Robust watermarking of H.264-encoded video: Extension to SVC

Peter Meerwald and Andreas Uhl
Dept. of Computer Sciences
University of Salzburg, Austria
{pmeerw, uhl}@cosy.sbg.ac.at

Abstract

In this paper we extend a framework for robust watermarking of H.264-encoded video to scalable video coding (SVC) as defined in Annex G of the standard. We focus on spatial scalability and show that watermark embedding in the base resolution layer of the video is insufficient to protect the decoded video of higher resolution. This problem is mitigated by a proposed upsampling technique of the base layer watermark signal when encoding the enhancement layer. We demonstrate blind watermark detection in the full- and low-resolution decoded video and, surprisingly, can report bit rate savings when extending the base layer watermark to the enhancement layer.

1. Introduction

Distribution of video content has become ubiquitous and targets small, low-power mobile to high fidelity digital television devices. The Scalable Video Coding (SVC) extension of the H.264/MPEG-4 Advanced Video Coding standard describes a bitstream format which can efficiently encode video in multiple spatial and temporal resolutions at different quality levels [7]. Scalability features have already been present in previous MPEG video coding standards. They came, however, at a significant reduction in coding efficiency and increased coding complexity compared to non-scalable coding. H.264/SVC employs inter-layer prediction and can perform within 10% bit rate overhead for a two-layer resolution scalable bitstream compared to coding a single layer with H.264.

In this work we extend a well-known robust watermarking framework proposed by Noorkami et al. [5, 6] for copyright protection and ownership verification applications of H.264-encoded video content. The aim is to provide a single scalable, watermarked bitstream which can be distributed to diverse clients without the need to re-encode the video material. Scalability is provided at the bitstream level. A bitstream with reduced spatial and/or temporal resolution can be obtained by discarding NAL units [7]. The watermark should be detectable in the compressed domain *and* the decoded video without reference to the original content.

In Section 2 we briefly review the H.264 watermarking framework and investigate its applicability for protecting resolution-scalable video encoded with H.264/SVC. We propose an upsampling step of the base-layer watermark signal in Section 3 in order to extend the framework to SVC. Experimental results are provided in Section 4 followed by discussion and concluding remarks in Section 5.

2. Watermarking of H.264-encoded video

Several strategies have been proposed for embedding a watermark in H.264-encoded video. Most commonly, the watermark signal is placed in the quantized AC coefficients of intra-coded macroblocks. Noorkami et al. [5] present a framework where the Watson perceptual model for 8×8 DCT coefficients blocks [9] is adapted for the 4×4 integer approximation to the DCT which is predominantly used in H.264. Other embedding approaches include the modification of motion vectors or quantization of the DC term of each DCT block [2], however, the watermark can not be detected in the decoded video sequence or the scheme has to deal with prediction error drift.

Figure 1 illustrates the structure of the watermarking framework integrated in the H.264 encoder; each macroblock of the input frame is coded using either intra- or inter-frame prediction and the difference between input pixels and prediction signal is the residual[1]. We denote by $r_{i,j,k}$ the coefficients of 4×4 residual block k with $0 \leq i, j < 4$ and similarly by $o_{i,j,k}$ and $p_{i,j,k}$ the values of the original pixels and the prediction signal, resp. Each block is transformed and quantized, T denotes the DCT and Q the quantization operation in the figure. Let $R_{i,j,k}$ represent the corresponding quantized DCT coefficients obtained by

[1] Other modes are possible, e.g. *PCM* or *skip* mode, but rarely occur or are not applicable for embedding an imperceptible watermark due to lack of texture.

