# A Capacity Estimation Technique for JPEG-to-JPEG Image Watermarking

Peter H. W. Wong, Member, IEEE, and Oscar C. Au, Senior Member, IEEE

Abstract—In JPEG-to-JPEG image watermarking (J2J), the input is a JPEG image file. After watermark embedding, the image is JPEG-compressed such that the output file is also a JPEG file. In this paper, we use the human visual system (HVS) model to estimate the J2J data hiding capacity of JPEG images, or the maximum number of bits that can be embedded in JPEG-compressed images. Watson's HVS model is modified to estimate the just noticeable difference (JND) for DCT coefficients. The amount of modifications on DCT coefficients is limited by JND in order to guarantee the invisibility of the watermark. Our capacity estimation method does not assume any specific watermarking method and thus would apply to any watermarking method in J2J framework.

Index Terms—Data hiding, JPEG watermarking, watermark capacity.

# I. INTRODUCTION

**N** OWADAYS, most digital images are stored in JPEG format, in digital cameras and the World-Wide Web alike. People are gradually motivated to embed watermark or information bits such as owner information, date, time, camera settings, event/occasion of the image, image title, or even secret messages in the JPEG images for value-added functionalities and possibly secret communication. In these applications, the input to the watermarking scheme is a JPEG image file and the output is also a JPEG image file. We call this kind of watermarking (or data hiding) scheme JPEG-to-JPEG (J2J) watermarking schemes. This paper is about J2J watermarking schemes.

There are many papers investigating the robustness of watermarks against JPEG compression such as [9]–[11]. Eggers *et al.* [12] analyzed the quantization effect on the detection of watermarks by considering the additive watermark signal as a dithering signal. Although many watermarking (or data hiding) algorithms were proposed to embed digital watermarks in uncompressed images, those algorithms may not be suitable for embedding watermarks in JPEG-compressed images (.jpg files). This is because the DCT coefficients in JPEG-compressed images have special statistical characteristics—they must be multiples of the corresponding quantization factors. These special characteristics reduce the degree of freedom for watermarking. If the output images are not JPEG compatible, the existence of the watermark may be detectable using steganalysis techniques

The authors are with the Department of Electrical and Electronic Engineering, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong.

Digital Object Identifier 10.1109/TCSVT.2003.815949

[13]. If the output images are JPEG-compatible which is the J2J framework, all DCT coefficients must be re-quantized after the watermark insertion, which further reduces the degree of freedom for watermarking.

There are a few existing schemes for J2J watermarking [2], [14]–[18]. Choi et al. [14] and Luo et al. [15] used inter-block correlation of selected DCT coefficients to embed the watermark bits, adding or subtracting an offset to the mean value of the neighboring DCT coefficients. Wong and Au proposed to hide bits by modifying the DC [16] and AC coefficients [17] in the block-based DCT domain. Hartung [2] used the spread spectrum technique (SST) [3] to embed watermarks in I-frames, P-frames or B-frames of MPEG-2 compressed video. Compression of I-frames is effectively the same as JPEG. Wong and Au proposed a robust watermark scheme using iterative SST [18]. These methods embedded different amount of watermark bits into JPEG images while maintaining good visual quality of the watermarked JPEG images. However, no one estimated the J2J data-hiding capacity, or the maximum amount of bits that can be embedded in JPEG image files.

There are some existing methods to estimate the data hiding capacity of digital images [19]-[28], though they are not JPEG images. Most of these methods apply the work of Shannon [7] and Costa [8]. Servetto et al. [19] used statistical models to analyze the robustness of the SST and estimated the watermarking capacity against the jamming noise. Barni et al. [20], [21] modeled each watermark channel by using Generalized Gaussian density to model the full frame DCT coefficients. Moulin *et al.* [22] modeled coefficients in different domains and estimated the data hiding capacity under mean square error (mse) constraints. Lin [37] estimated the zero-error capacity of images against JPEG attacks with largest applicable quantization step. Some papers combined the Information-Theoretic model [1] and perceptual models to estimate the capacity [23], [24]. Some [25], [26] focused on comparing the capacity among different transforms such as the identity transform (IT), discrete cosine transform (DCT), Karhunen–Loeve transform (KLT), and the Hadamard transform. Fei et al. [25] suggested that the coefficients in the Slant transform had the highest capacity while Ramkumar et al., [26] indicated that transforms with poor energy compaction property such as Hadamard transform tended to have higher capacity than those with higher energy compaction property such as DCT. Sugihara [27] estimated the capacity by taking robustness of the hidden data into account. Voloshynovskiy et al. [33] analyzed the security of the hidden data and suggested different modulation schemes for different purposes of data hiding. Kalker et al. [31] estimated the capacity of a particular data hiding area-lossless data embedding, first proposed by Fridrich et al. [32]. For lossless data

Manuscript received December 18, 2002; revised February 24, 2003. This work was supported by the Research Grants Council of the Hong Kong Special Administrative Region of China (HKUST6028/01E).



