An Improved Scalar Quantization-based Digital Video Watermarking Scheme for H.264/AVC

Adarsh Golikeri, Panos Nasiopoulos and Z. Jane Wang
Department of Electrical and Computer Engineering
University of British Columbia
Vancouver, BC, Canada V6T 1Z4
Email: {adarshg, panosn, zjanew}@ece.ubc.ca

Abstract—Digital video watermarking has attracted a great deal of research interest in the past few years in applications such as digital fingerprinting and owner identification. H.264/AVC is the latest and most advanced video coding standard, but to this date, there are very few watermarking schemes designed for it. This is mainly due to its complexity and compression efficiency which presents a major challenge for any video watermarking approach. We developed a new, quantization-based video watermarking scheme, which is designed to work with H.264/AVC. Our scheme offers constant robustness at all compression rates without affecting the overall bit rate and quality of the video stream. Experimental results show that, compared to existing methods, our scheme significantly outperforms existing methods under compression, transcoding, filtering, scaling, rotation and collusion attacks.

I. INTRODUCTION

In the past few years, the need for watermarking has gained significant attention due to the spread of illegal redistribution and unauthorized use of digital multimedia [1]. A great variety of watermarking schemes has been proposed in the literature. The Scalar Costa Scheme (SCS) is a reliable information embedding technique, which is based on Costa’s original, theoretical scheme ([2],[3]). SCS outperforms the related Dither Modulation (DM) techniques for low Watermark-to-Noise Ratios (WNR) and performs significantly better than the state-of-the-art blind Spread Spectrum (SS) watermarking [3]. SCS, however, has certain inherent limitations when used for video watermarking, such as fixed watermarking embedding strength and no Rate-Distortion optimization. Moreover, SCS does not provide any way of controlling the spatial and temporal distortions caused by the watermark insertion. Finally, the distribution of the watermark bits is not dependent on the regional visual importance of the input signal.

In this paper, we propose a new watermarking method which is specifically designed for video. Our method borrows ideas from SCS and is extremely robust to compression, transcoding, filtering, scaling, rotation and collusion. This is achieved by offering a locally adaptive watermark embedding strength and a scheme for optimum Rate-Distortion. A unique perceptual mask controls the levels of spatial and temporal distortion, while a built-in bit-rate control mechanism is used to ensure optimum watermark bit allocation. We designed our method to work within the H.264/AVC video standard, the latest and most advanced video coding standard [4]. This paper is organized as follows: Section II offers a brief overview of SCS. Section III describes our watermarking scheme. Section IV shows how our method is adapted to work with H.264/AVC. Experimental results are presented in Section V and conclusions in Section VI.

II. THE SCALAR COSTA SCHEME (SCS)

The watermarking process can be considered as a communications system with side-information at the encoder side (see Fig. 1) ([2],[3]). Henceforth, bold text (x) denotes a vector, while normal text (s) and italics (s) denote a scalar. Using a secure key K, the watermark message m is embedded into the host signal x of variance $\sigma_x^2$. The watermark is defined as $w = s - x$ and has a variance $\sigma_w^2$. The watermarked signal s is then transmitted over a channel, which introduces an additive white Gaussian noise (AWGN) v of variance $\sigma_v^2$, resulting in an attacked work r. The Watermark-to-Noise Ratio (WNR) is defined as $10 \times \log_{10}(\sigma_w^2/\sigma_v^2)$. The decoder receives signal r and, using the same key K which was used during embedding, extracts the watermark message estimate m’ (Fig. 1). The Scalar Costa Scheme (SCS) uses a structured codebook, constructed by a concatenation of scalar uniform quantizers [3]. For a Costa-type embedding of a watermark message m, SCS determines an intermediate sequence q which is nearly orthogonal to the cover work x. This message m is encoded into watermark letters d that belong to a D-ary alphabet (e.g., D=2 represents a binary alphabet). For embedding a watermark, the following sample-wise operation is performed:

$$q_n = Q_{\Delta}\{x_n - \Delta(d_n/D + k_n)\} - (x_n - \Delta(d_n/D + k_n)),$$  

(1)

where $q_n$, $x_n$, $d_n$ and $k_n$ are elements of the vectors q, x, d and k respectively, $Q_{\Delta}\{\cdot\}$ denotes scalar uniform quantization with step size $\Delta$, and $k_n \in [0,1)$ are the elements of a secure pseudo-random sequence k, derived.
from the watermark key K. The embedded watermark sequence is given by
\[ w = \alpha q, \]  
(2)

where \( \alpha (0 \leq \alpha \leq 1) \) is the watermark scale factor. SCS does not provide an analytical formula for determining the optimum value of \( \alpha \). Instead, an optimum \( \alpha \) (in the capacity sense) is derived by the following numeric expression:
\[ \alpha = \sqrt{\frac{\sigma_w^2}{\sigma_w^2 + 2.71 \sigma_v^2}}. \]  
(3)

