Watermarking of Dither Halftoned Images

Zachi Baharav, Doron Shaked
HP Laboratories Israel*
HPL-98-32
May, 1999

[zachi,dorons]@hpli.hpl.hp.com

watermarking,
halftone printing,
printing security,
dithering

This report concerns embedding of watermarks into dither halftoning. The basic idea is to use a sequence of two dither matrices (instead of one) to encode the watermark. Assuming a specific distribution of the input image leads to an optimal decoding algorithm, and an optimal dither matrix pair matching. Finally, an example is given, which demonstrates the ideas. The example is synthetic in the sense that it does not use printing and re-scanning of the image.

*HP Laboratories Israel, Technion City, Haifa 32000, Israel

1 Introduction

Watermarking of images has gained a lot of interest in the last few years [1, 2]. This is mainly due to the ease with which digital images are transfered through the Internet. Embedding a watermark in an image (also called watermarking), means embedding information in the image, in a manner which does not deteriorate the image quality to the viewer. Usually there are additional requirements from the watermark [1], like immunity to tempering, retrieval ease, and difficult duplication. Possible applications of watermarking are tracing of illegal image documents, authentication, etc.

A few works deal with watermarking of printed material, e.g. [3], where watermarking of printed text is discussed. The watermarking is done by modifying spaces in the printed output, like changing spaces between lines, words etc. However, this method will not work for printed images, since images are composed of pixels rather than lines of characters.

In this report we deal with watermarking of printed images. Printing an image is different from printing a text, since a halftoning process is involved, which converts an input image of continuous gray levels to a binary image of black and white. A similar work, dealing with watermarking of hard-copy images, is [4], however the solution there is different.

The problem of watermarking of hard-copy images can be formulated as follows: Given an input image and a watermark bit-stream, produce a halftoned version of the input image, with a seamless watermark.

In this report only a specific halftoning method is considered, that of dithering [5]. In dithering, each pixel in the input image is compared to a threshold in a dither-matrix, and is rendered accordingly. Usually one dither matrix of small size, say 16×16 pixels, is used, and duplicated to tile the whole image. The idea presented in this paper is to use a sequence of two (or more) different dither matrices in the halftoning process to encode the watermark. When two matrices are used their order is the binary representation of the watermark.

We mention in this stage that a crucial problem, not discussed in this report, is the problem of designing two dither matrices, that will tile seamlessly. Dither matrices are designed to be self-compliant in the sense that the matrix is assumed to be on a torus [9]. One possibility is to modify design procedures, so as to produce two dither matrices which are checked to be compliant. Another alternative might be the Quad-Fly algorithm [10], which is capable of producing numerous compliant dither matrices.

The organization of the report is as follows: Section 2 describes the terms, and formulates the problem as a communication problem. In Section 3 the image pixels are assumed to be independent and identically distributed (i.i.d.), and a theoretical analysis is made which leads to a closed form decoding algorithm. In Section 4 a characterization of an optimal matrix pair, in terms of the above decoding algortihm, is given. Finally, an example is given in Section 5. Summary and future directions are given in Section 6.

2 Problem Formulation

In order to facilitate notation, we order the image pixels as a one-dimensional vector. This is merely to avoid the double index notation in the sequel. In the same manner, the dither matrices are ordered to form one-dimensional vectors.

The problem of embedding a watermark in the halftoning process can be formulated as shown in Fig. 1:

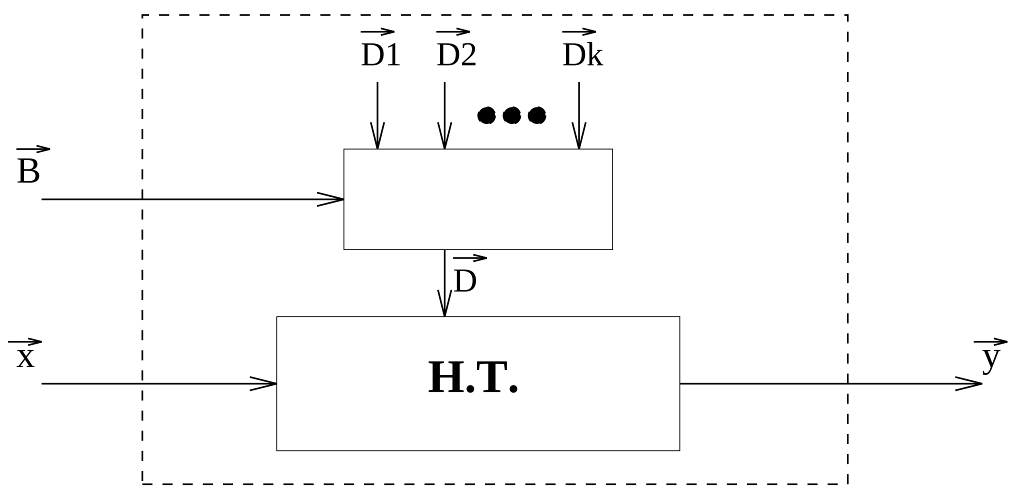


Figure 1: Schematic description of the encoding process.

- $\vec{x}$ - Input contone image, of size L:

$$\vec{x} = \{x_i\}_{i=1}^{L} \quad , \quad x_i \in [0,1] \quad . \tag{1}$$

- $\vec{y}$ - Output halftone image.

$$\vec{y} = \{y_i\}_{i=1}^{L} \qquad , \qquad y_i \in \{0,1\} \quad . \tag{2}$$

- $\vec{B}$ - The watermark data. We assume K possible symbols, and that there are M symbols to be marked:

$$\vec{B} = \{B_i\}_{i=1}^{M} \qquad , \qquad B_i \in \{1,\ldots,K\} \quad . \tag{3}$$

- $\vec{D}^1, \vec{D}^2 \ldots \vec{D}^K$ - Dither vectors corresponding to different symbols. The size of the dither matrices is denoted as N:

$$\vec{D^j} = \left\{D_i^j\right\}_{i=1}^{N} \qquad j = 1:K, \qquad D_i^j \in [0,1) \quad . \tag{4}$$

- $\vec{D}$ - The dither vector as selected according to $\vec{B}$. The selection can be changed once every N input-elements have been dithered.

$$\vec{D} \in \left\{\vec{D^j}\right\}_{j=1}^{K} \quad . \tag{5}$$

In what follows, we assume a binary watermark ($K = 2$) with equal symbol probability, and an input image $\vec{x}$ of size $L = N$, namely $M = 1$ (since we use $L = M \cdot N$).

The output halftone image $\vec{y}$ is determined according to the following rule:

$$y_i = \begin{cases} 0 & \text{if } x_i \leq D_i \ , \\ 1 & \text{if } x_i > D_i \ . \end{cases} \tag{6}$$

Now that we have seen how the output vector $\vec{y}$ is created, we can make the following observation about the dither matrices. Since the average output value should be similar to the average input value (see also discussion of mean-value reproduction rule in [8]), the dither matrix elements are permutation of a single set, namely

$$\forall j = 1 \ldots K \quad , \quad \vec{D^j} = \text{Permutation of } \left\{\frac{i}{(N+1)} \quad , \quad i = 1 \ldots N\right\} \quad . \tag{7}$$

One last word about notations: The notation $\Pr(\cdot)$ is context dependent and is used to describe both the probability of an event (discrete-case, e.g. y_i), and the probability-distribution-function of a random variable (continuous-case, e.g. x_i). The context will be clear from the discussion in the text.

2.1 Formulation as a communication channel

The following interpretation of the process as a communication channel is useful, and enables performance analysis of the system. In this model the data to be transmitted is $\vec{B}$, and the noise is created by the halftoning process, as induced by the input image $\vec{x}$. This is shown schematically in Fig. 2. The problem at hand is to find the best receiver for extracting the information.

