Adaptive quantization watermarking

Job Oostveen[a], Ton Kalker[a], Marius Staring[b]

[a]Philips Research, Prof. Holstlaan 4, 5656 AA Eindhoven, The Netherlands.
job.oostveen@philips.com, ton.kalker@ieee.org

[b]Image Science Institute, University Medical Center Utrecht, Heidelberglaan 100, 3584 CX Utrecht, The Netherlands. marius@isi.uu.nl

ABSTRACT

In this paper we study the use of an adaptive quantization step size, instead of a fixed one, for the Scalar Costa Scheme. We propose an adaptation method based on Weber's law. This allows for a more effective embedding, which is also shown to render the watermark robust against sample value scaling. A model for the bit error probability due to the estimation of the adaptive quantization step size at the detector is derived, which provides insight in the required precision of estimating the quantization step size in the detector.

Keywords: W atermarking; Quantization Index Modulation; Scalar Costa Scheme.

1. INTRODUCTION

Recent years have witnessed a new paradigm in watermarking research. A classical result by Max Costa[1] has changed the perspective on watermarking as just some form of communication but in an extremely noisy environment. Nowadays, watermarking is generally studied from the point of view of communication with side-information. As a result, information theory has become an important tool for studying watermarking systems. An important class of systems stemming from this approach is the class of Distortion Compensated Dither Modulation schemes, or more generally Quantization Index Modulation (QIM) schemes. These schemes were introduced in a paper by Chen and Wornell [2] and are a practical implementation of the 'Dirty Paper' coding techniques introduced by Costa. They result from maximizing the watermark capacity, subject to i.i.d. Gaussian host data and i.i.d. Gaussian channel noise and taking into account that the watermark embedder has full knowledge of the host signal.

Although the quantization index modulation schemes are optimal from an information theoretic capacity-maximization point of view, their robustness may be too restricted for widespread practical usage. Most papers assume that host signal samples are identically distributed from a single source distribution and there is no need for them to consider local adaptivity. In practice there may be however several reasons for doing so. Most importantly:

1. Due to varying signal statistics, the optimal values of parameters in the watermarking system (like, for instance, the quantization step-size Δ and the distortion compensation parameter α) need to vary, as well.
2. In practical signals, the perceptual masking properties may vary strongly over different regions of the signal. Therefore the watermark-induced distortion needs to be adapted to the local perceptual masking behavior of the host signal.
3. By using an adaptive quantizer, we may increase the robustness of the watermark against selected signal modifications. A good example is value scaling (e.g. contrast changes for visual content). If the quantization step-size scales in the same way as the signal, then the resulting system is rendered inherently robust to value scaling.

In this paper we study the question of how to adapt QIM schemes to local signal characteristics. In particular we take one specific watermarking system, the Scalar Costa Scheme, introduced by Eggers (see, e.g.,[2]) as a representative member of this general class.

In the Scalar Costa Scheme, the optimal values of the quantization step-size Δ and the distortion compensation parameter α are completely determined by the host signal variance σ_x^2, the Document-to-Watermark Ratio $DWR = \sigma_x^2/\sigma_w^2$ and the channel noise variance σ_n^2. If the host signal is not i.i.d. but can be can segmented into regions in which the signal

variance, the noise variance and the *DWR* are substantially constant, then the optimal value of the quantization step-size and the distortion compensation parameter can be chosen to vary from region to region. As a consequence however, the detector needs to re-compute these parameters as well. In practice, this proves to be problematic as processing and attacks may have a strong influence on the signal variance and watermark power (for instance if the signal has undergone lossy compression like MPEG-2 or JPEG). The detector will thus in general receive a signal whose signal variance and *DWR* is quite different from its variance during embedding, and it will not be able to reproduce the used value of Δ.

Therefore we propose in this paper to investigate other forms of adaptivity that depend on local signal characteristics. We study how the theoretical optimal system can be adapted to achieve practical levels of robustness and imperceptibility. We do this by choosing the basic watermark parameters on the basis of a perceptual model. As argued above an important aspect is the robustness of the statistic on which the adaptation rule is based. The detector needs to be able to accurately re-estimate the value of the parameters as used by the embedder, even in the presence of strong channel noise.