Fig. 1. JPEG-to-JPEG watermarking (J2J).

embedding, the original cover work can be restored at the decoder. This is particularly useful for many digital media such as medical images. Cohen *et al.* [30] analyzed the capacity for private and public (or blind) data hiding schemes [6] and the capacity under additive attacks. Instead of estimating the capacity, some proposed realizations to approach the theoretical limit of capacity such as [29], [33], [34]. Pérez-González *et al.* [29] suggested to use convolutional and orthogonal codes. Eggers *et al.* [34] proposed the scalar costa scheme (SCS) by considering the data hiding as the communication-with-side-information problem which has good performance at high watermark-to-noise ratio (WNR).

In this paper, we attempt to estimate the data hiding capacity of JPEG images in J2J watermarking schemes. To embed watermarks in JPEG-compressed images, the JPEG file needs to be partially or fully decoded. The level of decoding depends on the domain the watermark will be embedded in. If the watermark is embedded in the bitstream domain, only variable-length decoding is needed. If the watermark is embedded in 8-by-8 block-based DCT domain, inverse zigzag scanning and inverse quantization are necessary. If the watermark is embedded in spatial domain or other frequency domains, inverse DCT would be needed. The J2J model is shown in Fig. 1. In this paper, we make two assumptions. The first assumption is that the watermarked images will be JPEG-compressed using either the original quantization table  $\mathbf{Q}_{\mathbf{o}}$  extracted from the input JPEG file or a new quantization table  $Q_n$  defined by the user. With this assumption, we have the J2J framework. The second assumption is that the dimensions of the images are not changed in the watermark embedding. The J2J model makes no assumption on the domain the watermark is embedded in. There are four J2J cases as follows and this paper addresses most of the cases.

- 1) The original quantization table  $\mathbf{Q}_{\mathbf{o}}$  is used to compress the watermarked image, i.e.,  $\mathbf{Q}_{\mathbf{n}} = \mathbf{Q}_{\mathbf{o}}$ , and no other processing is applied to the image. An example is a watermarking command program.
- 2) The original quantization table  $\mathbf{Q}_{o}$  may or may not be used to compress the watermarked image, i.e.,  $\mathbf{Q}_{n} = \mathbf{Q}_{o}$ or  $\mathbf{Q}_{n} \neq \mathbf{Q}_{o}$ , and no other processing is applied to the image. An example is a watermarking command program with an option to choose a different quality factor (QF) or  $\mathbf{Q}_{n}$ .
- 3) The original quantization table  $\mathbf{Q}_{\mathbf{o}}$  is used to compress the watermarked image, and the image may be altered by some kind of processing before or after the water-

marking insertion. An example is image processing software such as Adobe Photoshop with watermarking functionality, and the user performs red-eye reduction or other filtering before or after the watermark insertion and then chooses the 'Save' (instead of 'Save as') command to save the image.

4) The original quantization table  $\mathbf{Q}_{\mathbf{0}}$  may not be used to compress the watermarked image, and the image may be altered by some kind of processing before or after the watermarking insertion. An example is the user performs red-eye reduction before or after the watermark insertion and then chooses the "Save As" command (instead of "Save") to save the image. In the "Save As" command, the user may choose different QF in the JPEG compression.

For case 1, we propose a method in Section II to estimate the data-hiding capacity of the JPEG images. The estimated capacity is the upper bound of the amount of bits that can be embedded in JPEG image files without causing visible artifacts. The estimated capacity can be passed to the watermark embedding module as a reference. For case 2, since the new quantization table  $Q_n$  is unknown to the watermark embedding module, the problem is similar to embedding a watermark against JPEG attack. As most quantization tables in JPEG encoders are obtained by scaling a reference quantization table (most probably the default quantization table recommended in the JPEG standard) with a QF, the typical  $\mathbf{Q}_{\mathbf{n}}$  corresponding to different QFs and the reference quantization table can be derived in J2J. For case 2, our proposed algorithm in Section II can be used to estimate the data hiding capacity for a wide range of OF. The resulting capacity curve can be passed to the watermark embedding module as a reference. For cases 3 and 4, if the modification is done before watermarking insertion, our proposed algorithm in Section II can be used to estimate data hiding capacity. If the modification is done after the watermarking insertion, the modification should be treated as attacks leading to lower capacity. Our algorithm is not designed to handle this case.

