RATIONAL DITHER MODULATION WATERMARKING USING A PERCEPTUAL MODEL

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

Quantization index modulation (QIM) is a computationally efficient method of informed watermarking. However, the original method is particularly sensitive to variations in the amplitude of the signal. Previously, we proposed using a modification of Watson's perceptual model to adaptively adjust the quantization index step size. This simultaneously improved both the robustness and fidelity of the watermarked image and, most importantly, provided invariance (to a large degree) to valumetric scaling. Contemporaneously, rational dither modulation was proposed as an alternative QIM with valumetric invariance. In this paper, we combine the two methods and compare the performance of the new algorithm with our previous results. Experimental results demonstrate that the new algorithm outperforms the previous algorithms over the entire range of valumetric scale factors, albeit at the expense of a small decrease in fidelity. However all algorithms have a superior performance and improved fidelity compared with QIM.

1. INTRODUCTION

Quantization index modulation is a method to perform watermarking that was originally proposed by Chen and Wornell [1]. It is a form of watermarking with side information [2, 3, 4] in which the cover Work, e.g. image, video, audio, is quantized by two sets of quantizers, one encoding a 1-bit and the other encoding a 0-bit. The resulting lattice codes exhibit very high capacity and computational simplicity. As a result, QIM has received significant attention within the watermarking community.

A significant limitation of the original algorithm is its extreme sensitivity to changes in the amplitude of the cover signal, e.g. volume changes to audio or brightness changes to images or video. This valumetric scaling is a very common occurrence and there has therefore been considerable work addressing this issue. Two different approaches have been pursued.

In the first approach, researchers have attempted to determine the valumetric scaling and then compensate for it. This approach was initially pioneered by Eggers *et al* [5] who proposed to estimate the valumetric scaling by "securely embedd[ing] SCS pilot watermark". However the need for a calibration signal may lead to security weaknesses. Lee *et al* [6] 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".

In the second approach, researchers proposed to adaptively modify the quantization step size, which was originally fixed. Oostveen *et al* [7] uses a simple perceptual model based on Weber's law. In [8] we proposed basing the quantization step size on the perceptual "slack", i.e. the amount a sample could be altered before introducing a just noticeable distortion (JND). The slack was determined using a modification of Watson's perceptual model [9]. The modification, described in Section 4 permits the slacks to scale linearly with valumetric scaling, thereby providing valumetric invariance. The embedding is performed in the Discrete Cosine Transform (DCT) domain, and the slacks are computed from the 8×8 DCT blocks. These blocks are modified during the watermark embedding procedure. However, the detector uses these modified DCT blocks to compute the slacks needed for detection. It is implicitly assumed that the slacks calculated during detection are unaffected by the embedding stage. Section 5 provides more detail.

Perez *et al* [10] proposed an alternative QIM method in which the quantization step size at time, k, is a function of the watermarked samples at earlier times. This algorithm is described in Section 3.

In this paper, we propose to calculate the slacks for the current block, k, based on the previous watermarked block. In so doing, we guarantee that the slacks used during detection are unaffected by embedding. This is described in Section 6.

Section 2 and 3 provide an introduction to quantization index modulation and rational dither modulation. Section 7 provides an experimental comparison between the new algorithm and our previous proposal. We also implement a version of rational dither modulation in the DCT domain (RDM-DCT) for comparison purposes. Experimental results show that the new algorithm has a superior bit error rate over the entire range of valumetric scale factors. This is achieved at the expense of a small decrease in fidelity Section 8 summarizes our results and discusses future work.

2. QUANTIZATION INDEX MODULATION

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

Quantization index modulation embeds a message by quantizing the host signal with the associated quantizer, which is selected based on the value of the message bit to be encoded.

Let Δ be the quantization step size, N represent the length of the host signal x and m represent the message (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

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

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

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

2.1. Soft detection with QIM

The above description embedded one bit per sample. In practice, we usually spread one message bit into a sequence of P samples. One way to achieve this is to use a rate $1/P$ repetition encoding. In the detector we can accumulate the two Euclidean distances for P samples and then determine the detected message bit, i.e.,

$$\begin{aligned} &\hat{m_n} = \underbrace{argmin}_{l \in 0,1} \textstyle\sum_{h=(n-1)P+1}^{nP} (r_h - S_h(r_h, l))^2, \\ &n = 1, 2, \ldots, N/P. \end{aligned} \tag{3}$$

where r is the received signal and r_h denotes the h-th sample of r. $S_h(r_h, 0)$ and $S_h(r_h, 1)$ are generated by embedding "0" and "1" into the received signal r separately. That is, during detection, the detector calculates two signals $S_h(r_h, 0)$ and $S_h(r_h, 1)$ by embedding "0" and "1" into the received signal r separately, in the same manner as Equation (2).