2. THE SCALAR COSTA SCHEME

Assume the host signal $\mathbf{x}$ to be an i.i.d. random vector of length N with power σ_x^2. No additional assumptions are made about the shape of the probability distribution of $\mathbf{x}$. A message vector $\mathbf{m}$ is embedded into the host signal $\mathbf{x}$. We assume that the message $\mathbf{m}$ is encoded such that it is a vector of length N (the same length as the host signal) over the alphabet $\mathcal{D} = \{0, 1, \ldots, D-1\}$. The resulting watermarked signal is denoted $\mathbf{y}$ and the watermark $\mathbf{w}$ is defined as the difference between the host signal and the watermark signal: $\mathbf{w} = \mathbf{y} - \mathbf{x}$. The detector receives the received vector $\mathbf{r}$, which is a noise-corrupted version of the watermarked signal $\mathbf{y}$: $\mathbf{r} = \mathbf{y} + \mathbf{n}$, where the only assumption made on the noise vector $\mathbf{n}$ is, that it is a zero mean random process with power σ_n^2.

The Scalar Costa Scheme works on a sample-by-sample basis. The value of the to-be-embedded symbol m determines the selection of a quantizer $Q_\Delta^m(\cdot)$, which quantizes according to

$$Q_\Delta^m(x) \quad = \quad \text{round}\left(\frac{x}{\Delta} - \frac{m}{D}\right)\Delta + \frac{m\Delta}{D}. \tag{1}$$

That is, $Q_\Delta^m(x)$ denotes rounding x to the nearest value $\left(k + \frac{m}{D}\right)\Delta$, where Δ is the (fixed) quantization step-size. For the binary case ($D = 2$) the quantization (1) is as simple as rounding x to the nearest odd or even multiple of $\frac{\Delta}{2}$, depending on the value of m.

Usually, the quantization should include the use of a dither signal. Among others, this would lead to a uniform distribution of the to-be-quantized signal within each quantization bin. As a result, the quantization error would be uniformly distributed on the interval $[\ -\Delta/2\ ,\ \Delta/2\ ]$.[3] In this paper, we will not explicitly include the dither in all formulas. However, we do assume that the host signal is uniformly distributed within each quantization bin.

The quantization error is given by $Q_\Delta^m(x) - x$. A message symbol m is embedded in the corresponding host signal sample x according to the following formula:

$$\begin{aligned} y \quad &= \quad (1-\alpha)x + \alpha Q_\Delta^m(x) \\ &= \quad x + \alpha\left(Q_\Delta^m(x) - x\right). \end{aligned} \tag{2}$$

In this formula, α is the distortion compensation parameter. It determines what fraction of the quantization error is embedded as watermark; $w = \alpha(Q_\Delta^m(x) - x)$. It can be used to trade-off embedding distortion and robustness, and as such it is used to maximize the capacity of the watermarking scheme. The optimal values of the distortion compensation parameter and the quantization step-size are completely determined by the host signal variance σ_x^2 and the channel noise variance σ_n^2, according to the following formulas (assuming Gaussian sources and a given document-to-watermark ratio $DWR = \sigma_x^2/\sigma_w^2$, where σ_w^2 is the variance of the watermark signal w):

$$\begin{aligned} \Delta^2 \quad &= \quad \frac{12(\sigma_x^2 + 2.7\sigma_n^2\, DWR)}{DWR}, \\ \alpha^2 \quad &= \quad \frac{\sigma_x^2}{\sigma_x^2 + 2.7\sigma_n^2\, DWR}. \end{aligned}$$

In the detector, the received value r is rounded to the nearest value $(k + \frac{m}{D})\Delta$, and the corresponding value $\hat{m}$ is the decoded message bit:

$$\hat{m} \quad = \quad \text{round}\left(\frac{rD}{\Delta}\right) \quad \text{mod } D. \tag{3}$$

3. THE CASE FOR ADAPTIVITY

In this paper, we propose a method for making the SCS adaptive to local signal characteristics. As said before, this adaptivity serves a number of purposes.

First of all, the theory of SCS assumes a stationary host signal, and the embedding parameters are tuned to the variance of this host signal. In practice, however, a real host signal of interest (audio, video or image) is never stationary. Therefore, to retain the optimality of the scheme one needs to choose the embedding parameters based on the *local* signal variance of the non-stationary signal.