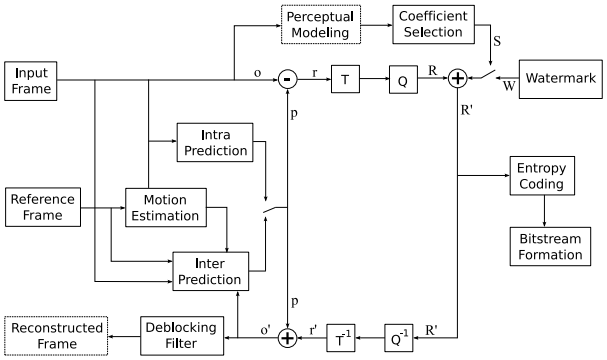

Figure 1. Watermarking 4×4 residual blocks

$R_k = \mathsf{Q}(\mathsf{T}(r_k))$. $R_{0,0,k}$ thus denotes the quantized DC coefficient of block k. After watermark embedding, described in the following paragraphs, and entropy coding, the residual information is written to the output bitstream.

For each block, a bipolar, pseudo-random watermark $W_{i,j,k} \in \{-1, 1\}$ with equiprobable symbols is generated and added to the residual block to construct the watermark block R',

$$R'_{i,j,k} = R_{i,j,k} + S_{i,j,k} \cdot W_{i,j,k} \tag{1}$$

where $S_{i,j,k} \in \{0, 1\}$ selects the embedding locations for block k. The design of S determines the properties of the watermarking scheme and differentiates between various approaches: in [5], embedding locations are selected based on the masked error visibility thresholds derived from the Watson perceptual model. Further, the number of locations is constrained to avoid error pooling and AC coefficients of large magnitude are preferred in the selection process.

The pixels of the reconstructed, watermarked video frame are given by $o'_{i,j,k} = p_{i,j,k} + r'_{i,j,k}$ where $r'_k = \mathsf{T}^{-1}(\mathsf{Q}^{-1}(R'_k)) = \mathsf{T}^{-1}(\mathsf{Q}^{-1}(R_k) + Q_k \cdot S_k \cdot W_k)$. For simplicity, we have dropped the coefficient indices i, j.

Watermark detection is performed *blind*, i.e. without reference to the original host signal, and can be formulated as a hypothesis test to decide between

$$\begin{aligned} \mathcal{H}_0 &: Y_l = O_l \text{ (no/other watermark)} \\ \mathcal{H}_1 &: Y_l = O_l + Q_l \cdot W_l \text{ (watermarked)} \end{aligned} \tag{2}$$

where O_l denotes the selected 4×4 DCT coefficients of the received video frames, Q_l the corresponding quantization step size and W_l the elements of the watermark sequence; l indicates the l^{th} selected coefficient or watermark bit to simplify notation. We adhere to the location-aware detection (LAD) scenario [6] where the embedding positions are known to the detector. For efficient blind watermark detection, accurate modeling of the host signal is required. We assume a Cauchy distribution of the DCT coefficients [1] and chose the Rao-Cauchy (RC) detector [4] whose detection statistic for the received signal Y_l of length L and the test against a detection threshold T are given by

$$\rho(Y_l) = \frac{8\hat{\gamma}^2}{L} \left[\sum_{l=1}^{L} \frac{Y_l \cdot W_l}{\hat{\gamma}^2 + Y_l^2} \right]^2 \quad \text{and} \quad \rho(Y_l) \gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} T. \tag{3}$$

$\hat{\gamma}$ is an estimate of the Cauchy PDF shape parameter which can be computed using fast, approximate methods [8]. According to [3], $\rho(Y_l)$ follows a χ^2_1 distribution with one degree of freedom under $\mathcal{H}_0$ and we can write the probability of false-alarm $P_f = \mathbb{P}(\rho(Y_l) > T | \mathcal{H}_0)$ as

$$P_f = 2\,\mathrm{Q}\left(\sqrt{T}\right) \quad \text{and express} \quad T = \left[\mathrm{Q}^{-1}\left(\frac{P_f}{2}\right) \right]^2 \tag{4}$$

where $Q(\cdot)$ denotes the Q-function of the Normal distribution. Under $\mathcal{H}_1$, the test statistic follows a non-central Chi-Square distribution $\chi^2_{1,\lambda}$ with one degree of freedom and non-centrality parameter λ. By estimating λ from experimental detection responses, the performance of the detector can be analyzed in terms of the probability of missing the watermark,

$$P_m = 1 - \mathbb{P}(\rho > T | \mathcal{H}_1) = 1 - \mathrm{Q}(\sqrt{T} - \sqrt{\lambda}) + \mathrm{Q}(\sqrt{T} + \sqrt{\lambda}). \tag{5}$$