The quantization step size for QIM is usually fixed. However, by adaptively selecting the step size based on a local neighbourhood of the content, it is possible to both improve fidelity and robustness, and provide invariance to valumetric scaling. Two adaptive methods are described in Sections 3 and 4.

3. RATIONAL DITHER MODULATION

Rational dither modulation (RDM) was first proposed by [10] and is intended to provide valumetric invariance to QIM. Given a host signal, $x = (x_1 \ldots x_N)$ and a watermarked signal, $y = (y_1 \ldots y_N)$, then the k-th bit of a watermark message, $m = (m_1 \ldots m_M)$, is embedded as

$$y_k = g\left(y_{k-L}^{k-1}\right) Q_{m_k}\left(\frac{x_k}{g(y_{k-L}^{k-1})}\right) \tag{4}$$

where y_{k-L}^{k-1} denotes the set of past signals, $(y_{k-L} \ldots y_{k-1}$ and the function, $g()$ maps its L-dimensional input vector to a real value and has the property that for any valumetric scaling factor $\rho > 0$,

$$g(\rho y) = \rho g(y) \tag{5}$$

This definition of RDM is intrinsically invariant to valumetric scaling. The function $g()$ can be chosen from a very large set of possible functions, including the l_p-norms, i.e.

$$g\left(y_{k-L}^{k-1}\right) = \left(\frac{1}{L}\sum_{i=1}^{L} |y_{k-i}|^p\right)^p \tag{6}$$

However, it is well-known that l_p-norms poorly model the human perceptual system. Thus, it is interesting to consider a function $g()$ that models properties of perception and satisfies the constraint of Equation (5). Such a function is described next.

4. A MODIFIED WATSON PERCEPTUAL MODEL

Watson's perceptual model [9] is a popular perceptual model that estimates the perceptibility to changes in individual terms of an image's block DCT. It is commonly used in digital watermarking.

For a given k-th block of the cover Work, c, Watson's model can be used to determine how much each of the DCT coefficients, $C[i,j,k], 0 \leq i,j \leq 7$, can be altered without introducing perceptible artifacts.

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 provided in [11]. Note that it is a constant value table.

Luminance Masking

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

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

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 refers to the reduction in visibility of a change in a frequency due to the energy present in that frequency. This results in a masking threshold, $s[i,j,k]$, given by

$$s[i,j,k] = \max\left(t_L[i,j,k], |C_o[i,j,k]|^{0.7} t_L[i,j,k]^{0.3}\right) \tag{8}$$

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*.

The slacks do *not* scale linearly with amplitude scaling of the image intensities. Thus, Watson's original perceptual model cannot be used to provide for valumetric invariance. In

[8] we proposed a modification to Watson's model in order to maintain a linear relationship between the perceptual slacks and valumetric scaling. In particular, we modified the luminance masking of Equation (7) to be t_L^M, given as:

$$\begin{aligned} t_L^M[i,j,k] &= t_L[i,j,k](C_{0,0}/128) \\ &= t[i,j](C_o[0,0,k]/C_{0,0})^{\alpha_T}(C_{0,0}/128) \end{aligned} \quad (9)$$

$C_{0,0}$ 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:

$$s^M[i,j,k] = \max\left(t_L^M[i,j,k], |C_o[i,j,k]|^{0.7} t_L^M[i,j,k]^{0.3}\right) \quad (10)$$

Thus after the modification, when the image is amplitude scaled by factor of ρ, the luminance masking and slack scale linearly with ρ.

5. QIM WITH MODIFIED WATSON DISTANCE

In [8], we used this modified slack to adaptively set the quantization step size and thereby provide robustness to valumetric scaling. Figure 1. is a block diagram of the system.