The final watermarked data is represented by:
\[ s = x + w = x + \alpha q. \]  
(4)

For watermark decoding, the received data \( r \) is quantized to the nearest codebook entry. The sample-wise extraction rule is:
\[ y_n = Q_{\Delta} \{ r_n - \Delta k_n \} - (r_n - \Delta k_n), \]  
(5)

where \( y_n, r_n \) and \( k_n \) are elements of \( y, r \) and \( k \), respectively. For binary SCS, \( |y_n| \) should be close to zero if \( d_n = 0 \) and close to \( \Delta/2 \) for \( d_n = 1 \). This is known as minimum distance decoding.

III. PROPOSED WATERMARKING SCHEME

Lately, Spread Transform coding [5] has been used in combination with traditional SCS (known as ST-SCS) to improve the bit-error rate of watermarking (Fig. 2). We have developed a new watermarking method which borrows ideas from Spread Transform coding and SCS and is specifically designed for video. In traditional Spread Transform Scalar Costa Scheme (ST-SCS) watermarking, the host signal \( x \) is projected onto a pseudo-random vector. The disadvantage of this approach is that it does not account for perceptual masking effects of the Human Visual System (HVS). In our scheme, we derive a unique perceptual mask sequence \( t \) (derived from the host signal \( x \) itself) in order to achieve imperceptibility.

The generation of \( t \) differs depending on which type of macroblock is used. In case of Intra macroblocks, spatial masking effects are considered. First, a Gaussian low-pass filter is applied to the macroblock in order to mitigate the effect of noise. Then, for each given macroblock, a perceptual mask is computed using the Watson’s model. Watson’s model estimates the perceptibility of changes in the coefficients of the block-based DCT of an image. This model consists of a frequency sensitivity function, luminance and contrast masking components [1]. The frequency sensitivity is a table defined by the model, with each table entry representing the smallest magnitude of the corresponding DCT coefficient in a block that can be perceived by the eye. We use the standard frequency sensitivity table used in [1]. Luminance masking accounts for the effect of the DC-component (i.e. the average brightness of the block) on the frequency sensitivity table. Contrast masking takes into account the effect of visibility of a change in one frequency due to the energy present in that particular frequency. After accounting for these effects, the thresholds or slacks for the individual DCT coefficients are obtained. These slacks represent the amounts by which the individual coefficients maybe changes, before resulting in a perceptible change in the block. These slacks represent the perceptual mask, denoted by \( p \). For Intra macroblocks the perceptual mask sequence \( t \) is given by:
\[ t = \frac{p}{|p|}. \]  
(6)

In case of Inter macroblocks, the effect of both spatial and temporal masking must be considered. Previous research has shown that watermark artifacts, such as “mosquito” effects and flicker, are visible in the fast moving regions of a frame [6]. These artifacts correspond to regions with a large motion vector values. For this reason, the strength of the watermark should be reduced in such regions. This is achieved by weighting the perceptual mask by the inverse of the motion vector magnitude. Thus, for Inter macroblocks, the perceptual mask is given by:
\[ t = \frac{p}{|mv|}, \]  
(7)

where \( p \) is the Watson’s perceptual mask and \( |mv| \) is the absolute magnitude of the motion vector. The elements of \( t \) in (7) are then normalized.

Once the perceptual mask \( t \) is generated, the projection of \( x \) onto \( t \) is found. This operation yields a scalar quantity:
\[ x' = x^T t. \]  
(8)

In our scheme, \( x \) represents the transform domain coefficients of Intra and Inter coded macroblocks. The watermark key \( K \) is used to generate the random scalar value \( k \in [0,1) \). For binary ST-SCS, the equation for embedding a '0' bit is obtained by putting \( d=0 \) and \( D=2 \) in equation (1):
Similarly, embedding a ‘1’ bit is possible by setting d=1 and D=2 in (1):

\[ s' = Q_{\alpha}(x' - \Delta k) - (x' - \Delta k). \]  

(9)

Equation (9) and (10) represent subtractive dithered quantization. The components of x that are orthogonal to t are equal to x-x'. These components are not altered during the embedding process and are for this reason they are added back to the watermark data. Therefore, the final watermarked data s is obtained by combining (4) with the orthogonal components:

\[ s = (x' + \alpha s')t + (x-x'). \]  

(11)

Traditional ST-SCS uses a fixed \( \alpha \) that is computed from global statistics (see equation (3)). In contrast, our method uses a locally adaptive value for \( \alpha \) which is computed in real-time from a combination of local and global statistics. As a result, we obtain stronger control over the watermark scale factor, which makes our watermark to adapt better to the host signal characteristics. As a result, our method is much more robust than traditional ST-SCS.

