A Novel Video Watermarking Approach Based on Implicit Distortions

Hannes Mareen, *Student Member, IEEE*, Johan De Praeter, Glenn Van Wallendael, and Peter Lambert

***Abstract*—Copyright-sensitive videos are commonly leaked or illegally distributed by so-called digital pirates. Video owners aim to prevent this by hiding a unique watermark in every video that contains information about the receiver. If the video is then illegally distributed, the copyright owner can extract the watermark and identify the malicious consumer. However, pirates may manipulate the video in the hope of destroying the embedded watermark. Although a variety of imperceptible and robust solutions exist, these introduce many artificial distortions to the video. Therefore, this paper proposes a novel video watermarking approach in which only a single encoder decision is explicitly changed. Then, the explicit change automatically propagates into a large collection of implicit distortions that represents the watermark. The implicit distortions resemble ordinary, encoder-created compression artifacts and hence are imperceptible. Additionally, they prove to be robust against video manipulations. Furthermore, the proposed scheme requires no modification of existing consumer electronic devices. Consequently, the proposed watermarking approach can be applied to help combat piracy without bothering innocent users with unnatural distortions.**

***Index Terms*—High efficiency video coding, watermarking, video security.**

I. Introduction

Consumers of video-on-demand services [1] or users of digital video recorders [2] often copy their requested content and upload it to torrent sites. Similarly, screeners of films and TV series are often leaked on the Internet before they are officially released [3]. Typical encryption-based security measures are not sufficient to protect video owners from copyright infringement [4], since so-called digital pirates [5] may legally acquire an encrypted video and, thus, decrypt the video without a problem. Then, the pirate can illegally distribute a copy of the decrypted video on the Internet.

Since traditional, encryption-based security [6]–[9] may fail, watermarking should be applied as an extra security measure. More specifically, a unique watermark that contains information about the receiver should be hidden in every video. If a malicious receiver illegally distributes his or her watermarked version of the video, he or she can be identified by extracting the watermark out of the video. Since malicious consumers may manipulate the video in an effort to destroy the watermark, it is important for the watermark to be robust, i.e., it should survive video manipulations. Additionally, it is often required that the watermark is hidden imperceptibly, such that honest consumers are not bothered by the security measure. Although a significant volume of research on watermarking was conducted in the past two decades, imperceptible and robust watermarking still remains a challenging problem [10].

This paper evaluates and extends a video watermarking technique that was first proposed briefly in previous work [11]. As its main novelty, the proposed technique introduces so-called implicit distortions into the video. These distortions are automatically generated by a video encoder, in contrast to artificial distortions that are used in the state-of-the-art. The embedding scheme is implemented using the High Efficiency Video Coding (HEVC) standard, but the underlying ideas are applicable to other video coding standards as well. Moreover, the scheme requires no modification of existing Consumer Electronic (CE) devices, since the videos can be decoded by a standard video decoder. In addition to the previously published work, this paper proposes a novel watermark extraction scheme based on outlier detection and extensively evaluates various properties of the proposed watermarking technique.

The rest of this paper is organized as follows. First, Section II gives a brief overview of state-of-the-art watermarking techniques. Next, the proposed watermarking scheme is explained in Section III. Then, Section IV evaluates the method by quantifying the amount of implicit distortions, the perceptibility, the robustness and the bit rate increase. Finally, the conclusion is drawn in Section V.

II. State-of-the-Art

Many watermarking algorithms apply a variation of least-significant-bit modification on certain pixels or (quantized) transform coefficients [3], [12]–[18]. That is because these least significant bits can easily be swapped with watermarking information without being very perceptible. However, as a consequence, attackers can easily delete the watermark since they can also change this bit plane without notably degrading the video quality.

Alternatively, watermarks can be represented as a relationship between certain bitstream components. For example,

Manuscript received March 1, 2018; revised May 30, 2018; accepted June 26, 2018. Date of publication July 2, 2018; date of current version September 24, 2018. This work was supported in part by IDLab (Ghent University—imec), in part by the Flanders Innovation and Entrepreneurship (VLAIO), and in part by the European Union. The work of H. Mareen was supported by the Research Foundation—Flanders under Grant 1S55218N. *(Corresponding author: Hannes Mareen.)*

The authors are with Ghent University—imec, IDLab, Department of Electronics and Information Systems, 9000 Ghent, Belgium (e-mail: hannes.mareen@ugent.be; johan.depraeter@ugent.be; glenn.vanwallendael@ugent.be; peter.lambert@ugent.be).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

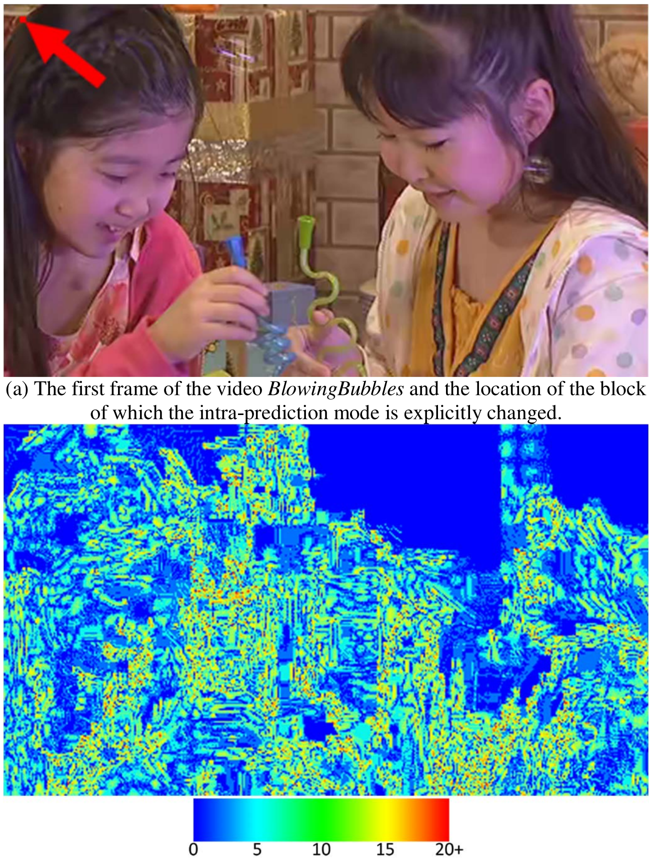
Digital Object Identifier 10.1109/TCE.2018.2852258

Gaj *et al.* [19] proposed to change the difference in number of non-zero coefficients of selected blocks in consecutive intra-frames. Similarly, Dutta and Gupta [20] suggested to change the relationship between the two first non-zero AC coefficients of selected blocks in consecutive intra-frames.