In Section II, we will describe the proposed capacity estimation algorithm for J2J watermarking. We will use the human visual system (HVS) model to determine the maximum allowable modification of block-based DCT coefficients. This maximum allowable modification is called the just noticeable difference (JND). In Section III, we make a simplifying assumption and derive the necessary conditions for the watermark signal to achieve data hiding capacity. In Section IV, we describe the JND model. While the JND is known to be difficult to estimate accurately, a commonly used JND model is the Watson's model [4] which estimates the JND of block-based DCT coefficients. In [24], Watson's model was used in part of the capacity estimation process. In our experiments, we observe that the Watson's model is not enough to guarantee the invisibility of watermarks. Thus in Section IV, we will propose a modified model based on Watson's model to estimate the JND. Experimental results and discussions are given in Section V.

# II. CAPACITY ESTIMATION FOR J2J WATERMARKING

As the cover work, or the input image, is assumed to be a JPEG-compressed image, the quantization table  $Q_0$  can be ex-

tracted from the input JPEG file header. We denote the  $ij^{\text{th}}$  quantized DCT coefficient of the  $k^{\text{th}}$  block as  $D_q(i, j, k)$ . The dequantized DCT coefficient  $D_o(i, j, k)$  is

$$D_o(i, j, k) = Q_o(i, j) \cdot D_q(i, j, k) \quad 1 \le i, j \le 8$$
 (1)

where  $Q_o(i, j)$  is the  $ij^{\text{th}}$  element of the quantization table  $\mathbf{Q}_{\mathbf{0}}$ . Suppose the corresponding JND of the DCT coefficient is  $J_o(i, j, k)$ , if any distortion of magnitude less than  $J_o(i, j, k)$  is added to the image, the distortion would be unnoticeable to the human eyes. It should be noted that  $J_o(i, j, k)$  is a nonnegative quantity by definition. To ensure perceptually invisible watermarking artifacts, the amount of modification on  $D_o(i, j, k)$  is limited to  $J_o(i, j, k)$  in the watermark embedding process. The model of JND will be described in detail in Section IV. Recall that we have made two assumptions. Based on our second assumption that the image dimensions do not change, the number of DCT coefficients does not change after watermark insertion. Let  $D_w(i, j, k)$  be the  $ij^{\text{th}}$  watermarked DCT coefficient of the  $k^{\text{th}}$  block. Based on our first assumption,  $D_w(i, j, k)$  is quantized with a quantization table  $\mathbf{Q}_{\mathbf{n}}$  to give the  $ij^{\text{th}}$  JPEG-decoded watermarked DCT coefficient of the  $k^{th}$  block

$$D_n(i,j,k) = Q_n(i,j) \cdot \text{round}\left(\frac{D_w(i,j,k)}{Q_n(i,j)}\right)$$
(2)

in the output JPEG image file, where  $Q_n(i, j)$  is the  $ij^{\text{th}}$  entry of  $\mathbf{Q_n}$ . The quantization matrix  $\mathbf{Q_n}$  used in the output JPEG file may or may not be equal to  $\mathbf{Q_o}$  used in the input JPEG file. To ensure high visual quality of the watermarked image, the quantized watermarked DCT coefficient  $D_n(i, j, k)$  should satisfy

$$|D_n(i,j,k) - D_o(i,j,k)| \le J_o(i,j,k)$$
(3)

which guarantees the invisibility of the watermark, given an accurate JND model. If  $J_o(i, j, k)$  is large enough, the maximum number of possible values of  $D_n(i, j, k)$  (or quantized values of  $D_w(i, j, k)$ ) within the allowable range is given by  $N_w(i, j, k)$  as follows:

$$N_{w}(i,j,k) = \left[\frac{D_{o}(i,j,k) + J_{o}(i,j,k)}{Q_{n}(i,j)}\right] - \left[\frac{D_{o}(i,j,k) - J_{o}(i,j,k)}{Q_{n}(i,j)}\right] + 1$$
(4)

where  $\lfloor x \rfloor$  denotes the largest integer smaller than or equal to x and  $\lceil x \rceil$  denotes the smallest integer greater than or equal to x. This is because we can use at most so many possible values of  $D_n(i, j, k)$  within the allowable range without going outside JND. Note that  $N_w(i, j, k) \ge 0$  because

$$x \ge y$$
 implies  $\lfloor x \rfloor - \lceil y \rceil + 1 \ge \lfloor y \rfloor - \lceil y \rceil + 1 \ge -1 + 1 = 0$ 

where  $x = [D_o(i, j, k) + J_o(i, j, k)]/Q_n(i, j)$  and  $y = [D_o(i, j, k) - J_o(i, j, k)]/Q_n(i, j)$ . The  $N_w(i, j, k)$  in (4) can be equal to zero when  $J_o(i, j, k)$  is less than both  $[Q_n(i, j) - (D_o(i, j, k) \mod Q_n(i, j))]$  and  $[D_o(i, j, k) \mod Q_n(i, j)]$ . In other words,  $N_w(i, j, k)$  can be equal to zero when the allowable JND range around  $D_o(i, j, k)$  does not contain any legitimate output JPEG-compatible

