Exploring effective coefficients in transform-domain perceptual watermarking

Chih-Wei Tang[1] and Hsueh-Ming Hang[2]
Department of Electronics Engineering
National Chiao-Tung University
Hsinchu, Taiwan, R.O.C.

ABSTRACT

An image watermark parameter optimization procedure is proposed for selecting the most effective DCT coefficients for watermark embedding. Using this set of coefficients improves the watermark robustness and reliability against attack while it maintains the transparency of the embedded watermark. With the aid of prior knowledge of attacks, the visual masking effect and the attack distortion on each (DCT) transform coefficient are pre-calculated so that a maximum strength watermark within visual threshold can be inserted. There are two stages in the design phase. First, taking into account the combined effect of watermark embedding and attack, we pick up the robust coefficients that resist a specific type of attacks and in the meanwhile we keep the distortion lower than the visual threshold. Although typically the watermark detection reliability increases with the increasing number of embedded coefficients, the less effective coefficients may degrade the overall detection performance. Thus, in the second stage, some initially selected coefficients are discarded by an iterative process to reduce the overall error detection probability. Since digital images are often compressed for efficient storage and transmission, we adopt JPEG compression as the attacking source. The simulation results show that the detection error probability is significantly reduced when the selected robust coefficients are in use. These coefficients with watermark embedded on them can also survive color reduction, Gaussian filtering, and frequency mode Laplacian removal (FMLR) attacks.

Keywords: Digital watermark, robust coefficients, watermark detection, watermark capacity, JPEG attack.

1. INTRODUCTION

Many digital watermarking schemes have recently been proposed for copyright protection and other applications due to the rapid growing demand for multimedia data distribution. Watermark designing issues include detection robustness, detection reliability, imperceptibility and capacity. Several works studying the watermark capacity issue have been published using the theoretical analysis approach [23][24][25]. There are tradeoffs between the achievable watermarking rate, allowable distortion for information hiding, and robustness against attacks [24]. It has been reported that the transform–domain watermarking techniques may offer a higher capacity under specific attacks (such as compression) [6][23]. Since our targeting watermarks should be invisible to human eyes, we are especially interested in the watermark capacity and robustness under the combined consideration of reliable detection and visual fidelity. The perceptual watermark capacity in different transform domains is analyzed in [10]. In [25], the capacity constrained by reliable statistical detection is calculated. In [12], the minimum number of coefficients in discrete wavelet domain with spread spectrum watermark embedding is theoretically analyzed using the human visual model and a probabilistic detection model.

These previous researches estimate the watermark capacity bound under certain assumptions, but the exact locations of the coefficients for watermark embedding are not identified.. In this paper, we develop a procedure that find these effective coefficients in natural images, which achieve both detection robustness and watermark invisibility. Since digital images are often compressed for efficient storage and transmission, in this study, JPEG compression is the attacking source although our method can be generalized to the other sources. There are two stages in selecting coefficients. In the first stage, deterministic analysis is applied to pick up the proper coefficients and decide the

[1] chihwei.ee88g@nctu.edu.tw
[2] hmhang@mail.nctu.edu.tw

Security and Watermarking of Multimedia Contents V, Edward J. Delp III, Ping Wah Wong, Editors, Proceedings of SPIE-IS&T Electronic Imaging, SPIE Vol. 5020 (2003)

watermark strength so that the attacked coefficients still bear the valid mark and the total distortion is within the visual threshold. However, the real attack may be somewhat different from the one in the design phase. Hence, we calculate the statistical features of images and attacks and discard the coefficients that reduce detection reliability (increase error probability). In Section 2, the robust and imperceptible coefficient selection process is developed. Section 3 describes the human visual masking model used in our example. Section 4 contains the description of the reliability improvement process. Section 5 and 6 cover the details of the watermark embedding and detection procedures. Simulation results summarized in Section 7 will show the performance of our scheme. Finally, section 8 concludes this presentation.

2. ROBUST AND IMPERCEPTIBLE COEFFICIENT SELECTION

Our goal is to achieve the maximum detection robustness while the watermark imperceptible property has to be retained. Several factors affect the watermark detection ability. In the case of transform-domain watermark embedding, the first item one may consider is to use more coefficients. However, some coefficients with low energy, say, may be inappropriate for carrying watermarks. Similar to the signal design process in digital communications through noisy channels, signals (now transform coefficients) have to be carefully selected to achieve the robustness goal. Increasing the magnitude of watermark generally increases the watermark robustness. But on the other hand, large-magnitude changes on coefficients may be perceptually visible. Also different types (and amount) of attacks produce different-levels of damages on the watermarks. The coefficients that can tolerate a specific type of attack can be identified with the aid of damage analysis on potential attacks [29]. The analysis on the visibility of embedded watermarks becomes quite complicated when both attacks and watermarks co-exist. In [11], the author suggests that the joint distortion due to watermarking and the attack (compression) on the original host data should be kept lower than the just noticeable difference (JND) of the human perceptual system. Therefore, the watermark capacity is also constrained by the human visual threshold.

We first assume both the attacking method and the watermark embedding method are known. In our experiment, JPEG compression is used as the attacking method. And as said earlier, we adopt the transform-domain warermarking embedding technique. Now, the robustness of watermark can be increased by either selecting proper coefficients or adjusting watermark embedding parameters. For example, a DCT coefficient is more robust if it is larger than half of the quantization step size before and after embedding [19]. If a DCT coefficient is modified to an integral multiple of a certain step size, which is larger than all the allowable quantization steps used in the JPEG compression attack, then the modified value of this watermarked coefficient can be correctly reconstructed after JPEG compression [9].

We adopt the DCT-domain additive embedding scheme for data hiding. An original DCT coefficient $x[i]$ is positively watermarked if the watermark bit $w[i]$ is +1, and it is negatively watermarked if $w[i]$ is –1. That is,

$$x'[i] = \begin{cases} x[i] + \alpha[i] \cdot w[i] = x[i] + \alpha[i], & \text{if } w[i] = +1 \\ x[i] - \alpha[i] \cdot w[i] = x[i] - \alpha[i], & \text{if } w[i] = -1 \end{cases},$$

where $\alpha[i]$ is the watermark strength for coefficient $x[i]$, and its value is decided by the visual threshold as stated below.

