ABSTRACT
A data capacity estimation method is proposed for image watermarking. The data capacity is maximum number of bits that can be embedded in an image such that the image has no perceptual loss. It is assumed that the input is a JPEG2000 image file and after watermark embedding, the image is JPEG2000-compressed using the same quantization factors. This is called JPEG2000-to-JPEG2000 (J2K2J2K) watermarking. A Human Visual Systems (HVS) model is used to estimate the Just Noticeable Difference (JND) of each Discrete Wavelet Transform (DWT) coefficients.

1. INTRODUCTION
JPEG2000 is a new international standard for image coding [1,2]. It is known that the JPEG2000 standard gives better image quality than the popular JPEG standard. It is expected that the JPEG2000 standard will become a popular image-coding standard in the coming future. People are motivated to embed watermark or information bits such as owner info, date, time, camera settings, event/occasion of the image, image title, or even secret message in the images for value-added functionalities and possibly secret communication. In these applications, if the input to the watermarking scheme is a JPEG2000 image file and the output is also a JPEG2000 image file, we call this kind of watermarking (or data hiding) scheme as JPEG2000-to-JPEG2000 (J2K2J2K) watermarking scheme. This paper is concern about J2K2J2K watermarking schemes.

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 JPEG2000-compressed images. This is because the DWT coefficients in JPEG2000-compressed images have special statistical characteristics – they must be multiples of the corresponding quantization factors. Not only do these special characteristics reduce the degree of freedom for watermarking, but also the modification in the watermarked images can be easily detected by steganalysis [4] if the output images are not JPEG2000-compatible. However, the resistance to steganalysis may be increased by designing watermarking schemes that output JPEG2000-compatible images, which motivates this J2K2J2K work. In J2K2J2K schemes, all DWT coefficients must be re-quantized after the watermark insertion which further reduces the degree of freedom for watermarking.

There are some existing methods to estimate the data hiding capacity of digital images [6-14], though they are not JPEG images. Servetto et al. [6] used statistical models to analyze the robustness of Spread Spectrum Technique and estimated the watermarking capacity against jamming noise. Barni et al. [7] modeled each watermark channel by using Generalized Gaussian density to model the full frame DCT coefficients. Moulin et al. [8] modeled coefficients in different domains and estimated the data hiding capacity under MSE constraints. Some papers combined the Information-Theoretic model and perceptual models to estimate the capacity [9-10]. Some [10-11] focused on comparing the capacity among different transforms such as Identity transform, discrete cosine transform (DCT), Karhunen-Loeve transform (KLT) and Hadamard transform. Sugihara [12] estimated the capacity by taking robustness of the hidden data into account. Cohen et al. [13] analyzed the capacity for private and public (or blind) data hiding schemes [5] and the capacity under additive attacks. Voloshynovskiy et al. [14] analyzed the security of the hidden data and suggested different modulation schemes for different purposes of data hiding.

In this paper, we attempt to estimate the data hiding capacity of JPEG2000 images in J2K2J2K watermarking schemes. By using a HVS model [3], the data capacity is estimated such that the watermarked image is perceptually lossless.

2. BACKGROUND OF JPEG2000
JPEG2000, as noted previously, is the new international standard for still image coding. In the following, we restrict the description to Part I of the standard that defines the core coding system. JPEG2000 is based on the DWT, scalar quantization, coefficient bit modeling, arithmetic coding and the post-compression rate control.

The DWT decomposes an image (or sub-image called
tile) into LL, HL, HH and HL subbands as shown in Fig. 1. The subscripts of the subband symbols indicate the decomposition levels \( l \) of the DWT decomposition. We use the term orientation \( \theta \) to denote different subband and \( \theta = 1,2,3,4 \) for LL, HL, LH, HH respectively. The subbands consist of coefficients that represent the horizontal and vertical spatial frequency characteristics of the image/tile. Each subband is then quantized and divided into rectangular blocks (called code blocks in JPEG2000) with size typically at 64x64. The quantized code block data is entropy coded using coefficient bit modeling and arithmetic coding. The coded data is organized by rate control and finally outputted to the code-stream in packets.

