Quantisation Effects and Watermarking Capacity

Stéphane Bounkong, David Saad and David Lowe
{bounkons,d.saad,d.lowe}@aston.ac.uk
Neural Computing Research Group
Aston University
Birmingham, B4 7ET, United Kingdom

Abstract

Digital watermarking aims at embedding information in digital data. The watermark is usually required to be imperceptible, unremovable and to have a high information content. Unfortunately, these three requirements are contradicting. For example, having a more robust watermark makes it either more perceptible or/and less informative. This paper investigates the relationship between the watermark information content and the induced distortion due to quantisation, such as lossy compression.

1 Introduction

Digital media have become very popular over the last decade. The development of efficient compression algorithms, such as MPEG [5], JPEG [7], or JPEG2000 [1] has made it easy to distribute data over the Internet but also increased their vulnerability to illicit distribution or retailing. Interest in watermarking techniques has grown significantly in the past few years, mainly due to the need to protect intellectual property rights of these products [3]. In this paper, we investigate the relation between the maximum information content that can be embedded and successfully retrieved after being transmitted over a quantised channel (an attack). Such a channel is typically encountered in lossy compression (JPEG, MPEG) methods, which are an essential tool in the transmission of digital media. Without loss of generality, here, the analysis is carried out over JPEG compression for images.

2 Watermarking Communication Channel

The basic problem of watermarking, is how to embed information, usually termed a watermark, in the data [3], with an imperceptible loss of quality and such that common processing will not remove the embedded watermark. Quantisation is essential for lossy compression, and for the transmission and storage of digital data. In this paper, the influence of quantisation on the watermarking transmission rate is investigated. Figure 1 depicts the communication channel studied (termed channel with side information in the information theory literature), where X is the original data, M the message to be embedded, $\hat{X}$ the watermarked data, QF the compression quality factor, Y the quantised data and $\hat{M}$ the message estimate after an attack.

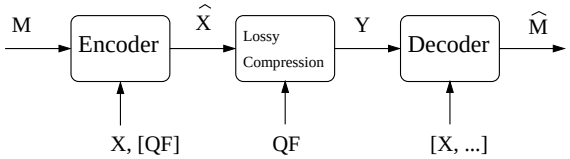

Figure 1: Watermarking communication channel.

2.1 Lossy Compression and Watermarking

Recent research has focused on evaluating the capacity of the watermarking channel [4, 6]. Typically, all processes studied so far are modelled using simple distributions, mostly Gaussian. However,

this modelling may be misleading for deterministic attacks, such as lossy compression. The standard deviation values used to model the quantisation noise lead to very low information transmission rate. Among other results, the present work will show the discrepancy between the transmission rate obtained using the Gaussian model and the true transmission rate.

In the present paper, quantisation, which is the heart of all lossy compression methods, is considered as a deterministic process. Quantisation reduces all states of the data within a quantisation bin to a unique state named the quantised state. This reduces the number of possible states for the data, watermarked or not, and therefore bounds the achievable information rate (IR) of the watermarking scheme. For a given allowed induced distortion, the maximum number of informative bits that can be encoded, when subject to such an attack is therefore given by the number of quantised states within the radius defined by the allowed distortion.

An illustration is given in Fig. 2, where '+' represent the quantised points and 'X' the original data point, the area associated with each quantised point is marked by plain lines. Each quantised point is associated with a certain information represented by the letters 'A' to 'D'. Therefore, in order to encode the letter, the induced distortion will be characterised by 'C1', 'C2', 0, and 'C3' respectively. So, given a quantisation step (width of the quantisation bin) δ, one can derive a relation between the IR and the induced distortion K.

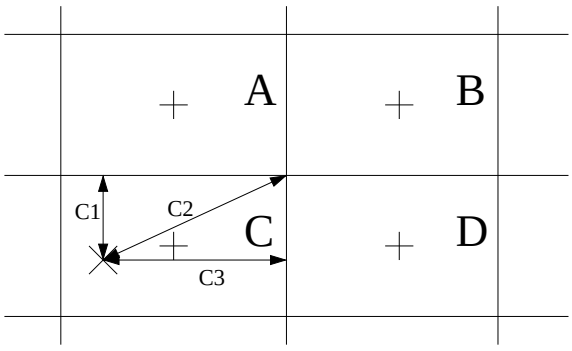

Figure 2: Quantisation and watermark.

