An efficient robust watermarking method integrated in H.264/SVC

Peter Meerwald* and Andreas Uhl

Dept. of Computer Sciences, University of Salzburg,
Jakob-Haring-Str. 2, A-5020 Salzburg, Austria
{pmeerw, uhl}@cosy.sbg.ac.at
http://www.wavelab.at

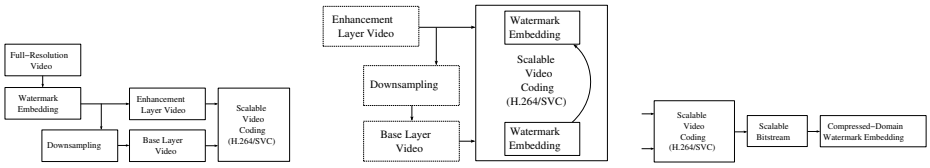
Abstract. In this article we investigate robust watermarking integrated with H.264/SVC video coding and address coarse-grain quality and spatial resolution scalability features according to Annex G of the H.264 standard. We show that watermark embedding in the base layer of the video is insufficient to protect the decoded video content when enhancements layers are employed. The problem is mitigated by a propagation technique of the base layer watermark signal when encoding the enhancement layer. In case of spatial resolution scalability, the base layer watermark signal is upsampled to match the resolution of the enhancement layer data.
We demonstrate blind watermark detection in the full- and low-resolution decoded video for the same adapted H.264/SVC bitstream for copyright protection applications and, surprisingly, can report bit rate savings when extending the base layer watermark to the enhancement layer. Further, we consider watermark detection integrated in the H.264/SVC decoder operating on the partially decoded residual data for copy control or authentication applications.

Keywords: Robust watermarking, blind detection, H.264/SVC, scalable video coding

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 bit stream format which can efficiently encode video in multiple spatial and temporal resolutions at different quality levels [16, 17]. 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.

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



(a) Watermark embedding before video encoding

(b) Integrated watermark embedding and coding

(c) Compressed-domain embedding after encoding

Fig. 1: Different embedding scenarios for watermarking resolution-scalable H.264/SVC video content.

In this work we investigate a well-known robust watermarking framework proposed by Noorkami et al. [11, 12] for copyright protection and ownership verification applications of H.264-encoded video content. The aim is to provide a single scalable, watermarked bit stream which can be distributed to diverse clients without the need to re-encode the video material. Scalability is provided at the bit stream level. A bit stream with reduced quality, spatial and/or temporal resolution can be efficiently obtained by discarding NAL units [16]. The watermark (i) should be detectable in the compressed domain *and* the decoded video without reference to the original content, and (ii) must be detectable in the decoded video at all scalability operation points, starting from the base layer.

In Fig. 1 we distinguish three embedding scenarios for producing a watermarked, scalable H.264/SVC bitstream: (a) embedding before encoding, (b) embedding integrated in the coding process, (c) altering the scalable bit stream (embedding in the compressed domain). The first embedding scenario offers little control over the resulting bitstream and thus makes detection in the compressed domain difficult. As watermark embedding takes place before video encoding, any robust video watermarking schemes can be applied. However, lossy compression and downsampling of the full-resolution video have an impact on the embedded watermark signal. Caenegem et al. [2] describe the design of a watermarking scheme resilient to H.264/SVC but treat the encoding only from a robustness point of view. In [19], Shi et al. propose a wavelet-domain embedding approach that exploits the transform's multi-resolution representation to cope with different resolution and quality layers. Both aforementioned techniques employ high-definition video frames (with HDTV and 4CIF resolution, respectively).

Finally, the third scenario appears to be overly complex from an implementation point of view given the intra-frame [6] and inter-layer prediction structure of H.264/SVC which necessitates drift compensation to minimize error propagation [11, 4]. Zou et al. [24, 23] propose a bitstream replacement watermark by altering H.264 CAVLC and CABAC symbols of HDTV video content several minutes long; scalability features are not addressed.