The quantization is performed by two stages in lossy compression. In the first stage, the transform coefficients are quantized using scalar quantization with a deadzone. Each of the transform coefficients \( D_q(x,y,l,\theta) \) of the subband \( (l,\theta) \) is quantized to the \( D_q(x,y,l,\theta) \) by

\[
D_q(x,y,l,\theta) = \text{sign}(D_q(x,y,l,\theta)) \cdot \left\lfloor \frac{|D_q(x,y,l,\theta)|}{\Delta_{l,\theta}} \right\rfloor \tag{1}
\]

where \( \Delta_{l,\theta} \) is the quantization stepsize and \( \left\lfloor x \right\rfloor \) denotes the largest integer smaller than or equal to \( x \). In the second stage, the quantized coefficients are encoded by bit-plane entropy coding where a second quantization is performed in bit-plane level.

In the bit-plane coding process (called Tier-1 coding in JPEG2000), each code block is coded from the most significant bit-plane (MSB) to the least significant bit-plane (LSB) by using three coding passes. The three coding passes are significance pass, refinement pass and clean-up pass. Each coefficient bit of a bit-plane is coded by one of the three coding passes. In the Tier-1 coding, each coefficient is essentially quantized using Eqn. 1 with stepsize \( 2^{\omega(x,y,l,\theta)} \) where \( \omega(x,y,l,\theta) \) is the number of unencoded bit-plane for that coefficient. The reconstructed transform coefficient \( D_r(x,y,l,\theta) \) is given by Eqn. 2,

\[
D_r(x,y,l,\theta) = \begin{cases} 
D_q(x,y,l,\theta) + \frac{1}{2} \cdot 2^{\omega(x,y,l,\theta)} \cdot \Delta_{l,\theta} & D_q(x,y,l,\theta) > 0 \\
D_q(x,y,l,\theta) - \frac{1}{2} \cdot 2^{\omega(x,y,l,\theta)} \cdot \Delta_{l,\theta} & D_q(x,y,l,\theta) < 0 \\
0 & D_q(x,y,l,\theta) = 0
\end{cases} \tag{2}
\]

where \( D_r(x,y,l,\theta) \) is the encoded quantization index.

### 3. CAPACITY ESTIMATION

There are two assumptions in J2K-J2K capacity estimation. The first assumption is that the watermarked images will be compressed to JPEG2000 compatible format using the same analysis filters, the same number of levels and using the same quantization factors for each wavelet coefficients. The second assumption is that the dimensions of the images do not change in the watermark embedding. The J2K-J2K model does not concern which domain the watermark is embedded in. Suppose the corresponding JND of the \( xy \)th wavelet coefficient in \( (l,\theta) \) subband is \( J(x,y,l,\theta) \), which is a non-negative quantity by definition. In other words, if any noise of magnitude less than \( J(x,y,l,\theta) \) is added to the coefficient, the noise would be unnoticeable to the human eyes. The model of JND will be described in Section 4. To ensure perceptually invisible watermarking artifacts, we use \( J(x,y,l,\theta) \) to limit the amount of data embedded in \( D_r(x,y,l,\theta) \) in the watermark embedding process. Recall that we have made two assumptions. Based on the assumptions that the image dimensions and the number of decomposition levels do not change, the number of transform coefficients does not change after watermark insertion. Let \( D_w(x,y,l,\theta) \) be the \( xy \)th watermarked wavelet coefficient in the \( (l,\theta) \) subband. Based on the first assumption, \( D_w(x,y,l,\theta) \) is quantized by the effective quantization factor

\[
Q_e(x,y,l,\theta) = 2^{\omega(x,y,l,\theta)} \cdot \Delta_{l,\theta} \tag{3}
\]

to give

\[
D_n(x,y,l,\theta) = \text{sign}(D_n(x,y,l,\theta)) \cdot \left\lfloor \frac{|D_n(x,y,l,\theta)|}{Q_e(x,y,l,\theta)} \right\rfloor \tag{4}
\]

in the output JPEG2000 image file. To ensure perceptually lossless of the watermarked image, the watermarked transform coefficient \( D_n(x,y,l,\theta) \) should satisfy

\[
|D_n(x,y,l,\theta) - D_r(x,y,l,\theta)| \leq J(x,y,l,\theta) \tag{5}
\]

The maximum number of possible values of \( D_n(x,y,l,\theta) \) (or quantized values of \( D_r(x,y,l,\theta) \) )
within the allowable range is calculated by counting the number of reconstruction level(s) which satisfies Equ. 5. This is because we can use at most so many possible values of $D_{\theta}(x, y, l, \theta)$ within the allowable range without going outside the JND.

Assume the maximum number of reconstruction level(s) is $N_w(x, y, l, \theta)$, the data hiding capacity or the maximum number of bits that can be embedded is given by

$$C(x, y, l, \theta) = \log_2[N_w(x, y, l, \theta)] \quad (7)$$

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

$$C_w = \sum_{\theta=1}^{4} \sum_{l=1}^{2} \sum_{\theta} \log_2[N_w(x, y, l, \theta)] \quad (8)$$

As this capacity estimation method does not assume any specific watermarking method, it should apply to any watermarking methods in the J2K2JK framework.