Another common watermark representation is an additive noise signal. For example, Kalker *et al.* [21] proposed a spread-spectrum technique in which a scaled version of the watermark is added as noise. Hartung and Girod [22] proposed a similar technique, in which they first added the watermark to a pseudorandom noise pattern before adding it to the original video. Lastly, Yamada *et al.* [23] used the watermark as a seed to generate a pseudorandom noise pattern that was added to the video, while taking the perceptibility of every pixel change into account.

When the watermark is represented by a noise signal, its presence is detected by using correlation-based techniques. That is, the observed noise in a possibly-watermarked video is compared to the noise that represents a watermark. If the resulting correlation value exceeds a certain, predefined threshold, then the watermark is said to be detected. Unfortunately, it is impossible to define a universal threshold, because the correlation values will vary significantly when different attacks are applied [24]. Section III-B explains this problem in more detail and proposes a novel extraction algorithm as an alternative.

In general, all existing robust watermarking techniques explicitly add many artificial distortions to the video. Even though these distortions are usually imperceptible, it is questionable whether adding such unnatural distortions is desirable. Therefore, this paper contributes to the-state-of-art by introducing a new type of imperceptible distortions that are added in a more natural fashion.



(a) The first frame of the video *BlowingBubbles* and the location of the block of which the intra-prediction mode is explicitly changed.

(b) Visualization of the resulting implicit distortions. Blue means that the corresponding luma pixel value is not changed, whereas red signifies a pixel difference of 20 or higher. A luma pixel value is represented by 8 bits.

Fig. 1. By explicitly changing a single coding decision of a single block (a), many distortions are implicitly created (b).

III. Proposed Scheme

A. Watermark Embedding

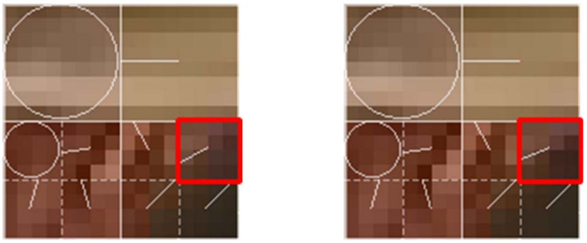
As an alternative to existing watermark embedding methods, this section proposes a novel approach that adds distortions in an *implicit* way, automatically during video encoding.

A video encoder creates a compressed video bit stream that consists of coding information and a residual signal. The coding information describes the structure of the video and attempts to predict every region of the video based on other, surrounding regions. This prediction is usually not perfect, hence the residual signal has to correct the prediction errors. However, in order to achieve stronger compression of the video data, the residual signal is quantized, leading to a loss of information and resulting in compression artifacts.

In normal circumstances, when compressing a video, the coding information consists of coding decisions that were made by the encoder. The proposed watermark embedding scheme explicitly modifies a single coding decision, while preserving all others as when encoding the video without a watermark. When a block's coding decision is explicitly changed, the corresponding region will often be predicted differently and will therefore usually result in a different quantized residual signal. Consequently, this region will produce slightly different compression artifacts than when the optimal coding decisions were used. In other words, the explicit change introduces so-called implicit distortions.

When an implicitly distorted block is used for prediction by other blocks, these blocks will be predicted differently as well. Thus, more implicit distortions will be generated. Similarly, those other blocks will be subsequently used for prediction, meaning further propagation of the implicit distortions. As a result, implicit distortions will be spread over the whole video due to intra- and inter-frame propagation. Fig. 1 illustrates the creation of implicit distortions by showing the first frame of the video sequence *BlowingBubbles*, the location of a block in that frame of which the intra-prediction mode was explicitly changed, and the resulting (imperceptible, implicit) distortion map. Additionally, Fig. 2 zooms in on the area around the explicitly changed block in both the unwatermarked and watermarked frame. In the figure, one can observe that the intra-prediction mode of the indicated block is slightly changed.

Note that, since the explicit change is made during the encoding process and hence the encoding loop is always closed, no drift-error propagation occurs. Instead, the implicit distortions are assumed to be imperceptible because the encoder still tries to resemble the original video as closely as possible. As a result, the introduced distortions are similar to ordinary encoding artifacts that always occur during lossy



(a) Unwatermarked decisions (b) Watermarked decisions

Fig. 2. Comparison of the unwatermarked encoder decisions (a) and the watermarked decisions (b). The intra-prediction mode of the indicated block was explicitly changed from 5 to 6 (or from 67.5° to 75°).

video compression, contrary to the artificial distortions created in state-of-the-art algorithms.

Assuming every explicit change results in different implicit distortions, a watermark can be represented by a unique explicit change, i.e., by the resulting (large) set of implicit distortions. However, it is possible that the quantized residual signal of a block creates identical compression artifacts with or without an explicit change. In this case, the explicit change does not create any implicit distortions and thus cannot be used to represent a unique watermark. Additionally, even when an explicit change does create different compression artifacts in the corresponding block, it is possible that the implicit distortions are not spread sufficiently over the video. This happens when the block is only rarely used as a reference for intra-and inter-frame prediction, and the implicit distortions can thus not be propagated sufficiently. As a result, there are only few implicit distortions, making the corresponding watermark not very robust.

As a solution to the above-described problems, explicit changes that create only few or no implicit distortions should not be used to represent a watermark. That is, the amount of implicit distortions should be quantified during embedding. When this amount is not sufficiently high, another explicit change should be used to represent the watermark instead. Section IV-C evaluates the effect of eliminating such explicit changes on the robustness.

The proposed embedding scheme was implemented using version 16.5 of the HEVC reference Model (HM). However, note that the underlying ideas are applicable to other standards as well. The implementation used to evaluate the proposed embedding scheme explicitly changes the luma intra-prediction mode, which can take 35 different values in HEVC (Planar, DC and 33 angles). If the intra-prediction mode is Planar, it is changed to DC and the other way around. If the mode is an angle, the angle is changed to an adjacent angle. More specifically, the angle mode is reduced by one in the case of an intra-direction of 17 or 34 and incremented by one in all other cases. The explicit changes are only made in intra-frames, in order to assure that there are many intra-blocks available to create many unique watermarks. In other words, the watermark is embedded once every intra-period, which is usually every few seconds.

B. Watermark Extraction

Because a watermark is represented by a unique collection of implicit distortions, one can consider these distortions to be a noise pattern that was added to the video. Such unique noise patterns are usually detected by using correlation-based techniques. However, this section explains why traditional correlation-based techniques fail and proposes an alternative, novel watermark extraction scheme.