Integrated H.264/SVC video encoding and watermarking as shown in Fig. 4 offers control over the bitstream; for example the watermark can be placed ex-

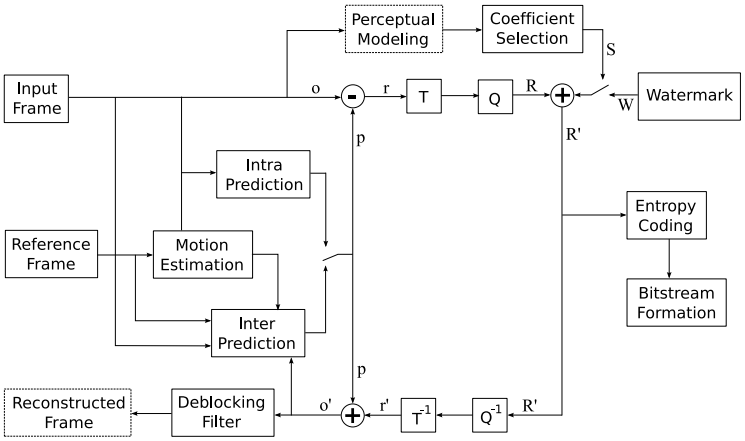


Fig. 2: Watermark embedding in quantized 4×4 DCT residual blocks.

clusively in non-zero quantized residual coefficients [12]. Further, the embedding operation can efficiently be implemented in the same transform domain as used by the encoder. A combined encryption and watermarking-based authentication method for H.264/SVC encoding is proposed by Park and Shin [13]. Authentication information is encoded in the bits signalling the intra prediction mode, but can not be verified on the decoded video. Many proposals for H.264 integrated watermarking have been put forward using spread-spectrum or replacement techniques for authentication and copyright protection (e.g. [15, 22, 18, 9]), however, watermarking of a scalable bitstream and the bitrate overhead is not considered. Recently, Park and Shin [14] put forward a method altering the DC coefficient of intra-coded blocks for copyright protection of H.264/SVC encoded content. The observed bit rate increase of over 10 % for certain sequences prompts for a more efficient solution.

The present work is an extension of [10]. A robust watermark is embedded in intra-coded frames during H.264/SVC encoding and detectable in the bitstream and decoded frames. In Section 2 we briefly review the H.264 watermarking framework [11] and investigate its applicability for protecting resolution-scalable video encoded with H.264/SVC. We propose a propagation step of the base-layer watermark signal in Section 3 in order to extend the framework to H.264/SVC, including resolution and quality scalability. 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. [11] present a framework where the Watson perceptual model for 8×8 DCT coefficients blocks [21]

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 [3], however, the watermark can not be detected in the decoded video sequence or the scheme has to deal with prediction error drift.

2.1 Watermark embedding

Figure 2 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 spatial domain 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, respectively. 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 $R_k = \mathsf{Q}(\mathsf{T}(r_k))$. $R_{0,0,k}$ 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 [11], 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

$$r'_k = \mathsf{T}^{-1}(\mathsf{Q}^{-1}(R'_k)) = \mathsf{T}^{-1}(\mathsf{Q}^{-1}(R_k) + Q_k \cdot S_k \cdot W_k). \tag{2}$$

For simplicity, we have dropped the coefficient indices i, j.

2.2 Blind watermark detection

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

$$\begin{aligned} \mathcal{H}_0 &: Y_l = O_l \text{ (no/other watermark)} \\ \mathcal{H}_1 &: Y_l = O_l + Q_l \cdot W_l \text{ (watermarked)} \end{aligned} \tag{3}$$

[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.

