective measure does not always agree with the subjective one
- → Objective measure based on Human Visual Perception Model is the preferred one.



- The peak signal-to-noise ratio (PSNR) is a popular distortion measure used in image and video processing.



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

- Where M and N are the dimensions of the image, and P is the maximum peak-to-peak signal swing, e.g., $P=255$ for 8-bit image



- For Binary Images, 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



(a)



(b)



	Natural Images	Binary Images
levels	Continuous multiple-tune	Two level
objects	Physical objects	Invented symbols and/or signals (such as documents)

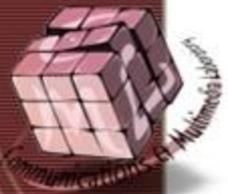


- The perception of distortion in (document) binary images is very different from that in natural images
- In a particular “Language”, such as English, people know very well what a certain alphabetic character should look like.
- →
 - (i) Distortion in document images could be more obtrusive than distortion in natural images
 - (ii) The distortion measures proposed for color/gray-level images are not often applicable to binary images



II. Distance Reciprocal Distance Measure

- A number of single-letter images are used to study distortion in binary document images.
- Each single-letter image is converted from a letter typed in Microsoft 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).



Observation

- For a binary document image, the distance between two pixels play a major role in their mutual interference perceived by human eyes.
- Since readers are so familiar with alphabetic characters that even a single-pixel distortion can be perceived easily.
- →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 the field of view of the pixel is 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.



- The distortion of a processed image $g(x,y)$ compared with the original image $f(x,y)$ is measured by using a weighted matrix with each of its weights determined by the reciprocal of a distance measured from the center pixel. → Distance-Reciprocal Distortion Measure.
- Assume the weight matrix W_m is of size $m \times m$, where $m = 2n + 1$ and $n \in \mathbb{N}$. The center element of this matrix is at (i_c, j_c) where $i_c = j_c = (m + 1)/2$



$$W_m(i, j) = \begin{cases} 0, & \text{for } i = i_c \text{ and } j = j_c \\ \frac{1}{\sqrt{(i-i_c)^2 + (j-j_c)^2}}, & \text{otherwise.} \end{cases} \quad (2)$$

- The normalized weight matrix is defined as

$$W_{N_m}(i, j) = \frac{W_m(i, j)}{\sum_{i=1}^m \sum_{j=1}^m W_m(i, j)} \quad (3)$$



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



- Suppose that there are S flipped pixels in $g(x,y)$. Each pixel will have a distance-reciprocal distortion DRD_k , $k=1,2,\dots,S$.
- For the k -th flipped pixel at $(x,y)_k$ in the output image $g(x,y)$, the resulted distortion is calculated from an $m \times m$ block B_k in $f(x,y)$ that is centered at $(x,y)_k$.
- The Distortion DRD_k measured for this flipped pixel $g[(x,y)_k]$ is given by

$$DRD_k = \sum_{i,j} [D_k(i,j) \times W_{N_m}(i,j)] \quad (4)$$



- Where the (i,j) -th element of the different matrix D_k is given by

$$D_k(i, j) = |B_k(i, j) - g[(x, y)_k]| \quad (5)$$

- Thus, DRD_k equals to the weighted sum of the pixels in the block 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 W_{N_m} 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

$$DRD = \frac{\sum_{k=1}^S DRD_k}{NUBN} \quad (6)$$



- Where $NUBN$ is to estimate the valid (nonempty) 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 numbers $M \times N$ 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.



III. Experimental Results

- The original image is of size 198x109 and there are 122 ($\approx 39\%$) non-uniform 8x8 blocks out of total 312 blocks

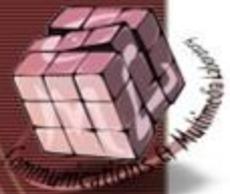
These *are* testing.
Thank *You!*
12.48, 3576?