There are two stages in our proposed scheme. In the first robustness and imperceptibility coefficient selection stage, the attack (JPEG quantization) effect on the watermarked coefficients are examined for robustness and imperceptibility. One DCT coefficient is declared robust if both its positive watermark embedding and negative watermark embedding can survive the attack.

On the other hand, since the human eyes are rather sensitive to low-frequency coefficient variations, we do not embed the watermark in the DC coefficients. The robustness and imperceptibility of all AC coefficients are examined. Two criterions are used to select DCT coefficients: (i) the embedded watermark bit can still be detected correctly after JPEG quantization, and (ii) the joint effect of watermark embedding and JPEG quantization is perceptually invisible.

The flowchart of this coefficient selection process is shown in Fig. 1. Initially, the embedded watermark strength $\alpha[i]$ of the ith AC coefficient $x[i]$ is set to be the visual masking threshold $hvs[i]$ of this coefficient (details to be described in Section 3). Let the quantization step size of JPEG compression with quality factor q be Δ_q. The JPEG compression is applied to both the positively watermarked and the negatively watermarked values of the same DCT

coefficient. As a result, the distortion between the unwatermarked AC coefficient $x[i]$ and its quantized positively watermarked coefficient $x'[i]$ is $e_{posw}[i]$, where

$$e_{posw}[i] = \begin{cases} \Delta_q \cdot Round\left\lfloor \dfrac{\Delta_q / 2 + (x[i] + hvs[i])}{\Delta_q} \right\rfloor - x[i], \text{if } (x[i] + hvs[i]) \geq 0 \\ \Delta_q \cdot Round\left\lfloor \dfrac{-\Delta_q / 2 + (x[i] + hvs[i])}{\Delta_q} \right\rfloor - x[i], \text{if } (x[i] + hvs[i]) < 0 \end{cases}, \tag{1}$$

and the distortion between the original AC coefficient $x[i]$ and its quantized negatively watermarked coefficient $x'[i]$ is $e_{negw}[i]$, where

$$e_{negw}[i] = \begin{cases} \Delta_q \cdot Round\left\lfloor \dfrac{\Delta_q / 2 + (x[i] - hvs[i])}{\Delta_q} \right\rfloor - x[i], \text{if } (x[i] - hvs[i]) \geq 0 \\ \Delta_q \cdot Round\left\lfloor \dfrac{-\Delta_q / 2 + (x[i] - hvs[i])}{\Delta_q} \right\rfloor - x[i], \text{if } (x[i] - hvs[i]) < 0 \end{cases}. \tag{2}$$

A DCT coefficient is retained in the candidate set if both its positively and negatively watermarked values satisfy the aforementioned two criterions. First, the robustness criterion requires that the watermark bit is correctly detected after JPEG compression; that is, the sign of $e_{posw}[i]$ is +1 and the sign of $e_{negw}[i]$ is –1. Second, the imperceptibility criterion requires that the watermarked and JPEG compressed coefficients do not depart from their original values by the visual threshold; that is, both of $e_{posw}[i]$ and $|e_{negw}[i]|$ should not be greater than $hvs[i]$. If the second criterion is not satisfied, we adjust the watermark strength $\alpha[i]$ to meet the above criterions if it is achievable. Let

$$e_{org}[i] = \begin{cases} \Delta_q \cdot Round\left\lfloor \dfrac{\Delta_q / 2 + x[i]}{\Delta_q} \right\rfloor - x[i], \text{if } x[i] \geq 0 \\ \Delta_q \cdot Round\left\lfloor \dfrac{-\Delta_q / 2 + x[i]}{\Delta_q} \right\rfloor - x[i], \text{if } x[i] < 0 \end{cases}. \tag{3}$$

If either the sign of $e_{posw}[i]$ or the sign of $e_{negw}[i]$ is equal to that of $e_{org}[i]$, the watermark strength $\alpha[i]$ is set to $e_{org}[i]$ since JPEG quantization error $e_{org}[i]$ on $x[i]$ is visually tolerable in JPEG compression.

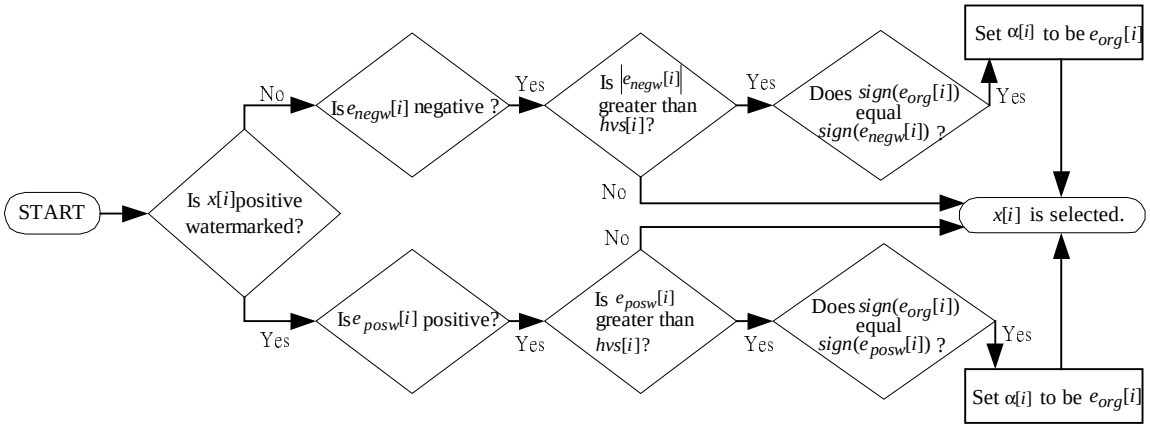

Figure 1: Robust and imperceptible coefficient selection stage.

The sets of robust coefficients under JPEG compression corresponding to different quality factors are thus produced. Since Δ_q decreases as the quality number of the JPEG compression increases, a coefficient that survives the JPEG compression with a low quality factor (such as 50) usually also survives the JPEG compression with a higher quality factor (such as 90). Therefore, the number of the coefficients passing a low-quality-factor JPEG attack is smaller than that passing a higher quality-factor attack. These selected coefficients will be further screened in Section 4.