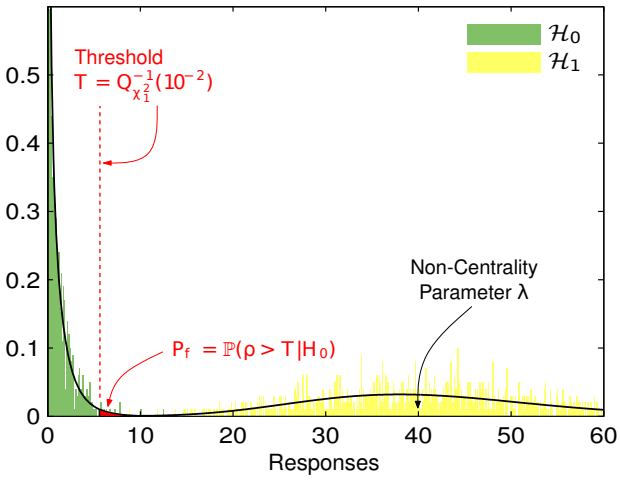


Fig. 3: Illustration of χ^2_1 detection response statistics under $\mathcal{H}_0$ and $\mathcal{H}_1$.

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 [12] 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 [7] whose detection statistic for the received signal Y_l of length L and the test against a detection threshold T are given by

$$\rho(Y_l) = \frac{8\hat{\gamma}^2}{L}\left[\sum_{l=1}^{L}\frac{Y_l\cdot W_l}{\hat{\gamma}^2+Y_l^2}\right]^2 \quad \text{and} \quad \rho(Y_l)\gtrless_{\mathcal{H}_0}^{\mathcal{H}_1} T. \tag{4}$$

$\hat{\gamma}$ is an estimate of the Cauchy probability density function shape parameter which can be computed using fast, approximate methods [20]. According to [5], $\rho(Y_l)$ follows a Chi-Square distribution (χ^2_1) 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

$$P_f = 2\,\mathrm{Q}\left(\sqrt{T}\right) \quad \text{and express} \quad T = \left[\mathrm{Q}^{-1}\left(\frac{P_f}{2}\right)\right]^2 \tag{5}$$

where $Q(\cdot)$ denotes the Q-function of the Normal distribution (exploiting the relation $Q_{\chi^2_1}(x) = 2Q(\sqrt{x})$ with the Q-function of the χ^2_1 distribution). Note that no parameters need to be estimated to establish the detection threshold. The Rao-Cauchy test is a constant false-alarm rate detector [5] which simplifies the experimental setup. 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,

$$P_m = 1 - \mathbb{P}(\rho > T|\mathcal{H}_1) = 1 - \mathrm{Q}(\sqrt{T} - \sqrt{\lambda}) + \mathrm{Q}(\sqrt{T} + \sqrt{\lambda}). \tag{6}$$

Figure 3 shows a histogram of the detection responses obtained by Eq. 4 under both hypothesis as well as the probability density functions of the χ_1^2 distributions which fit the observed data. The non-centrality parameter λ can be estimated from detector responses $\rho(Y_l)_p$, $1 \leq p \leq P$ obtained from P experiments

$$\hat{\lambda} = \left(\sum_{p=1}^{P} \sqrt{\rho(Y_l)_p} \right)^2 \tag{7}$$

and plugged into Eq. 6 to derive the *estimated* probability of missing the watermark. The measure obtained can be immediately interpreted and compared in a more meaningful way than – for example – a correlation coefficient or signal-to-noise ratio. However, some caution is in place with regard to the absolute value as usually only a small number of experiments (e.g. $P = 1000$) can be performed.

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 [16] 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 colocated reference block. See Fig. 4 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.