- The design criteria for generating independent test images:
 - the number of flipped pixels is the same in each test image, test image generated through this distortion generator should have a wide variety in term of “how noticeable the flipping is”
- There are 1763 black pixels in Fig.2.
- We flip 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) Forty black pixels out of 1763 are randomly chosen using a random number generator with uniform distribution.
- 3) For each black pixel, one pixel is flipped in its neighboring area. As shown in Fig. 3, the pixel to be flipped is randomly selected from the Band 1 pixel (the black pixel itself), or eight Band 2 pixels, or 16 Band 3 pixels, with probability of p_1 , p_2 , and p_3 , respectively, and $p_1 + p_2 + p_3 = 1$. For the Band 2 and Band 3 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 p_1 , p_2 , and p_3 , with $p_3=0, 0.2, 0.4, 0.6, 0.8$ and 1 , $p_1=(1-p_3)/10$, and $p_2=9 \cdot p_1$.
- 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.



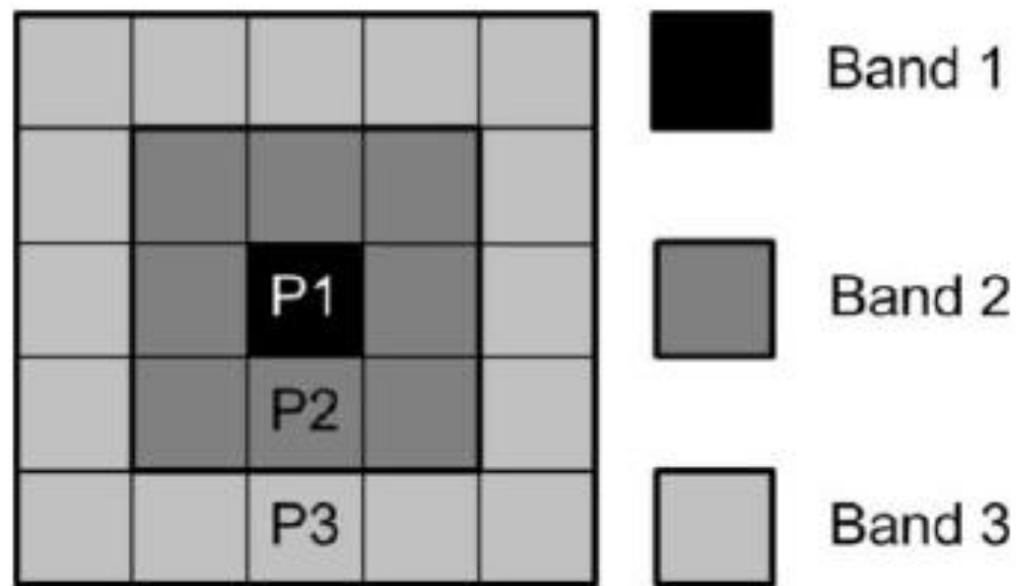
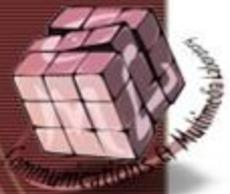


Fig. 3. Distortion generation for subjective testing.



- 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 27.33 dB according to (1).
- One set of the test images generated is shown in Fig.4.



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

(a)

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

(b)

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

(c)

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

(d)

Fig. 4. One set of test images. (a) PSNR = 27.32 dB, DRD = 0.1869. (b) PSNR = 27.32 dB, DRD = 0.1557. (c) PSNR = 27.32 dB, DRD = 0.2098. (d) PSNR = 27.32 dB, DRD = 0.2461.



- 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. The observers are asked 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 4 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.

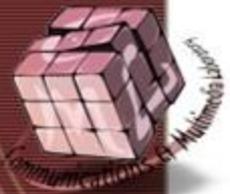


TABLE II
EXPERIMENTAL RESULTS

Test Images	Mean Subjective Rank	PSNR (dB)	Average <i>DRD</i> (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



- The distribution of the subjective ranking scores for each group is shown in a subfigure in Fig.5.
- In each subfigure, the abscissa represents 4 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 4 sets of test images, there are 240 scores in total for each group.