In the following, the distortion cost K of transmitting information over a known uniform quantised channel is investigated. A mean square error metric in the image space is used to measure the latter cost, $MSE = \frac{1}{hw}\sum_{i,j}(X_{ij} - \hat{X}_{i,j})^2$, where h is the height of the image, w the width of the image, i the vertical index of the pixel, and j the horizontal index of the pixel.

2.2 Lossy Compression: JPEG

In this section, a brief presentation of the JPEG standard, on which the analysis is based, is given. More details can be found in [7].

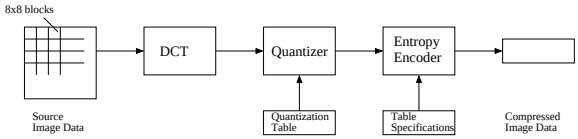

Figure 3: JPEG Lossy Compression Block Diagram.

Figure 3 depicts JPEG compression for still images. First, the image is divided into contiguous patches of 8×8 pixels. The DCT of all patches is taken, the obtained coefficients are quantised according to

$$\hat{X} = Qu(X, \delta) = \left[\frac{X}{\delta}\right]\delta \,, \qquad (1)$$

where $\hat{X}$ is the quantised data X, δ is the quantisation step and where '[.]' stands for the *fix* rounding operator (rounding towards '0').

The quantisation step δ is computed from the JPEG quality factor parameter QF (Eq. 2 and 3) and the predefined quantisation table Q (Appendix, Fig. 9), which provides the different values for each coefficient in a patch. Finally, the quantised coefficients are encoded using a lossless compressor.

$$\delta = kQ \,, \qquad (2)$$

$$k = \begin{cases} 50/QF & \text{if } QF < 50 \,, \\ \frac{(200-2QF)}{100} & \text{if } QF \geq 50 \,. \end{cases} \qquad (3)$$

3 Information Rate and Induced Distortion

In this section, we investigate the maximum induced distortion required to achieve a certain in-

formation rate for a given level of quantisation (attack) at the encoder. A general analysis will be carried out, followed by a particular example (JPEG).

3.1 Analysis

Assume n parallel channels, subject to a quantisation attack, as defined in Eq. 1, with different quantisation levels δ_i ($i = 1 \dots n$). In the worst case, the induced distortion K can be expressed as follows

$$K = \sum_{i|n_i>0} 2^{2(n_i-1)}\delta_i^2 \ , \quad \text{with} \quad N = \sum_i n_i \ , \qquad (4)$$

where N is the number of bits we want to embed, n_i is the number of bits encoded using the ith channel coefficient (e.g. DCT coefficient).

Our aim is therefore to find the best distribution of n_i, which minimises K for a given N. By introducing a Lagrange multiplier in Eq. 4 and solving the system of equations given by the first order condition of optimality, we get

$$n_i = \frac{N}{n} + \frac{1}{n}\sum_j \log_2 \delta_j - \log_2 \delta_i \ . \qquad (5)$$

Then, the validity of the obtained solutions n_i has to be verified for each of the channels. For instance, each n_i has to be a positive integer. If not, it means that the solution lies on the system boundaries. The problem can also be overcome using classical methods, for example a Langrange multiplier for each non verified constraint changing them into an equality if needed.

3.2 Gaussian Model Analysis

In this section, we present a short analysis on the Gaussian model mentioned earlier and as it is usually presented in the literature. In this framework, the quantisation noise is modelled by a centred Gaussian noise. Its standard deviation is estimated from the distortion introduced by the quantisation. Assuming a flat distribution for the source to be quantised, the standard deviation of the noise is $\sigma = \delta/\sqrt{12}$. Furthermore, for a channel with additive Gaussian noise, the capacity is given by [2, 4]

$$C = \frac{1}{2}\log_2\left(1 + \frac{\sigma_w^2}{\sigma^2}\right) \ , \qquad (6)$$

$$= \frac{1}{2}\log_2\left(1 + \frac{12\sigma_w^2}{\delta^2}\right) \ , \qquad (7)$$

where σ_w is the standard deviation of the introduced watermark. In our problem, n parallel channels, indexed by i, are considered, and for these the best distribution has to be found for a given information rate and set of attack strengths δ_i. This gives rise to a similar optimisation problem as in Sec. 3.1, with

$$K_g = \sum_i \sigma_{wi}^2 \ , \qquad (8)$$

to be used as a cost to minimise under the constraint

$$N = \sum_i C_i = \frac{1}{2}\log_2\left(1 + \frac{\sigma_{wi}^2}{\sigma_i^2}\right) . \qquad (9)$$