3. THE VISUAL MASKING EFFECT IN THE TRANSFORM DOMAIN

From the watermarking viewpoint, we are particularly interested in the masking properties of the human visual system (HVS) [4]. Masking effect means that the visibility of one (image) signal is changed due to the existence of the other (image) signal. Several visual masking effects have been identified such as spatial masking, luminance masking, and contrast masking. In watermarking applications, a watermark may become invisible due to the existence of the original image as predicted by the masking effect.

The inclusion of human perceptual characteristics into the watermarking design process helps maintaining the watermark imperceptibility,. Another advantage of this approach is that if the watermarked image spectra is similar in shape to the spectrum of the original image then the attackers cannot easily identify the embedded watermark by using some prior knowledge on the image's statistics [16][17][27][28]. Examples of image adaptive visual watermarking schemes can be found in [5][7][22]. Details about perceptual masking effects can be found in [1][3][5].

In the following procedure, we adopt the visual masking model in the DCT domain since both the JPEG compression and our watermark embedding process are conducted in the DCT domain. The visual masking thresholds are calculated only for AC coefficients because the DC coefficients are not marked. Therefore, the Watson's DCT-based visual model is employed to calculate the contrast masking threshold e_{mnk} of the AC coefficient $x[i]$ at the 2-D frequency index (m,n) of block k. The visual threshold $hvs[i]$ is then set to e_{mnk} and it is used to adjust the watermark embedding strength $\alpha[i]$ as described in Section 2. The contrast masking often has the strongest impact (masking) on the visual effect.

In our experiment, the values of the parameters for contrast masking threshold computation are the same as those in the Checkmark package [30]. This set of setting is decided through subjective tests and is widely adopted in research. More details can be found in [26] and [30]. Here, we briefly describe its computational steps.

1. Set W_x to $(180/\pi)\times(v/(344\times d_v))$, W_y to $(180/\pi)\times(u/(342\times d_v))$, where v is the vertical screen image size, u is the horizontal screen image size, and d_v is the viewing distance. In our experiments, v is 8.8, u is 9.4, and d_v is 72.
2. Calculate $f_{mn}=\frac{1}{16}\sqrt{(m/W_x)^2+(n/W_y)^2}$, where (m,n) is the 2-D frequency index of the 8×8 DCT block.
3. Calculate $r_{mn}=r+(1-r)\cos^2\theta_{mn}$, where $\theta_{mn}=\sin^{-1}\frac{2f_{m0}f_{0n}}{f_{mn}^2}$ and $r=0.7$.
4. Calculate $T_{\min}=\begin{cases}\frac{L}{S_0} & \text{if } L>L_T\\ \frac{L}{S_0}(\frac{L_T}{L})^{1-a_t} & \text{if } L\le L_T\end{cases}$, where $L=10^{7/(255\times 128)-4)}$, $L_T=13.45$, $S_0=94.7$, and $a_t=0.649$. L is the mean luminance of the 8×8 blocks.
5. Calculate $k=\begin{cases}k_0 & \text{if } L>L_k\\ k_0(\frac{L}{L_k})^{a_k} & \text{if } L\le L_k\end{cases}$, where $L_k=300$, $k_0=3.125$, and $a_k=0.0706$.
6. Calculate $f_{\min}=\begin{cases}f_0 & \text{if } L>L_f\\ f_0(\frac{L}{L_f})^{a_f} & \text{if } L\le L_f\end{cases}$, where $L_f=300$, $a_f=0.182$, and $f_0=6.78$.
7. Calculate $\log_{10}t_{mn}=\log_{10}\frac{T_{\min}}{r_{mn}}+k(\log_{10}f_{mn}-\log_{10}f_{\min})^2$.
8. Calculate the luminance masking threshold at frequency index (m,n) for block k: $t_{mnk}=t_{mn}(\frac{c_{00k}}{c_{00}})^{a_t}$, where $c_{00}=128\times 8$, and c_{00k} is the DC coefficient of block k.

9. Calculate the contrast masking threshold: $e_{mnk} = \max[t_{mnk}, |c_{mnk}|^{w_{mn}} \cdot t_{mnk}^{1-w_{mn}}]$, where w_{mn} is chosen experimentally. We set $w_{00} = 0$ and $w_{mn} = 0.7$ for all the other coefficients. Note that c_{mnk} is the (m,n) AC coefficient of block k.

4. DETECTION RELIABILITY IMPROVEMENT

A watermarking system can be viewed as a communication system with, possibly, side information [13]. Thus, if the watermark detector is known and the type of attacks is also known in advance, the coefficients that have higher detection error probability can be predicted and dropped to improve the overall detection reliability. The so-called error probability includes both the false positive probability P_{FP} and false negative probability P_{FN}. The probability that an unwatermarked image is wrongly declared watermarked by the detector is P_{FP}. On the other hand, the error probability of undetected watermark is P_{FN}. The average error probability is $P_{error} = (P_{FP} + P_{FN})/2$ if we assume an image is equally likely marked or unmarked. Let H_0 denote the state that an image is unwatermarked and H_1 denote the watermarked state.

In the second stage that DCT coefficients are further screened to enhance the detection reliability, we need an attack distortion model. In other words, how the detection error probability of a particular coefficient is affected by the JPEG compression. We first collect statistics from the real data. We apply the (JPEG) attack to the watermarked image. Then, the error probability is estimated based on a statistical model of the distorted watermarked coefficients. In [14][15], a theoretical model for additive watermarks under JPEG quantization effect is proposed based on the dither quantization theory [20]. The pseudo-noise watermark and the original image are assumed to be independent. It was shown that the JPEG quantization distortion on individual coefficient cannot be approximated by an AWGN channel. In particular, the distributions of fine quantization errors and coarse quantization errors are different. Therefore, we do not apply the normal distribution model to the individual coefficient. Instead, we adopt the approach based on the central limit theorem [15]. That is, the mean value of the normalized correlation sum is approximated by the normal distribution. This model can be extended to other attacking sources.

The candidate coefficients that have passed the robustness (and imperceptibility) process in Section 2 are examined against the reliability test at the reliability improvement stage. Only one coefficient is discarded in each iteration. The coefficient discarding process is repeated until the overall error probability cannot be reduced further more. At the beginning of one iteration, if there are N remaining coefficients, N candidate sets are formed by deleting one coefficient alternatively in this N-coefficient set. Consequently, there are N-1 coefficients in each candidate set. Then, the statistics based on the Bayes' decision rule for watermark detection for each candidate set is calculated individually. The set with the lowest error detection probability is retained if the overall detection error probability decreases monotonically. The derivation is described below.