3. Extension to H.264/SVC

H.264/SVC resorts to several coding tools in order to predict enhancement layer data from the base layer representation [7] and exploit the statistical dependencies: (a) inter-layer intra prediction can adaptively use the (upsampled) reconstructed reference signal of intra-coded macroblocks, (b) macroblock partitioning and motion information of the base layer is carried over via inter-layer motion prediction for inter-coded macroblocks, and (c) inter-layer residual prediction allows to reduce the residual energy of inter-coded macroblocks in the enhancement layer by subtracting the (upsampled) transform domain residual coefficients of the collocated reference block. See Fig. 3 for an illustration.

In this work we focus on watermark embedding in intra-coded macroblocks of an H.264-coded base layer using the method reviewed in Section 2. In case a spatial enhancement layer with twice the resolution in each dimension is to be coded for SVC spatial scalability, the watermarked base-layer representation is used for predicting the enhancement layer. In inter-layer intra prediction mode, the transform-domain enhancement layer residual of a 4×4 block k^E collocated with reference layer block k^B is given by

$$R'^E_{k^E} = \mathsf{Q}(\mathsf{T}(o^E_{k^E} - \mathsf{H}(o'^B_{k^B}))) \tag{6}$$

and the reconstructed, full-resolution video pixels are obtained by

$$o'^E_{k^E} = \mathsf{H}(o'^B_{k^B}) + \mathsf{T}^{-1}(\mathsf{Q}^{-1}(R'^E_{k^E})). \tag{7}$$

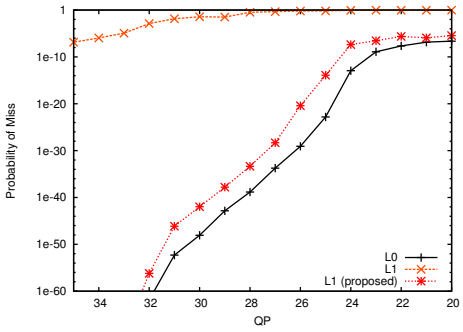

Figure 2. Transfer of base-layer watermark to spatial enhancement layer

H denotes the normative H.264/SVC upsampling operation and superscripts ^{B} and ^{E} indicate base and spatial enhancement layer data, resp. Apparently, the first right-hand term of Eq. (7) represents the upsampled, watermarked base-layer signal and the second term the quantized difference to the full-resolution, original video. Depending on the quantization parameter used to code the enhancement layer, the base-layer watermark can propagate to the decoded enhancement-layer video. Coarse quantization preserves a stronger watermark signal as illustrated in Fig. 2.

Watermarking only the base layer data is clearly not effective in protecting the full-resolution video. Not only does the watermark fade away, but also the bit rate for the enhancement layer increases, see Table 2, due to the added independent watermark signal which increased energy of the residual $R'^E_{k^E}$. To remedy these shortcomings, we propose to upsample the base layer watermark signal

$$W^E_{k^E} = \mathsf{Q}(\mathsf{T}(\mathsf{H}(\mathsf{T}^{-1}(Q_{k^B} \cdot S_{k^B} \cdot W^B_{k^B})))) \qquad (8)$$

and add the resulting enhancement layer watermark $W^E_{k^E}$ to the residual blocks $R'^E_{k^E}$ to form *compensated* residual blocks

$$R''^E_{k^E} = R'^E_{k^E} + W^E_{k^E}. \qquad (9)$$

Watermark detection is always performed with the base-layer watermark W, the full-resolution video is downsampled for detection.