3.1 Resolution scalability

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 can be adaptively 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 colocated with reference layer block k^B is given by

$$R'^E_{k^E} = \mathsf{Q}(\mathsf{T}(o^E_{k^E} - \mathsf{H} \uparrow (o'^B_{k^B}))) \tag{8}$$

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

$$o'^E_{k^E} = \mathsf{H} \uparrow (o'^B_{k^B}) + \mathsf{T}^{-1}(\mathsf{Q}^{-1}(R'^E_{k^E})). \tag{9}$$

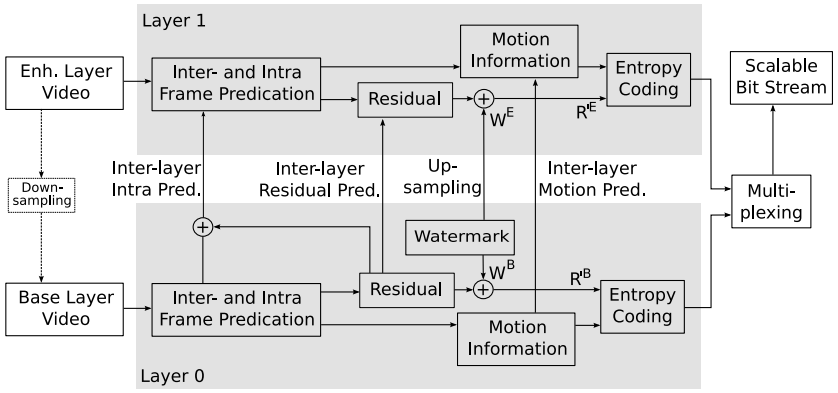


Fig. 4: Simplified H.264/SVC encoding and watermarking structure for two spatial resolution layers.

$\mathsf{H}\uparrow$ denotes the normative H.264/SVC upsampling operation and superscripts ^{B} and ^{E} indicate base and spatial enhancement layer data, respectively. Apparently, the first right-hand term of Eq. (9) 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 Figure 5 (a).

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 3, 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

$$W^E_{k^E} = \mathsf{Q}(\mathsf{T}(\mathsf{H}\uparrow(\mathsf{T}^{-1}(Q_{k^B} \cdot S_{k^B} \cdot W^B_{k^B})))) \tag{10}$$

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

$$R''^E_{k^E} = R'^E_{k^E} + W^E_{k^E}. \tag{11}$$

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

3.2 Quality scalability

In Fig. 5 (b) we plot the watermark transfer between two QCIF coarse-grain scalability (CGS) quality layers for a range of coding quantization parameters. The quality enhancement layer is coded using $QP - 3$ with respect to the base layer. It can be seen that the base layer watermark is effectively overshadowed by

the enhancement layer video data coded with finer quantization. Simply adding the same watermark in the enhancement layer restores the watermark signal.

4 Experimental results

Experiments have been performed using the Joint Scalable Video Model (JSVM) reference software version 9.19.9. Source code for the watermarking schemes investigated in this article will become available at `http://www.wavelab.at/sources`. All experiments have been performed on widely-available test video sequences[2] in CIF (352×288) and QCIF (176×144) resolution; QCIF sequences have been obtained by downsampling using the JSVM tools. Experiments are repeated 1000 times with different watermark seeds to estimate the detection performance.

The watermark is embedded in the base layer as described in Section 2 with an average target PSNR in the luminance channel of 40 dB between the original and the coded and watermarked video. We opt for always selecting the first 4×4 DCT AC coefficient in zig-zag order as the embedding location when it is non-zero; formally

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

The upsampled watermark signal is added to the quantized, transform-domain enhancement layer residuals as proposed in Section 3.1 with a target PSNR of 40 dB. The resulting watermarked, resolution-scalable bitstream can be decoded into QCIF and CIF video sequences.

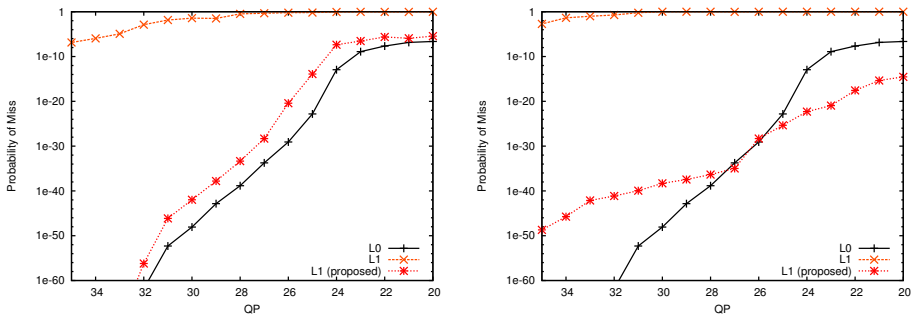
For video quality scalability (cf. Section 3.2), the base-layer watermark is propagated to the enhancement layer and the coarse-grain quality scalable bitstream can be decoded into two differently quantized representations.