During video encoding, several coding parameters such as macroblock prediction modes, motion vectors and transform coefficient quantization levels have to be determined. Since natural video has widely varying spatial and temporal (motion) content, the selection of different coding options for different parts of the image becomes necessary. Therefore, the task of the video coder is to find a set of coding parameters so that a trade-off between the video bit-rate and distortion (R-D) is achieved. This means that for a given video bit-rate, the encoder has to find the combination of coding options that minimizes the distortion. Lagrangian bit-allocation techniques for R-D coding have been widely accepted in recent video codec development, due to their effectiveness and simplicity. Adding a watermark to a video stream may also affect the bit rate and quality of the image. It is, therefore, highly desirable that the watermark embedding procedure incorporates R-D optimized coding in order to compute the optimum watermark for different regions of a video frame.

To this end, we use the Lagrangian multiplier technique to compute the locally optimum value of \( \alpha \) at the macroblock level. The simplified Lagrangian cost function for a particular value of \( \alpha \) is:

\[ J_\alpha = D_\alpha + \lambda_w E_\alpha, \]  

(12)

where \( D_\alpha \) is the distortion (sum of squared differences or SSD) between the host signal x and the watermarked work \( s' \), \( \lambda_w \) is the Lagrangian parameter and is dependent on the choice of the video standard used for encoding [7], and \( E_\alpha \) is the decoding error = \( ||D_\alpha - D_e|| \). We define \( D_\alpha \) as the decoded distance and \( D_e \) as the expected distance. \( D_\alpha \) is equal to 0 if the embedded message bit \( d = 0 \), and it is equal to \( \pm \Delta/2 \) if \( d = 1 \). To obtain \( D_\alpha \), the watermarked data \( s \) is projected onto the perceptual mask \( t \), which results in the scalar \( e' \). The quantization of \( e' \) yields \( D_e \):

\[ D_e = Q_{\alpha}(e' - \Delta k) - (e' - \Delta k). \]  

(13)

For each macroblock, we compute the value of \( \alpha \) which minimizes the Lagrangian cost function (11). This value of \( \alpha \) is the Rate-Distortion optimum watermark scale factor which we use in our method.

Watermarked video may require many more bits than unwatermarked video, especially at low video bit-rates. Therefore, it is desirable to have a bit-rate control scheme in order to find the optimum trade-off between the fidelity of the watermark and that of the host signal. This scheme should determine the best allocation of available watermark bits between different watermarked macroblocks. In video coding, the most important factor for controlling the bit-rate is the residual signal coding fidelity, which is controlled by choosing a suitable quantization step-size for the transform coefficients. Our scheme is designed in such a way that it achieves watermark bit-rate control simply by changing the quantization step \( \Delta \) which is used to embed the watermark in (9), (10) and (13). Therefore, our scheme has a built-in mechanism for watermark bit-rate control, through the parameters \( \Delta \) and \( \lambda_w \). This is an important advantage over existing schemes, such as [6], which require an explicit bit-rate controller. In Section 4, we explain how bit-rate control is achieved when we implement our scheme on H.264/AVC.

Another important advantage of our method is that the watermark can be decoded from the partially decompressed video bit stream, since the watermark is embedded in the transform coefficients. Decoding of the watermark requires knowledge of the secure key K, which is needed to generate the pseudorandom scalar \( k \). The perceptual mask \( t \) is computed for this macroblock as explained at the beginning of this section. The transform coefficients are projected onto \( t \) to obtain the scalar projection \( y' \). This projection is then quantized using (13) and simple hard decision decoding is used to extract the message \( m' \).