The detection error probability is calculated based on the watermark detection rule. Here, the watermark detection rule is designed to minimize the average cost using the Bayes' rule. The binary hypotheses of watermark detection for a received image are [13][19]

$$\begin{aligned} \mathrm{H}_0 &: y[i] = (x[i] + e_{\mathrm{H}_0}[i]) - x[i] = e_{\mathrm{H}_0}[i] \\ \mathrm{H}_1 &: y[i] = (x[i] + d[i] + e_{\mathrm{H}_1}[i]) - x[i] = d[i] + e_{\mathrm{H}_1}[i] \end{aligned},$$

where $y[i]$ is the difference between the received coefficient and the unwatermarked coefficient $x[i]$, $d[i]$ is the embedded watermark, $e_{\mathrm{H}_0}[i]$ is the attack distortion on the original coefficient, and $e_{\mathrm{H}_1}[i]$ is the attack distortion on the watermarked coefficient.

Let $c[i]$ be the normalized correlation value between $y[i]$ and $d[i]$, and C is the mean value of the normalized correlation sum. Let c_{10} be the cost of the false positive decision, c_{01} be the cost of the false negative decision, c_{00} be the cost of detecting watermark correctly, and c_{11} be the cost of detecting the absence of watermark correctly. Then, the Bayes' decision rule is choosing H_1 if [21]

$$\mathrm{H}_1 : \quad \frac{P(C \mid \mathrm{H}_1)}{P(C \mid \mathrm{H}_0)} > K = \frac{(c_{10} - c_{00}) \cdot P(\mathrm{H}_0)}{(c_{01} - c_{11}) \cdot P(\mathrm{H}_1)}, \tag{4}$$

where C is an estimated value of $\mathrm{E}\{c\}$,

$$\mathrm{E}\{c\} \approx C = \frac{1}{M}\sum_{i=1}^{M} c[i] = \frac{1}{M}\sum_{i=1}^{M} \frac{y[i]\cdot d[i]}{\sigma_d^2} \quad \text{and} \quad d[i] = w[i]\cdot\alpha[i]\,. \tag{5}$$

As described earlier that $w[i]$ is the watermark signature with antipodal signaling {-1,1}, and $\alpha[i]$ is the adjustable watermark embedding strength of $x[i]$ (described in Section 2). In (5), M is the number of the watermarked coefficients, and σ_d^2 is the variance of embedded watermark signals. Since in the denominator, σ_d^2, rather than $\frac{1}{M}\sqrt{\sum_{i=1}^{M} y^2[i]\cdot\sum_{i=1}^{M} d^2[i]}$, is in use, C is not bounded to [-1, 1] and could be any value in $(-\infty,\infty)$. When M is sufficiently large, the probability distribution of C can be approximated by the Gaussian distribution according to the central limit theorem [2].

The variance of C is

$$\mathrm{Var}\{C\} = \frac{1}{M}\mathrm{Var}\{c\}\,. \tag{6}$$

Therefore, the left hand side of the decision rule (4) becomes

$$\mathrm{H}_1: \quad \frac{(2\pi\,\mathrm{Var}\{c\,|\,\mathrm{H}_1\}/M)^{-1/2}\exp\left(-\frac{(C-\mathrm{E}\{c\,|\,\mathrm{H}_1\})^2}{2\mathrm{Var}\{c\,|\,\mathrm{H}_1\}/M}\right)}{(2\pi\,\mathrm{Var}\{c\,|\,\mathrm{H}_0\}/M)^{-1/2}\exp\left(-\frac{(C-\mathrm{E}\{c\,|\,\mathrm{H}_0\})^2}{2\mathrm{Var}\{c\,|\,\mathrm{H}_0\}/M}\right)} > K\,. \tag{7}$$

Equivalently,

$$\begin{aligned}\mathrm{H}_1: \quad & 2\log K + \log\left(\frac{\mathrm{Var}\{c\,|\,\mathrm{H}_1\}}{\mathrm{Var}\{c\,|\,\mathrm{H}_0\}}\right) + \frac{\mathrm{E}^2\{c\,|\,\mathrm{H}_1\}}{\mathrm{Var}\{c\,|\,\mathrm{H}_1\}/M} - \frac{\mathrm{E}^2\{c\,|\,\mathrm{H}_0\}}{\mathrm{Var}\{c\,|\,\mathrm{H}_0\}/M} \\ & < \left(\frac{1}{\mathrm{Var}\{c\,|\,\mathrm{H}_0\}/M} - \frac{1}{\mathrm{Var}\{c\,|\,\mathrm{H}_1\}/M}\right)\cdot C^2 + 2\left(\frac{\mathrm{E}\{c\,|\,\mathrm{H}_1\}}{\mathrm{Var}\{c\,|\,\mathrm{H}_1\}/M} - \frac{\mathrm{E}\{c\,|\,\mathrm{H}_0\)}{\mathrm{Var}\{c\,|\,\mathrm{H}_0\}/M}\right)\cdot C\end{aligned}\,. \tag{8}$$

Finally, the maximum-likelihood (ML) detector is obtained with K=1 in (4) assuming that (i) $c_{10}-c_{00}=c_{01}-c_{11}$, and (ii) $P(H_0)=P(H_1)$. Then, (8) can be simplified and expressed as $(C-x_1)(C-x_2)>0$. The detection threshold x_c is either x_1 or x_2 as its value should locate between $\mathrm{E}\{c\,|\,\mathrm{H}_0\}$ and $\mathrm{E}\{c\,|\,\mathrm{H}_1\}$. As a result, the image is declared watermarked if $C > x_c$. Consequently,

$$P_{FP} = \int_{x_c}^{\infty} \frac{\exp\left(-\frac{(x-\mathrm{E}\{c\,|\,\mathrm{H}_0\})^2}{2\,\mathrm{Var}\{c\,|\,\mathrm{H}_0\}/M}\right)}{\sqrt{2\pi\mathrm{Var}\{c\,|\,\mathrm{H}_0\}/M}}dx = \frac{1}{2}erfc\left(\sqrt{M}\,\frac{x_C-\mathrm{E}\{c\,|\,\mathrm{H}_0\}}{\sqrt{2\,\mathrm{Var}\{c\,|\,\mathrm{H}_0\}}}\right), \tag{9}$$

