ZERO-ERROR INFORMATION HIDING CAPACITY OF DIGITAL IMAGES

Ching-Yung Lin[*] *and Shih-Fu Chang*[**]

[*]: IBM T. J. Watson Research Center, 30 Saw Mill River Rd., Hawthorne, NY 10532
[**]: Department of Electrical Engineering, Columbia University, New York, NY 10027
cylin@watson.ibm.com, sfchang@ee.columbia.edu

ABSTRACT

In this paper, we derive a theoretical capacity for digital image watermarking with zero transmission errors. We present three different discrete memoryless channel models to represent the watermarking process. Given the magnitude bound of noise set by applications and the acceptable watermark magnitude determined by the just-noticeable distortion, we estimate the zero-error capacity by applying Shannon's adjacency-reducing mapping technique. The capacity we estimate here corresponds to a deterministic guarantee of zero error, different from the traditional theorem approaching zero error asymptotically.

1. INTRODUCTION

In watermarking schemes, multimedia data is considered as a communication channel to transmit messages. An important theoretical issue of watermarking is: how much information can be reliably transmitted as watermarks without causing noticeable quality losses?

Theoretical capacity issues of digital watermarking have not been fully understood. Most of the previous works on watermarking capacity [1, 8, 9] directly apply Shannon's well-known channel capacity bound,

$$C = \frac{1}{2} log_2(1 + \frac{P}{N}), \tag{1}$$

which provides a theoretic capacity bound of an analog-value time-discrete communication channel in a static transmission environment, *i.e.*, where the (codeword) signal power constraint, P, and the noise power constraint, N, are constants [10]. Transmitting message rate at this bound, the probability of decoding error can approach zero if the length of codeword approaches infinite, which implies that infinite transmission samples are expected.

Considering multimedia data, we found there are difficulties if we directly apply Eq. (1). The first is the number of channels. If the whole image is a channel, then this is not a static transmission environment because the signal power constraints are not uniform throughout the pixels, based on the human vision properties. If the image is a composition of parallel channels, then this capacity is meaningless because there is only one or few sample(s) in each channel. The second difficulty is the issue of digitized values in the multimedia data. Contrary to floating point values which have infinite states, integer value has only finite states. This makes a difference in both the applicable embedding watermark values and the effect of noises. The third obstacle is that we will not know how large the watermark signals can be without an extensive study of human vision system models, which is usually ignored in most previous watermarking researches, perhaps because of its difficulties and complexity. The fourth hurdle is related to noise modeling. Despite the existence of various distortion/attack, additive noises might be the easiest case. Other distortions may be modeled as additive noises if the distorted image can be synchronized/registered. There are other issues such as private or public watermarking and questions as to whether noise magnitudes are bounded. For instance, Eq. (1) is a capacity bound derived for Gaussian noises and is an upper bound for all kinds of additive noises. However, in an environment with finite states and bounded noises, transmission error can actually be zero, instead of approaching zero as in Eq. (1). This motivated a research of zero-error capacity initialed by Shannon in 1956 [11]. Quantization, if an upper bound on the quantization step exists, is an example of such a noise. We can find the zero-error capacity of a digital image if quantization is the only source of distortion such as in JPEG.

A braod study of theoretical watermarking capacity based on the above four obstacles can be found in [5]. In [6], we showed the watermarking capacity based on multivariant capacity analysis and four HVS models. In this paper, we focus on the zero-error capacity of digital images. Shannon defined the zero-error capacity of a noisy channel as the least upper bound of rates at which it is possible to transmit information with zero probability of error [11]. In contrast, here we will show that rather than a probability of error approaching zero with increasing code length, the probability of error can be actually zero under the conditions described above. This property is especially needed in applications that no errors can be tolerated. For instance, in multimedia authentication, it is required that no false alarm occur under manipulations such as JPEG compression. In some applications, we need to correctly retrieve all the hidden information in the watermarked image within a pre-selected range of acceptable compression quality factors.

In this paper, we will show that the semi-fragile watermarking method that we proposed in [4] is, in fact, one way of achieving the zero-error capacity. We will also show two sets of curves that represent the zero-error capacity. Although most of our discussion will focus on image watermarking subject to JPEG manipulation, the zero-error capacity we showed here can be applied to other domains as long as the noise magnitude is contrained.