A second reason for introducing adaptivity, is the fact that the watermark variance σ_w^2 does not give a good indication of the perceptibility of the watermark. The local signal characteristics determine the sensitivity of human perception to changes in the signal. Several models exist to quantize this masking property of a signal. The more realistic models are specific to a modality (auditory, visual, ...) and moreover, they are usually very complex. One very simple, generic perceptual model is Weber's Law. Because of its simplicity, we will use is as the basis for our adaptive scheme.

According to Weber's law, sensitivity human perception is inversely proportional to the intensity of the stimulus. For instance, the human eye is less sensitive to brightness changes for high luminance values, than it is for low luminance values: It is more difficult to see the difference between 100 and 101 candles, than it is between 1 and 2. Therefore, it is possible to create relatively large distortions in samples with high luminance values, without compromising the perceptual quality.

Another simple perceptual model on which the adaptivity could be based would be the local signal variance: The masking capability of a signal is proportional to its variance. This would lead to a watermarking scheme where locally $\sigma_w^2 = \beta\sigma_x^2$. This model is actually being used for spread spectrum watermarking.[4] Note that it would would fit very well with the formula's for SCS. $\sigma_w^2 = \beta\sigma_x^2$ implies that the *DWR* is constant, but α and Δ are varying because of variations in local signal variance σ_x^2. The major drawback of this adaptation law is that the local signal variance is very sensitive to many attacks. Most particularly, lossy compression and noise addition have a very strong impact on the signal variance. For that reason, the detector would be unable to re-estimate quantization step-size used during embedding.

The third reason for adaptivity is to realize robustness to value scaling (like contrast changes in video content). Such changes are modelled by scalar multiplication $\tilde{x} = \beta x$. Currently, the fixed quantization step-size leads to high decoding error rates, when the signal sample values have been scaled with a scaling factor β which is sufficiently different from 1. Introduction of an adaptation rule which scales together with the scaling of the signal samples would introduce a high level of robustness to value scaling attacks. All adaptation rules which are linear in the signal sample values do exhibit this improvement. The Weber-based adaptive scheme, as introduced in the following section, is an example of such a linear adaptation rule.

4. WEBER BASED ADAPTIVE SCS

We have explained in the previous section why it is a good idea to develop an adaptive version of SCS, based on Weber's law. That is, an SCS scheme where the quantization step-size is proportional to the local signal sample values.

A general adaptive scheme would look like the scheme in figure 1. We will now discuss the contents of the two blocks for determining the quantization step-size.

As a first guess, one could think that it is appropriate to use the adaptation rule $\Delta(x) = \gamma x$ for some positive number γ. However, this is not a good choice, as it does not allow the retrieval of embedded information. This can be seen from

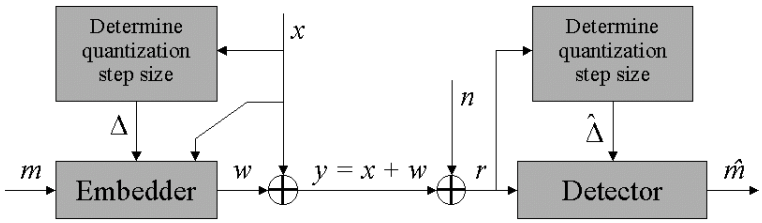

Figure 1. The adaptive quantization step-size Δ is derived from the host signal x and is used in the embedder by applying Equation (2). At the detector, the adaptive quantization step-size is estimated from the received signal r, which results in an estimate $\hat{\Delta}$. $\hat{\Delta}$ is used for the watermark detection given by Equation (3).

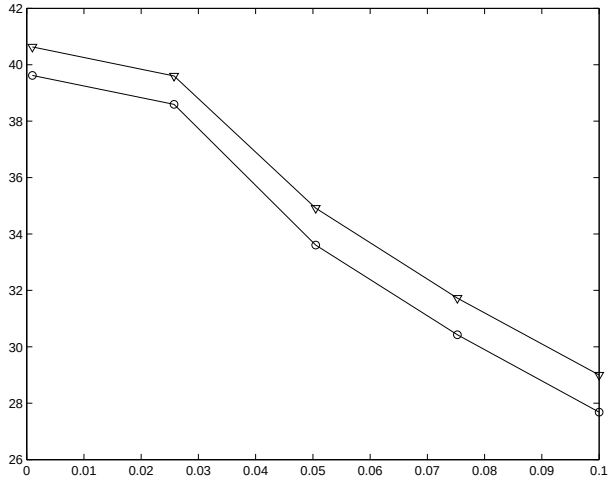

Figure 2. The Document-to-watermark ratio (in dB) as a function of γ for two different images. Upper curve is for the "peppers" image, the lower curve for "Lena".