$$P_{FN} = \int_{-\infty}^{x_c} \frac{\exp\left(-\frac{(\mathrm{E}\{c\,|\,\mathrm{H}_1\}-x)^2}{2\,\mathrm{Var}\{c\,|\,\mathrm{H}_1\}/M}\right)}{\sqrt{2\pi\mathrm{Var}\{c\,|\,\mathrm{H}_1\}/M}}dx = \frac{1}{2}erfc\left(\sqrt{M}\,\frac{\mathrm{E}\{c\,|\,\mathrm{H}_1\}-x_C}{\sqrt{2\,\mathrm{Var}\{c\,|\,\mathrm{H}_1\}}}\right). \tag{10}$$

To estimate P_{FP} and P_{FN} in (9) and (10), the statistics $\mathrm{E}\{c\,|\,\mathrm{H}_0\}$, $\mathrm{E}\{c\,|\,\mathrm{H}_1\}$, $\mathrm{Var}\{c\,|\,\mathrm{H}_0\}$, and $\mathrm{Var}\{c\,|\,\mathrm{H}_1\}$ are derived from the image data. We assume a coefficient is equal likely being positively or negatively watermarked. Then, $\mathrm{E}\{c\,|\,\mathrm{H}_0\}$, $\mathrm{E}\{c\,|\,\mathrm{H}_1\}$, $\mathrm{Var}\{c\,|\,\mathrm{H}_0\}$, and $\mathrm{Var}\{c\,|\,\mathrm{H}_1\}$ are calculated by the following equations:

$$\mathrm{E}\{d\} = \frac{1}{M}\sum_{i=1}^{M} d[i],\ where\ \ d[i] = \alpha[i]\cdot w[i] \tag{11}$$

$$\sigma_d^2 = \mathrm{Var}\{d\} = \frac{1}{M-1}\cdot\sum_{i=1}^{M}(d[i]-\mathrm{E}\{d\})^2 = \left(\frac{1}{M-1}\sum_{i=1}^{M} d^2[i]\right) - \left(\frac{M}{M-1}\mathrm{E}^2\{d\}\right) \tag{12}$$

$$\mathrm{E}\{c\,|\,\mathrm{H}_0\} = \frac{1}{M}\sum_{i=1}^{M} c_{\mathrm{H}_0}[i] = \frac{1}{M}\sum_{i=1}^{M}\frac{y_{\mathrm{H}_0}[i]\cdot d[i]}{\sigma_d^2} = \frac{1}{M}\sum_{i=1}^{M}\frac{e_{\mathrm{H}_0}[i]\cdot d[i]}{\sigma_d^2} \tag{13}$$

$$\mathrm{E}\{c\,|\,\mathrm{H}_1\} = \frac{1}{M}\sum_{i=1}^{M} c_{\mathrm{H}_1}[i] = \frac{1}{M}\sum_{i=1}^{M}\frac{y_{\mathrm{H}_1}[i]\cdot d[i]}{\sigma_d^2} = \frac{1}{M}\sum_{i=1}^{M}\frac{(d[i]+e_{\mathrm{H}_1}[i])\cdot d[i]}{\sigma_d^2} \tag{14}$$

$$\mathrm{Var}\{c \mid \mathrm{H}_0\} = \frac{1}{M-1} \cdot \sum_{i=1}^{M} (c_{\mathrm{H}_{01}}[i] - \mathrm{E}\{c \mid \mathrm{H}_0\})^2 \tag{15}$$

$$= \frac{1}{M-1} \sum_{i=1}^{M} \left(\frac{e_{\mathrm{H}_0}[i] \cdot d[i]}{\sigma_d^2}\right)^2 - \left(\frac{M}{M-1} \mathrm{E}^2\{c \mid \mathrm{H}_0\}\right)$$

$$\mathrm{Var}\{c \mid \mathrm{H}_1\} = \frac{1}{M-1} \cdot \sum_{i=1}^{M} (c_{\mathrm{H}_1}[i] - \mathrm{E}\{c \mid \mathrm{H}_1\})^2 \tag{16}$$

$$= \frac{1}{M-1} \sum_{i=1}^{M} \left(\frac{(d[i] + e_{\mathrm{H}_1}[i]) \cdot d[i]}{\sigma_d^2}\right)^2 - \left(\frac{M}{M-1} \mathrm{E}^2\{c \mid \mathrm{H}_1\}\right)$$

The above parameters are computed from the candidate coefficient sets. That is, in each iteration, the average error probability based on the average of (9) and (10) are computed for every candidate set. And the best one is selected as described earlier. This iteration process continues until the average error does not decrease any more. Note that the coefficients sets associated with different JPEG compression quality numbers are different as discussed in Section 2.

5. WATERMARK EMBEDDING SCHEME

The watermark embedding process is outlined in Fig. 2. At beginning, an original image is converted to some transform domain and the visual masking thresholds of all transform coefficients are calculated. The robust and imperceptible coefficient selection process (Robustness stage in Section 2) generates robust coefficients within the visual fidelity. We then perform the reliability detection improvement process (Reliability stage in Section 4) on the candidate coefficients iteratively to decrease the detection error probability. Next, watermarks are embedded on the selected coefficients in the DCT domain. Finally, the watermarked DCT-domain image is converted back to the spatial domain.

In our experiment, the original image is transformed by 8×8 2-D DCT and the contrast masking thresholds are calculated for all AC coefficients. After the locations of robust coefficient are determined by the Robustness and the Reliability stages, multiple watermark sequences are embedded to the selected coefficients subject to different JPEG compression quality factors. Let the DCT coefficients of the same 2-D frequency index constitute one sub-channel and there are thus totally 63 sub-channels. Typically, the AC coefficients in a subchannel can be modeled as generalized Gaussian distribution source[18]. Then, sub-channels containing the selected coefficients are watermarked in the raster-scan order of 8×8 blocks. The watermarked coefficient is generated by $x'[i] = x[i] + \alpha[i] \cdot w[i]$. The watermark strength $\alpha[i]$ of this watermark bit is determined by (4) and (5) in Section 2. Finally, the watermarked 8×8 blocks of coefficients are converted back to the spatial domain by 2-D IDCT.