In Section 2, we discuss the meaning and classification of channels in an image. A theoretical derivation of zero-error capacity of a discrete memoryless channel and an image is discussed in Section 3. The capacity curves and some experiments results are shown in Section 4. In Section 5, we show a conclusion of this paper.

2. NUMBER OF CHANNELS IN AN IMAGE

Here we consider the case that the maximal acceptable level of lossy compression is pre-determined. In JPEG, maximum distortion of each DCT coefficient is determined by the quantization step size. Since JPEG uses the same quantization table in all blocks, maximum distortion just depends on the position in the block and is the same for all coefficients from different blocks but at the same position. If we define a pre-selected lower bound of acceptable compression quality factors, then all the quantization step size at any specific position of blocks will be smaller than or equal to the quantization step size from the selected lowest quality factor [4].

Assume a digital image $\mathbf{X}$ has $M \times N$ pixels that are divided into B blocks. Here, in the blocked-based DCT domain, $\mathbf{X}$ may be considered as

- Case 1: a variant-state discrete memoryless channel (DMC). Transmission utilizes this channel for $M \times N$ times.
- Case 2: a product of 64 static-state DMCs, in which all coefficients in the same position of blocks form a DMC. Each channel can be at most transmitted B times. In other words, the maximum codeword length is B for each channel.
- Case 3: a product of $M \times N$ static-state DMCs, in which each coefficient forms a DMC. Each channel can be at most transmitted once.

In most information theory research works, channel is usually considered invariant in time and has uniform power and noise constraint, which is usually valid in communication. Time variant cases have been addressed (*e.g.*, [2]), called Arbitrarily Varying Channel (AVC). However, such a work on AVC may not be adequate to the watermarking problem because the channel does not vary in a statistically arbitrary way. We think that Case 2 is the best candidate for the capacity analysis problem if the image is only manipulated by JPEG. However, assuming no error correction codes are used in this zero-error environment, the codes in Case 2 will be sensitive to local changes. Any local changes may cause loss of the whole transmitted information in each channel. In applications that information bits have to be extracted separately from each block, Case 3 may be the best candidate. For instance, in the authentication case, some blocks of the image may be manipulated. By treating each coefficient as a separate channel (as in Case 3), we can detect such manipulations in a local range.

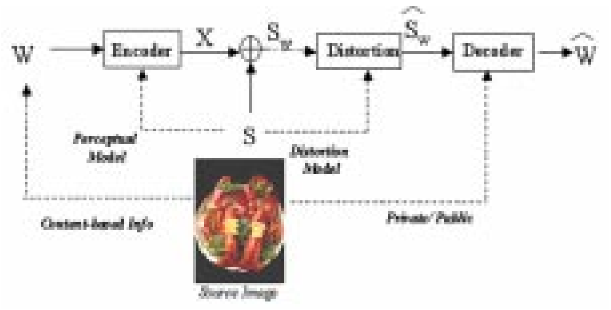

Figure 1: Watermarking: multimedia data as a communication channel

A general watermarking model is shown in Figure 1. Here, a message, W, is encoded to X which is added to the source multimedia data, S. The encoding process may apply some perceptual model of S to control the formation of the watermark codeword X. The resulted watermarked image, S_W, can always be considered as a summation of the source image and a watermark X. At the receiver end, this watermarked image may have suffered from some distortions, *e.g.*, additive noise, geometric distortion, nonlinear magnitude distortion, *etc.* The decoder uses the received watermarked image, $\hat{S}_W$, to reconstruct message, $\hat{W}$. In general, we call the watermarking method "private" if the decoder needs the original source image S, and "public" or "blind" if S is not required in the decoding process. Watermarking capacity refers to the amount of message bits in W that can be reliably transmitted.

