USING PERCEPTUAL MODELS TO IMPROVE FIDELITY AND PROVIDE INVARIANCE TO VALUMETRIC SCALING FOR QUANTIZATION INDEX MODULATION WATERMARKING

Qiao Li, Ingemar J. Cox

Departments of Computer Science and Electronic and Electrical Engineering,University College London
Torrington Place, London, WC1E 7JE, England. Email: q.li@ee.ucl.ac.uk, ingemar@ieee.org

ABSTRACT

Quanitization index modulation (QIM) is a computationally efficient method of watermarking with side information. This paper proposes two improvements to the original algorithm.

First, the fixed quantization step size is replaced with an adaptive step size that is determined using Watson's perceptual model. Experimental results on a database of 1000 images illustrate significant improvements in both fidelity and robustness to additive white Gaussian noise.

Second, modifying the Watson model such that it scales linearly with valumetric (amplitude) scaling, results in a QIM algorithm that is invariant to valumetric scaling. Experimental results compare this algorithm with both the original QIM and an adaptive QIM and demonstrate superior performance.

1. INTRODUCTION

Digital watermarking can be best modeled as communication with side information [1]. The available side information is the original cover Work or host signal, which is entirely known to the watermark embedder (analogous to the transmitter). Earlier models of watermarking were also communications based, but the cover Work was modeled as one of two unknown noise sources. The second noise source modeled the distortions that occur after embedding but before detection. Costa showed [2] that the channel capacity of a communications channel with two noise sources, one of which is entirely known to the transmitter, but both unknown to the receiver, is equivalent to a channel in which the known noise source is absent. It is now recognized that this research has very important implications for digital watermarking. However, Costa's paper did not provide a practical method of implementation.

Quantization index modulation (QIM), first proposed by Chen and Wornell [3], uses a structured lattice code to provide a computational efficient watermarking algorithm with high data capacity. While dithering was proposed to improve the performance and reduce perceptual distortion, the quantization step size is fixed. However, it is well-known that improvements in image fidelity and robustness can be achieved by adapting the watermark strength based on the local perceptual characteristics of the cover Work.

The most serious disadvantage of QIM has been its extreme sensistivity to valumetric scaling. Even small changes in the brightness of an image, or the volume of a song, can result in dramatic increases in the bit error rate.

Several papers [4, 5, 6] have addressed this issue. Eggers *et al* [4] proposed to estimate the valumetric scaling by "securely embedd[ing] SCS pilot watermark". However the need for a calibration problem may lead to security weaknesses. Lee *et al* [5] proposed estimating the global scaling factor using an EM algorithm, which does not need a pilot watermark. However, they note that the "complexity could be impractical". The closest work to ours is that of Oostveen *et al* [6] which uses a simple perceptual model based on Weber's law. We compare our methods with theirs and demonstrate improved performance.

Section 2 discusses the embedding and detecting procedures for QIM. Section 3 describes Watson's perceptual model and Section 4 shows how this model can be used to adaptively set the quantization step size. Section 4.1 experimentally demonstrates significant improvements. Section 5 then describes a modification to Watson's model such that it scales linearly with valumetric changes. This modified perceptual model is then used to choose the quantization step size and Section 5.1 provides experimental verification of the valumetric invariance properties. Experimental comparison is also provided between the new algorithm and the original algorithm of [7] and the valumetric invariant method of [6]. Section 6 summarizes our results.

2. QUANTIZATION INDEX MODULATION

Watermarking with side information is modeled by the communication system shown in Figure 1. The message, m and the cover Work or host signal, x, (i.e. image or song), are input into the watermark embedder, which outputs a watermark, w that is added to the cover Work to produce the watermarked Work, y. The watermarked Work then undergoes a number of distortions that are modelled as an unknown noise source, v. The watermark detector receives a distorted, watermarked Work, r, i.e. $r = x + w + v$, and decodes a message $\hat{m}$.