We first address watermark detection on the decoded video frames. Then watermark detection directly on the residual transform coefficients after arithmetic decoding is considered, e.g. for copy control or authentication applications.

4.1 Watermark detection on the decoded video frames

Figure 5 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 resolution enhancement layer have been coded with the same QP (cf. Fig. 5a), the coarse grain quality enhancement layer (cf. Fig. 5b) with $QP - 3$ relative to the base layer. The watermark can be reliably detected in the decoded

[2] The uncoded test sequences are available for download from `http://media.xiph.org/video/derf/`.



(a) Resolution scalability (b) Coarse grain quality scalability

Fig. 5: Transfer of the base-layer watermark to a (a) spatial resolution, and (b) coarse-gain quality enhancement layer for different quantization parameters (QP).

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, detection 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 the 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 (column *L1 BL WM*), the watermark is not detectable using the decoded enhancement layer CIF resolution video since the base layer watermark does not propagate to the higher resolution layer. The fourth column (*L1 indep. WM*) lists the detection results for an independent watermark embedded in the enhancement layer. As the host signal is now four times larger, the probability of miss is drastically reduced. When the upsampled watermark signal is added to the enhancement layer residual (column *L1 proposed*) as presented in Section 3, the watermark can be reliably detected from the decoded CIF video sequence.

4.2 Watermark detection on the partially decoded bitstream

Table 2 shows the watermark detection results directly on the residual transform coefficients after arithmetic decoding of the bitstream, i.e. detection integrated in the H.264/SVC decoder. The encoder settings are the same as above.

Sequence	Probability of Miss ($P_f = 10^{-3}$)			
	L0	L1 (BL WM)	L1 (indep. WM)	L1 (proposed)
Foreman	$2.3 \cdot 10^{-25}$	0.81	~ 0.0	$3.2 \cdot 10^{-17}$
Soccer	$2.6 \cdot 10^{-69}$	1.0	~ 0.0	$1.1 \cdot 10^{-49}$
Bus	$1.0 \cdot 10^{-8}$	1.0	$2.5 \cdot 10^{-316}$	$6.2 \cdot 10^{-8}$
Container	$5.2 \cdot 10^{-119}$	0.44	~ 0.0	$1.1 \cdot 10^{-91}$
Coastguard	$9.8 \cdot 10^{-133}$	0.68	~ 0.0	$5.2 \cdot 10^{-97}$
Stefan	$8.5 \cdot 10^{-30}$	0.91	~ 0.0	$3.2 \cdot 10^{-23}$

Table 1: Detection results on the decoded base (L0) and resolution enhancement layer (L1) video frames.

First, we observe than detection performance is higher when operating on the residual coefficients compared to detection on the decoded video frames (cf. Tables 1 and 2, column *L0*). This is expected as H.264 spatial prediction acts as an additional additive noise source on the residual signal where the watermark is embedded.

Second, the base layer watermark can be reliably detected in the partially decoded enhancement layer bitstream (column *L1 BL WM*). For watermark detection, the enhancement layer residual blocks $R^E_{k^E}$ are downsampled

$$R'^B_{k^B} = \mathsf{Q}(\mathsf{T}(\mathsf{H} \downarrow (\mathsf{T}^{-1}(\mathsf{Q}^{-1}(R^E_{k^E}))))) \tag{12}$$

so that the base layer watermark W_k can be correlated with the colocated blocks $R'^B_{k^B}$ of base layer resolution.