As mentioned in Section II, correlation-based methods correlate the observed noise pattern in a possibly-watermarked video with the noise pattern that represents the watermark. Several correlation measures exist, such as the correlation coefficient (z_{cc}), which is an extension of the normalized correlation (z_{nc}) [3]. These measures are defined in (1), in which o and w are vectors of pixels, representing the observed and watermarked video, respectively. Additionally, $|o|$ and $|w|$ represent the Euclidean length of o and w, respectively, and $\bar{o}$ and $\bar{w}$ represent the mean of o and w, respectively.

$$\begin{aligned} z_{nc}(o, w) &= \sum_i \frac{o[i]}{|o|} \cdot \frac{w[i]}{|w|}, \\ z_{cc}(o, w) &= z_{nc}(o - \bar{o}, w - \bar{w}) \end{aligned} \tag{1}$$

In traditional correlation-based detection methods, the presence of the watermark is detected by comparing the calculated correlation value to a certain, pre-defined threshold. That is, if the correlation value is higher than the threshold, the watermark is detected. However, the correlation value is significantly influenced by the type of attack that is performed on the video. As a consequence, the correlation value between a severely attacked watermarked video and the corresponding watermark can be lower than the correlation value between a weakly attacked watermarked video and an incorrect watermark.

This problem is illustrated in Fig. 3: all 2433 blocks in the first frame of the sequence *BlowingBubbles* are explicitly changed, resulting in 2433 different watermarked videos. Then, the watermarked video corresponding to block no. 1000 is attacked twice: once by re-encoding it with a Quantization Parameter (QP) of 20, and once with a QP of 30. Re-encoding the video with a QP of 30 degrades the quality more than re-encoding it with a QP of 20. In Fig. 3, the correlation values between the two attacked videos and all 2433 watermarked videos are shown. In both groups of correlation values, the correlation with watermark no. 1000 is clearly higher than all other correlations within that same group. However, the highest correlation value in the lower group (QP 30) is smaller than *all* correlation values in the upper group (QP 20). In other words, if it is desired that the watermark can be detected for an attack with a QP of 30, the correlation threshold should be set to approximately 0.989. But if an attack is performed with a QP of 20, the correlation values of *all* watermarks are higher than this threshold and would thus all be detected. In other words, it is impossible to find a threshold that correctly detects the watermark both in case it is attacked with a QP of 20 and 30.

As a proposed solution, all watermarks should be taken into account when performing watermark extraction, instead of only a single one. That is, the correlation is calculated with all watermarked videos. Then, the watermark corresponding to the highest correlation value is detected. In order to have a notion of confidence, the extraction scheme should also take

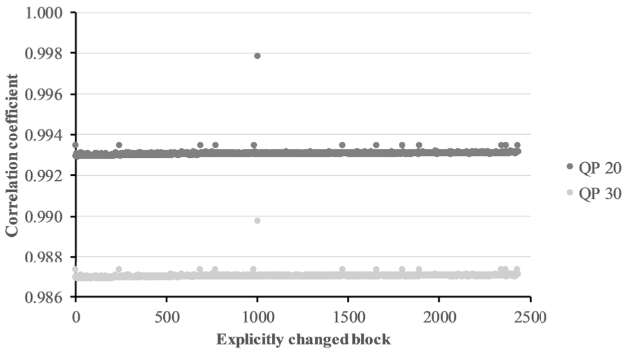


Fig. 3. Two groups of correlation values between all 2433 watermarks and an attacked watermarked video. In the upper group, watermarked video no. 1000 is attacked by re-encoding it with a QP of 20, whereas in the other group, it is re-encoded with a QP of 30. In both groups, the correlation with watermark no. 1000 is clearly higher than all other correlations within that same group. However, the highest correlation value in the lower group is smaller than *all* correlation values in the upper group.

into account how much the highest correlation value differs from the other correlation values. This is proposed to be done by performing outlier detection. That is, for every calculated correlation value, the distance is calculated to the distribution of all other correlation values. Then, the maximum distance (i.e., the outlier) corresponds to the detected watermark, with the distance as a confidence measure. In the proposed scheme, a modified version of the normalized Euclidean distance is used as a distance measure and is given in (2). In this equation, x represents the investigated correlation value, and μ and σ represent the mean and standard deviation of the distribution of all other correlation values, respectively.

$$d(x, \mu, \sigma) = \frac{x - \mu}{\sigma} \tag{2}$$

In order to provide robustness against collusion attacks, in which multiple attackers blend their watermarked versions, multiple watermarks should be extracted such that all colluders can be identified [3], [6]. In such a case, there will be multiple outliers in the distribution of correlations. Therefore, in the proposed scheme, not only the watermark corresponding to the maximum distance is detected, but also those corresponding to distances that are very close to the maximum. Concretely, the watermarks corresponding to highest p% of the distances, i.e., distances in the range given in (3), are detected. In this range, p represents the desired percentage as a decimal number, *max* the maximum distance, and *mean* the mean of the distribution of distances.

$$range = \left[max - p \cdot (max - mean), max\right] \tag{3}$$

Detecting multiple watermarks instead of just a single one, increases the probability of a watermark being selected. Thus, such a scheme will generally perform better in terms of true positive detections, i.e., watermarks being correctly detected. However, this also means that more watermarks can be detected that should *not* be detected, i.e., false positive detections. Increasing p will hence allow more true positive detections, but also more false positive detections. Thus, p regulates a trade-off between the number of true positive and false positive detections. For the evaluation of the proposed scheme in Section IV-C, p is chosen to be 1%, as experimental tests demonstrated that this value produces good results.

In summary, in order to extract the watermark, every watermarked video is correlated with the observed, potentially-attacked video. Then, the distance between every correlation value and the distribution of other correlation values is calculated. The maximum distances (i.e., the outliers) then correspond to the extracted watermarks.

C. Complexity Analysis

This section discusses the complexity of the proposed watermarking scheme, and therefore its applicability to CE devices. Since the watermark embedding is performed on the server side, there is no complexity increase on the clients' devices. In fact, the watermark requires no modification of existing CE devices, since the videos can be decoded by a standard video decoder.

Although the watermark embedding is performed during video encoding, the scheme is scalable. That is because this encoding process can be sped up by not recalculating the optimal coding decisions. More specifically, the high-complexity encoder runs only once, encoding the video without a watermark. Then, the watermarks are embedded using low-complexity encoders that use the coding decisions made by the high-complexity encoder. As a result, these encoders only need to calculate the residual signal. In conclusion, the watermark embedding is performed server-side with a low complexity, and the amount of resources required by the CE devices are not increased by the watermarking process.