3. ZERO-ERROR CAPACITY OF A DISCRETE MEMORYLESS CHANNEL AND A DIGITAL IMAGE

The zero-error capacity of discrete memoryless channel can be determined by applying adjacency-reducing mapping on the adjacency graph of the DMC (Theorem 3 in [11]). For a discrete-value channel, Shannon defined that two input letters are adjacent if there is a common output letter which can be caused by either of these two [11]. Here, in the JPEG cases, a letter means an integer value within the range of the DCT coefficient. Adjacency-reducing mapping means a mapping of letters to other letters, $i \rightarrow \alpha(i)$, with the property that if i and j are not adjacent in the channel (or graph) then $\alpha(i)$ and $\alpha(j)$ are not adjacent. In other words, it tries to reduce the number of adjacent states in the input based on the adjacency of their outputs. Adjacency means that i and j can be mapped to the same state after transmission. We should note that the problem of determining such a mapping function for an arbitrary graph is still wide open. Also, it is sometimes difficult to determine the zero-error capacity of even some simple channels [3].

Fortunately, we can find an adjacency-reducing mapping and the zero-error capacity in the JPEG case. Assume the just-noticeable-change on a DCT coefficient is $\frac{1}{2}Q_w$ [1] and

[1] Note that Q_w can be assumed to be uniform in all coefficients

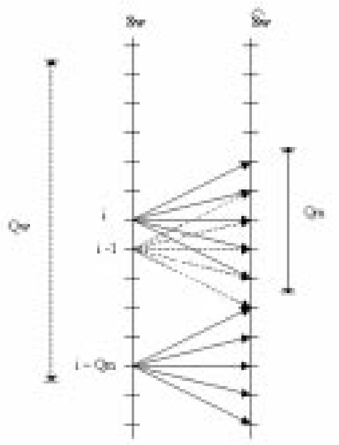

Figure 2: Adjacency-reducing mapping of discrete values given bounded quantization noise

assume the largest applicable JPEG quantization step to this coefficient is Q_m, then the zero-capacity of this channel will be

$$C(Q_w, Q_m) = \log_2(\lfloor \frac{Q_w}{Q_m} \rfloor + 1) \tag{2}$$

Eq. (2) can be proved by using the adjacency-reducing mapping as in [11]. Figure 2 shows an example to reduce adjacency points. Given a Q_m, which is the maximum quantization step that may be applied to the watermarked coefficient S_w, then the possible value $\hat{S}_w$ at the receiver end will be constrained in a range of Q_m possible states. According to Shannon's adjacency-reducing mapping, we can find that the non-adjacent states have to separate from each other for a minimum of Q_m. For instance, assume the source coefficient value is i, then its closest non-adjacency states of i are $i + Q_m$ and $i - Q_m$. To find out the private watermarking capacity, we assume that all the states within the just-noticeable range of $[i - \frac{1}{2}Q_w, i + \frac{1}{2}Q_w)$ are invisible. Therefore, there are Q_w candidate watermarking states in this range. Since we have shown that the non-adjacent states have to separate from each other by Q_m, then there will be $\lfloor \frac{Q_w}{Q_m} \rfloor + 1$ applicable states in the Q_w ranges that can be used to represent information without noticeable change. Therefore, from the information theory, we can get the capacity of this channel in Eq. (2). For instance, in Figure 2, $Q_w = 11$ and $Q_m = 5$. Using Eq. (2), we can obtain the capacity rate to be 1.59 *bits/sample*.

Eq. (2) is a bound for private watermarking with known source values in the receiver. However, in the public watermarking cases, i is unknown at the receiver end. In this case, we can fix the central position of the applicable states in the $\hat{S}_w$ axis. Then, the number of applicable states in the just-noticeable range, $[i - \frac{1}{2}Q_w, i + \frac{1}{2}Q_w)$, will be either $\lfloor \frac{Q_w}{Q_m} \rfloor + 1$ or $\lfloor \frac{Q_w}{Q_m} \rfloor$ if $Q_w \geq Q_m$, or only 1 state if $Q_w < Q_m$. The number of state can be represented as $\lfloor \frac{max(Q_w - Q_m, 0)}{Q_m} \rfloor + 1$. Therefore, we can get the minimum capacity of public watermarking,

$$\tilde{C}(Q_w, Q_m) = \log_2(\lfloor \frac{max(Q_w - Q_m, 0)}{Q_m} \rfloor + 1). \tag{3}$$