Note that the $R'^B_{k^B}$ actually contain the inverted watermark $-W$ when H.264/SVC inter-layer intra prediction is used since

$$r^E_{k^E} = o^E_{k^E} - \mathsf{H} \uparrow (o'^B_{k^B}) \tag{13}$$

where $o'^B_{k^B}$ contains the additive base layer watermark W. However, the Rao-Cauchy detector employed is agnostic to the sign of the watermark due to the square operation in Eq. 4.

The use of inter-layer intra prediction for enhancement layer coding can be an adaptive decision of the H.264/SVC encoder. For the experiments, the adaptive decision was enabled in the JSVM reference implementation and it is observed that the majority of intra-coded blocks are in fact coded with inter-layer prediction.

4.3 Bit rate assessment of watermarked scalability layers

In Table 3 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

Sequence	Probability of Miss ($P_f = 10^{-3}$)	
	L0	L1 (BL WM)
Foreman	$3.8 \cdot 10^{-48}$	$1.6 \cdot 10^{-44}$
Soccer	$1.1 \cdot 10^{-107}$	$1.5 \cdot 10^{-84}$
Bus	$5.4 \cdot 10^{-35}$	$1.1 \cdot 10^{-32}$
Container	$4.7 \cdot 10^{-198}$	$8.0 \cdot 10^{-95}$
Coastguard	$2.2 \cdot 10^{-210}$	$2.4 \cdot 10^{-88}$
Stefan	$8.1 \cdot 10^{-77}$	$1.7 \cdot 10^{-41}$

Table 2: Detection results on partially decoded base (L0) and resolution enhancement layer (L1) coefficients.

watermarks. For reference, the second column (*L1 no WM*) lists the bit rates for coding the sequences without any watermark. The third column (*L1 BL WM*) contains the bit rate when watermarking the base layer only. We notice an increase of about 3% on average due to the added watermark signal. The fourth column (*L1 indep. WM*) lists the bit rate when independent watermarks are added to the base and enhancement layer; two independent watermarks in the two layer produces the highest bitrate. The rightmost column (*L1 proposed*) presents the results when adding the upsampled watermark to the enhancement layer residual as proposed. Surprisingly, the bit rate can be reduced compared to the previous two columns and is lower than having no watermark in the decoded enhancement layer at all.

The reason follows from the observations made in Section 4.2: in case inter-layer intra prediction is employed by the H.264/SVC encoder, the enhancement layer bitstream codes the difference between the upsampled base layer reconstruction (the prediction) and the full-resolution video signal. An additional, uncorrelated watermark signal increases that difference and hence the enhancement layer bitrate as shown in column *L1 BL WM*. When the upsampled watermark is added to the enhancement layer video, the prediction (containing the base layer watermark) better matches the enhancement layer data to be coded and hence reduces the bitrate.

Table 4 lists the bit rates in Kbit/s for the coarse-grain quality (CGS) enhancement layer. The QCIF base layer is coded with $QP = 30$ and the enhancement layer of the same resolution with $QP = 24$. We can observe that watermarking the enhancement layer with the same watermark as the base layer (column *L1 proposed*) slightly reduces the bit rate over the case where the enhancement does not carry a watermark (column *L1 BL WM*) and only the base layer (BL) is watermarked, or – to a larger extent – when a different watermark (column *L1 indep. WM*) is embedded in the two quality scalability layers.

H.264/SVC also supports so-called medium-grain scalability (MGS) to enable quality adaptation without the need to code separate layers. MGS is realized

Sequence	Bit rate (Kbit/s)			
	L1 (no WM)	L1 (BL WM)	L1 (indep. WM)	L1 (proposed)
Foreman	883.1	939.5	1018.9	924.5
Soccer	1188.0	1239.1	1303.8	1227.0
Bus	1693.0	1732.0	1779.0	1721.0
Container	906.6	957.7	982.1	944.7
Coastguard	1506.6	1557.8	1572.6	1534.2
Stefan	1621.4	1657.0	1715.0	1651.0

Table 3: Bit rate of the resolution enhancement layer (L1).