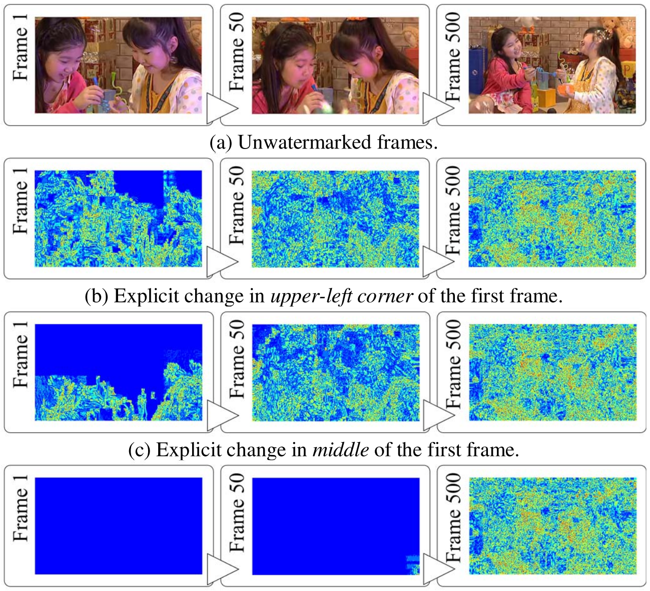
The complexity of the watermark extraction increases linearly with the number of video consumers. This is because the proposed extraction scheme calculates the correlation with all watermarks, in contrast to state-of-the-art watermark detection algorithms that calculate only a single correlation. However, it is expected that the number of malicious consumers is very low compared to the total number of consumers. Therefore, a high extraction complexity is not considered a problem, since the extraction process happens infrequently. Additionally, a single correlation only has to be calculated on a short video segment of several seconds, as discussed in Section IV-C, thus it is calculated quickly.

IV. Evaluation

In this section, the amount of implicit distortions, the perceptibility, the robustness, and the bit rate increase of the proposed watermarking scheme are evaluated.

All test sequences are encoded and watermarked using HM v16.5 with the low-delay-main configuration, meaning that only the first frame is an intra-frame and all other frames are inter-frames (B-frames). Since the watermark is embedded in every intra-frame, every sequence contains the watermark only a single time. By only inspecting the first f frames, intra-periods of length f are simulated.

In each section, the video *BlowingBubbles* is extensively analyzed first. This sequence has a resolution of 416x240 and a length of 500 frames. Furthermore, it is encoded



(a) Unwatermarked frames.

(b) Explicit change in *upper-left corner* of the first frame.

(c) Explicit change in *middle* of the first frame.

(d) Explicit change in *lower-right corner* of the first frame.

Fig. 4. Visualization of implicit distortions by explicitly changing different blocks. Changing a block in the upper-left corner creates more implicit distortions than a block in the lower-right corner. However, given enough frames for inter-frame propagation, they all generate many implicit distortions.

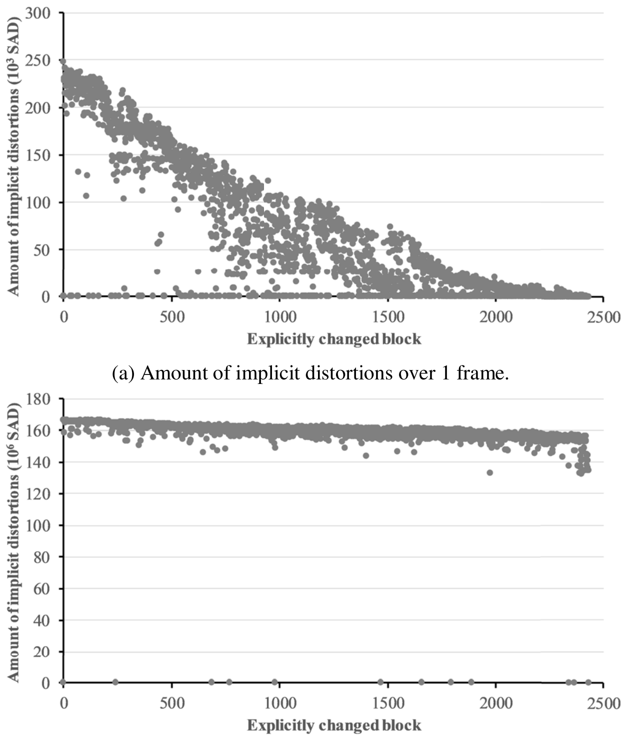
with a QP of 32 and contains 2433 blocks in its first frame. These blocks are numbered in zig-zag order from 0 to 2433. Subsequently, results from the sequences *Traffic*, *BasketballDrive*, *BasketballDrill*, and *Johnny* are briefly summarized. The results of those sequences are analyzed using 100 randomly chosen watermarks, instead of using all possible explicit changes as watermarks. This is because the computational complexity of the robustness test is quadratic in the number of used watermarks, as is further explained in Section IV-C.

A. Amount of Implicit Distortions

Recall that a watermark is represented by the implicit distortions created by an explicit change. This section quantifies the collection size by the Sum of Absolute Differences (SAD) between the unwatermarked and watermarked video.

Fig. 4 provides some illustrative examples of the effect of the position of the explicit change and of the number of used frames. In Fig. 4 (a), unwatermarked frames 1, 50, and 500 of the sequence *BlowingBubbles* are given. In Fig. 4 (b), (c), and (d), the corresponding implicit distortions are visualized after explicitly changing a block in the upper-left corner, middle, and lower-right corner of the first frame, respectively.

The following observations can be made from Fig. 4. First, there are more implicit distortions present in later frames than in earlier frames. This is because of inter-frame propagation. That is, if more frames are available for inter-frame prediction, the probability that implicit distortions are propagated is larger. Secondly, Fig. 4 shows that an explicit change in the upper-left corner contains considerably more implicit distortions in the first frame(s) than explicit changes in the lower-right corner. This is because it is more likely for a block in the upper-left corner to be used for intra-frame prediction than a block in the lower-right corner, because blocks in the upper-left corner are encoded earlier in the encoding process.



(a) Amount of implicit distortions over 1 frame.

(b) Amount of implicit distortions over 500 frames.

Fig. 5. Amount of implicit distortions of all watermarks, represented as the SAD and calculated over (a) 1 frame, and (b) 500 frames. Over 1 frame, not all explicit changes create many implicit distortions, whereas over 500 frames, almost all explicit changes create many implicit distortions. However, there are 17 blocks that do not create any implicit distortions at all.

