2016 24th European Signal Processing Conference (EUSIPCO)

Detection of Double AVC/HEVC Encoding

Andrea Costanzo, *Mauro Barni*, *Fellow, IEEE*

Dept. of Information Engineering and Mathematics, University of Siena, Italy
E-mail: `andreacos82@gmail.com`, `barni@dii.unisi.it`

***Abstract*—New generation video codecs are designed to improve coding efficiency with respect to previous standards and to support the latest hardware and applications. High Efficiency Video Coding (HEVC) is the successor of Advanced Video Coding (AVC), which is by far the most adopted standard worldwide. To promote the new standard, producers are re-releasing recent movies in HEVC format. In such a scenario, a fraudulent provider that does not own the original uncompressed data could sell old, lower quality AVC content re-encoded as if it were natively HEVC. Furthermore, with several hundred hours of video content uploaded every minute, it is not unlikely that re-edited low quality clips are labelled as HEVC to increase popularity and revenues from advertising. We tackle with these and similar issues by proposing a forensic technique to detect whether a HEVC sequence was obtained from an uncompressed sequence or by re-encoding an existing AVC sequence.**

***Index Terms*—Multimedia Forensics, Video Forensics, Video re-encoding, HEVC, AVC.**

I. INTRODUCTION

High Efficiency Video Coding (HEVC) [1], [2] is the new generation video compression standard jointly developed by the ISO/IEC Moving Picture Experts Group and the ITU-T Visual Coding Experts Group. HEVC has been developed to overcome the limitations of the Advanced Video Coding (AVC) standard [3], which despite its ten years of age is still today the *de-facto* reference for video compression. The HEVC standardization effort was driven by the growing demand for efficiency, in order to enable a set of diverse new applications including Ultra HD resolutions (e.g. 4K and 8K), 3D and multiview display.

Following its third release in early 2015, HEVC has begun replacing AVC thanks to the growing support from hardware producers and leading companies in movie entertainment, mobile communications and Internet services. During this transition period, existing video content is being re-released into the market in HEVC format to promote the diffusion of the latest standard. New security issues raise in such a scenario. For instance, a fraudulent service that does not own the original uncompressed data could attempt to sell old, lower quality AVC content re-encoded as if it were natively HEVC. Additionally, several hundred hours worth of video content are uploaded every minute to sharing platforms such as YouTube, which totalises more than four billion views overall each day[1]. It is reasonable to hypothesize that a share of such contents are low quality AVC videos, re-encoded as HEVC and wrongly (or deliberately) tagged as HEVC to increase popularity and, consequently, revenues from advertising.

The Multimedia Forensics research community has devoted a significant effort to the authentication and integrity verification of video content [4]. Several techniques have been devised to identify footprints left by the acquisition device [5]–[7], the employed codec and its parameters [8]–[13], multiple encodings [14], [15] and forgeries such as frame or object removal or duplication [16], [17].

In this paper we contribute to the Multimedia Forensics mission with a method to detect double AVC/HEVC encoding, under the assumption that the former compression has a lower quality than the latter compression. The main idea behind the technique we have developed stems from the observation that the first AVC encoding tends to alter the statistical properties of a video sequence hindering the flexibility of the subsequent HEVC's quad-tree partitioning. More specifically, we argue that the first AVC encoding influences the decision strategy of the subsequent HEVC encoding regarding the motion prediction modes in bi-directionally predicted frames. Based on such an observation, the goal of this paper is to present a method to reveal double AVC/HEVC encoding and to estimate the first AVC quantization parameter.

The rest of the paper is organized as follows. In Sec. II, we briefly introduce the HEVC codec. In Sec. III, we describe the footprint we rely on to detect double AVC/HEVC encoding. In Sec. IV, we validate our technique and we show how to use it to estimate the quantization parameter of the former AVC compression.

II. HIGH EFFICIENCY VIDEO CODING (HEVC)

Due to space limitations, only the key-features of HEVC required to understand the method presented in Sec. III are briefly discussed. The interested reader is referred to [2] for a thorough description.

Like the majority of prior video coding standards, HEVC processes all pictures of the input sequence in block units, each of which is compared to a reference block that is computed either from previously decoded pictures (inter-prediction) or previously decoded samples from the same picture (intra-prediction). The reference block is then subtracted from the input block to obtain the residuals, which are transformed (usually in the frequency domain), quantized and entropy coded. Specifically, HEVC partitions each picture into *Coding Tree Units* (CTU), whose fixed size can be either 16×16, 32×32 or, more commonly, 64×64. Each CTU consists of

---

[1] Source: http://expandedramblings.com/index.php/youtube-statistics/.

three *Coding Tree Blocks* (CTB): one for the luma samples and two for the chroma samples in YCbCr color space. The size of luma CTB is the same as the CTU size, whereas the chroma CTBs' size depends on the adopted sampling scheme[2].

To better adapt the coding parameters to the local content, each CTB can be further sub-divided according to a quad-tree structure, the same for luma and chroma, into smaller *Coding Blocks* (CB) down to a minimum of 8×8 for luma and 4×4 for chroma. The luma CB and the corresponding chroma CBs are logically grouped into a *Coding Unit* (CU). The prediction type (i.e. inter or intra) is decided at the CU level. To improve the accuracy of the prediction, each CU can be sub-divided into smaller square or rectangular *Prediction Units* (PU) depending on the temporal/spatial predictability of the content. Once the prediction is made at the PU level, each CU is split into *Transform Units* according to a second quad-tree that may not be aligned to the first one. Each TU is transformed (integer DCT), quantized and entropy coded [2].