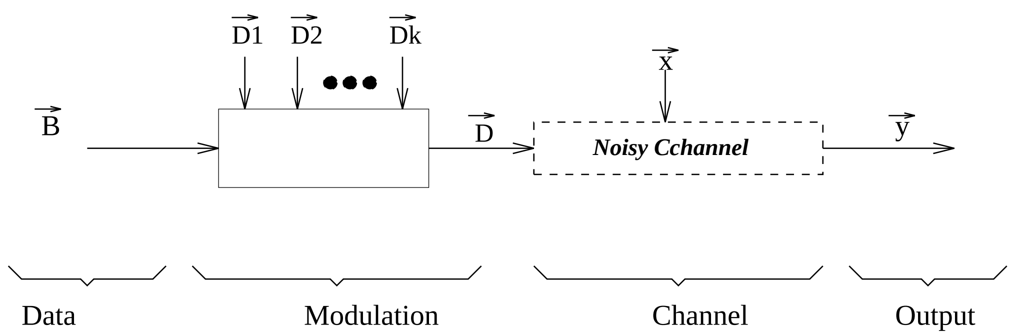


Figure 2: Schematic description of the watermarking process as a communication channel.

Using a MAP (Maximum a Posteriori) based receiver, the decision rule is [6]:

$$\Lambda(\vec{y}) = \frac{\Pr(\vec{y}/\vec{D}^1)}{\Pr(\vec{y}/\vec{D}^2)} \quad \begin{cases} > 1 & \Rightarrow \text{Decode } \vec{D}^1 \ , \\ < 1 & \Rightarrow \text{Decode } \vec{D}^2 \ . \end{cases} \tag{8}$$

(Cases of $\Lambda(\vec{y}) = 1$ can be decided randomly).

Let us now define the error 'e' as either '1' or '0', describing whether an error has occurred or not. The error probability, incurred by this decision rule, can be written as

$$\Pr(e = 1/\vec{y}) = \frac{\min\left\{\Pr(\vec{y}/\vec{D}^1)\, ,\, \Pr(\vec{y}/\vec{D}^2)\right\}}{\Pr(\vec{y}/\vec{D}^1)\, +\, \Pr(\vec{y}/\vec{D}^2)} \quad . \tag{9}$$

The terms in the likelihood ratio can be written explicitly, as follows:

$$\Pr(\vec{y}/\vec{D}^1) = \int_{\vec{x}} \Pr(\vec{y}/\vec{D}^1\vec{x}) \Pr(\vec{x})\, d\vec{x} \ . \tag{10}$$

The probability $\Pr(\vec{y}/\vec{D}^1\vec{x})$ is either 1 or 0, as can be seen from (6):

$$\Pr(\vec{y}/\vec{D}^1\vec{x}) = 1 \qquad \text{iff the following holds}$$

$$\forall i = 1, \ldots, N \qquad \begin{cases} (y_i = 0) & \text{and } x_i \in [0, D_i^1] \\ (y_i = 1) & \text{and } x_i \in (D_i^1, 1] \end{cases} \ . \tag{11}$$

Inserting it into (10) leads to:

$$\Pr(\vec{y}/\vec{D}^1) = \left| \int_{x_N = D_N^1}^{y_N} \cdots \int_{x_1 = D_1^1}^{y_1} \Pr(\vec{x})\, d\vec{x} \right| \ . \tag{12}$$

Equation (12) characterizes the communication channel described in Fig. 2. In the sections to follow, we assume a specific probability distribution of $\vec{x}$, and describe in a closed form the decoding algorithm, evaluate the error-probability, and characterize the pair of dither matrices which minimize it.