Fig. 5 plots the amount of implicit distortions for all 2433 watermarks that one can create by explicitly changing a single block in the first frame of the sequence *BlowingBubbles*. In Fig. 5 (a), the SAD is calculated over only the first frame, whereas in Fig. 5 (b), it is calculated over 500 frames. Fig. 5 (a) illustrates again that not all explicit changes result in many implicit distortions, and blocks in the upper-left corner generally create more implicit distortions than blocks in the lower-right corner. Fig. 5 (b) illustrates that, given enough frames for intra-prediction, almost all explicit changes result in many implicit distortions, just as all three examples in Fig. 4 generate many implicit distortions. However, one can also observe that there are some explicit changes that do not create any implicit distortions at all, i.e., the resulting SAD is 0. As mentioned in Section IV-A, these explicit changes should not be used to represent a watermark.

Table I shows the fraction of watermarks that do not generate any implicit distortions at all for some other sequences. Similarly to *BlowingBubbles*, one can observe that these values are always relatively low, namely between 0% and 4%.

In conclusion, the amount of implicit distortions is dependent on the location of the explicit change and the number of available frames for inter-frame propagation. In addition, a fraction of the explicit changes do not create any implicit distortions at all.

TABLE I
EVALUATION RESULTS OF SEVERAL WATERMARKED SEQUENCES (ENCODED USING A QP OF 32)

Test sequence	Resolution & length	Fraction of SAD < 100 000	Avg. Δ SSIM	Avg. bit rate increase	Detection rate [b]			False positive rate [b]		
					QP 20	QP 30	QP 40	QP 20	QP 30	QP 40
(A) Traffic [a]	2560x1600, 150 frames	0%	0.2%	3.0%	100%	100%	100%	0%	0%	0%
(B) BasketballDrive [a]	1920x1080, 500 frames	4%	0.1%	3.6%	100%	100%	100%	0%	0%	0%
(C) BasketballDrill [a]	832x480, 500 frames	1%	0.3%	2.9%	100%	100%	100%	0%	0%	0%
(D) BlowingBubbles	416x240, 500 frames	0.5%	0.6%	4.9%	100%	100%	100%	0.08%	0.08%	0.08%
(E) Johnny [a]	1280x720, 600 frames	0%	0.1%	3.7%	100%	100%	100%	2%	2%	2%

[a] Tested on a random subset of 100 watermarks.

[b] Using only watermarks with many implicit distortions, i.e., (iii) $\text{SAD} \geq 100\,000$.

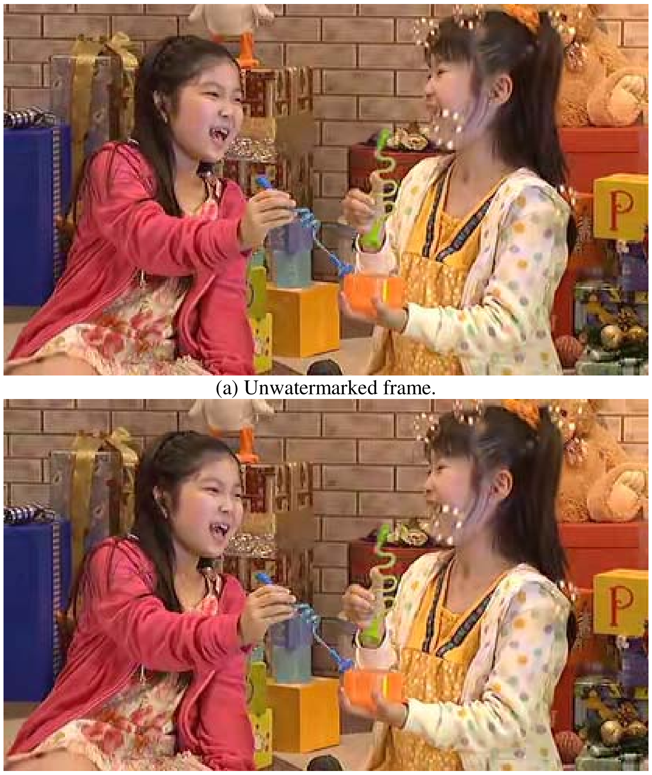
B. Perceptibility

A watermark is embedded by explicitly changing few coding decisions during the encoding process. These explicit changes generate many implicit distortions, that are assumed to be imperceptible because they are ordinary compression artifacts that are automatically generated by the video encoder. However, the set of coding decisions is not optimal anymore, since they do not take the explicit change into account. Therefore, the video quality is expected to decrease and the watermark may thus be perceptible. This subsection quantifies this quality decrease.

The Peak Signal-to-Noise Ratio (PSNR) is a widely-used metric as an objective video quality measure, but does not perform well on measuring perceptual quality. Therefore, in this paper, the Structural SIMilarity (SSIM) index is used instead, because it takes the structural information into account [25]. More specifically, the SSIM between the original and unwatermarked video are calculated on the one hand, and between the original and a watermarked video on the other. Then, the difference of these two results (Δ SSIM) represents the quality decrease that results from adding the watermark. In addition, the relative differences are calculated for a more accurate representation. For example, when watermarking *BlowingBubbles*, the Δ SSIM is approximately 0.005 or, relatively, 0.6%, meaning the objective quality decrease is low.

In addition to evaluating objective quality measures, Fig. 6 enables a subjective quality evaluation. The figure visualizes the watermarked frame that exhibited the highest objective quality decrease over all 500 frames of all 2433 watermarked versions of *BlowingBubbles*, and its unwatermarked variant. When comparing the unwatermarked and watermarked image, it can be observed that there are some regions with perceptible differences. Fig. 7 (a) and (c) zoom in on the throat of the left girl, showing that the watermarked version of the throat contains more perceptible artifacts than the unwatermarked version. However, one only notices these differences when pausing the videos and comparing the unwatermarked and watermarked version side by side. This is not considered a problem, since pirates will never have access to the unwatermarked version.

Additionally, note that some regions exhibit less perceptible compression artifacts in the watermarked version. For example, Fig. 7 (b) and (d) zoom in on the orange box on the floor, showing that the watermarked version of the box has a better quality than the unwatermarked version. Hence, even when comparing a watermarked and unwatermarked video, it is very hard – if not impossible – to tell which one is which.


(a) Unwatermarked frame.

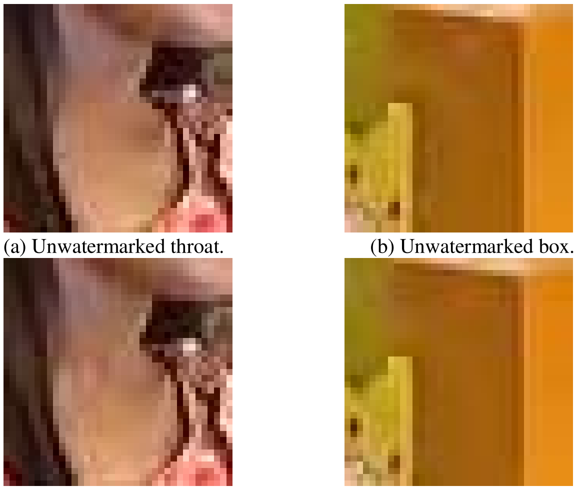
(b) Watermarked frame.