The problem is solved in the same way. The optimal allocation is then given by

$$\sigma_{wi}^2 = \frac{1}{12}\left\{2^{\frac{2N}{n}}\left(\prod_j \delta_j^2\right)^{\frac{1}{n}} - \delta_i^2\right\} , \qquad (10)$$

then the validity of the solution has to be verified with respect to the positivity condition.

3.3 Practical Case

In this section, a JPEG attack is assumed and presented as a case study. A brief description of the algorithm used in a practical case is also given. In Sec. 3.1, we presented and solved the problem for the worst-case source data. In practice, the original value of the source data is of high relevance as it has a significant influence on the introduced distortion for low information rate. Furthermore, it is important to notice that the quantisation bin centred around 0 is double the size of all others, (Eq. 1).

A simple way to tackle the problem is to use a greedy algorithm. The latter searches for the lowest distortion to embed one more bit per iteration. This basically means that the number of reacheable states with the allowed distortion has to double for one of the parallel channels considered at each iteration. The channel with the lower cost is selected and the cost associated to it is added to the total distortion cost at the previous iteration. Then, the cost related to further use of the selected channel is updated. The description of the algorithm is given by the following steps.

1. The distortion C_i for first use of each of the channels is evaluated by

$$C_i = \begin{cases} (\delta_i - r_i)^2 & \text{if } q_i = 0, \\ \min(r_i^2, (\delta_i - r_i)^2) & \text{if } q_i \neq 0, \end{cases} \quad (11)$$

$$q_i = \left\lfloor \frac{c_i}{\delta_i} \right\rceil \ , \ \text{ and } \ r_i = c_i - q_i \delta_i \ , \quad (12)$$

where c_i is the original (real) value of the source, the original costs are also referred to as O_i.

2. The lower distortion is selected and added to the previous distortion.

$$K = K + \min_i C_i \ , \ \text{ and } \ n_i = n_i + 1 \ . \quad (13)$$

3. The selected marginal distortion C_i for the selected channel is then updated using

$$C_i = \begin{cases} (2^{n_i}\delta_i - O_i)^2 - C_i \ , \\ \qquad \text{if } |q_i| - n_i < 0 \ ; \\ ((2^{n_i} - 1)\delta_i + O_i)^2 - C_i \ , \\ \qquad \text{if } |q_i| - n_i \geq 0 \ . \end{cases} \quad (14)$$

4. Steps 2 and 3 are repeated until $N = \sum_i n_i$.

3.4 Experiments

Following, the previous analysis in Sec. 3, we implement an algorithm evaluating the maximum distortion introduced in order to transmit an N bit message, when the watermarked data is subject to a quantisation attack (lossy compression) of known strength at the encoder. The results are shown for the JPEG standard for different strengths. The experiments are given for the worst host data (monochrome black picture) and for a practical case using the well-known Lena picture. For the latter, only the average maximum distortion induced is reported for different quantisation strengths.

3.5 Results

Figure 4 represents the distortion introduced to encode N bits given a QF in a black image (worst case), while Fig. 5 is the average cost of encoding N bits per patch of 8×8 pixels in the Lena picture. In both plots, quantisation is treated as a deterministic process. While, Fig. 6 represents the distortion

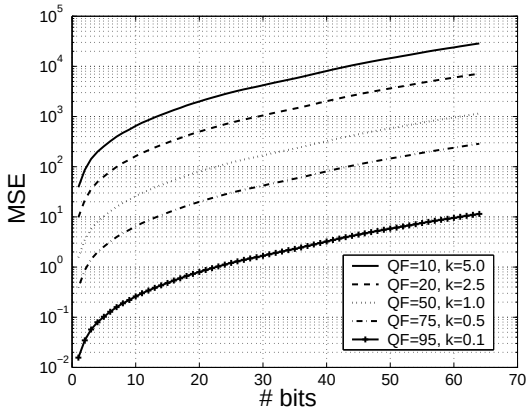

Figure 4: Maximum distortion versus IR for a JPEG attack for a known QF.