the detection formula. In the detector, the quantization step-size will be estimated as $\hat{\Delta} = \gamma r$, which results in the detected message

$$\begin{aligned} \hat{m} &= \text{round}\left(\frac{rD}{\hat{\Delta}}\right) \mod D \\ &= \text{round}\left(\frac{D}{\gamma}\right) modD. \end{aligned}$$

Clearly, $\hat{m}$ is completely independent of the embedded message m, so no information can be transmitted in this way.

Instead, we propose to choose the adaptive step-size to be proportional to a local average of the signal samples:

$$\Delta(x) = \frac{\gamma}{L}\sum_{i=1}^{L} x_i, \tag{4}$$

where $x_1, \ldots, x_L$ is a neighborhood of size L of the sample x. γ is a scalar parameter which controls the overall embedding distortion σ_w^2. In Figure 4 the watermark to noise ratio *DWR* (in dB) is plotted as a function of γ for two different images.

Detection is done with estimated adaptive quantization step-size

$$\begin{aligned} \hat{\Delta}(r) &= \frac{\gamma}{L}\sum_{i=1}^{L}(x_i + w_i + n_i) \\ &= \Delta(x) + \frac{\gamma}{L}\sum_{i=1}^{L}(w_i + n_i). \end{aligned}$$

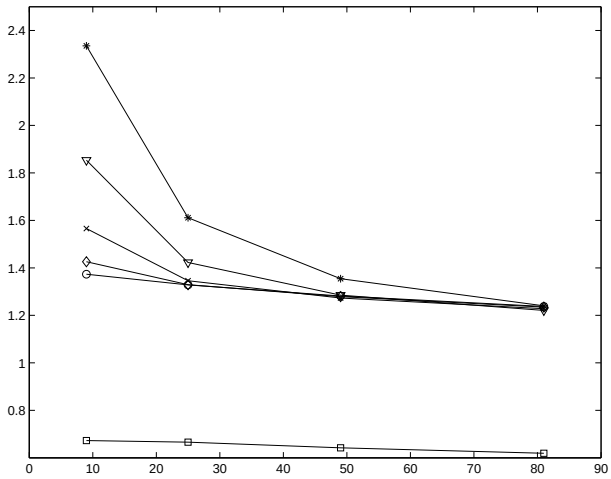

Figure 3. The percentual relative estimation error for the "peppers" image. Horizontal axis: Neighborhood sizes (L). Vertical axis: Relative Δ estimation error (in $\%$). Noise variances ranging from 0 to 32.

As $\mathrm{E}w = \mathrm{E}n = 0$, the estimation error $\epsilon = \Delta - \hat{\Delta}$ satisfies

$$\mathrm{E}(\epsilon) = 0, \tag{5}$$

$$\mathrm{E}(|\epsilon|) = \frac{\sigma}{\sqrt{2\pi}}, \tag{6}$$

$$\mathrm{var}(\epsilon) = \frac{\gamma^2}{L}(\sigma_w^2 + \sigma_n^2), \tag{7}$$

which tends to 0 for increasing values of L. Therefore, L determines the robustness of the estimation of Δ. This allows for a trade-off between the locality of the adaptation (i.e., a more accurate match to the perceptual model) and the error rate of the detector. To give an indication of the typical bahaviour, we computed the percentual relative estimation error $100 * \epsilon/\Delta$ for different values of the noise variance σ_n^2 and the neighborhood size L. In figure 3 the result have been plotted for the "peppers" image.