Fig. 6. The (a) unwatermarked and (b) watermarked frame that correspond to the highest observed objective quality. Both frames contain compression artifacts and exhibit perceptible differences in a few regions, although these are only noticeable when pausing the video and closely inspecting the videos side by side.

It should be stressed that Fig. 6 and 7 show a frame extracted from the video when it was encoded and watermarked with a QP of 32. When encoding the video with a lower QP, the resulting compression artifacts will be less perceptible.

Table I shows the average observed relative Δ SSIM values for other sequences. Similarly to *BlowingBubbles*, the other sequences have very low Δ SSIM values, varying between 0.1% and 0.3%, and averaging to 0.26%.

Unfortunately, state-of-the-art techniques do not use the Δ SSIM as a perceptibility metric. Instead, they often use the Δ PSNR, or the SSIM between the unwatermarked and watermarked video. Therefore, Table II shows the Δ PSNR



(a) Unwatermarked throat.

(b) Unwatermarked box.

(c) Watermarked throat.

(d) Watermarked box.

Fig. 7. Zoomed-in areas of the (a, b) unwatermarked and (b, c) watermarked frame of Fig. 6. The watermarked version of the throat (c) contains more perceptible compression artifacts than the unwatermarked version (a). However, the watermarked version of the box (d) contains less perceptible compression artifacts than the unwatermarked version (b).

TABLE II
STATE-OF-THE-ART COMPARISON

Algorithm	SSIM	Δ PSNR	Detection rate (QP 30)	FPR (QP 30)	Bit rate increase
Proposed method	0.966	0.14	100%	0.4%	3.6%
Jiang *et al.* [16]	0.989	–	78%	–	0.1%
Chen *et al.* [17]	–	0.06	76%	–	0%
Ma *et al.* [18]	–	1.25	–	–	3.6%

Non-available values are represented by a dash (–).

and SSIM of the proposed method, averaged over all tested sequences. Additionally, the table compares the obtained values with state-of-the-art techniques that have a similar computational complexity [16]–[18]. It can be observed that the perceptibility results of the proposed method are slightly worse than the values reported by Jiang *et al.* and Chen *et al.*, whereas they are slightly better than the method by Ma *et al.* However, it should be stressed that the distortions of the proposed method are generated implicitly by the video encoder, in contrast with the state-of-the-art techniques that explicitly distort the video.

In conclusion, there is a small objective quality decrease in watermarked videos. However, the resulting different compression artifacts are only perceptible upon close inspection and when comparing the watermarked frame directly to the unwatermarked frame, to which users do not have access.

C. Robustness

This section discusses the robustness of the proposed scheme and its relation to the amount of implicit distortions. This is done using the detection rate and False Positive Rate (FPR). First, *all* watermarks are generated using every possible block in the first frame of the video. Then, the watermarked videos are all attacked and subsequently extracted. Since the extraction complexity is linear in the number of watermarks, the total robustness evaluation complexity is quadratic in the number of watermarks. The detection rate is calculated by dividing the number of true positive detections with the total number of watermarks. A true positive detection happens when an embedded watermark is correctly detected. Similarly, the FPR is calculated by dividing the number of watermarked videos with false positive detections by the total number of watermarks. A false positive detection happens when a watermark is detected in a video in which it was not embedded.

TABLE III
DETECTION AND FALSE POSITIVE RATES (%) OF BLOWINGBUBBLES

Frames	Detection rate			False positive rate		
	QP 20	QP 30	QP 40	QP 20	QP 30	QP 40
(i) SAD ≥ 0						
1 frame	100	99.88	57.67	28.52	29.14	46.90
50 frames	100	100	94.08	2.14	2.10	7.89
500 frames	100	100	100	0.58	0.58	0.58
(ii) SAD ≥ 1						
1 frame	100	99.88	57.98	28.14	28.80	46.65
50 frames	100	100	94.55	1.65	1.61	7.43
500 frames	100	100	100	0.08	0.08	0.08
(iii) SAD ≥ 100 000						
1 frame	100	100	99.21	5.16	5.16	5.56
50 frames	100	100	95.37	0.67	0.67	6.55
500 frames	100	100	100	0.08	0.08	0.08

As discussed in Section IV-A, not all explicit changes generate an equal amount of implicit distortions. First, some explicit changes generate many implicit distortions from the first frame until the last. Secondly, some explicit changes result in very few distortions in the first frame(s), but propagate to many implicit distortions in later frames. Thirdly and lastly, some explicit changes do not generate any implicit distortions at all. As mentioned in Section IV-A, explicit changes that do not contain a sufficiently high amount of implicit distortions should not be used as a watermark. In other words, a threshold should be chosen for the minimum amount of implicit distortions. This section analyzes some thresholds, represented as the SAD: (i) SAD$\geq$ 0, meaning that all explicit changes are used as a watermark, (ii) SAD$\geq$ 1, meaning only explicit changes that generate no implicit distortions are eliminated, and (iii) SAD$\geq$ 100000, meaning that only explicit changes with a very high amount of implicit distortions are retained.

Table III shows the detection and false positive rates that were obtained by attacking and extracting all 2433 watermarked versions of *BlowingBubbles*, and using the three above-described thresholds. The robustness is evaluated against a re-encoding attack with a QP of 20, 30, and 40. Recall that a lower QP results in a better quality than a higher QP. Additionally, the results are analyzed for sequences of 1 frame, 50 frames and 500 frames. In the table, cells with a darker background represent worse results than cells with a lighter background. Note that the value p in (3) is set to 1%.

In part (i) of the table (i.e., SAD$\geq$ 0), it can be observed that using only a single frame is not robust. That is, the detection rate is very low (e.g., 57.67% for an attack QP of 40) and the

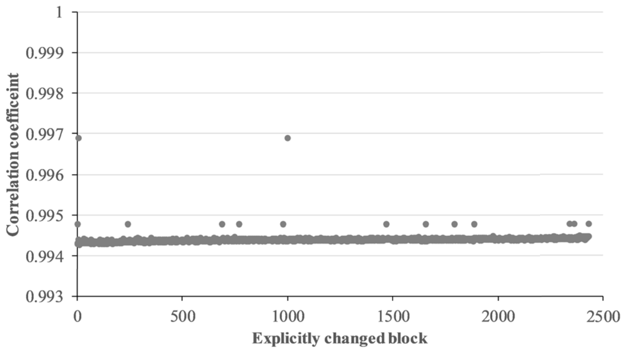


Fig. 8. Correlation values with all watermarks after applying a collusion attack by averaging block no. 7 and no. 1000, followed by a re-encoding with a QP of 20. The values corresponding to the colluding blocks are clearly outliers.