DCT coefficient values (multiples of  $Q_n(i,j)$ ). When this happens, the distortion of the JPEG recompression would already introduce perceptually detectable artifacts, in the absence of watermarking. Thus, when  $N_w(i,j,k) = 0$ in (4), no data should be embedded in this DCT coefficient to prevent the situation from getting worse, and the output DCT coefficient would simply be the nonwatermarked value  $Q_n(i,j) \cdot \text{round}[D_o(i,j,k)/Q_n(i,j)]$ . Thus, there is only one way to choose the output DCT coefficient and we force  $N_w(i,j,k) = 1$  to reflect it. In this way,  $N_w(i,j,k)$  would always be greater than or equal to 1 such that  $\log_2[N_w(i,j,k)] \ge 0$ .

With this, the data hiding capacity or the maximum number of bits that can be embedded in the  $ij^{\text{th}}$  DCT coefficient of  $k^{\text{th}}$ block is given by C(i, j, k)

$$C(i, j, k) = \log_2 [N_w(i, j, k)]$$
  
= 
$$\log_2 \left[ \left\lfloor \frac{D_o(i, j, k) + J_o(i, j, k)}{Q_n(i, j)} \right\rfloor - \left\lceil \frac{D_o(i, j, k) - J_o(i, j, k)}{Q_n(i, j)} \right\rceil + 1 \right].$$
 (5)

Since each DCT coefficient can be considered as an independent channel, the total data-hiding capacity of the image is given by  $C_w$ 

$$C_w = \sum_{k=1}^{K} \sum_{j=1}^{8} \sum_{i=1}^{8} \log_2\left[N_w(i, j, k)\right]$$
(6)

where K is the number of blocks in the image. Our estimated capacity depends only on JND model and the quantization factor. Regardless of which domain the watermark is embedded in, the amount of embedded information bits is bounded by (5) as long as the JND constraint in (3) is satisfied. As this capacity estimation method does not assume any specific watermarking method, it should apply to any watermarking method in the J2J framework.

We observe that our capacity estimation method agrees somewhat with the works of Costa [8]. For the  $ij^{\text{th}}$  DCT coefficient of the  $k^{\text{th}}$  block, if we assume the watermark signal W to be uniformly distributed in  $[-J_o(i, j, k), J_o(i, j, k)]$ , the power of the watermark signal  $\sigma_w^2$  should be  $[J_o(i, j, k)]^2/3$ . Also assuming the quantization noise is uniformly distributed in  $[-Q_n(i, j)/2, Q_n(i, j)/2]$ , the power of quantization noise  $\sigma_n^2$  should be  $[Q_n(i, j)]^2/12$ . Equation (5) can be re-written approximately as

$$C(i, j, k) = \log_2 \left[ \left\lfloor \frac{D_o(i, j, k) + J_o(i, j, k)}{Q_n(i, j)} \right\rfloor - \left\lceil \frac{D_o(i, j, k) - J_o(i, j, k)}{Q_n(i, j)} \right\rceil + 1 \right]$$
$$\approx \log_2 \left( \left\lceil \frac{2 \cdot J_o(i, j, k)}{Q_n(i, j)} \right\rceil \right)$$
$$= \frac{1}{2} \cdot \log_2 \left( \left\lceil \frac{2 \cdot J_o(i, j, k)}{Q_n(i, j)} \right\rceil^2 \right)$$
$$\approx \frac{1}{2} \log_2 \left\lceil \frac{4 \cdot [J_o(i, j, k)]^2}{[Q_n(i, j)]^2} \right\rceil$$

$$\approx \frac{1}{2} \log_2 \left( 1 + \frac{4 \cdot [J_o(i, j, k)]^2}{[Q_n(i, j)]^2} \right)$$
  
=  $\frac{1}{2} \log_2 \left( 1 + \frac{\frac{[J_o(i, j, k)]^2}{3}}{\frac{[Q_n(i, j)]^2}{12}} \right)$   
=  $\frac{1}{2} \log_2 \left( 1 + \frac{\sigma_w^2}{\sigma_n^2} \right)$  (7)

where the approximation  $\lceil x \rceil^2 \approx \lceil x^2 \rceil \approx x^2 + 1$  is used. An interesting observation is that this expression is similar to the AWGN channel capacity obtained by Costa in [8].