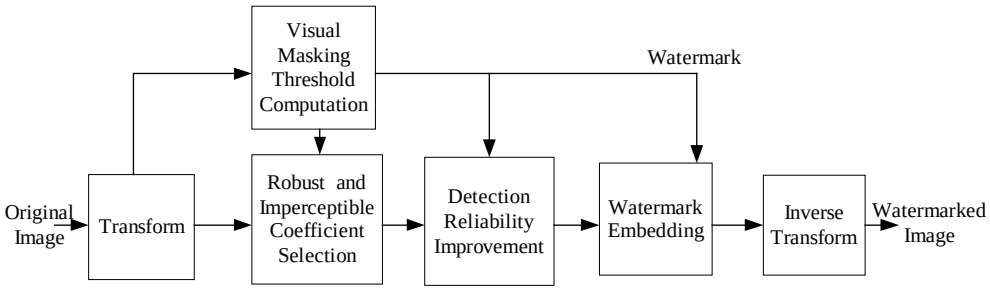

Figure 2: Watermark embedding scheme.

6. WATERMARK DETECTION SCHEME

Figure 3 shows the block diagram of our watermark detection scheme. The robust coefficients of an original image are first extracted in transform domain. Since the locations of watermarked coefficients are image-dependent, they have to be found with the aid of the original image during watermark detection. The Robustness and the Reliability stages are the same as those at watermark embedding. The visual masking thresholds are determined based on the original image. In fact, the watermark information in embedding such as the number and locations of embedded coefficients can be recorded for detection purpose. Finally, watermark sequences are extracted from the received image and correlate with the original watermark for binary hypothesis testing and decision.

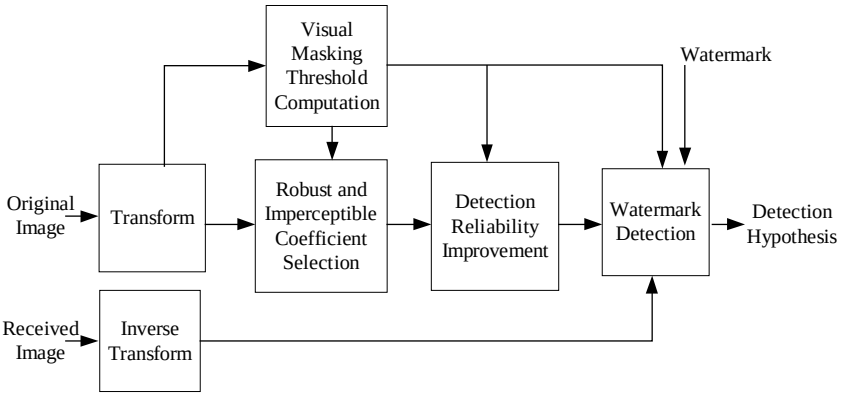

Figure 3: Watermark detection scheme.

Multiple watermarks designed for different JPEG compression quality may be inserted. For each watermark sequence, the hypothesis decision is performed based on the Bayes' decision rule. The mean value of the normalized correlation sum C is computed by

$$C = \frac{1}{M}\sum_{i=1}^{M} c[i] = \frac{1}{M}\sum_{i=1}^{M} \frac{y[i]\cdot d[i]}{\sigma_d^2} \text{ with } d[i] = w[i]\cdot\alpha[i], \tag{17}$$

where $y[i]$ is the DCT coefficient in the received image, $\alpha[i]$ is the watermark strength of coefficient $x[i]$, and $w[i]$ is the watermark signature. Then, C is compared against the threshold x_c defined by (8). Finally, the binary detection hypothesis corresponding to each watermark sequences is conducted. The presence of the watermark is declared if at least one watermark sequence (among multiple watermarks) is successfully detected.

7. SIMULATION RESULTS

We tested the proposed watermarking scheme on the 512×512 images, Lena and Baboon. Due to the limitation of space, the experiment results are listed here mainly for the image Lena. The JPEG compression quantization step size defined in StirMark 3.1[31] is used in the Robustness and the Reliability stage, and it is defined by

$$Scale = \begin{cases} 5000/quality & \text{, if quality} < 50 \\ 200 - quality\times 2 & \text{, otherwise.} \end{cases} \tag{18}$$

$$QuanStepSize[i] = (BasicQuanMatrix[i]\times Scale + 50)/100.$$

Sets of robust coefficients are generated corresponding to JPEG compression quality factors range from 50 to 85 with step size 5. The difference images between the original images and the watermarked images in the spatial domain are magnified by a factor of 30 and are shown in Figs. 4 (a) and (b). The watermark mainly spreads over the visual significant portions as we expect.

The properties of the selected coefficients corresponding to different JPEG compression quality factor after two processing stages for image Lena are shown in Table 1. The number of dropped coefficients and the improved error detection probability for higher JPEG compression quality factors are usually larger than those for lower JPEG compression quality factors. This is partially because there are more candidate coefficients surviving higher JPEG compression quality factors in the Robustness stage. However, the estimated error probability of the selected robust coefficients is relatively small because the detection is done with the original image and the attacking source is assumed known in the design phase.

The estimated statistics for selected coefficients corresponding to different JPEG compression quality factors after detection reliability improvement stage but before watermark embedding is shown in Table 2 for image Lena. The experimental results show that the variance of the embedded watermark strength $\mathrm{Var}\{d \mid \mathrm{H}_1\}$ is larger for lower JPEG compression quality factors since the attack produced by a lower JPEG compression quality factor will cause higher

quantization distortion. We also observe that $\mathrm{Var}\{c \mid \mathrm{H}_0\}$ is not equal to $\mathrm{Var}\{c \mid \mathrm{H}_1\}$ for real images. $\mathrm{E}\{c \mid \mathrm{H}_0\}$ is not equal to zero but it is usually very close to zero. $\mathrm{E}\{c \mid \mathrm{H}_1\}$ depends on the distortion model (attack) and it is approximately 0.74 for different JPEG compression quality factors in our case. The detection threshold x_c is computed from (8), and it is roughly near the average value of $\mathrm{E}\{c \mid \mathrm{H}_0\}$ and $\mathrm{E}\{c \mid \mathrm{H}_1\}$ as we expect.