FPR is very high (e.g., 46.90% for an attack QP of 40). This is because many explicit changes did not propagate sufficiently in the first frame yet. However, the robustness increases when more frames are used. For example, when 500 frames are used, the detection rate is 100% and the FPR is only 0.58%.

In part (ii) of the table (i.e., SAD≥ 1), a small increase in robustness can be observed compared to part (i). For example, the detection rate when using 50 frames and an attack QP of 40 increases from 94.08% to 94.55%. Similarly, the corresponding FPR decreases from 7.69% to 7.43%. This is due to the elimination of the few explicit changes that did not generate any implicit distortions at all.

In part (iii) of the table (i.e., SAD≥ 100000), all detection rates further increased, although using only 1 or 50 frames still does not result in a detection rate of 100% when a re-encoding attack with a QP of 40 is performed. The FPRs are also further decreased, although it is still relatively high for 1 frame. For 500 frames, the false positive detection rate remains equal compared to part (ii), being 0.08%. These false positive detections are due to two explicit changes that both generate many implicit distortions, but whose implicit distortions are equal. Future work should be dedicated to eliminating such explicit changes and hence achieving a FPR of 0%.

Note that one can exponentially increase the detection rate and decrease the FPR by using multiple segments for the detection. For example, in the case that 500 frames are used in combination with threshold (iii), the FPR is 0.0008 (0.08%). When using n such segments, the FPR decreases to 0.0008^n.

Attackers do not limit themselves to re-encoding attacks. For example, they can perform an average collusion attack, in which multiple watermarked videos are averaged pixel by pixel, in an effort to deceive the detection algorithm [3], [6]. Traditional, collusion-secure algorithms solve this by inserting watermarked bit sequences with certain special characteristics. However, such techniques cannot be applied to the proposed watermarking approach in this paper, since there is no explicit control over the implicit distortions that represent the watermarked bit sequence. Therefore, the average collusion attack is briefly evaluated. Fig. 8 shows the correlation values after averaging the watermarked versions of *BlowingBubbles* corresponding to block no. 7 and no. 1000, followed by a re-encoding with a QP of 20. Two outliers are observed, corresponding to the blocks that were used for the attack. Thus, this suggests that the proposed watermarking algorithm is robust against a collusion attack, although future studies should do a more extensive evaluation.

Table I shows the detection rates and FPRs for the re-encoding attack for some other sequences, when all frames (i.e., 150, 500, or 600 frames, depending on the test sequence) are used in combination with threshold (iii). Recall that only 100 random watermarks are used for these other sequences. One can observe that the detection rate is always 100% and the FPR is always 0%, except for the sequence *Johnny*, which has a FPR of 2%. These results are in line with those of *BlowingBubbles*.

Table II compares the observed robustness results with those of several state-of-the-art techniques. Even when considering the FPR, the robustness of the proposed method is much higher than the algorithms by Jiang *et al.* and Chen *et al.* However, it should be noted that the detection rate is calculated differently in the state-of-the-art. That is, the number of detected bits is divided by the total number of embedded bits. As a result, they can use error correction codes to compensate for the lower detection rate [3]. However, the proposed method cannot apply such codes because the watermark is created automatically by the encoder, i.e., it is not embedded bit by bit.

In conclusion, when many frames (e.g., 150 frames or more) are used for watermark embedding and extraction, and explicit changes that do not generate any implicit distortions are eliminated, detection rates of 100% are obtained for a re-encoding attack up to a QP of 40. The corresponding false positive rate is relatively close to 0% and can be further increased by using multiple segments.

D. Bit Rate Increase

As mentioned in Section IV-B, the set of coding decisions is not optimal anymore due to the explicit change. Therefore, it will also have a negative effect on the bit rate. That is, when one watermarks a video using the same QP as when not watermarking the video, the bit rate is expected to increase. Table I shows the average observed bit rate increase for several sequences, varying between 2.9% and 4.9%.

These bit rate increases are similar to those obtained in the algorithm by Ma *et al.*, as shown in Table II. However, watermarking techniques that preserve the bit rate have been developed as well, because a bit rate increase poses a problem for applications with strict bandwidth requirements. For example, the methods of Jiang *et al.* and Chen *et al.* report a bit rate increase of approximately 0%. In the proposed scheme, rate control could be applied while encoding and watermarking the video such that the bit rate increase would be nonexistent or negligible. However, it is expected that applying rate control would also result in an additional quality decrease. Therefore, future research should address this impact.

V. CONCLUSION

This paper proposed a novel watermarking approach, based on so-called implicit distortions. These distortions are automatically generated by an encoder in which only a single coding decision is explicitly changed. Additionally, existing correlation-based detection techniques are extended by applying outlier detection. Several properties of the proposed scheme were evaluated. First, most explicit changes generate a very large collection of implicit distortions when many frames are available for inter-frame propagation. Secondly, although a small objective quality decrease is observed, the implicit distortions are considered subjectively imperceptible. Thirdly, for video segments of several seconds long, a detection rate of 100% was obtained in combination with a false positive rate close to 0%, when eliminating watermarks that do not generate a sufficient amount of implicit distortions. Lastly, a small increase in bit rate was observed.

In conclusion, the proposed watermarking technique can be applied to help combat piracy. That is, the introduced watermark can identify malicious consumers of video services that illegally distribute a video, even when they significantly lower the quality in an effort to delete the watermark. Moreover, the proposed scheme requires no modification of existing CE devices.

Acknowledgment

The computational resources (STEVIN Supercomputer Infrastructure) and services used in this work were kindly provided by Ghent University, the Flemish Supercomputer Center, the Hercules Foundation and the Flemish Government department EWI.

References

[1] H. Joo, "Private and fair pay-per-view scheme for Web-based video-on-demand systems," *IEEE Trans. Consum. Electron.*, vol. 49, no. 2, pp. 403–407, May 2003.

[2] J. Son, R. Hussain, H. Kim, and H. Oh, "SC-DVR: A secure cloud computing based framework for DVR service," *IEEE Trans. Consum. Electron.*, vol. 60, no. 3, pp. 368–374, Aug. 2014.

[3] I. J. Cox, M. L. Miller, J. A. Bloom, J. Fridrich, and T. Kalker, *Digital Watermarking and Steganography*, 2nd ed. San Francisco, CA, USA: Morgan Kaufmann, 2008.

[4] G. Van Wallendael, A. Boho, J. De Cock, A. Munteanu, and R. Van De Walle, "Encryption for high efficiency video coding with video adaptation capabilities," *IEEE Trans. Consum. Electron.*, vol. 59, no. 3, pp. 634–642, Aug. 2013.