## **III. NECESSARY CONDITIONS TO ACHIEVE CAPACITY**

We derive the necessary conditions for the distribution of the watermark signal W for achieving the upper bound of the data hiding capacity in (5). Here we assume  $D_o$  is known at the watermark encoder. As we treat each DCT coefficient as an independent channel, we only consider the  $ij^{\text{th}}$  DCT coefficient of the  $k^{\text{th}}$  block and  $D_o$  can be used as the side information of the channel. For simplicity, we denote C(i, j, k),  $D_o(i, j, k)$ ,  $D_n(i, j, k)$ ,  $Q_n(i, j)$  and  $J_o(i, j, k)$  as C,  $D_o$ ,  $D_n$ ,  $Q_n$  and  $J_o$ , respectively. We also denote the probability density function of W as f(w), the probability mass function of  $D_o$ , and  $D_n$  as  $p(D_o)$  and  $p(D_n)$ , respectively. According to [35], the capacity of a communication channel with side information is given by

$$C_W = \max\left[I(U, D_n) - I(U, D_o)\right] \tag{8}$$

where U is a finite alphabet auxiliary random variable,  $I(U, D_n)$  is the mutual information of U and  $D_n$ , and the maximum is over all joint distributions of the form  $p(D_o)$ ,  $p(u, w|D_o)$ ,  $p(D_n, w|D_o)$  and  $p(D_n|w, D_o)$ . According to [36], for a deterministic channel,  $U = D_n$ 

$$C_W = \max \left[ H(D_n) - I(D_n, D_o) \right] = \max \left[ H(D_n) - \{ H(D_n) - H(D_n | D_o) \} \right] = \max \left[ H(D_n | D_o) \right]$$
(9)

where  $H(D_n)$  is the entropy of  $D_n$ .

As  $D_n$  is a discrete random variable in  $[D_o - J_o, D_o + J_o]$  with  $N_w$  possible values, the maximum of  $[H(D_n|D_o)]$  is achieved when  $D_n$  is equally probable in  $[D_o - J_o, D_o + J_o]$  when  $D_o$  is known

$$p(D_n) = \begin{cases} \frac{1}{N_w}, & D_o - J_o \le D_n \le D_o + J_o \\ 0, & \text{otherwise} \end{cases}$$
(10)

and the equivalent necessary conditions for the probability density function of W are

$$\binom{(m+\frac{1}{2})\cdot Q_n}{\int} f(w)dw = \int_{(m+\frac{1}{2})\cdot Q_n} f(w)dw = \dots$$
$$\binom{(m+\frac{1}{2})\cdot Q_n}{(m+N_w - \frac{1}{2})\cdot Q_n}$$
$$= \int_{(m+N_w - \frac{3}{2})\cdot Q_n} f(w)dw = \frac{1}{N_w} (11)$$

where  $m = \lceil ((D_o - J_o)/Q_n) \rceil$  and the corresponding capacity is  $C_W = \log_2 N_w$  as given in (5).

# **IV. JND ESTIMATION**

To estimate the JND of the block-based DCT coefficients, we choose the well-known Watson's model [4]. After evaluating the Watson's model, we observe that the Watson's model is not sufficient to ensure the invisibility of watermarks in our experiments. In the experiments, we modify all DCT coefficients by adding distortions with magnitude equal to the corresponding JND. Obviously, this is the worst situation that could happen. Some visible artifacts are observed, especially at regions with edges and regions with textures. We modify the Watson's model slightly to avoid these visible patterns.

#### A. The Watson's Model

The model consists of three parts: the sensitivity table, the luminance masking, and the contrast masking. The sensitivity table is derived in [5]. It specifies the JND of a DCT coefficient without considering any masking, and its  $ij^{\text{th}}$  component is denoted as  $J_s(i, j)$ . Taking the background luminance masking effect into account, the luminance masking uses  $J_s(i, j)$  to estimate the JND of a DCT coefficient. As a bright background can mask more noise than a dark background, the JND for luminance masking of the  $ij^{\text{th}}$  DCT coefficient of the  $k^{\text{th}}$  block  $J_L(i, j, k)$  is given by

$$J_L(i,j,k) = J_s(i,j) \cdot \left[\frac{D_o(1,1,k)}{\overline{D}}\right]^{\alpha}$$
(12)

where  $D_o(1, 1, k)$  is the DC value of the  $k^{\text{th}}$  block,  $\alpha$  is a constant, and  $\overline{D}$  is the mean intensity of the background or expected intensity of images. The suggested value of  $\alpha$  by Watson is 0.649. The third part of the Watson's model is contrast masking. Based on  $J_L(i, j, k)$ , it estimates the JND more accurately by considering the noise masking property due to the presence of AC energy in the corresponding DCT coefficient. Taking contrast masking into account, the JND for the  $ij^{\text{th}}$  DCT coefficient of the  $k^{\text{th}}$  block is given by

$$J_{o}(i,j,k) = \max \left\{ J_{L}(i,j,k), |D_{o}(i,j,k)|^{v(i,j)} \\ \cdot J_{L}(i,j,k)^{1-v(i,j)} \right\}$$
(13)

where v(i, j) is a constant between 0 and 1. Watson suggested to give v(i, j) a value of 0.7 for all i, j.