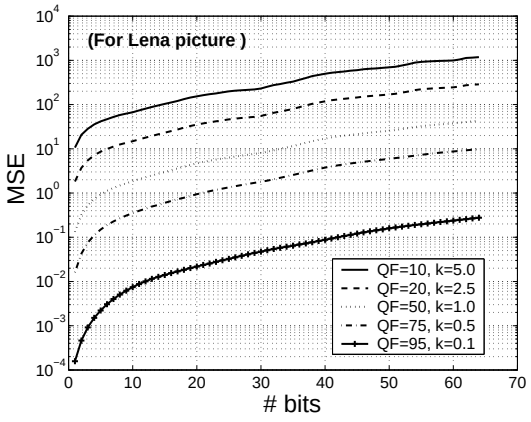

Figure 5: Distortion versus IR for a JPEG attack for a known QF.

introduced by encoding N bits given QF in a black image modelling quantisation as a Gaussian noise.

The results show that Gaussian modelling overestimates the distortion needed in all studied ranges of quantisation strength and transmitted number of bits. This also explains why some schemes in the literature using the Gaussian model, designed with the full knowledge of the compression standard, achieve better results than expected. In our analysis the quantisation strength is assumed to be known at the encoder, which might not be the case in most practical cases. When quantisation is treated as a deterministic process as above, one can easily show that if the attack has not the expected strength, even if weaker, this may introduce

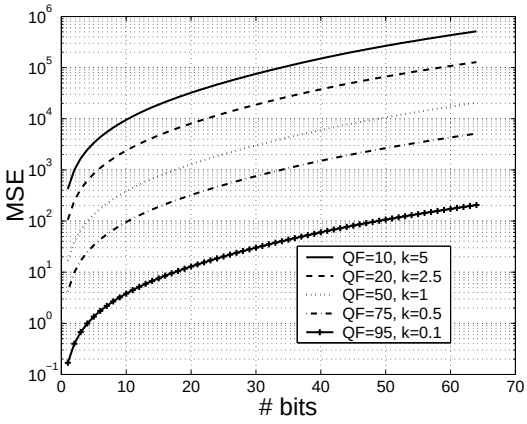

Figure 6: Maximum distortion versus IR for JPEG attack for a known QF, modelled as an additive Gaussian noise.

significant errors in the decoding. When Gaussian modelling is used, the reported distortion is still valid if the real quantisation strength is not greater than the one for the informed case. In the following section, we discuss the amendment needed to our approach, when the maximum strength of the attack is only known at the encoder.

4 Unknown Quality Factor

In this section, the δ is assumed to be unknown at the encoder but bounded from above by δ_m which is known. We are interested in the relation between the IR and the distortion to introduce to achieve it. Results obtained in Sec. 3 where for the case of known δ at the encoder.

4.1 Ambiguity Problem

In the case described in Sec. 3, embedding some information only requires moving the data to the correct hypercube. However, when δ is unknown, the problem lies in the uncertainty introduced by overlapping bins, when different δ values can be used. The problem is explained in Fig. 7. Encoding a given letter refers to moving the data to the appropriate bin, assuming a quantisation strength of δ_m (Fig. 7, line 1). If the quantisation attack results in $\delta = a$ (Fig. 7, line 2); a received data $Y = a$ can come from intervals associated with either A or B (Fig. 7, line 1). This creates ambiguity in the previous scheme. To amend it, first we notice that values are always quantised toward the value '0', which is the only fixed point. The areas of ambiguity for each bin can be found using the maximum remainder over δ of the Euclidean division of $i\,\delta_m$ by δ (Fig. 7, line 2 to 3, the areas of ambiguity are underlined by arrows), which can be expressed formally as follows, where $i \in \mathbb{N}$ denotes the index of the bin from the bin centre at '0',

$$R_{\delta_m}(i) = \max_{\delta} i\,\delta_m - \delta\left[\frac{i\,\delta_m}{\delta}\right], \quad (15)$$

$$= \frac{i\delta_m}{i+1} + \epsilon\ ,\ 0 < \epsilon \ll \delta_m\ . \quad (16)$$

The intervals of ambiguity are therefore of the form $[iR_{\delta_m}(i); (i+1)R_{\delta_m}(i)]$.