3 Optimal decoding scheme

In this section we assume that $\vec{x}$ is composed of N independent random variables, each with a uniform distribution in $[0, 1]$. Therefore, the probability of $\vec{x}$ is

$$\Pr(\vec{x}) = \prod_{i=1}^{N} \Pr(x_i) = 1 \ , \tag{13}$$

The first equality follows from the i.i.d. of $\vec{x}$, and the second follows from the fact that x_i is uniformly distributed as described above.

Substituting into (12), we arrive at

$$\Pr(\vec{y}/\vec{D}^1) = \left| \int_{x_N = D_N^1}^{y_N} \Pr(x_N)\, dx_N \cdots \int_{x_1 = D_1^1}^{y_1} \Pr(x_1)\, dx_1 \right| = \prod_{i=1}^{N} \left| y_i - D_i^1 \right| \ . \tag{14}$$

Another helpful point-of-view is the geometric one, where each of $\vec{D}^1$, $\vec{D}^2$, and $\vec{y}$, is described as a point in R^N. Moreover, $\vec{y}$ is a vertex of the unit cube $[0,1]^N$, and $\vec{D}^1, \vec{D}^2$ are interior points in that cube. The decision rule then reduces to selecting the dither-matrix-point 'closest' to $\vec{y}$, where the distance $d\left(\vec{y}, \vec{D}\right)$ is defined as

$$d(\vec{y}, \vec{D}) \triangleq \prod_{i=1}^{N} |y_i - D_i| \quad , \tag{15}$$

namely the N-dimensional volume of the sub-cube defined by $\vec{y}$ and $\vec{D}$.

4 Matrix pair characterization

In this section we define an error criterion, and characterize the pair of matrices $\vec{D}^1, \vec{D}^2$ which minimize it. Recall that the two matrices are permutation of the same set, and are therefore permutations of each other. Therefore, the characterization describes the permutation relating the two matrices.

Let $\vec{D}^1$ be a dither vector,

$$D_i^1 = \text{Permutation of} \left\{ \frac{i}{(N+1)} \quad , \quad i = 1 \ldots N \right\} \quad . \tag{16}$$

Let the output vector be denoted as $\vec{y_j} \in \{0,1\}^N$, and let J denote the set of 2^N such vectors.

Let us now compute the average error:

$$\begin{aligned} E\{e\} &= \sum_{e=\{0,1\}} e \cdot \Pr(e) && (17)\\ &= \sum_{e=\{0,1\}} e \cdot \sum_{j \in J} \Pr(e/\vec{y_j}) \Pr(\vec{y_j}) && (18)\\ &= \sum_{j \in J} \Pr(e = 1/\vec{y_j}) \Pr(\vec{y_j}) \quad . && (19)\end{aligned}$$

Writing explicitly,

$$\Pr(\vec{y_j}) = \Pr(\vec{y_j}/\vec{D}^1) \Pr(\vec{D}^1) + \Pr(\vec{y_j}/\vec{D}^2) \Pr(\vec{D}^2) \tag{20}$$

$$= \frac{1}{2}\left[\Pr(\vec{y_j}/\vec{D}^1) + \Pr(\vec{y_j}/\vec{D}^2)\right] \quad . \tag{21}$$

Substituting equations (9) and (21) into (19), we get

$$\begin{aligned} E\{e\} &= \sum_{j\in J} \frac{\min\left\{\Pr(\vec{y_j}/\vec{D}^1)\,,\,\Pr(\vec{y_j}/\vec{D}^2)\right\}}{\left[\Pr(\vec{y_j}/\vec{D}^1)\,+\,\Pr(\vec{y_j}/\vec{D}^2)\right]} \cdot \frac{1}{2}\left[\Pr(\vec{y_j}/\vec{D}^1)\,+\,\Pr(\vec{y_j}/\vec{D}^2)\right] && (22) \\ &= \frac{1}{2}\sum_{j\in J} \min\left\{\Pr(\vec{y_j}/\vec{D}^1)\,,\,\Pr(\vec{y_j}/\vec{D}^2)\right\} && (23) \\ &= \frac{1}{2}\sum_{j\in J} \min\left\{d(\vec{y_j},\vec{D}^1)\,,\,d(\vec{y_j},\vec{D}^2)\right\} \quad , && (24) \end{aligned}$$

where the last equality holds, using (15), for the i.i.d. case.