#### B. Modification of Watson's Model

We modify the Watson's model to prevent the over-estimation of JND at the edge blocks and texture blocks. For each input JPEG image block, the median value and the mean value of the  $ij^{\text{th}}$  DCT coefficient and the  $ij^{\text{th}}$  DCT coefficients of the eight surrounding blocks are computed, and the minimum between the median and the mean is chosen and denoted as  $D_m(i, j, k)$ . For the blocks at the edges of the image, replicated blocks are used outside the edges. This  $D_m(i, j, k)$  is used to replace  $D_o(i, j, k)$  in (12) and (13) to compute the final JND



Fig. 2. Magnified image using Watson's model.



Fig. 3. Magnified image using modified Watson's model.

TABLE I Default Quantization Table of JPEG

16	11	10	16	24	40	51	64
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

 $J_o(i, j, k)$ . The difference between the Watson's model and our modified Watson's model is shown in Figs. 2 and 3.

## V. EXPERIMENTAL RESULTS AND DISCUSSIONS

To demonstrate the use of the proposed capacity estimation algorithm, the J2J data hiding capacities for two common test images are estimated and reported in this paper. They are the 512  $\times$  512 gray-scale images, "Lena" and "Pepper". Only the luminance component is used in the experiments. To control the compression ratio in JPEG, a scaling factor (SF) is used to scale the default quantization table in JPEG, shown in Table I. In J2J model, there are two quantization tables,  $Q_0$  and  $Q_n$ , which



Fig. 4. Compression ratio versus SF for "Lena".



Fig. 5. Compression ratio versus SF for "Pepper".

depend on three parameters: 1) the JPEG default quantization table; 2) the SF<sub>before</sub>; and 3) the SF<sub>after</sub>. The SF<sub>before</sub> is the SF used to scale the  $Q_o$  for the input JPEG-compressed image, and the Sf<sub>after</sub> is the used to scale  $Q_n$  for the output JPEGcompressed image. All the entries in the scaled  $Q_o$  and  $Q_n$  table are then rounded to integers. Most research papers specify the JPEG quality by the QF. The SF is related to the QF by

$$SF = \begin{cases} \frac{50}{QF}, & QF \le 50\\ 2 - \frac{QF}{50}, & 50 < QF < 100 \end{cases}$$
(14)

In our simulations, JPEG compression with  $\mathbf{Q}_{o}$  is applied to the testing images. These are the input JPEG images to the J2J model. To study the worst-case visual quality of J2J watermarked images, the JND value  $J_o(i, j, k)$  is computed and is added as distortion to the DCT coefficients  $D_o(i, j, k)$  for all i, jand k. The images are then JPEG-compressed using  $\mathbf{Q}_n$  to give the output JPEG images, which we call the JND watermarked images. The compression ratio in terms of bits-per-pixel (bpp) is shown in Fig. 4 and 5 for "Lena" and "Pepper" respectively. The PSNR of the JPEG-compressed images is shown in Fig. 6 and 7. The estimated data hiding capacity using (6) is shown in Fig. 8 and 9. Various values of SF<sub>before</sub> and SF<sub>after</sub> are simulated in our experiments. Four possible cases of SF<sub>before</sub> are shown, namely SF<sub>before</sub> =1, 2, 3, 4. Eight possible cases for SF<sub>after</sub> are shown, namely SF<sub>after</sub> = 0.5, 1, 1.5, 2, 2.5, 3, 3.5, 4. As a



Fig. 6. PSNR of JPEG-compressed images and JND watermarked images of "Lena".



Fig. 7. PSNR of JPEG-compressed images and JND watermarked images of "Pepper".



Fig. 8. Estimated capacity of "Lena".

control, the results for the input JPEG image (without any watermarks) are also shown and are labeled as "JPEG only". The horizontal axis in Figs. 4-9 is SF<sub>after</sub>.

When  $SF_{after}$  is equal to  $SF_{before}$ , which is a very probable situation, the maximum increases in bpp due to J2J watermark for Lena in Fig. 4 and Pepper in Fig. 5 are only 0.0030 and 0.0032 bpp, respectively, which are negligible. In Figs. 6



Fig. 9. Estimated capacity of "Pepper".

and 7, the worst-case PSNR losses due to J2J watermarking are 0.2768 and 0.3556 dB for Lena and Pepper, respectively, which are small. In Fig. 8 and 9, the data hiding capacity of Lena and Pepper both decrease rapidly with  $SF_{after}$  and drop below 15 when  $SF_{after}$  is 4 for all cases. These suggest that the capacity can be significantly affected by the  $SF_{after}$ . However, Watson's model was designed to estimate the JND for uncompressed images. It may not be able to estimate accurately JND for compressed images. The same may also be true for our proposed modified Watson's model.