IV. WATERMARKING OF H.264/AVC VIDEO

One of the challenges for designing watermarking schemes for H.264 is that even the Intra-frames consist mainly of residual data. This means that adding a watermark without affecting the picture quality or the bit rate is extremely difficult. H.264 achieves bit rate reduction for the Intra-frames by using spatial prediction for Intra macroblocks, a major departure from the previous coding standards. Intra macroblocks can be predicted using Intra_16x16 (entire macroblock predicted from top and left neighbors) or Intra_4x4 (each 4x4 luma block separately predicted from its neighbors) modes. Inter macroblocks use variable block-size motion compensation. The different sizes include 16x16, 16x8, 8x16 and 8x8. The 8x8 partition can be further divided into 8x4, 4x8 or 4x4 blocks. Thus a motion vector is transmitted for each partition, adding more complexity to the coding scheme.
Another challenge that H.264 poses to watermarking is that it uses an integer transform. This is a major problem for traditional Spread Spectrum watermarking schemes which embed watermarks drawn from a Gaussian distribution.

Our algorithm addresses all the above issues and is designed to work efficiently within H.264. First, in order to take into account the variety of motion vectors while deriving the perceptual mask sequence \( t \) for a subdivided macroblock in (7), we divide the Watson’s perceptual mask \( p \) by the motion vectors of the corresponding regions.

We consider watermarking the luma components of both the Intra and Inter macroblocks, operating on the integer transform coefficients of the macroblock residual data. This is possible since we designed our method without making any assumptions about the nature of the host signal \( x \). For macroblocks predicted using the Intra_16x16 mode, the Hadamard coefficients are watermarked, because they contain most of the energy of the macroblock. For those predicted using the Intra_4x4 mode and all the Inter macroblocks, we watermark the integer transform coefficients.

Let \( Q_{P,H.264} \) denote the Quantization Parameter (QP) value used in H.264, \( Q_{step} \) denote the H.264 quantization step size, \( Q_{P,w} \) denote our watermark quantization parameter and \( \Delta \) denote our watermark quantization step-size. The \( Q_{P,H.264} \) and \( Q_{step} \) are related as follows:

\[
Q_{step} = 0.6282 \times \exp(Q_{P,H.264} \times 0.1155), \quad 0 \leq Q_{P,H.264} \leq 51. \quad (14)
\]

The behavior of (14) is such that for values of \( Q_{P,H.264} \) in the range 0-30, the corresponding \( Q_{step} \) changes by very small amounts. However, for \( Q_{P,H.264} \) between 31-51 the \( Q_{step} \) increases very rapidly. Our main objective is to control the distribution of the watermark bits in such a way that minimum Bit-Error Rate (BER) is achieved, independent of the video bit-rate. We achieve this by establishing a relationship between our watermark step size \( Q_{P,w} \) and \( Q_{P,H.264} \). Evaluations over a large set of video sequences indicated that the minimum BER is obtained when:

\[
Q_{P,w} = 48, \quad 0 \leq Q_{P,H.264} \leq 30, \quad \text{and} \quad (15)
\]

The relationship between \( Q_{P,H.264} \) and \( \Delta \) is exactly the same as that between \( Q_{P,H.264} \) and \( Q_{step} \). Therefore, equivalently,

\[
\Delta = 160, \quad 0 \leq Q_{P,H.264} \leq 30, \quad \text{and} \quad (17)
\]

\[
\Delta = 0.6882 \times \exp(Q_{P,H.264} \times 0.1686), \quad 31 \leq Q_{P,H.264} \leq 51. \quad (18)
\]

Thus, there is a close relationship between (14) and (18). Both equations represent exponential curves, with an initial slow ascent between 0-30 and then a rapid increase in the range 31-51. This guarantees that our watermark robustness is constant at all different compression rates.

Our watermark is embedded on the transform coefficients before the Quantization process. In H.264, bit-rate control is achieved through proper selection of the \( Q_{P,H.264} \). It has been shown in [7] that there is a strong relationship between the Lagrangian parameter \( \lambda \) used for R-D coding and \( Q_{P,H.264} \):

\[
\lambda = 0.85 \times \text{pow}(2, (Q_{P,H.264} - 12)/3). \quad (19)
\]

Therefore, bit-rate control in H.264 is conducted by controlling \( Q_{P,H.264} \) and accordingly adjusting the value of \( \lambda \). Similarly, in our method, the watermark bit allocation is controlled by choosing the step size of the quantizer \( Q_{P,w} \) (or equivalently, \( \Delta \)) and adjusting the value of \( \lambda_{w} \) used in (11). The Lagrangian parameter \( \lambda_{w} \) is given by:

\[
\lambda_{w} = 0.85 \times \text{pow}(2, (Q_{P,w} - 12)/3). \quad (20)
\]

We then use this \( \lambda_{w} \) in (12) to determine the locally optimum watermark scale factor \( \alpha \). By selecting this value of \( \alpha \), we ensure that our watermark will not affect the R-D optimized coding decisions of the H.264 encoder. When the H.264 encoder varies \( Q_{P,H.264} \) to achieve the desired overall video bit-rate, \( Q_{P,w} \) changes proportionally since it is related to \( Q_{P,H.264} \) through (15) and (16). Thus, the watermark bits are allocated in proportion to the H.264 encoder’s bit-rate control algorithm. This ensures that the overall video bit-rate is not adversely affected.