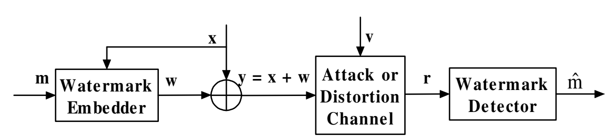

Fig. 1. Watermarking as a communication system

2.1. Embedding of QIM

A quantizer is a function that maps a value to the nearest point belonging to a class of pre-defined discontinuous points. Here, function $round(.)$ denotes rounding value to the nearest integer and the standard quantization operation with step size Δ is defined as $Q(x, \Delta) = round(\frac{x}{\Delta})\Delta$

QIM embeds a message by first modulating an index or sequence of indices with a message to be embedded and then quantizing the host signal with the associated quantizer or sequence of quantizers. Let Δ be the quantization step size and L represent the length of the host signal x and the message m (we embed one bit per sample). If dithering is used, then we choose $d[n, 0]$ pseudo-randomly with a uniform distribution over $[-\Delta/2, \Delta/2]$. and

$$d[n,1] = \begin{cases} d[n,0] + \Delta/2, & d[n,0] < 0. \\ d[n,0] - \Delta/2, & d[n,0] > 0. \end{cases} \quad n = 1, 2, \ldots, L. \tag{1}$$

Here $d[n, 0]$ or $d[n, 1]$ is used for embedding message bit "0" or "1" respectively. The watermarked signal is given by:

$$y_n(x_n, m_n) \quad = Q(x_n + d[n, m_n], \Delta) - d[n, m_n] \tag{2}$$

2.2. Detecting with QIM: Hard and Soft Decision

During detection, the detector calculates two signals $S_r(n, 0)$ and $S_r(n, 1)$ by embedding "0" and "1" into the received signal r separately, in the same manner as Equation (2). The detected message bit is then determined by judging which of these two signals has the minimum Euclidean distance to the received signal r

$$\hat{m_n} = \underbrace{argmin}_{l \in 0,1} (r_n - S_r(n,l))^2 \tag{3}$$

The above description embedded one bit per sample. In practice, we usually spread one message bit into a sequence of N samples. One way to achieve this is to use a rate $1/N$ repetition encoding. Detection can still be performed on a one bit per sample basis. Rate $1/N$ repetition decoding is finally employed to obtain the message. We refer to this as *Hard Decision* detection.

Alternatively, in the detector we can accumulate the two Euclidean distance for N samples and then determine the detected message bit, i.e.,

$$\begin{aligned} &\hat{m_n} = \underbrace{argmin}_{l \in 0,1} \textstyle\sum_{h=(n-1)N+1}^{nN} (r_h - S_r(h,l))^2, \\ &n = 1, 2, \ldots, L/N. \end{aligned} \tag{4}$$

The code rate is also $1/N$ but this *Soft Decision* decoding usually outperforms hard decision decoding.

For non-adaptive QIM, the quantization step size is independent of the content. However, it is well known that the ability to perceive a change depends on the content. For example, the human visual system is much less sensitive to changes in heavily textured regions and much more sensitive to changes in uniform regions. To account for this, we propose using a perceptual model to automatically select the quantization step size at each sample.

3. WATSON PERCEPTUAL MODEL

In this section, we describe the Watson's perceptual model [8]. This model estimates the perceptibility to changes in individual terms of an image's block DCT. We denote one term of the kth block of the cover Work, C, by $C[i, j, k], 0 \le i, j \le 7$. $C[0, 0, k]$ is the DC term, i.e.,the mean pixel intensity in the block. Watson's model consists of a sensitivity function, two masking components based on luminance and contrast masking, and a pooling component.

Sensitivity

The model defines a frequency sensitivity table, t. Each table entry, $t[i, j]$, is approximately the smallest magnitude of the corresponding DCT coefficient in a block that is discernible in the absence of any masking noise. The resulting frequency sensitivity table is shown in [9]. Note that it is a constant value table.

Luminance Masking

Luminance adaptation refers to the fact that a DCT coefficient can be changed by a larger amount before being noticed if the average intensity of the 8×8 is brighter. The luminance-masked threshold, $t_L[i, j, k]$, is given by

$$t_L[i,j,k] = t[i,j](C_o[0,0,k]/C_{0,0})^{\alpha_T} \tag{5}$$

where α_T is a constant with a suggested value of 0.649, $C_o[0, 0, k]$ is the DC coefficient of the kth block in the original image, and $C_{0,0}$ is the average of the DC coefficient in the image. Alternatively, $C_{0,0}$ may be set to a constant value representing the expected intensity of images.