StirMark 3.1[31] is used to test the robustness of the watermark. The data shown in Tables 3 to 6 are each averaged over 10 different watermark pseudo random sequences. As shown in Tables 3 and 4 for images Lena and Baboon, respectively, our scheme can survive JPEG compression at different quality factors. Attacks of JPEG compression at other quality factors are applied to a watermark sequence designed for a different compression quality factor and the detection results sometimes are even better.

Although our scheme is designed as a JPEG-robust watermarking scheme, it can resist many other signal processing attacks including color reduction, Gaussian filtering, median filtering and frequency mode Laplacian removal (FMLR). The results are shown in Tables 5 and 6. Our watermark can survive several combinations of attacks such as JPEG compression together with Gaussian filtering and JPEG compression together with FMLR attacks. However, the combined attacks of JPEG compression with 4×4 median filtering may fail our scheme. The reason may be due to the common lowpass filtering property of the Gaussian filtering and JPEG compression. Therefore, essentially, the attacked images are heavily lowpass filtered. In fact, this attack damages image quality.

8. CONCLUSIONS

In this paper, a selection procedure is designed to identify the most effective DCT coefficients for watermarking purpose. The target is to improve both the watermark robustness and its detection reliability under the JPEG compression attack. There are essentially two steps in the design phase. Candidate coefficients and watermark signal (strength) are first picked to achieve both robustness against JPEG attack and perceptual invisibility. Then, we examine the error probability of using these candidate coefficients. The weak ones that lower the detection probability are discarded. Finally, we obtain a set of robust coefficients, which are both picture and attack dependent. However, because many attacks have similar statistical characteristics, our designed watermark can survive many other types of attacks too. Our simulations show that the proposed watermarking scheme performs very well in achieving high detection probability while maintaining the transparency of the embedded watermark. The methodology presented here for finding the most effective coefficients can be extended to the other types of attacks and/or watermarking techniques.

(a) (b)

Figure 4: The (absolute) difference image between the original image and the watermarked image. The magnitude in display is amplified by a factor of 30. (a) Lena and (b) Baboon.

JPEG Quality Factor	No. of Selected Coefficients after Stage 1	No. of Selected Coefficients after Stage 2	Estimated P_{error} after Stage 1	Estimated P_{error} after Stage 2
50	293	289	4.801553e-040	1.399886e-041
55	254	251	1.379195e-031	1.211739e-033
60	302	291	5.086452e-031	1.327997e-036
65	436	403	1.079740e-039	9.773688e-051
70	706	643	6.579702e-059	4.829724e-078
75	909	789	2.441071e-065	8.627588e-084
80	1735	1453	4.439007e-091	2.905769e-151
85	2793	2531	8.608640e-133	3.578031e-211

Table 1: The properties of the selected coefficients corresponding to different JPEG compression quality factor after the robustness and imperceptibility stage (stage 1) and after the reliability improvement stage (stage 2) for image Lena.

JPEG Quality Factor	$E\{c \mid H_0\}$	$Var\{c \mid H_0\}$	$E\{c \mid H_1\}$	$Var\{c \mid H_1\}$	$E\{d \mid H_1\}$	$Var\{d \mid H_1\}$	Detection Threshold x_c
55	-0.01831	0.37449	0.74550	0.15477	0.05499	57.53850	0.44606
65	-0.01842	0.40716	0.74495	0.15064	0.02267	33.56159	0.45583
75	-0.03309	0.39831	0.73687	0.23630	-0.00372	21.52627	0.40172
85	-0.06030	0.37031	0.73288	0.46117	-0.00635	11.07714	0.31457

Table 2: The estimated statistics of the selected coefficients corresponding to different JPEG compression quality factors after the reliability improvement stage before watermark embedding/detection for image Lena.

		Quality Factor of JPEG in Design Phase							
		50	55	60	65	70	75	80	85
JPEG	50	0.75828	1.08891	1.04426	1.02631	1.00508	1.15911	0.93469	0.97188
Attack	60	0.99471	1.01038	0.76611	1.02535	1.02814	0.98421	1.05550	0.97467
Quality	70	0.97866	1.02394	0.99183	0.98873	0.78478	1.09376	1.00775	1.07304
Factor	80	0.98994	0.99587	0.97218	1.00782	0.99724	1.00715	0.78528	1.01181
	90	0.99550	0.99733	0.99988	1.00381	1.00815	1.00796	0.98741	0.96969

Table 3: The mean of the normalized correlation sum C of the JPEG robust watermark sequences corresponding to different JPEG compression quality factors for image Lena.

		Quality Factor of JPEG in Design Phase							
		50	55	60	65	70	75	80	85
JPEG	50	0.76344	1.13343	1.02823	0.99543	1.02616	1.25157	0.94913	0.95039
Attack	60	0.99721	1.07403	0.76942	1.02157	0.92880	1.02635	0.97016	1.03013
Quality	70	0.98575	0.97904	0.98298	0.98360	0.76418	1.07157	1.05176	1.09250
Factor	80	0.98615	0.98364	0.97140	1.02317	0.99243	0.98069	0.77840	1.01349
	90	0.99654	0.99814	0.99935	1.00039	1.00022	1.00090	0.99916	0.99564

Table 4: The mean of the normalized correlation sum C of the JPEG robust watermark sequences corresponding to different JPEG compression quality factors for image Baboon.

	4x4 Median Filtering+ JPEG90	Gaussian Filtering+ JPEG90	FMLR+JPEG90	Color Reduction
JPEG 55	0.91803	0.84612	0.72875	0.97436
JPEG 65	-0.01058	0.64457	0.50121	0.98607
JPEG 75	0.21487	0.48922	0.52089	0.98635
JPEG 85	0.03617	0.48588	0.56394	0.96424

Table 5: The mean of the normalized correlation sum C of JPEG robust watermark sequences under various signal processing attacks for image Lena.