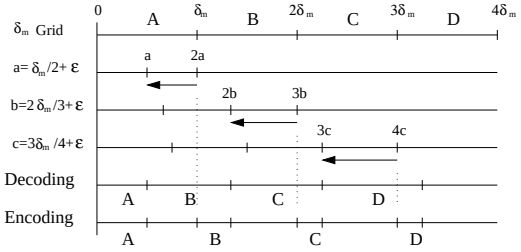

Figure 7: Maximum uncertainty generated by multiple quantisers. Encoding and decoding strategy.

Amending the algorithm involves modifying the boundaries of both encoding and decoding bins. The area between a and $2a$ (Fig. 7, line 2) should be associated with the information B at the decoder, but at present data can be from either A or B. The encoder will be modified so that no information is encoded in this area. Respectively, the area between $2b$ and $3b$ (Fig. 7, line 3) has to be associated with the information C and so on (Fig. 7, line 5). This also defines the bounds for the encoding process (Fig. 7, line 6); for example to encode B, the watermarked data has to lie between δ_m and $2b$ (Fig. 7, line 3). If the modified data is below δ_m, using a quantiser with $\delta = R_{\delta_m}(1) - \epsilon$, with an appropriate value ϵ, will automatically bring it below a and lead to a bad decoding. If the data is greater than $2b$, no quantisation will automatically lead to a bad decoding to C. Once these boundaries are established, a similar algorithm to the one described in Sec. 3.4 can be applied.

4.2 Results

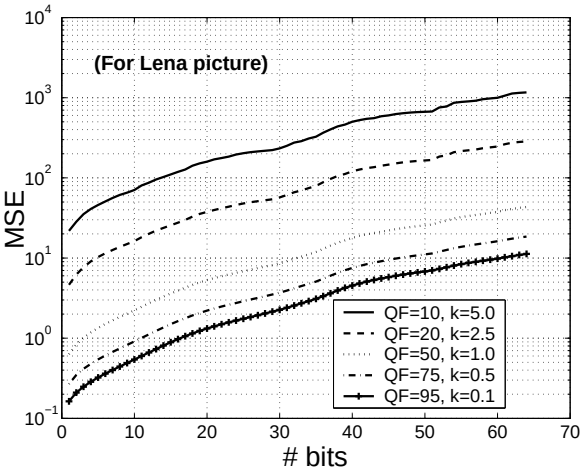

Figure 8: Distortion vs IR for JPEG attack for a known δ_m.

Figure 8 shows the average cost over the patches to transmit N bits per patch. As expected the cost increases significantly only for low IR, since the marginal cost remains the same, equal to δ_m, for every coefficient. From Fig. 8, it can be seen that current state of the art methods are still far below the maximal information rate. For example, if $QF_m = 20$, and the distortion cost allowed is about 3, at least 1 bit can be embedded reliably per patch, which means more than 4096 bits for a 512×512 pixels, Lena picture; this is 4 to 40 times greater than the performances reported in the literature.

5 Conclusion

This paper provides a clear framework for computing the relationship between the IR and distortion introduced by a watermark under a quantisation attack. An example on a typical picture commonly used by research community is also provided. The results show that current watermarking schemes are still far below the maximal IR of this channel. Further research will include evaluation of the IR under various types of attacks.

Acknowledgement: Support from EPSRC research grant GR/N63178 is acknowledged.

A JPEG Quantisation Table

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	89	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

Figure 9: JPEG quantisation table: Q.

References

[1] M. D. Adams. The JPEG-2000 still image compression standard. Technical Report JTC 1/SC 29/WG1N 2412, ISO/IEC, September 2001.

[2] T. M. Cover and J. A. Thomas. *Elements of information theory*. John Wiley and Sons, New York, 1991.

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

[4] J. J. Eggers and B. Girod. Quantization effects on digital watermarks. *Signal Processing*, 81(2):239–263, 2001.

[5] ISO/IEC JTC1/SC29/WG11. Coding of moving pictures and audio. Technical Report ISO/IEC-11172 and ISO/IEC-13818 and ISO/IEC-14496, ISO/IEC, 1988. http://mpeg.telecomitalialab.com/standards.htm.

[6] P. Moulin and J. A. O'Sullivan. Information-theoretic analysis of information hiding. *IEEE Transactions on Information Theory*, March 2003.

[7] G. K. Wallace. The JPEG still picture compression standard. *Communications of the Association for Computing Machinery*, 34(4), April 1991.