Claim 1 *The permutation $\vec{D}^2 = \mathcal{P}(\vec{D}^1)$ for which*

$$D_i^2 = 1 - D_i^1 \qquad\qquad \forall i \;\; , \tag{25}$$

minimizes $E\{e\}$.

Proof of Claim 1

Before dwelling on the proof, we need the following Lemma.

Lemma 1

$$d(\vec{y_j},\vec{D}) = d\left(\mathcal{P}(\vec{y_j}),\mathcal{P}(\vec{D})\right) \qquad \forall j \quad \textit{and} \quad \forall \mathcal{P} \;\; . \tag{26}$$

Proof of Lemma 1 *The proof of the lemma is immediate due to the fact that $d(\cdot,\cdot)$ is a separable operation, as described in (15). Thus, a permutation only changes the multiplication order in (15).* ∎

To simplify notations, let us note that the permutation described in (25) is self invertible, namely $\vec{D}^1 = \mathcal{P}(\vec{D}^2)$ and $\vec{D}^2 = \mathcal{P}(\vec{D}^1)$.

Following the lemma, rather than looking for a permutation which minimizes (24), we minimize

$$\sum_{j \in J} \min\left\{d(\vec{y}_j, \vec{D}^1)\, ,\, d(\mathcal{P}(\vec{y}_j), \vec{D}^1)\right\} \quad . \tag{27}$$

Now, let us examine the term

$$d_j \triangleq d(\vec{y}_j, \vec{D}^1) \quad , \tag{28}$$

as a function of j. There are 2^N vertices $\vec{y}_j$. There are therefore 2^N values which d_j can acquire (some of which may be equal). Let us order these values in a descending order. Each of these values appears twice in

$$\min\left\{d(\vec{y}_j, \vec{D}^1)\, ,\, d(\mathcal{P}(\vec{y}_j), \vec{D}^1)\right\} \quad , \tag{29}$$

once as the value of $d(\vec{y}_j, \vec{D}^1)$ for some j, and once as the value of $d(\mathcal{P}(\vec{y}_j), \vec{D}^1)$ for a possibly different j. Thus, (27) is minimal iff none of the $\frac{2^N}{2}$ highest values of d_j is the result of (29). Therefore, $\mathcal{P}$ should associate vertices which lead to one of the $\frac{2^N}{2}$ top values of d_j, with vertices which leads to one of the $\frac{2^N}{2}$ lower values of d_j.

It is important to note that not every association of vertices is a valid permutation. For example, there is no permutation that associates vertex $\{1, 1, 1, 1\}$ to $\{0, 0, 0, 0\}$.

Lemma 2 *The association of vertices*

$$\vec{y}_j \Longleftrightarrow \vec{1} - \vec{y}_j \quad , \tag{30}$$

where $\vec{1}$ is a vector of all ones, leads to the following pair, $d(\vec{y}_j, \vec{D})$ and $d\left(\vec{1} - \vec{y}_j, \vec{D}\right)$, where one of them is from the lower half values of d_j, and the second is from the top half values of d_j.