4. Results

Experiments have been performed using the Joint Scalable Video Model (JSVM) reference software version 9.19.6. Source code for the watermarking schemes investigated will become available at `http://www.wavelab.at/sources`. All experiments have been performed on widely-available test video sequences in CIF and QCIF resolution; QCIF sequences have been obtained by downsampling. The watermark is embedded in the base layer as described in Section 2; we opt for always selecting the first

Table 1. Detection results (P_m) on base (L0) and enhancement layer (L1)

Sequence	L0	L1	L1 (proposed)
Foreman	$2.3 \cdot 10^{-25}$	0.81	$3.2 \cdot 10^{-17}$
Soccer	$2.6 \cdot 10^{-69}$	1.0	$1.1 \cdot 10^{-49}$
Bus	$1.0 \cdot 10^{-8}$	1.0	$6.2 \cdot 10^{-8}$
Container	$5.2 \cdot 10^{-119}$	0.44	$1.1 \cdot 10^{-91}$
Coastguard	$9.8 \cdot 10^{-133}$	0.68	$5.2 \cdot 10^{-97}$
Stefan	$8.5 \cdot 10^{-30}$	0.91	$3.2 \cdot 10^{-23}$

Table 2. Enhancement layer bit rate (Kbit/s)

Sequence	L1 (no WM)	L1	L1 (proposed)
Foreman	883.1	939.5	924.5
Soccer	1188.0	1239.1	1227.0
Bus	1693.0	1732.0	1721.0
Container	906.6	957.7	944.7
Coastguard	1506.6	1557.8	1534.2
Stefan	1621.4	1657.0	1651.0

4×4 DCT AC coefficient in zig-zag order as the embedding location when it is non-zero; formally

$$S_{i,j,k} = \begin{cases} 1 & i = 0, j = 1 \wedge R_{0,1,k} \neq 0 \\ 0 & \text{otherwise} \end{cases} \quad \forall k. \qquad (10)$$

The upsampled watermark signal is added to the quantized, transform-domain enhancement layer residuals as proposed in Section 3. The resulting watermarked, resolution-scalable bitstream can be decoded into QCIF and CIF video sequences. Watermark detection is performed on the decoded video.

Figure 2 shows the watermark detection performance for the *Foreman* sequence in terms of probability of miss (P_m) as a function of the H.264/SVC quantization parameter QP varying from 20 to 35. In the experiment, the false-alarm rate (P_f) is set to 10^{-3} and detection is performed on the first frame only; base layer and spatial enhancement layer have been coded with the same QP. The watermark can be reliably detected in the decoded base layer video (L0). Detection performance increases with coarser quantization as the watermark signal gets stronger relative to the host – remember that we added ± 1 to the quantized residual. We observe that the watermark embedded in the base layer is hardly detectable in the enhancement layer (L1). Only for coarse quantization ($QP \geq 28$) when no residual information is coded for most L1 blocks and solely the inter-layer intra prediction signal is available for reconstruction, de-

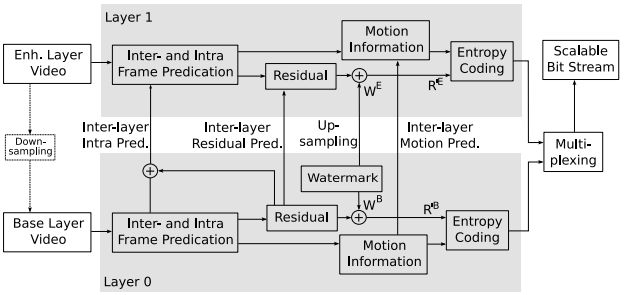

Figure 3. H.264/SVC encoding and watermarking structure for two resolution layers

tection becomes possible. However, using the upsampled base layer watermark, watermark detection performance in the enhancement layer is substantially improved (L1 proposed) and mostly restored to the level of the base layer watermark. Table 1 provides watermark detection results for six resolution-scalable H.264/SVC video sequences coded with $QP = 25$. The second column (L0) shows the probability of missing the watermark (P_m) for the decoded video in base layer QCIF resolution. When the watermark is embedded just in the base layer (L1), the watermark is not detectable using the decoded enhancement layer CIF resolution. When the upsampled watermark signal is added to the enhancement layer residual (L1 proposed), the watermark can be reliably detected from the decoded CIF video sequence.