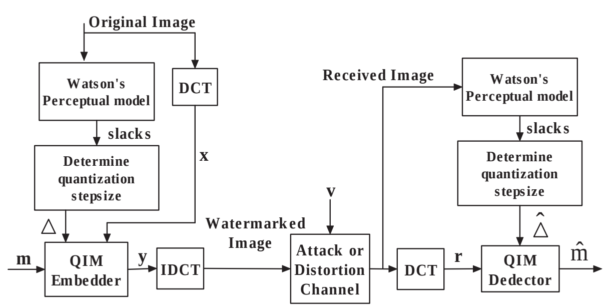

Fig. 1. Adaptive QIM watermarking system based on modified Watson model.

Notice that the quantization step size is determined by a local neighbourhood around the host sample, x_k, and that this neighborhood is altered during the embedding of the watermark. Thus, during detection, we must rely on the fact that these alterations are small, and hope that the slacks based on the modified local neighborhood are the same as those determined during embedding. While this is often true, rational dither modulation suggests an alternative approach, in which the perceptual slack at time k is based on a nearby neighborhood of previously watermarked samples. Clearly, there may be some degradation in perceptual quality since a perceptual estimate made in a nearby neighborhood, is not guaranteed to be perceptually relevant. We denote this earlier algorithm as QIM-MW.

In the next section, we briefly describe an implementation of RDM using the modified Watson model. The new algorithm is denoted RDM-MW. Experimental results in Section 7 examine the perceptual impact.

6. IMPLEMENTATION

Before describing the implementation of rational dither modulation with a modified Watson perceptual model, we first describe an implementation of RDM in the DCT-domain. This algorithm, denoted RDM-DCT is used for comparison.

Our implementation of RDM-DCT quantizes the 62 DCT coefficients of each 8×8 block (excluding the DC and highest frequency terms). We use the DC coefficient from the previous 8×8 block to determine the quantization step size. Thus, the window size is 64. And the function $g()$ is equivalent to the average intensity of the block, i.e. we use an L_1-norm in Equation (6). All quantization step sizes are scaled by a global constant that is chosen so that the document-to-watermark ratio (DWR) averaged over all watermarked images equals a desired value.

We embed one bit in 31 DCT coefficients. These coefficients are randomized as described below.

We now describe a number of implementation issues related to the new algorithm, RDM-MW:

1. The DCT coefficients are randomized. In [8] this randomization was over all the coefficients in the image. Here, because we require access the neighboring block in order to computer the slacks, we randomize the DCT coefficients from 32 blocks. See below.

2. Each image is (i) divided into 32 disjoint regions, (ii) a random block in each region is assigned as the start block, (iii) blocks are indexed in a zig zag order in both the positive and negative directions, as depicted in Figure 2. The perceptual slack at block k is computed from its neighbouring block $k-1$. The scan initially precedes in each partition from a random block 1 towards the right. On reaching the end of the partition, the scan then precedes in direction left of the initial block. At any iteration k, the DCT coefficients from the 32 corresponding blocks of the 32 partitions are randomized.

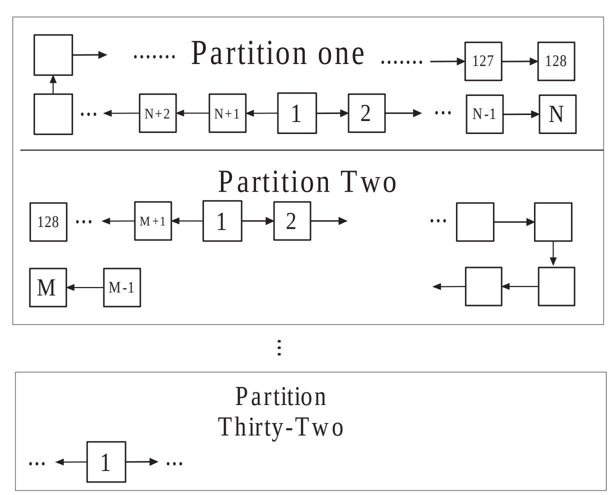

Fig. 2. Order of Scaning Blocks for Embedding

To embed a bit m_i, 31 random coefficients from the 32 blocks are assigned to the message bit and quantized as in Equation (2). The quantization step size for each coefficient is determined by the slack associated with the corresponding coefficient from the previous block. These slacks are multiplied by a global constant that is chosen to provided a desired DWR for the watermarked images.

7. EXPERIMENTAL RESULTS

We watermarked 1,000 images from the Corel database. These images were normalized such that the minimum intensity value was set to 20 and the maximum value did not exceed 235. This normalization was performed in order to effectively eliminate the clipping noise which occurs after watermarking the DCT coefficients and performing the inverse DCT. We acknowledge that such a normalization may not be possible in real-world applications. However, for the purposes of this experiment we did not want to complicate the interpretation of results by the addition of clipping noise.