As explained before, another important advantage of the proposed scheme using the adaptation rule given by Equation (4) is its robustness with respect to sample value scaling. Let us do the analysis for our specific scheme. In the case of scaling the received signal with a parameter β, the received value equals $r = \beta(x + w)$, from which the adaptive quantization step-size

$$\hat{\Delta}_{\mathrm{s}}(r) = \frac{\gamma}{L}\sum_{i=1}^{L}\beta(x_i + w_i)$$

is estimated. This is equal to β times the adaptive quantization step-size Δ_{ns} in the case of no sample value scaling, so $\hat{\Delta}_{\mathrm{s}} = \beta\hat{\Delta}_{\mathrm{ns}}$. The detected message bit

$$\begin{aligned}\hat{m} &= \mathrm{round}\left(\frac{\beta(x+w)}{\Delta_{\mathrm{s}}/D}\right) \mod D \\ &= \mathrm{round}\left(\frac{x+w}{\Delta_{\mathrm{ns}}/D}\right) \mod D,\end{aligned}$$

which is exactly the detected message bit in the case of no sample value scaling. The robustness to sample value scaling is confirmed by experiments, see Figure 4. The errors occurring for the adaptive quantization scheme at smallest and largest scales are mainly due to effects of quantization and clipping of the scaled image samples to 8-bit values.

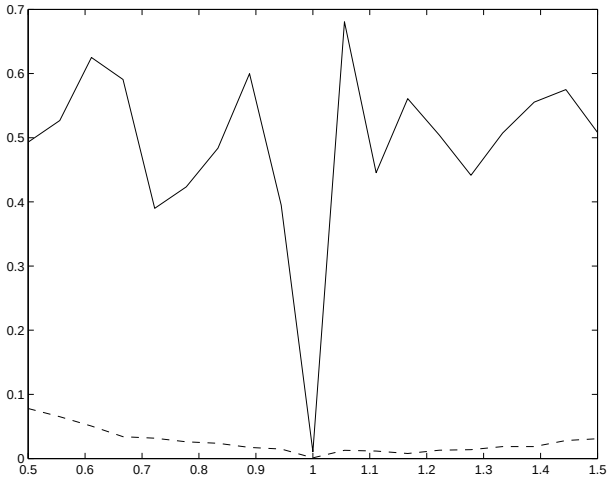

Figure 4. Adaptive quantization versus fixed quantization for the "Lena" image. The measured bit error probability as a function of the brightness scaling factor. We used rate 1/32 repetition coding and a Document-to-Watermark-Ratio equal to 37 dB.

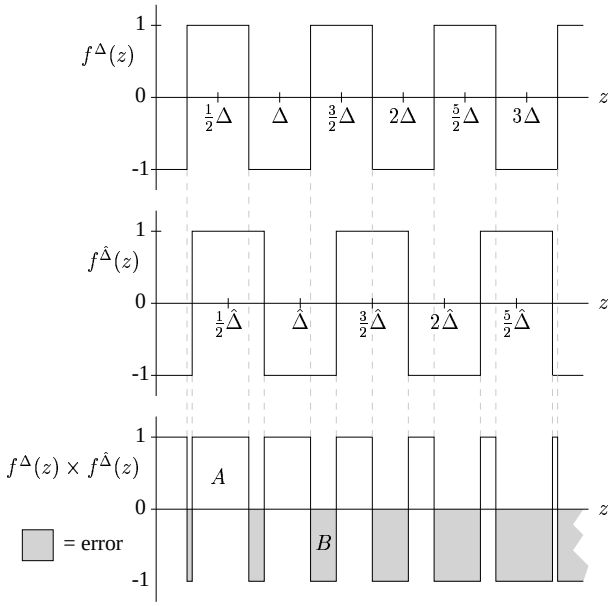

Figure 5. The detection error made due to the errors in the estimation of the quantization step-size.

5. A MODEL FOR THE BIT ERROR PROBABILITY

As can be seen from Figure 1, the estimation of the adaptive quantization step-size from the received value r, introduces errors in the detected message $\hat{m}$. In this section we derive a model for determining the Bit Error Probability due to the use of an adaptive quantization step-size. We assume binary messages ($D = 2$) and the absence of noise ($\mathbf{n} = 0$).

Consider Figure 5. The horizontal axis represents the sample value, the vertical the detected message bit (where -1 represents 0). The upper plot displays the detection response as a function of the received value r with a correctly estimated quantization step-size Δ and the middle plot detection with an estimated adaptive quantization step-size $\hat{\Delta}$ slightly larger than Δ. An error is made when the signs of the two function are not equal, or when the product is negative. The detection function is given by

$$f^{\Delta}(r) \quad = \text{sign}\left(-\cos\left(\frac{\pi}{\frac{1}{2}\Delta}r\right)\right). \tag{8}$$

The lower plot of Figure 5 shows the sample values where an error is made.