V. EXPERIMENTAL RESULTS

The performance of our scheme was tested on 10 video sequences (Carphone, Coastguard, Football, Foreman, Flower Garden, Mother Daughter, News, Paris, Tempete and Tennis) that represent all aspects of content. These are: Under the same picture quality (PSNR and bit rate), we compared the robustness of our method against the traditional ST-SCS scheme for several different attacks. We used the H.264/AVC reference software version JM9.3 for our implementation [8].

Fig. 3 shows the Bit Error Rates (BER) caused by H.264 compression at various bit-rates, for 2 representative streams. We observe that our method significantly outperforms ST-SCS. Moreover, the robustness of our method is constant at all the bit-rates, and results in dramatic improvements over ST-SCS. This indicates that the watermark bit-rate control in (15) and (16) and the Rate-Distortion optimization in (20) are performing very well by adapting the watermark step size \( \Delta \) and the watermark scale factor \( \alpha \) to the compression rate. On the other hand, the performance of traditional ST-SCS suffers in spite of using the same watermark power \( \sigma_{w}^{2} \) during embedding.

Fig. 4 shows the Watermark-to-Noise Ratios (WNRs) against BERs (at 512kb/s) for two extreme cases. On average, our scheme requires about 2 decibels less WNR to achieve minimum BER. We also observe that, for the same WNR, the BER achieved by our scheme is about 2 orders of magnitude lower than ST-SCS.

For the same picture quality (PSNR), Table I shows the results for transcoding (sequences recompressed at the same bit rate but with a different GOP structure), Gaussian filtering (3x3, variance 0.5), 75% scaling, 5° rotation with
bilinear sampling, and collusion attacks (5 different watermarked copies were averaged at 512 kb/s). We observe that, on average, our scheme is 3 times better than ST-SCS when transcoding is performed. Under Gaussian low-pass filtering, downscaling and rotation attacks, our scheme yields BERs less than half those of ST-SCS. Finally, the average BER obtained by our scheme after collusion attacks is less than 1/4th that of ST-SCS. All in all, our scheme significantly outperforms ST-SCS in all the above attack categories.

VI. CONCLUSION

We have presented a new video watermarking scheme, which is designed to work for H.264/AVC. Our scheme consists of a locally adaptive, R-D optimized watermark which is inserted in the transform coefficients of macroblock residuals. A unique perceptual mask limits distortion, while a built-in bit-rate control mechanism ensures optimum watermark bit allocation. In the category of compression-decompression, our scheme yields bit-error rate improvements of more than two orders of magnitude compared to traditional ST-SCS. Our scheme achieves the same BER as ST-SCS, using 2 decibels less Watermark-to-Noise (WNR). In addition, our scheme significantly outperforms ST-SCS after geometric and collusion attacks.

REFERENCES


TABLE I. BIT ERROR RATES AFTER VARIOUS WATERMARK ATTACKS

<table>
<thead>
<tr>
<th></th>
<th>Transcoding</th>
<th>Gaussian filtering</th>
<th>75% downscaling</th>
<th>5° rotation</th>
<th>Collusion</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>BER x 10^-3</td>
<td>BER x 10^-3</td>
<td>BER x 10^-3</td>
<td>BER x 10^-3</td>
<td>BER x 10^-3</td>
</tr>
<tr>
<td>Football</td>
<td>Proposed</td>
<td>Existing</td>
<td>Proposed</td>
<td>Existing</td>
<td>Proposed</td>
</tr>
<tr>
<td>Football</td>
<td>6.5724</td>
<td>48.5791</td>
<td>4.1212</td>
<td>40.6060</td>
<td>2.0471</td>
</tr>
<tr>
<td>News</td>
<td>15.6296</td>
<td>36.8014</td>
<td>26.6126</td>
<td>46.3030</td>
<td>16.7137</td>
</tr>
</tbody>
</table>

Figure 3. Bit error rates after H.264 compression attack at different bit-rates

Figure 4. Bit error rates at different Watermark-to-Noise-Ratios