Proof of Lemma 2 *The proof is immediate since from (15)*

$$d(\vec{y}_j, \vec{D}) \cdot d\left(\vec{1} - \vec{y}_j, \vec{D}\right) = Const \qquad \forall j \in J \quad . \tag{31}$$

■

To prove the claim, it is sufficient to show that using $\mathcal{P}$ in (25), and given a vertex j, the following holds

$$d\left(\mathcal{P}(\vec{y}_j), \vec{D}\right) = d\left(\vec{1} - \vec{y}_j, \vec{D}\right) \quad . \tag{32}$$

We start by computing the left hand-side of (32):

$$\begin{aligned} d\left(\mathcal{P}(\vec{y}_j), \vec{D}\right) &= d\left(\vec{y}_j, \mathcal{P}(\vec{D})\right) & (33) \\ &= \prod_{i/\vec{y}_j(i)=0} (1 - D_i) \cdot \prod_{i/\vec{y}_j(i)=1} D_i \quad . & (34) \end{aligned}$$

And the right hand-side is

$$d\left(\vec{1} - \vec{y}_j, \vec{D}\right) = \prod_{i/\vec{y}_j(i)=0} (1 - D_i) \cdot \prod_{i/\vec{y}_j(i)=1} D_i \quad , \tag{35}$$

and this completes the proof. ■

Evaluating the average error for various dither matrix sizes, N, gives a Rate-Distortion graph, since the size of the dither matrix is inversely proportional to the number of symbols we can encode in a given image. The rate-distortion graph, for a pair of matrices determined according to (25), is described in Fig. 4. It is important to recall that N is the total size of the dither matrix. Therefore, a value of $N = 3$, amounts for a dither matrix of size 1×3.

It is interesting to note that using the dither matrix pair according to (25), leads to an equal probability of y_i being either $'1'$ or $'0'$. Explicitly:

$$\begin{aligned} \Pr(y_i = 0) &= \Pr(y_i = 0/\vec{D}^1)\Pr(\vec{D}^1) + \Pr(y_i = 0/\vec{D}^2)\Pr(\vec{D}^2) \\ &= D_i^1 \frac{1}{2} + D_i^2 \frac{1}{2} = \frac{1}{2} \quad . & (36) \end{aligned}$$

5 Numerical examples

We start by describing the exact process by which the watermark was implemented, and then analyze the results.

An 8 bit gray level image of size 1230×726 pixels was used as a source. This image was dithered using dither-matrices of size 256×256. The dither employed here was random