	4x4 Median Filtering+ JPEG90	Gaussian Filtering+ JPEG90	FMLR+JPEG90	Color Reduction
JPEG 55	0.41853	0.58843	0.88425	0.96525
JPEG 65	0.28093	0.57276	0.85765	0.97251
JPEG 75	0.33408	0.57476	0.85326	0.97441
JPEG 85	0.00109	0.44087	0.78815	0.95511

Table 6: The mean of the normalized correlation sum C of JPEG robust watermark sequences under various signal processing attacks for image Baboon.

ACKNOWLEDGMENTS

This work is partially supported by the Lee & MTI Center for Networking Research at National Chiao Tung University, Taiwan.

REFERENCES

1. A. B. Watson, "Visual optimization of DCT quantization matrices for individual images," *Proceedings, AIAA Computing in Aerospace 9*, San Diego, CA, American Institute of Aeronautics and Astronautics, pp. 286-291, 1993.
2. A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, McGraw-Hill, New York, 3rd edition, 1991.
3. A. J. Ahumada and H. A. Peterson, "Luminance-model based DCT quantization for color image compression," *SPIE Proceedings*, 1666, 365-374, 1992.
4. B. A. Wandell, *Foundations of Vision*, Sunderland, MA: Sinuaer, 1995.
5. C. D. Vleeschouwer, J.-F. Delaigle, and B. Macq, "Invisibility and Application Functionalities in Perceptual Watermarking – An Overview," *Proceedings of the IEEE*, vol. 90, no. 1, January 2002.
6. C. Fei, D. Kundur and R. Kwong, "The choice of watermark domain in the presence of compression," *Proceedings International Conference on Information Technology: Coding and Computing*, pp. 79-84, 2001.
7. C. I. Podilchuk, and W. Zeng, "Image-adaptive watermarking using visual models," *IEEE Journal on Selected Areas in Communications*, vol. 16, No. 4. May 1998.
8. C. S. Lu, H. Y. Liao and M. Kutter, "Denoising and copy attacks resilient watermarking by exploiting prior knowledge at detector," *IEEE Transactions on Image Processing*, vol. 11, no.3, March 2002.
9. C.-Y. Lin and S.-F. Chang, "Semi-fragile watermarking for authenticating JPEG visual content," *Security and Watermarking of Multimedia Contents II, Proceedings of SPIE*, vol. 3971, 2000.
10. C.-Y. Lin and S.-F. Chang, "Watermarking capacity of digital images based on domain-specific masking effects," *IEEE International Conference on Information Technology: Coding and Computing*, Las Vegas, NV, April 2001.
11. D. Kundur, "Water-filling for watermarking," *Proceedings IEEE International Conference On Multimedia and Expo*, New York City, New York, pp. 1287-1290, August 2000.
12. I. Donescu and E. Nguyen, "Combining visual and detection models in spread-spectrum watermarking," *Proceedings IEEE International Conference On Image Processing*, 2000.
13. I. J. Cox, M. L. Miller, and A. L. Mckellips, "Watermarking as communications with side information," *Proceedings of the IEEE*, vol. 87, no. 7, July 1999.
14. J. J. Eggers and B. Girod, "Watermark detection after quantization attacks," *Proc. of 3rd Workshop on Information Hiding*, Germany, September/October 1999.
15. J. J. Eggers and B. Girod, "Quantization effects on digital watermarks," *Signal Processing*, vol. 81, no. 2, p. 239-263, December 2000.
16. J. K. Su and B. Girod, "On the robustness and imperceptibility of digital fingerprints," *Proceedings IEEE International Conference On Multimedia Computing and Systems*, vol. 2, pp. 530-535, 1999.
17. J. K. Su and B. Girod, "Power-spectrum condition for energy-efficient watermarking," *Proceedings IEEE International Conference on Image Processing*, 1999.
18. K. A. Birney and T. R. Fischer, "On the modeling of DCT and subband image data for compression," *IEEE Transactions on Image Processing*, vol. 4, no. 2, February 1995.
19. K. Kamijo, "Optimizing watermarking to improve the robustness without affecting the fidelity," *IEEE International Conference on Information Technology: Coding and Computing*, 2001.
20. L. Schuchman, "Dither signals and their effect on quantization noise," *IEEE Transaction on Communication Technology (COM)*, vol. 12, pp. 162-165, December 1964.
21. M. D. Srinath, P. K. Rajasekaran, and R. Viswanathan, *Introduction to statistical signal processing with applications*, Prentice Hall Englewood Cliffs, New Jersey 07632.
22. M. Kutter and Stefan Winkler, "A Vision-based masking model for spread-spectrum image watermarking," *IEEE Transactions on Image Processing*, vol. 11, no. 1. January 2002.
23. M. Ramkumar, A. N. Akansu and A. A. Alatan, "On the choice of transforms for data hiding in compressed video," *Proceedings, IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3049-3052, 1999.
24. P. Moulin and J. A. O'Sullivan, "Information-theoretic analysis of watermarking," Proceedings, *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6, pp. 3630-3633, 2000.

25. R. Sugihara, "Practical capacity of digital watermark as constrained by reliability," *IEEE Proceedings, International Conference on Information Technology: Coding and Computing*, pp. 85-89, 2001.
26. S. Pereira, S. Voloshynovskiy, M. Madueno, S. Archand-Maillet and T. Pun, "Second generation benchmarking and application oriented evaluation," *In Information Hiding Workshop*, Pittsburgh, PA. USA, April 2001.
27. S. Voloshynovskiy, S. Pereira, T. Pun, J.J. Eggers and J. K. Su, "Attacks on digital watermarks: classification, estimation-based attacks and benchmarks," *IEEE Communications Magazine* (Special issue on digital watermarking for copyright protection: a communications perspective), 39, 8, pp. 118-127, 2001.
28. S. Voloshynovskiy, A. Herrigel, N. Baumgaertner and T. Pun, "A stochastic approach to content adaptive digital image watermarking," *In International Workshop on Information Hiding*, vol. LNCS 1768 of Lecture Notes in Computer Science, pp. 212-236, 1999.
29. T. S. Liang and J. J. Rodriguez, "Robust watermarking using robust coefficients," *Security and Watermarking of Multimedia Contents II, Proceedings of SPIE*, vol. 3971, 2000.
30. Checkmark, http://watermarking.unige.ch/Checkmark/index.html
31. Stirmark, http://www.cl.cam.ac.uk/~fapp2/watermarking/stirmark/