## VI. CONCLUSION

In this paper, we propose a method to estimate the data hiding capacity of digital images in the JPEG-to-JPEG watermarking framework, using a HVS model. A modified Watson HVS model is used. As our capacity estimation does not assume any specific watermarking method, it should apply to any watermarking method in the J2J framework.

#### REFERENCES

- T. Cover and J. Thomas, *Elements of Information Theory*. New York: Wiley, 1991.
- [2] F. Hartung, Digital Watermarking and Fingerprinting of Uncompressed and Compressed Video. Aachen, Germany: Shaker Verlag, 2000.
- [3] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*. New York: Morgan Kaufmann, 2002.
- [4] A. B. Watson, "DCT quantization matrices optimized for individual images," in *Proc. SPIE Human Vision, Visual Processing, and Digital Display IV*, 1993, pp. 202–216.
- [5] A. J. Ahumada and H. A. Peterson, "Luminance-model-based DCT quantization for color image compression," in *Proc. SPIE*, vol. 1666, 1992, pp. 365–374.
- [6] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," in *Proc. SPIE Security and Watermarking of Multimedia Contents*, Jan. 1999, pp. 226–239.
- [7] C. E. Shannon, "A mathematical theory of communication," *Bell Syst. Tech. J.*, vol. 27, pp. 373–423 and 623–656, 1948.
- [8] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. IT-29, pp. 439–441, May 1983.
- [9] I. Cox, J. Kilian, T. Leighton, and T. Shamoon, "Secure spread spectrum watermarking for multimedia," *IEEE Trans. Image Processing*, vol. 6, pp. 1673–1687, Dec. 1997.
- [10] M. Kutter, F. Jordan, and F. Bossen, "Digital signature of color images using amplitude modulation," in *Proc. SPIE Storage and Retrieval for Image and Video Databases*, vol. 3022-5, Feb. 1997, pp. 518–552.
- [11] M. Kutter and F. A. P. Petitcolas, "A fair benchmark for image watermarking systems," in *Proc. SPIE Security and Watermarking of Multimedia Contents*, vol. 3657, Jan. 1999, pp. 226–239.

- [12] J. J. Eggers and B. Girod, "Quantization effects on digital watermarks," EURASIP Signal Processing, vol. 81, no. 2, pp. 239–263, 2001.
- [13] J. Fridrich, M. Goljan, and R. Du, "Steganalysis based on JPEG compatibility," in *Proc. SPIE Multimedia Systems and Applications IV*, Aug. 2001, pp. 275–280.
- [14] Y. Choi and K. Aizawa, "Digital watermarking using inter-block correlation: extension to JPEG coded domain," in *Proc. IEEE Int. Conf. Information Technology: Coding and Computing*, Mar. 2000, pp. 133–138.
- [15] W. Luo, G. L. Heileman, and C. E. Pizano, "Fast and robust watermarking of JPEG files," in *Proc. IEEE 5th Southwest Symp. Image Analysis and Interpretation*, Apr. 2002, pp. 158–162.
- [16] P. H. W. Wong and O. C. Au, "Data hiding and watermarking in JPEG compressed domain by DC coefficient modification," in *Proc. SPIE Security and Watermarking of Multimedia Contents*, vol. 3971, Jan. 2000, pp. 237–244.
- [17] —, "Data hiding technique in JPEG compressed domain," in Proc. SPIE Security and Watermarking of Multimedia Contents, vol. 4314, Jan. 2001, pp. 309–320.
- [18] —, "A blind watermarking technique in JPEG compressed domain," in *Proc. IEEE Int. Conf. Image Processing*, vol. 3, Sept. 2002, pp. 497–500.
- [19] S. D. Servetto, C. I. Podilchuk, and K. Ramchandran, "Capacity issues in digital image watermarking," in *Proc. IEEE Int. Conf. Image Processing*, vol. 1, Oct. 1998, pp. 445–449.
- [20] M. Barni, F. Bartolini, A. D. Rosa, and A. Piva, "Capacity of the watermark channel: how many bits can be hidden within a digital images," in *Proc. SPIE Security and Watermarking of Multimedia Contents*, vol. 3657, Jan. 1999, pp. 437–448.
- [21] —, "Capacity of full frame DCT image watermarks," *IEEE Trans. Image Processing*, vol. 9, pp. 1450–1455, Aug. 2000.
- [22] P. Moulin and M. K. Mihçak, "A framework for evaluating the datahiding capacity of image sources," *IEEE Trans. Image Processing*, vol. 9, pp. 1450–1455, Aug. 2000.
- [23] D. Kundur, "Implication for high capacity data hiding in the presence of lossy compression," in *Proc. IEEE Int. Conf. Information Technology: Coding and Computing*, Mar. 2000, pp. 16–22.
- [24] C. Y. Lin and S. F. Chang, "Watermarking capacity of digital images based on domain-specific masking effects," in *Proc. IEEE Int. Conf. Information Technology: Coding and Computing*, Apr. 2001, pp. 90–94.
- [25] C. Fei, D. Kundur, and R. Kwong, "The choice of watermark domain in the presence of compression," in *Proc. IEEE Int. Conf. Information Technology: Coding and Computing*, Apr. 2001, pp. 79–84.
- [26] M. Ramkumar and A. N. Akansu, "Capacity estimates for data hiding in compressed images," *IEEE Trans. Image Processing*, vol. 10, pp. 1252–1263, Aug. 2001.
- [27] R. Sugihara, "Practical capacity of digital watermark as constrained by reliability," in *Proc. IEEE Int. Conf. Information Technology: Coding* and Computing, Apr. 2001, pp. 85–89.
- [28] C. Candan and N. Jayant, "A new interpretation of data hiding capacity," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing*, vol. 3, May 2001, pp. 1993–1996.
- [29] F. Pérez-González, J. R. Hernández, and F. Balado, "Approaching the capacity limit in image watermarking: a perspective on coding techniques for data hiding applications," *Elsevier Signal Processing*, vol. 81, no. 6, pp. 1215–1238, June 2001.
- [30] A. S. Cohen and A. Lapidoth, "The gaussian watermarking game," *IEEE Trans. Inform. Theory*, vol. 48, pp. 1639–1667, June 2002.
- [31] T. Kalker and F. M. J. Willems, "Capacity bounds and constructions for reversible data-hiding," in *Proc. IEEE Int. Conf. Digital Signal Pro*cessing, vol. 1, Jul. 2002, pp. 71–76.