### 4. THE HVS MODEL

We adopt the Watson’s model [3] to estimate the JND of transform coefficients. Watson's works are intended to obtain an optimized quantization matrix for DWT coefficients. CDF 9/7 kernels is chosen for the DWT which is the only irreversible kernels supported by JPEG2000 Part I [1]. The optimal quantization matrix is given by:

$$Q_{l, \theta} = \frac{2}{A_{l, \theta}} a \cdot 10^{b \log(2 f_{sh}(r))} \quad (9)$$

where $A_{l, \theta}$ is the maximum absolute amplitude of the $(l, \theta)$ subband basis function. The suggested values of $a$, $k$, $f_0$, $g_1$, and $g_2$ is shown in Table 1 and $g_2$ and $g_3$ are equals to 1. The parameter $r$ is the display resolution. The optimal quantization factors for $r = 32$ pixels/degree and normalized CDF 9/7 kernels in grayscale domain are given in Table 2.

### Table 1: Parameters suggested for Eqn. 9.

<table>
<thead>
<tr>
<th>$a$</th>
<th>$k$</th>
<th>$f_0$</th>
<th>$g_1$</th>
<th>$g_2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.495</td>
<td>0.466</td>
<td>0.401</td>
<td>1.501</td>
<td>0.534</td>
</tr>
</tbody>
</table>

### Table 2: Optimal quantization factor.

<table>
<thead>
<tr>
<th>Orientation $\theta$</th>
<th>Level $l$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>7.025</td>
</tr>
<tr>
<td>2</td>
<td>11.515</td>
</tr>
<tr>
<td>3</td>
<td>11.515</td>
</tr>
<tr>
<td>4</td>
<td>29.38</td>
</tr>
</tbody>
</table>

In [3], it is assumed that the JND is the same for DWT coefficients in the same subband and the JND for subband $(l, \theta)$ is $Q_{l, \theta}/2$, we slightly modify the JND to

$$J(x, y, l, \theta) = \min\{Q_{l, \theta}/2, |D_{\theta}(x, y, l, \theta)|\} \quad (10)$$

to ensure that the noise amplitude is always less than the signal amplitude which is true in quantization case.

### 5. EXPERIMENTAL RESULTS

The capacity estimation results for common test image ‘Lena’ which has 512x512 pixels is shown in this paper. Only luminance component is used in the experiments. The testing image is initially compressed by JPEG2000 at different bitrate and the bitrate is measured in bit per pixel (bpp). The compressed images are then decoded and the watermark data is hidden in the decoded DWT coefficients. As the maximum number of possible values of $D_{\theta}(x, y, l, \theta)$ is $N_w(x, y, l, \theta)$, watermarking is achieved by randomly modified the $D_{\theta}(x, y, l, \theta)$ to one of the possible $D_{\theta}(x, y, l, \theta)^{'}$, with each $D_{\theta}(x, y, l, \theta)^{'}$ has the probability of $1/N_w(x, y, l, \theta)$. An example is shown in Fig. 2 and 3. Fig. 2 shows the JPEG2000 compressed ‘Lena’ at 1 bpp and the PSNR is 40.39dB. Fig. 3 shows the JND watermarked ‘Lena’ at 1bpp and the PSNR is 39.01dB. We observed that these two images are visually indistinguishable. The file size and the PSNR of the JPEG2000 compressed image and the watermarked images are shown in Fig. 4 and 5 respectively. The estimated capacities are shown in Fig. 6, the capacity decrease with the bitrate and drops to zero when bitrate is smaller than or equals to 0.25bpp.

### 6. CONCLUSIONS

In this paper, we propose a method to estimate the data hiding capacity of digital images in JPEG2000-to-JPEG2000 watermarking using a human visual system (HVS) model. As our capacity estimation does not assume any specific watermarking method, it should apply to any watermarking methods in the J2K2JK framework.

### ACKNOWLEDGEMENT

This work was supported by Research Grants Council of Hong Kong SAR (HKUST6028/01E).

### REFERENCES


Fig. 2. 'Lena' compressed at 1 bpp.

Fig. 3. JND watermarked 'Lena' at 1 bpp.

Fig. 5. PSNR of compressed images.

Fig. 6. Estimated capacity Vs target bit rate.