In Case 2, information is transmitted through B parallel channels, whose capacity can be summed up [11]. The total zero-error capacity of an image surviving JPEG compression is, therefore,

$$C = \lfloor B \times \sum_{\nu \in V} \tilde{C}_\nu(\mathbf{Qw}, \mathbf{Qm}) \rfloor \tag{4}$$

where V is a subset of $\{1..64\}$. Intuitively, V is equals to the set of $\{1..64\}$. However, in practical situation, even though the changes are all within the JND of each coefficient, the more coefficients changed the more possible the changes are visible. Also, not all the 64 coefficients can be used. We found that $V = \{1..28\}$ is an empirical reliable set that all coefficients are quantized as recommended in the JPEG standard by using some commercial software such as Photoshop and xv.[2] Therefore, we suggest to estimate the capacity based on this subset. An empirical solution of Q_w is Q_{50}, as recommended as invisible distortion bound in the JPEG standard. Although practical invisible distortion bounds may vary depending on viewing conditions and image content, this bound is considered valid in most cases [7]. Figure 3 shows the zero-error capacity of a gray-level 256×256 image.

In Case 3, we want to extract information through each transmission channel. Because the transmission can only be used once in this case, the information each channel can transmit is therefore $\lfloor \tilde{C} \rfloor$. Similar to the previous case, summing up the parallel channels, then we can get the zero-error capacity of public watermarking in Case 3 to be

$$C = B \times \sum_{\nu \in V} \lfloor \tilde{C}_\nu(\mathbf{Qw}, \mathbf{Qm}) \rfloor \tag{5}$$

A figure of Eq. (5) is shown in Figure 4. These bits can be restored independently at each utilized coefficient. In other words, changes in a specific block would only affect its hidden information in that block.

4. FIGURES OF ZERO-ERROR CAPACITY CURVE OF DIGITAL IMAGES

In Figure 3 and Figure 4, we show the zero-error capacity of any 256×256 gray level image. Three different just-noticeable changes in the DCT coefficients are used. The curve $Q_w = Q_{50}$ is the just-noticeable distortion suggested by JPEG. In Figure 3, we can see that if the image is quantized by a JPEG quality factor larger or equal to 75, (*i.e.*, $Q_m \leq Q_{75} = \frac{1}{2}Q_{50}$) then the zero-error capacity of this image is at least 28672 *bits*, which is equal to 28 *bit/block*. We can notice that when $75 < m \leq 72$, the capacity is not zero because some of their quantization steps in the quantization table are still the same as Q_{75}.

in the same DCT frequency position, or they can be non-uniform if we adopt some human perceptual properties. For Case 2, we assume the uniform property, while whether Q_w is uniform or non-uniform does not affect our discussion in Case 3.

[2] Some application software may discard all the {29..64}th DCT coefficients regardless of their magnitudes.

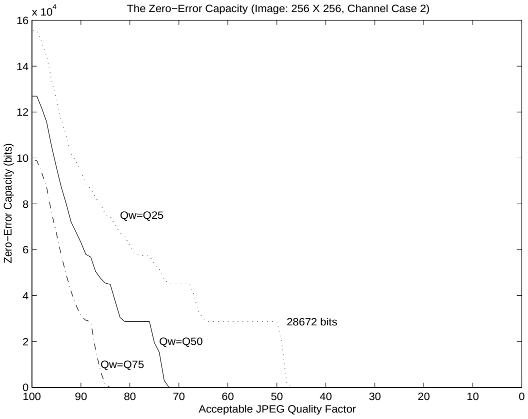

Figure 3: The Zero-Error Capacity of a 256×256 gray-level image for channel case 2

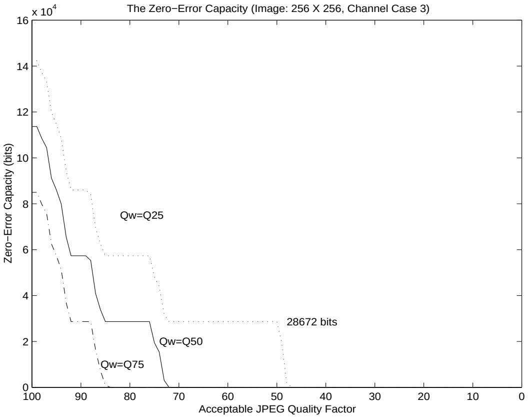

Figure 4: The Zero-Error Capacity of a 256×256 gray-level image for channel case 3

Comparing Eq. (5) with Theorem 1 in [4], we can see the watermarking technique proposed in [4] is one way of utilizing the zero-error capacity. The only difference is that, in [4], we fixed the ratio of $Q_w = 2Q_m$ and embed one or zero bit in each channel.