In Table 2 we examine the bit rate (in Kbit/s) of the resolution-scalable bitstream for the first 32 frames of six test sequences coded with $QP = 25$ and inter-layer prediction. Results have been averaged over 10 test runs with different watermarks. For reference, the second column (L1 no WM) lists the bit rates for coding the sequences without a watermark. The third column (L1) contains the bit rate when watermarking the base layer. We notice an increase of about 3% on average due to the added watermark signal. The rightmost column (L1 proposed) presents the results when adding the upsampled watermark to the enhancement layer residual. Surprisingly, the bit rate can be reduced compared to the previous column. An independent watermark could have been embedded in the enhancement layer using the same method as used for the base layer. However, this would have further increased the bit rate, making the scalable bitstream less attractive.

5. Discussion and Conclusion

In this work, we considered the application of a robust H.264-integrated watermarking method [5] in the context of H.264/SVC. A watermark embedded in the base layer data of a resolution-scalable bitstream is not detectable in the full-resolution decoded video sequence. We can resolve the issue by adding a compensation watermark signal to the enhancement layer residual. Note that the base layer watermark can be detected in the decoded video *and* the compressed domain, i.e. after entropy decoding. In contrast, the enhancement layer watermark can be either detected in the compressed domain residual data, *or* the decoded video due to inter-layer prediction of H.264/SVC. The aim of this work is to achieve the latter which seems more relevant for robust watermarking.

Upsampling the watermark cannot be easily extended to support several resolution enhancement layers as the watermark signal looses its high-pass characteristic; on the other hand, multi-layer H.264/SVC bitstreams have increasingly higher bit rate compared to non-scalable coding and are not likely to be adopted. Evaluation with regards to coarse-grain scalability (CGS) layers for quality adaptation is subject to further work.

Acknowledgment

Supported by Austrian Science Fund (FWF) project P19159-N13.

References

[1] Y. Altunbasak and N. Kamaci. An analysis of the DCT coefficient distribution with the H.264 video coder. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '04*, volume 3, pages 177–180, Montreal, Canada, May 2004. IEEE.

[2] X. Gong and H.-M. Lu. Towards fast and robust watermarking scheme for H.264 video. In *Proceedings of the IEEE International Symposium on Multimedia, ISM '08*, pages 649–653, Berkeley, CA, USA, Dec. 2008. IEEE.

[3] S. Kay. *Fundamentals of Statistical Signal Processing: Detection Theory*, volume 2. Prentice-Hall, 1998.

[4] R. Kwitt, P. Meerwald, and A. Uhl. A lightweight Rao-Cauchy detector for additive watermarking in the DWT-domain. In *Proceedings of the ACM Multimedia and Security Workshop (MMSEC '08)*, pages 33–41, Oxford, UK, Sept. 2008. ACM.

[5] M. Noorkami and R. M. Mersereau. A framework for robust watermarking of H.264 encoded video with controllable detection performance. *IEEE Transactions on Information Forensics and Security*, 2(1):14–23, Mar. 2007.

[6] M. Noorkami and R. M. Mersereau. Digital video watermarking in P-frames with controlled video bit-rate increase. *IEEE Transactions on Information Forensics and Security*, 3(3):441–455, Sept. 2008.

[7] H. Schwarz and M. Wien. The scalable video coding extension of the H.264/AVC standard. *IEEE Signal Processing Magazine*, 25(2):135–141, Mar. 2008.

[8] G. Tsihrintzis and C. Nikias. Fast estimation of the parameters of alpha–stable impulsive interference. *IEEE Transactions on Signal Processing*, 44(6):1492–1503, June 1996.

[9] A. B. Watson. DCT quantization matrices visually optimized for individual images. In *Proceedings of SPIE, International Conference on Human Vision, Visual Processing and Display*, pages 202–216, San Jose, CA, USA, Feb. 1993. SPIE.