Contrast Masking

Contrast masking, i.e., the reduction in visibility of a change in one frequency due to the energy present in that frequency, result in a masking threshold, $s[i, j, k]$, given by

$$s[i,j,k] = \max\left(t_L[i,j,k], |C_o[i,j,k]|^{0.7} t_L[i,j,k]^{0.3}\right) \tag{6}$$

The final threshold,$s[i, j, k]$, estimates the amounts by which individual terms of the block DCT may be changed before resulting in one JND. We refer to these thresholds as *slack*.

4. ADAPTIVE QIM BASED ON WATSON MODEL

We can use the slacks of Equation (6) to adaptively select the quantization step size. The adaptive QIM watermarking system is schematically shown in Figure 2. The cover Work is converted to the DCT domain and the coefficients serve as the host signal x. The slacks from Watson model are used to determine Δ. The message m is embedded by the QIM embedder to obtain watermarked signal y. The watermarked work can be retrieved by inverse DCT of y. After transmission, the received Work is used to generate a received signal r and estimated $\hat{\Delta}$. Finally, The message $\hat{m}$ is detected by QIM detector.

Note that we use the original Work to compute the quantization step size for each sample during embedding, and we use the distorted watermarked Work to compute the quantization step size for each sample during detection. If these two step sizes are not the same, then a bit error is very likely to occur. In practise, very good correspondence is achieved.

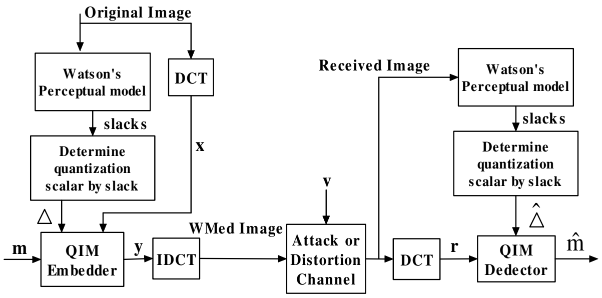

Fig. 2. Adaptive QIM watermarking system based on Watson model.

The quantization step size for QIM is determined by the slacks and a global constant, G, which can be adjusted to alter the watermark strength.

4.1. Experimental results

Figure 3 shows the bit-error-rate (BER) for as a function of the additive white Gaussian noise strength. Results for both our adaptive method and the original algorithm of [3] are provided. In both cases, we adjusted the watermark strength such that the *document to watermark ratio*(DWR) is $35dB$. For fixed quantization step size, $\Delta = 2.1$. Here, $DWR = 10\log_{10}(\frac{\sigma_x^2}{\sigma_w^2})$ and $w = y - x$. Each point on a curve is the

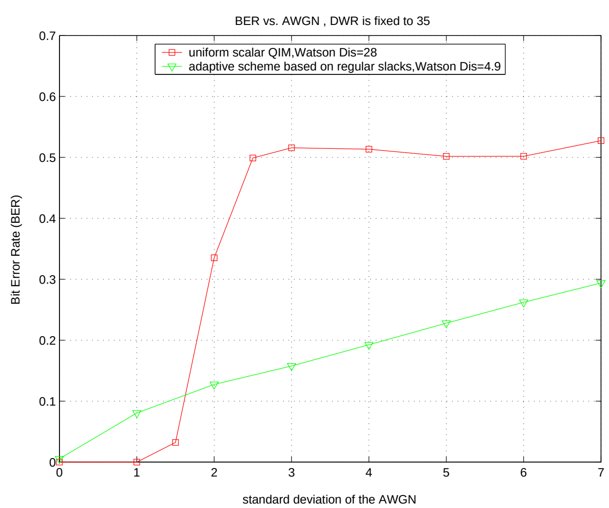

Fig. 3. Robustness versus Gaussian noise

BER averaged over 1000 images from the Corel database.

The original QIM algorithm has superior performance for low noise. We believe this is due to discrepancies between the corresponding estimated quantization step sizes at the embedder and the decode.

However, when the standard deviation in the noise exceeds 1.5, the adaptive method is clearly superior. Conversely, as the standard deviation of the noise approaches the fixed quantization step size, the performance of the original algorithm degrades rapidly. Note also, that the superior performance of our algorithm is achieved with a very low Watson distance of 4.9 (i.e. very high fidelity) compared with the orginal method which has a Watson distance of 28. Thus, improved robustness and improved fidelity have been simultaneously achieved.