Our experiments have shown that the estimated capacity bound described in this paper can be achieved in realistic applications. We tested 9 images by embedding 28 bits in each block based on [4]. Given $Q_w = Q_{50}$, these message can be reconstructed without any error if the image is compressed by JPEG with quality factor larger or equal to 75 using xv. Given $Q_w = 2 \times Q_{67}$, these messages can be totally reconstructed after JPEG compression using Photoshop 5.0 quality scale $10 - 4$. Fig. 5 is an example.

5. CONCLUSION

We derived and demonstrated the zero-error capacity for private and public watermarking in environments with magnitude-bounded noise. Because this capacity can be realized without using the infinite codeword length and can actually accomplish zero error, it is very useful in real applications.

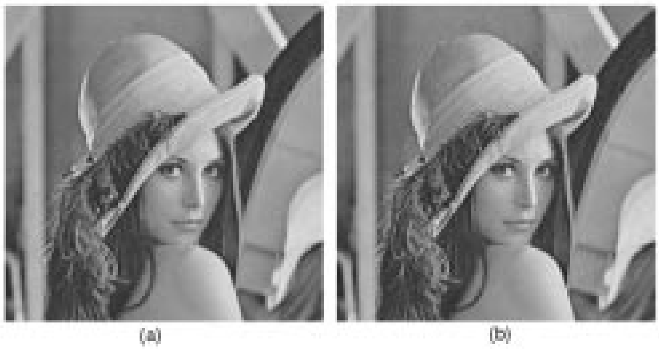

Figure 5: (a) The original 256×256 Lenna image. (b) The watermarked image with 28672 hidden bits that survive JPEG larger than or equal to Quality Factor = 75.

REFERENCES

[1] M. Barni, F. Bartolini, A. De Rosa, and A. Piva, "Capacity of the Watermark-Channel: How Many Bits Can Be Hidden Within a Digital Image?," *Proc. of SPIE*, Vol. 3657, January 1999.

[2] I. Csiszar and P. Narayan, "Capacity of the Gaussian Arbitrarily Varying Channel," *IEEE Trans. on Information Theory*, Vol. 37, No. 1, pp. 18-26, Jan 1991.

[3] J. Korner and A. Orlitsky, "Zero-Error Information Theory," *IEEE Trans. on Information Theory*, Vol. 44, No. 6, Oct 1998.

[4] C.-Y. Lin and S.-F. Chang, "Semi-Fragile Watermarking for Authenticating JPEG Visual Content," *Proc. of SPIE*, Vol. 3971, Jan 2000.

[5] C.-Y. Lin, "Watermarking and Digital Signature Techniques for Multimedia Authentication and Copyright Protection," *Ph.D. Thesis*, Columbia University, 2000.

[6] C.-Y. Lin and S.-F. Chang, "Watermarking Capacity of Digital Images based on Domain-Specific Masking Effects," *IEEE Intl. Conf. on Info. Tech: Coding and Computing*, Las Vegal, Apr 2001.

[7] W. B. Pennebaker and J. L. Mitchell, "JPEG: Still Image Data Compression Standard," *Van Nostrand Reinhold*, Tomson Publishing Inc., New York, 1993.

[8] M. Ramkumar and A. N. Akansu, "A Capacity Estimate for Data Hiding in Internet Multimedia," *Symposium on Content Security and Data Hiding in Digital Media*, NJIT, Jersey City, May 1999.

[9] S. D. Servetto, C. I. Podilchuk and K. Ramchandran, "Capacity Issues in Digital Image Watermarking," *IEEE Intl. Conf. on Image Processing*, Oct 1998.

[10] C. E. Shannon, "A Mathematical Theory of Communication," *Bell System Technical Journal*, Vol. 27, pp. 373-423, 623-656, 1948.

[11] C. E. Shannon, "The Zero-Error Capacity of a Noisy Channel," *IRE Trans. on Information Theory*, IT-2: 8-19, 1956.