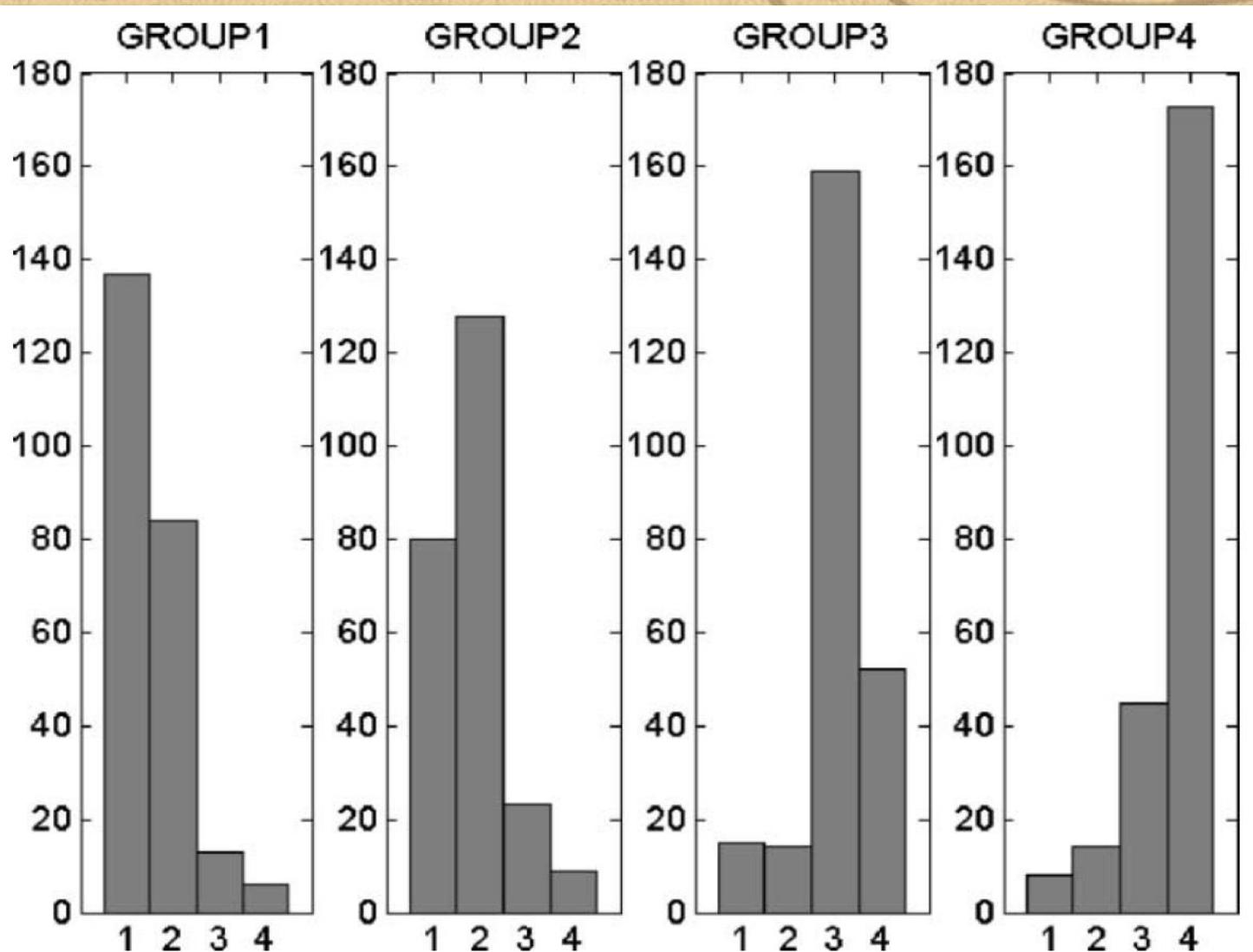


Fig. 5. Distribution of subjective ranking scores.