Unfortunately, despite the new method's superior performance under additive white Gaussian noise conditions, it remains vulnerable to *amplitude scaling*. When the amplitude of image is scaled by factor of β, the resulted luminance-masked threshold (marked as $\widehat{t_L}[i,j,k]$) is calculated as:

$$\widehat{t_L}[i,j,k] = t[i,j](\frac{\beta C_o[i,j,k]}{\beta C_{0,0}})^{\alpha_T} = t_L \tag{7}$$

i.e. t_L does not scale linearly with amplitude scaling. In fact, referring to equation (6), the slack and $\hat{\Delta}$ are not proportional to scaling factor β.

5. ADAPTIVE QIM WATERMARKING BASED ON MODIFIED WATSON MODEL

To be robust to valumetric scaling, we want the estimated $\hat{\Delta}$ to be multiplied by β when the amplitude of the signal is scaled by β. To this end, we modify the luminance masking in Equation (5) to be t_L^M, given as:

$$\begin{aligned} t_L^M[i,j,k] &= t_L[i,j,k](C_{0,0}/128) \\ &= t[i,j](C_o[0,0,k]/C_{0,0})^{\alpha_T}(C_{0,0}/128) \end{aligned} \tag{8}$$

$C_{0,0}$ is the average of DC components of the image, we have chosen it to be divided by 128 (the mean pixel brightness). The modified *slack* is then given by:

$$s^M[i,j,k] = \max\left(t_L^M[i,j,k], |C_o[i,j,k]|^{0.7} t_L^M[i,j,k]^{0.3}\right) \tag{9}$$

Thus after the modification, when the image is amplitude scaled by factor of β, the luminance masking and slack scale linearly with β. The modified slack can then used to determine the step size Δ_n^M. When the image is scaled by factor of β, the estimated quantization step size $\hat{\Delta}_n^M$ is theoretically also scaled by β. This provides an adaptive QIM algorithm that is invariant to valumetric scaling.

5.1. Experimental Results

For evaluation purposes, we again use a database of 1,000 images, each of dimension 512×512. A binary message of length 8192 bits is embedded into each image. We extract 62 DCT coefficients from each 8×8 block. The entire sequence of 62×4096 coefficients are then pseudo-randomized and each bit of the message is embedded in 31 random coefficients. This is equivalent to embedding two bits in each block of the image.

The watermarking schemes evaulated are marked as:

(A) The original non-adaptive QIM scheme of [7] using soft-decision detection

(B) The adaptive QIM scheme of [6] using hard decision detection

(C) Adaptive QIM based on regular Watson model, soft decision detection

(D) Adaptive QIM based on modified Watson model, hard decision detection

(E) Adaptive QIM based on modified Watson model, soft decision detection

To compare the performance of the different schemes, we fix DWR to be $35dB$. However, while the DWR is the same for images watermarked with the five algorithms, their average Watson distance between watermarked and original image differ considerably as shown in Table 1.

Scheme	A	B	C	D	E
Watson Distance	28	43	4.9	6.8	6.8

Table 1. Average Watson distance for different methods.

Table 1 shows that the three adaptive schemes proposed here have very much lower perceptual distortion as measured by Watson's distance. Importantly, the modification to the Watson distance used in methods *D* and *E* to provide robustness againsts valumetric scaling, produces only a small degradation in image quality and remains much better than methods *A* or *B*.

The performance of scaling robustness test for all schemes are shown in Figure 4 We observe that for very small changes

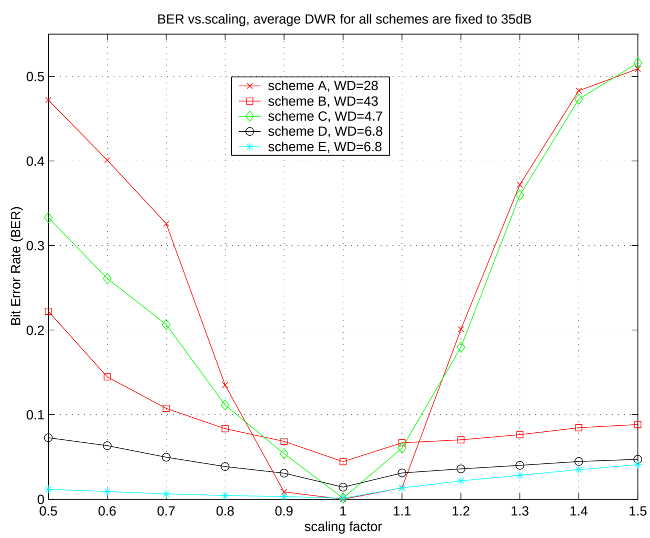

Fig. 4. Robustness versus amplitude scaling

in scale, $0.9 \leq \beta \leq 1.1$, the original algorithm *A* performs as well or better than the others. Our method *C* has poorer performance in this range, but for larger scale changes, it has similar or superior performance. It is also important to note that this is achieved with a perceptual distortion of less than 20% of that of method *A* (see Table 1).