We embedded 2 bits in each 8×8 block, i.e. 1 bit per 31 sample coefficients. In all experiments, the average document-to-watermark ratio (DWR) was fixed at 35dB.

The bit error rates (BER) as a function of valumetric scaling is shown in Figures 3. The average Watson distance of the watermarked images for each of the three algorithms is 6.8 for QIM-MW, 7.3 for RDM-MW and 11.2 for RDM-DCT.

The results clearly indicate that the new algorithm has the lowest BER over the entire range of valumetric scale factors. This is achieved at the expense of slightly reduced perceptual quality. The RDM-MW has an average perceptual distance of 7.3, compared a perceptual distance of 6.8 for our earlier algorithm (QIM-MW). The RDM-DCT algorithm has both the worse perceptual distance and the worse perceptual distortion, 11.2. However, it should be noted that this perceptual distortion is still better than that achieved with the original QIM algorithm which has an average Watson distance of 28 [8].

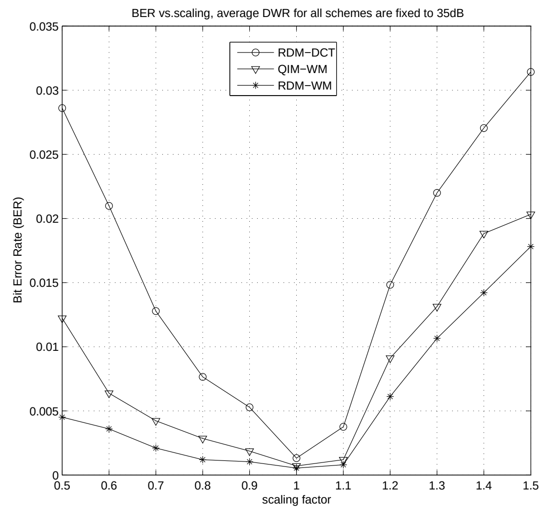

Fig. 3. Bit error rate as a function of valumetric scaling for an embedding rate of one bit per 31 samples (two bits per 8×8 block). The average perceptual distance, as measured by Watson's model, is 6.8 for QIM-MW, 7.3 for RDM-MW and 11.2 for RDM-DCT.

8. CONCLUSION

We have described a new algorithm that incorporates a perceptual model within the framework of rational dither modulation. The new algorithm provides approximate invariances to valumetric scaling over the range of 0.5 to 1.5.

Comparison of this algorithm (RDM-MW) with an earlier algorithm (QIM-MW) and an implementation of rational dither modulation (RDM-DCT) revealed superior performance over the entire range of valumetric scale factors. For a scale factor of 0.5, RDM-MW is approximately 7 times better than RDM-DCT and 3 times better than QIM-MW.

The is achieved at the cost of a small degradation in image fidelity. The average Watson distance of the RDM-MW was 7.3, compared with 6.8 for QIM-MW. The RDM-DCT algorithm had a perceptual distance of 11.2 which is still better than the original QIM algorithm of 28.

In the future, we intend to determine whether we can improve the fidelity further by determining whether the neighboring block is sufficiently correlated to rely on its perceptual estimate. We also intend to examine techniques to more realistically handle clipping issues.

9. REFERENCES

[1] 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.

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

[3] 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.

[4] Jim Chou, S. Sandeep Pradhan, and Kannan Ramchandran, "On the duality between distributed source coding and data hiding," *Thirty-third Asilomar conference on signals, systems, and computers*, vol. 2, pp. 1503–1507, 1999.

[5] 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, January 2002, pp. 387–398.

[6] 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.

[7] 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, January 2004, vol. 5306, pp. 37–39.

[8] Qiao Li and Ingemar J. Cox, "Using perceptual models to improve fidelity and provide invariance to valumetric scaling for quantization index modulation watermarking," in *Int. Conf. on Acoustics,Speech and Signal Processing(ICASSP)*, Philadelphia, USA, March 2005.

[9] 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.

[10] F. Perez-Gonzalez, M. Barni, A. Abrardo, and C. Mosquera, "Rational dither modulation: a novel data-hiding method robust to value-metric scaling attacks," in *IEEE International Workshop on Multimedia Signal Processing*, Siena, Italy, Sep. 2004.

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