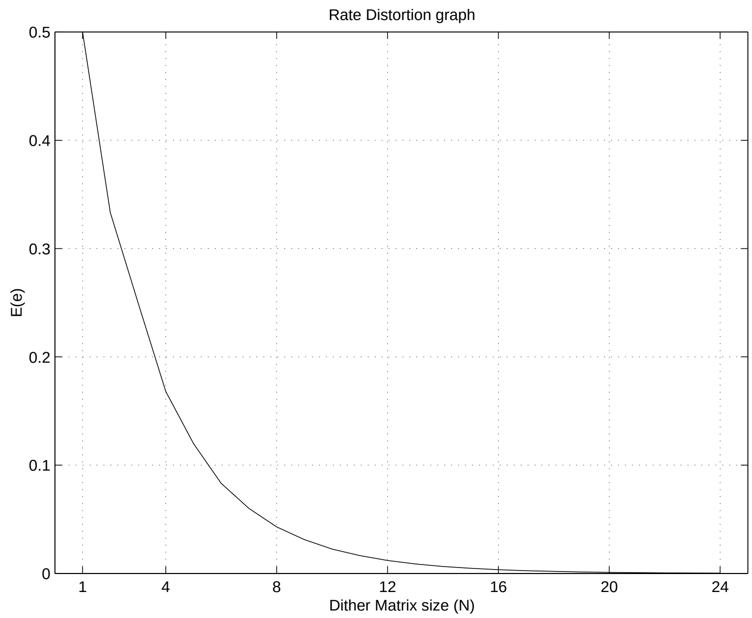


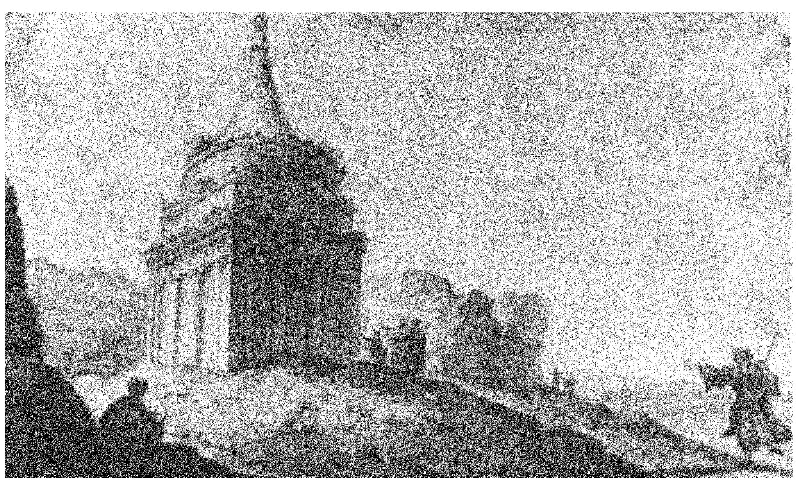
Figure 3: Rate-Distortion graph. The abscia describes the dither matrix size, N, which is inversely proportional to the number of symbols one can encode in a given image, and the ordinate is the average error.

dither, which is known to be inferior to other dither matrices [5, 9]. It was preferred due to the fact that currently there is no easy method for designing two dither matrices which tile seamlessly. For the case of halftoning with no watermarking, one dither matrix was used. For watermarking, its optimal pair (25) was employed as a second dither matrix. A total of 15 symbols were embedded in the image. The resulting halftoned image was then processed to extract the watermark. *No printing* and *no scanning* were involved[1].

The resulting images appear in Figure 4. In this case (as expected by the large size of the dither-matrices), all the 15 bits of data were extracted with no error. The rate-distortion performance will have to be examined much farther if a practical system is to be involved, which means having to cope with noise due to the printing and scanning of the image.

[1] The halftoned image $\vec{y}$ was written to a file by the watermark encoder, and then read by the watermark decoder.

(a)



(b)

Figure 4: (a)Halftoned image, no watermarking. (b)Halftoned image with watermark, using two dither matrices.

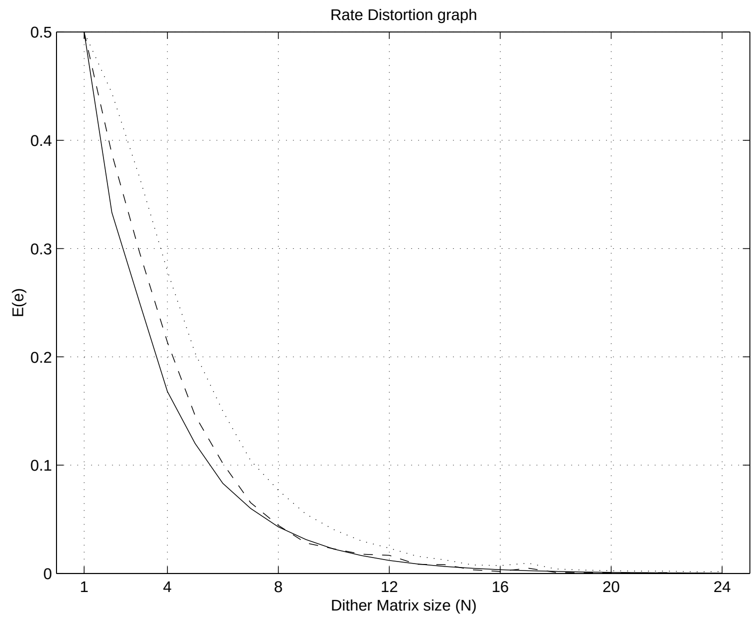


Figure 5: Rate-Distortion graph. The abscia describes the dither matrix size, N, which is inversely proportional to the number of symbols one can encode in a given image, and the ordinate is the average error. Three cases are considered: (a) Theoretical minimum for the i.i.d. case (solid-line). (b) Performance on a real-image, using the optimal permutation as in (25) (dashed-line). (c) Performance on a real-image, using two randomly permuted matrices (dotted-line).