Sequence	Bit rate (Kbit/s)			
	L1 (no WM)	L1 (BL WM)	L1 (indep. WM)	L1 (proposed)
Foreman	287.4	330.9	342.9	320.2
Soccer	342.7	380.6	401.3	371.6
Bus	463.8	500.0	507.4	490.5
Container	258.6	307.8	315.8	298.2
Coastguard	359.5	396.0	404.8	387.0
Stefan	483.1	525.8	536.0	517.5

Table 4: Bit rate of the coarse-grain quality layer (L1).

by grouping the DCT coefficients in zig-zag order and allowing to discard the endmost coefficient groups. Since the watermark in this work is embedded in the first AC coefficient, MGS does not impair the watermark detection results unless all AC coefficients are discarded.

5 Discussion and Conclusion

In this work, we considered the application of a robust H.264-integrated watermarking method [11] 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 which also reduces teh bitrate. The base layer watermark can be detected in the decoded video frames *and* the compressed domain, i.e. after entropy decoding of the residual data. With respect to the enhancement layer, the base layer watermark can be either detected in the compressed domain residual data, *or* the decoded video frames due to inter-layer prediction of H.264/SVC when employing the proposed

compensation technique. Li et al. [8] discuss watermarking of a scalable audio bitstream and focus on the first case.

We provide detection results for base- and enhancement layer watermarking and consider the bitrate of the resulting watermarked scalable bitstream. Contrary to other recent approaches [19, 2] we focus on the particularities of watermarking integrated in the H.264/SVC encoding step as opposed to watermarking before scalable bitstream formation.

The 8×8 DCT which is more efficient for coding high-resolution frames can be permitted for coding the enhancement layer, only the base layer watermark is constrained to embedding in the prevalent 4×4 transform blocks since the watermark detector is blind and has no information on the H.264/SVC mode decisions. 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.

Further work includes an assessment of different embedding strategies (incorporating perceptual shaping of the watermark) with regard to the bitrate of the watermarked bitstream and a comparison of blind detection approaches adapted to the quantized host signal coefficients.

References

1. Altunbasak, Y., Kamaci, N.: 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. vol. 3, pp. 177–180. IEEE, Montreal, Canada (May 2004)
2. van Caenegem, R., Dooms, A., Barbarien, J., Schelkens, P.: Design of an H.264/SVC resilient watermarking scheme. In: Proceedings of SPIE, Multimedia on Mobile Devices 2010. vol. 7542. SPIE, San Jose, CA, USA (Jan 2010)
3. Gong, X., Lu, H.M.: Towards fast and robust watermarking scheme for H.264 video. In: Proceedings of the IEEE International Symposium on Multimedia, ISM '08. pp. 649–653. IEEE, Berkeley, CA, USA (Dec 2008)
4. Hartung, F., Girod, B.: Watermarking of uncompressed and compressed video. Signal Processing 66(3), 283–301 (May 1998)
5. Kay, S.M.: Fundamentals of Statistical Signal Processing: Detection Theory, vol. 2. Prentice-Hall (1998)
6. Kim, D.W., Choi, Y.G., Kim, H.S., Yoo, J.S., Choi, H.J., Seo, Y.H.: The problems in digital watermarking into intra-frames of H.264/AVC. Image and Vision Computing 28(8), 1220–1228 (Aug 2010)
7. Kwitt, R., Meerwald, P., Uhl, A.: A lightweight Rao-Cauchy detector for additive watermarking in the DWT-domain. In: Proceedings of the ACM Multimedia and Security Workshop (MMSEC '08). pp. 33–41. ACM, Oxford, UK (Sep 2008)
8. Li, Z., Sun, Q., Lian, Y.: Design and analysis of a scalable watermarking scheme for the scalable audio coder. IEEE Transactions on Signal Processing 54(8), 3064–3077 (Aug 2006)