The area for which no error is made is called A, the area for which an error is made B. After scaling the variables r such that its range becomes $[0, 1]$ and scaling Δ and $\hat{\Delta}$ accordingly, we obtain

$$\begin{aligned} A + B &= 1, \\ A - B &= \int_0^1 \left(f^{\Delta}(r) \times f^{\hat{\Delta}}(r) \right) p_r(r)\, \mathrm{d}r, \end{aligned}$$

where r is the scaled version of the sample values and $p_r(r)$ the pdf of r. From this, we compute the bit error probability as a function of Δ and $\hat{\Delta}$:

$$\begin{aligned} P_{\mathrm{BE}}(\Delta, \hat{\Delta}) &= \frac{1}{2} - \frac{1}{2} \int_0^1 \Big(f^{\Delta}(r) \times f^{\hat{\Delta}}(r) \Big) p_r(r)\, \mathrm{d}r \\ &= \frac{1}{2} - \frac{1}{2} \int_0^1 \mathrm{sign}\Big(\cos\Big(\frac{\pi}{\Delta} r \Big) \cos\Big(\frac{\pi}{\hat{\Delta}} r \Big) \Big) p_r(r)\, \mathrm{d}r. \end{aligned}$$

This integral is hard to solve analytically for general distributions of the sample values $p_r(r)$. In Figure 6, a plot is shown of P_{BE} as a function of the relative Δ estimation error (in %) for different values of Δ ($\Delta = 1,\ 10,\ 20,\ 30,\ 40,\ 50$ assuming a uniform distribution on $[0, 255]$ for r. Given the maximum allowable bit error probability, it gives an indication of the necessary precision in estimating Δ. By Equation (7) this determines a minimum required value for L.

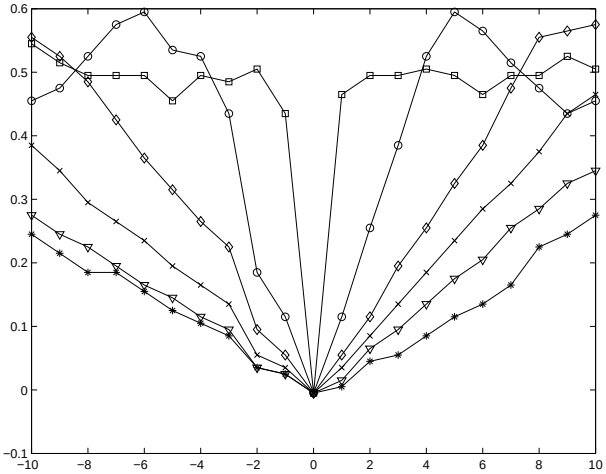

Figure 6. A plot of the Bit Error Probability P_{BE} as a function of the relative percentage Δ estimation error $100 * (\Delta - \hat{\Delta})/\Delta$, for various values of Δ.

6. CONCLUSIONS

We have introduced a quantization watermarking scheme with adaptive quantization step-size. The step-size for a sample is taken proportional to the average value of a number of neighboring samples. This allows for perceptual shaping of the watermark according to Weber's law. Moreover it leads to robustness against value scaling. We developed a model for the effect of step-size estimation errors on the Bit Error Probability. This model can be used to determine the needed precision of estimation of the adaptive quantization step-size.

REFERENCES

1. M. Costa, "Writing on dirty paper," *IEEE Transactions on Information Theory* **29**(3), pp. 439–441, 1983.
2. J. Eggers, R. Bäuml, R. Tzschoppe, and B. Girod, "Scalar Costa scheme for information embedding," *IEEE Transactions on Signal Processing* **51**(4), pp. 1003 – 1019, 2003.
3. S. Lipschitz, R. Wannamaker, and J. Vanderkooy, "Quantization and dither: a theoretical survey," *Journal of the audio engineering society* **40**(5), pp. 355–375, 1992.
4. M. Maes, T. Kalker, J. Linnartz, J. Talstra, G. Depovere, and J. Haitsma, "Digital watermarking for DVD copy protection," *IEEE Signal processing Magazine* **17**(5), pp. 47–57, 2000.