- [32] J. Fridrich, M. Coljan, and R. Du, "Lossless data embedding—new paradigm in digital watermarking," EURASIP J. Appl. Signal Processing—Special Issue on Emerging Applications of Multimedia Data Hiding, vol. 2002, no. 2, pp. 185–196, Feb. 2002.
- [33] S. Voloshynovskiy and T. Pun, "Capacity-security analysis of data hiding technologies," in *Proc. IEEE Int. Conf. Multimedia and Expo*, vol. 2, Aug. 2002, pp. 477–480.
- [34] J. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod. Scalar costa scheme for information embedding. *IEEE Trans. Signal Processing*
- [35] S. I. Gel'fand and M. S. Pinsker, "Coding for channel with random parameters," *Prob. Control and Inform. Theory*, vol. 9, no. 1, pp. 19–31, 1980.
- [36] W. Yu. Channel capacity with side information. [Online][Online]. Available: http://www.comm.toronto.edu/~limtj/seminars/WeiYu.ppt
- [37] C. Y. Lin, "Watermarking and digital signature techniques for multimedia authentication and copyright protection," Ph.D. dissertation, Columbia Univ., New York, 2000.



lence in 1998.

**Peter H. W. Wong** (M'01) received the B.Eng. degree (with first-class honors) in computer engineering from the City University of Hong Kong in 1996 and the M. Phil. degree in electrical and electronic engineering in 1998 from Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, where he is currently working toward the Ph.D. degree.

His research interests include time-scale modification, motion estimation, and digital watermarking.

Mr. Wong received the Schmidt Award of Excel-



**Oscar C. Au** (S'87–M'90–SM'01) received the B.A.Sc. degree from the University of Toronto, Toronto, ON, Canada, in 1986, and the M.A. and Ph.D. degrees from Princeton University, Princeton, NJ, in 1988 and 1991, respectively.

After a postion as a postdoctoral Researcher with Princeton Univ. for one year, he joined the Department of Electrical and Electronic Engineering, Hong Kong University of Science and Technology (HKUST), Clear Water Bay, in 1992, where he is now an Associate Professor and Director of the

Computer Engineering Program. His main research contributions include video and image coding, watermarking and steganography, speech and audio processing. His topics of research are fast motion estimation for MPEG-1/2/4 and H.261/3/L, fast rate control, transcoding, post-processing, JPEG/JPEG2000 and halftone image data hiding, etc. His fast motion estimation algorithm was recently accepted by the ISO/IEC 14 496-7 MPEG-4 Standard. He is currently applying for five patents on his signal processing techniques. He has published over 100 technical journal and conference papers.

Dr. Au is an Associate Editor of the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS-I: FUNDAMENTAL THEORY AND APPLICATIONS. He is a member of the Technical Committee on Multimedia Systems and Applications and the Technical Committee on Video Signal Processing and Communications of the IEEE Circuits and Systems Society. He served on the organizing committee of the IEEE International Symposium on Circuits and Systems in 1997 and the IEEE International Conference on Acoustics, Speech and Signal Processing in 2003.