Both algorithms *A* and *C* are not designed to be invariant to valumetric scaling. Bit error rates of greater than 10% occur for $\beta < 0.8$ and $\beta > 1.1$. In contrast, Oostveen *et al*'s method *B* and our methods *D* and *E* show much better robustness to scale changes. Clearly method *E* outperforms all others with a BER that never exceeds 5% over the range of β tested. To ensure that this performance was not due to soft decoding alone, method *D*, while performing worse than method *E* is still superior to Oostveen *et al*'s method (which also uses hard decoding).

Finally, we again note that while the perceptual distortion introduced by methods *D* and *E* is greater than for method C, the modification to the Watson model has only resulted in a small degradation in quality. Importantly, methods *D* and *E* have considerably higher quality than previous algorithms *A* and *B*.

6. CONCLUSION

We have proposed two modifications to QIM. First we use Watson's perceptal model to adaptively change the quantization step size. Experimental results confirm that for the same DWR, the perceptual distortion is reduced by over 80%.

Second, we modified Watson's perceptual model so that the adaptive QIM scheme is invariant to valumetric scaling. Experimental results demonstrate that using soft decision decoding, the BER does not exceed 5% over a scale range of 0.5 to 1.5. The perceptual distortion introduced by this method is much lower than previous methods.

Finally, we note that since the adaptive step size is computed locally, we would expect the algorithm to also be robust to spatially-variant scale changes. Further work needs to be done to verify this.

Acknowledgement

The authors thank J. C. Oostveen, T. Kalker and M. Staring for providing the code to their algorithm and European Office of Aerospace Research & Development(EOARD) for financial support. This research was sponsored by the Air Force Office of Scientific Research, Air Force Material Command, USAF, under grant number FA8655-03-1-3A46. The U.S. Government is authorized to reproduce and distribute reprints for Government purpose notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the author is and should be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of the Air Force Office of Scientific Research or the U.S. Government.

7. REFERENCES

[1] I. J. Cox, M. L. Miller, and A. McKellips, "Watermarking as communications with side information," *Proc. IEEE*, vol. 87, no. 7, pp. 1127–1141, 1999.

[2] M. Costa, "Writing on dirty paper," *IEEE Trans. Inform. Theory*, vol. 29, pp. 439–441, 1983.

[3] B.Chen and G.Wornell, "An information–theoretic approach to the design of robust digital watermarking systems," in *Int. Conf. on Acoustics,Speech and Signal Processing*, Phoenix, USA, March 1999.

[4] J.J. Eggers, R. Bauml, and B. Girod, "Estimation of amplitude modifications before scs watermark detection," in *Proc. SPIE: Security, Steganography, and Watermarking of Multimedia Contents*, San Jose, California, USA, 2002, pp. 387–398.

[5] Kiryung Lee, Dong Sik Kim, Taejeong Kim, and Kyung Ae Moon, "Em estimation of scale factor for quantization-based audio watermarking," in *Digital Watermarking,Second International Workshop,IWDW 2003*, Seoul, Korea, Oct. 2003.

[6] J.C. Oostveen, A.A.C. Kalker, and M. Staring, "Adaptive quantization watermarking," in *Proc. of SPIE:Security, Steganography, and Watermarking of Multimedia Contents VI*, San Jose, California, USA, 2004, vol. 5306, pp. 37–39.

[7] B. Chen and G. W. Wornell, “Quantization index modulation methods for digital watermarking and information embedding of multimedia,” *Journal of VLSI Signal Processing Systems for Signal, Image, and Video Technology, Special Issue on Multimedia Signal Processing*, pp. 7–33, Feb 2001.

[8] Andrew B. Watson, “DCT quantization matrices optimized for individual images,” *Human Vision, Visual Processing, and Digital Display IV*, vol. SPIE-1913, pp. 202–216, 1993.

[9] I. J. Cox, M. L. Miller, and J. A. Bloom, *Digital Watermarking*, Morgan Kaufmann, 2001.