To examine the importance of the best-permutation as compared to random-permutation, we compared the performance of two randomly generated matrices, to that of an optimal pair according to (25). The rate-distortion, mentioned in Sec. 4, was measured on a real-image, as described in Fig. 5. The curve for randomly chosen permutations was obtained by averaging the error over a very large ensemble of randomly selected permutations. The advantage of using the optimal permutation pair is evident from the graph.

6 Summary and Future Directions

This work introduced the idea of embedding a watermark in an halftoned image. The specific halftoning process discussed was the dithering method. The basic idea is to use different dither matrices to encode different symbols. An i.i.d. image model was analyzed, leading to a closed form decoding algorithm. Minimizing the corresponding average error, resulted in a characterization of optimal dither-matrix pair. Finally, numerical examples were given, which demonstrate the method.

Future directions include both theoretical extensions, and practical implementation considerations:

1. Full system considerations - A full system will include the scanning of a printed image, processing the image, and then extraction of the data. This system will depend heavily on the type of printer involved, and on the scanner quality.

2. Design of dither-matrices - The problem of designing dither matrices which tile seamlessly should be faced.

3. Theoretical analysis of measurement noise.

4. Introduction of error-correcting codes is important in order to cope with measurement noise. This will have to be done while keeping in mind the typical measurement errors that may occur.

5. Extending the characterization of optimal matrix-pair to a metric on dither matrices. This will enable us to use more than two-matrices for the encoding process. Of course, this definition will depend on the assumption as to the probability distribution of the input image, and on the error-definition.

Acknowledgments

We would like to thank Prof. Nery Merhav for pointing out the interpretation of the communication channel. Prof. Merhav is with the faculty of engineering, Technion-Israel Institute of Technology, and a HPLI consultant.

7 References

[1] W. Bender, D. Gruhl, N. Morimoto, and A. Lu, *Techniques for Data Hiding*, IBM Systems Journal, Vol. 35, Nos 3&4, 1996.

[2] I.J. Cox, M.L. Miller, *Review of watermarking and the importance of perceptual modeling* , Proceedings of the SPIE - The International Society for Optical Engineering, vol.3016, p. 92-9, 10-13 Feb. 1997, San Jose, CA, USA;

[3] J. T. Brassil, S. Low, N. F. Maxemchuk, L. O'Gorman, *Electronic marking and identification techniques to discourage document copying*, IEEE J. on Selected Areas in Communications, Vol. 13, No. 8, October 1995.

[4] K. Matsui, K.Oka, *Embedding a watermark to binary pictures in a hardcopy system*, Memoirs of the National Defense Academy, vol.36, no.2, p. 13-20, Natl. Defense Acad., March 1997, Japan (in English).

[5] R. Ulichney, *Digital Halftoning*, Cambridge, Ma: MIT Press,1987.

[6] R. O. Duda, P. E. Hart, *Pattern classification and scene analysis*, New York, John Wiley & Sons, Inc.

[7] N. Arad, D. Shaked, Z. Baharav, and Q. Lin, *Barycentric Screening: Dithering in the World of Color*, HPL Technical Report HPL-97-103, August 1997.

[8] D. Shaked, Z. Baharav, and Q. Lin, *G/M Dither or Converting a B/W Dither Matrix to Color Dither*, HPL Technical Report HPL-97-113, September 1997.

[9] Q. Lin, Super Smooth Dither, HPL Technical Report HPL-94-64, 1994.

[10] M. McGuire, *Halftone Dithering with and without a Mask (Quad-Fly)*, Patent application, HPL docket # 10960209-1, 1997.