[5] M. Buer and J. Wallace, "Integrated security for digital video broadcast," *IEEE Trans. Consum. Electron.*, vol. 42, no. 3, pp. 500–503, Aug. 1996.

[6] S. Lian and Z. Liu, "Secure media content distribution based on the improved set-top box in IPTV," *IEEE Trans. Consum. Electron.*, vol. 54, no. 2, pp. 560–566, May 2008.

[7] A. Boho *et al.*, "End-to-end security for video distribution: The combination of encryption, watermarking, and video adaptation," *IEEE Signal Process. Mag.*, vol. 30, no. 2, pp. 97–107, Mar. 2013.

[8] I. Spaliaras and S. Dokouzyannis, "A novel key refreshment scheme increasing the security of conditional access systems in digital satellite pay-TV," *IEEE Trans. Consum. Electron.*, vol. 59, no. 3, pp. 521–527, Aug. 2013.

[9] A. De Santis and C. Soriente, "A blocker-proof conditional access system," *IEEE Trans. Consum. Electron.*, vol. 50, no. 2, pp. 591–596, May 2004.

[10] M. Asikuzzaman and M. R. Pickering, "An overview of digital video watermarking," *IEEE Trans. Circuits Syst. Video Technol.*, to be published. [Online]. Available: https://ieeexplore.ieee.org/document/7938666/, doi: 10.1109/TCSVT.2017.2712162.

[11] H. Mareen, J. De Praeter, G. Van Wallendael, and P. Lambert, "A novel video watermarking approach based on implicit distortions," in *Proc. IEEE Int. Conf. Consum. Electron.*, Las Vegas, NV, USA, 2018, pp. 543–544.

[12] S. D. Lin and C.-F. Chen, "A robust DCT-based watermarking for copyright protection," *IEEE Trans. Consum. Electron.*, vol. 46, no. 3, pp. 415–421, Aug. 2000.

[13] C.-F. Wu and W.-S. Hsieh, "Digital watermarking using zerotree of DCT," *IEEE Trans. Consum. Electron.*, vol. 46, no. 1, pp. 87–94, Feb. 2000.

[14] C.-T. Hsu and J.-L. Wu, "DCT-based watermarking for video," *IEEE Trans. Consum. Electron.*, vol. 44, no. 1, pp. 206–216, Feb. 1998.

[15] G. C. Langelaar, I. Setyawan, and R. L. Lagendijk, "Watermarking digital image and video data. A state-of-the-art overview," *IEEE Signal Process. Mag.*, vol. 17, no. 5, pp. 20–46, Sep. 2000.

[16] X. Jiang, T. Sun, Y. Zhou, W. Wang, and Y.-Q. Shi, "A robust H.264/AVC video watermarking scheme with drift compensation," *Sci. World J.*, vol. 2014, p. 13, Feb. 2014. [Online]. Available: https://www.hindawi.com/journals/tswj/2014/802347/

[17] W. Chen, Z. Shahid, T. Stütz, F. Autrusseau, and P. Le Callet, "Robust drift-free bit-rate preserving H.264 watermarking," *Multimedia Syst.*, vol. 20, no. 2, pp. 179–193, 2014.

[18] X. Ma, Z. Li, H. Tu, and B. Zhang, "A data hiding algorithm for H.264/AVC video streams without intra-frame distortion drift," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 20, no. 10, pp. 1320–1330, Oct. 2010.

[19] S. Gaj, A. Sur, and P. K. Bora, "A robust watermarking scheme against re-compression attack for H.265/HEVC," presented at the 5th Nat. Conf. Comput. Vis. Pattern Recognit. Image Process. Graph., Patna, India, 2015, pp. 1–4.

[20] T. Dutta and H. P. Gupta, "A robust watermarking framework for high efficiency video coding (HEVC)—Encoded video with blind extraction process," *J. Vis. Commun. Image Represent.*, vol. 38, pp. 29–44, Jul. 2016.

[21] T. Kalker, G. Depovere, J. Haitsma, and M. J. Maes, "Video watermarking system for broadcast monitoring," in *Proc. SPIE Security Watermarking Multimedia Contents*, vol. 3657. San Jose, CA, USA, 1999, pp. 103–112. [Online]. Available: https://www.spiedigitallibrary.org/conference-proceedings-of-SPIE/3657.toc

[22] F. Hartung and B. Girod, "Watermarking of uncompressed and compressed video," *Signal Process.*, vol. 66, no. 3, pp. 283–301, 1998.

[23] T. Yamada, M. Maeta, and F. Mizushima, "Video watermark application for embedding recipient ID in real-time-encoding VoD server," *J. Real Time Image Process.*, vol. 11, no. 1, pp. 211–222, 2016.

[24] R. Dugad, K. Ratakonda, and N. Ahuja, "A new wavelet-based scheme for watermarking images," in *Proc. Int. Conf. Image Process.*, Chicago, IL, USA, 1998, pp. 419–423.

[25] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, pp. 600–612, Apr. 2004.

Hannes Mareen received the M.Sc. degree in computer science engineering from Ghent University, Belgium, in 2017. Since then, he has been an SB Ph.D. Fellow with imec, IDLab, Ghent University, with the financial support of the Research Foundation—Flanders (FWO). His main areas of interest are multimedia security and video coding.

Johan De Praeter received the M.Sc. degree in computer science engineering and the Ph.D. degree from Ghent University, Ghent, Belgium, in 2013 and 2017, respectively. Since 2013, he has been a Post-Doctoral Researcher with imec, IDLab, Ghent University. His main research interests include reducing the computational complexity of video compression, with a focus on high-efficiency video coding, transcoding, and simultaneous encoding.

Glenn Van Wallendael received the first M.Sc. degree in applied engineering from the University College of Antwerp, Belgium, in 2006 and the second M.Sc. degree in engineering from Ghent University, Belgium, in 2008, where he is currently pursuing the Ph.D. degree with IDLab, with the financial support of the Research Foundation—Flanders (FWO) and he is currently a Post-Doctoral Researcher. His main topics of interest are video compression including scalable video compression and transcoding.

Peter Lambert received the master's degrees in science (mathematics) and in applied informatics and the Ph.D. degree in computer science from Ghent University, Belgium, in 2001, 2002, and 2007, respectively, where he has been a full-time Associate Professor with imec, IDlab, since 2013. In 2009, he became a Technology Developer with Ghent University, which he combined with a part-time Assistant Professorship with IDLab since 2010. His research interests include (mobile) multimedia applications, multimedia coding and adaptation technologies, and 3-D graphics.