9. Lin, S., Chuang, C.Y., Meng, H.C.: A video watermarking in H.265/AVC encoder. In: Proceedings of the 5th International Conference on Intelligent Information Hiding and Multimedia Signal Processing, IIH-MSP '09. pp. 340–343. Kyoto, Japan (Sep 2009)
10. Meerwald, P., Uhl, A.: Robust watermarking of H.264/SVC-encoded video: quality and resolution scalability. In: Kim, H.J., Shi, Y., Barni, M. (eds.) Proceedings of the 9th International Workshop on Digital Watermarking, IWDW '10. Lecture Notes in Computer Science, vol. 6526, pp. 159–169. Springer, Seoul, Korea (Oct 2010)
11. Noorkami, M., Mersereau, R.M.: 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)
12. Noorkami, M., Mersereau, R.M.: Digital video watermarking in P-frames with controlled video bit-rate increase. IEEE Transactions on Information Forensics and Security 3(3), 441–455 (Sep 2008)
13. Park, S.W., Shin, S.U.: Combined Scheme of Encryption and Watermarking in H.264/Scalable Video Coding (SVC), pp. 351–361. Studies in Computational Intelligence, Springer (2008)
14. Park, S.W., Shin, S.U.: Authentication and copyright protection scheme for H.264/AVC and SVC. Journal of Information Science and Engineering 27, 129–142 (2011)
15. Qiu, G., Marziliano, P., Ho, A.T.S., He, D., Sun, Q.: A hybrid watermarking scheme for H.264/AVC video. In: Proceedings of the 17th International Conference on Pattern Recognition, ICPR '04. pp. 865–868. IEEE, Cambridge, UK (Aug 2004)
16. Schwarz, H., Wien, M.: The scalable video coding extension of the H.264/AVC standard. IEEE Signal Processing Magazine 25(2), 135–141 (Mar 2008)
17. Segall, C.A., Sullivan, G.J.: Spatial scalability within the H.264/AVC scalable video coding extension. IEEE Transactions on Circuits and Systems for Video Technology 17(9), 1121–1135 (Sep 2007)
18. Shahid, Z., Meuel, P., Chaumont, M., Puech, W.: Considering the reconstruction loop for watermarking of intra and inter frames of H.264/AVC. In: Proceedings of the 17th European Signal Processing Conference, EUSIPCO '09. pp. 1794–1798. EURASIP, Glasgow, UK (Aug 2009)
19. Shi, F., Liu, S., Yao, H., Liu, Y., Zhang, S.: Scalable and credible video watermarking towards scalable video coding. In: Proceedings of the Pacific Rim Conference on Multimedia, PCM '10. Lecture Notes in Computer Science, vol. 6297, pp. 697–708. Springer, Shanghai, China (Sep 2010)
20. Tsihrintzis, G., Nikias, C.: Fast estimation of the parameters of alpha–stable impulsive interference. IEEE Transactions on Signal Processing 44(6), 1492–1503 (Jun 1996)
21. Watson, A.B.: DCT quantization matrices visually optimized for individual images. In: Proceedings of SPIE, International Conference on Human Vision, Visual Processing and Display. pp. 202–216. SPIE, San Jose, CA, USA (Feb 1993)
22. Zhang, J., Ho, A.T.S., Qiu, G., Marziliano, P.: Robust video watermarking of H.264/AVC. IEEE Transactions on Circuits and Systems 54(2), 205–209 (Feb 2007)
23. Zou, D., Bloom, J.: H.264 stream replacement watermarking with CABAC encoding. In: Proceedings of the IEEE International Conference on Multimedia and Expo, ICME '10. Singapore (Jul 2010)
24. Zou, D., Bloom, J.: H.264/AVC substitution watermarking: a CAVLC example. In: Proceedings of the SPIE, Media Forensics and Security. vol. 7254. SPIE, Jan Jose, CA, USA (Jan 2009)