In HEVC and AVC quantization is based on a scalar quantizer whose step is derived from an index called Quantization Parameter (QP) assuming integer values in $[0, 51]$. The quantization step size doubles with every increment of 6 in QP according to a logarithmic structure. Larger QPs achieve higher compression rates at the expense of visual quality.

Typically, CTUs are grouped into *slices*. The number of slices needs to be constant across the whole video sequence, whereas the number of CTUs in each slice does not. Even though slices can be useful to recover from data losses at the cost of compression efficiency, it is common practice in many HEVC applications to use a single slice for each picture. A slice can belong to three different categories, called I-, P- and B-, defining the prediction types available to that slice: CUs in I- slices can only be intra-predicted; CUs in P- slices can either be intra or uni-predicted, i.e. inter-predicted from a single reference picture; CUs in B- slices can either be inter-predicted or bi-predicted, i.e. inter-predicted by resorting to more than one previously encoded reference picture.

To further increase coding efficiency in P- and B- slices, HEVC can resort to the skip mode. A Skip CU consists of a single PU whose motion data is derived from neighbouring CUs. No residuals are transmitted in Skip mode.

III. THE PROPOSED FOOTPRINT

HEVC improves AVC in nearly all aspects. Arguably, the most significant leap forward consists of the frame partitioning scheme. AVC sub-divides each picture into blocks of size 16×16, which can be further split only into 4 blocks of size 8×8 or into 16 blocks of size 4×4. Conversely, HEVC chooses block sizes in a more flexible and dynamic fashion according to a quad-tree structure which depends on the statistical properties of the content. The more (the less) the motion in a region, the smaller (the larger) the size of the CUs covering the region. In this way, HEVC significantly boosts motion prediction accuracy and either provides much better perceptual quality at the same bit-rate or higher compression efficiency at the same quality. In fact, HEVC bit-rate reduction with respect to AVC is about 50% according to [2].

A. Footprint Description

Suppose that a video sequence is obtained from an uncompressed source by encoding it twice, first with AVC and then with HEVC. Suppose also that the AVC quantization parameter is larger than the one adopted by HEVC, i.e. the latter compression is higher quality than the former. We hypothesize that the former AVC alters the statistical properties of the video sequence enough to break the flexibility of the subsequent HEVC's quad-tree partitioning, even when the two QPs are close (and thus the perceptual quality is similar) and no visual clues of double encoding can be noticed. If this were true, then in principle it would be possible to distinguish between such a sequence and a sequence obtained by encoding the same source once with HEVC.

To verify our hypothesis, we focused on the motion prediction modes chosen for CUs by the HEVC encoder in P- and B- frames (I- frames have no motion information). More specifically, we checked the frequency of motion prediction modes over all B- frames of video sequences that were compressed once with HEVC and twice according to the sequence AVC/HEVC. Interestingly, we observed that in B- frames the frequency of bi-predicted CUs is significantly higher than that of uni-predicted CUs when the HEVC sequence is obtained by starting from an uncompressed clip. Conversely, when the HEVC sequence is obtained by re-encoding a lower quality AVC sequence, P- CUs represent the large majority of all the units. We believe that this happens because the previous lower quality AVC encoding reduces the temporal differences between blocks, thus preventing the HEVC codec from using the more accurate bi-prediction in favour of the coarser uni-prediction. An example of such a behaviour is given in Fig. 1 for all the B- frames of the `Tennis` sequence[3]: frequencies of the single (resp. double) compressed CUs are shown on the left (resp. right). As it can be seen, P- CUs significantly increase if the sequence was originally compressed with AVC.

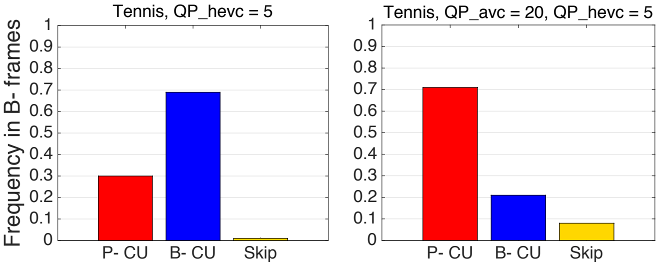

Fig. 1. Frequency of motion prediction modes in B- frames (`Tennis`).

---

[2] For example, in the widely used 4:2:0 sampling scheme, each 64×64 CTU consists of one 64×64 luma CTB and two 32×32 chroma CTBs.

[3] This is a High Definition (HD 1920×1080p) sequence belonging to the Class B benchmarks of the JCT-VC (Joint Collaborative Team on Video Coding). It should not be confused with the 532×288 homonym sequence.

In the following, we will refer to the footprint of a video sequence V as $\mathcal{F}_V = [P, B, Skip]$, which is the array with the frequencies of each type of motion prediction.

B. Detection of AVC/HEVC compression sequence

In order to exploit the footprint described in the previous section to detect a sequence of AVC/HEVC encoding, we rely on the concept of video compression idempotency first introduced in [12], [18]. Idempotency of lossy coding states that whenever a video coding scheme is applied twice with the same parameters to a video, it produces approximately the same output as if it were applied once. Hence, assuming that the codec is known, one can re-encode the sequence under analysis with different parameters and verify whether the result is similar to the input sequence. If that is the case, then most likely the parameters used for re-encoding correspond the parameters used to originally code the video.

We now apply the above ideas to solve the following problem: given a sequence V whose last QP_{hevc} is known, determine whether V has been compressed once with HEVC or twice according to the sequence AVC/HEVC with unknown $QP_{avc} > QP_{hevc}$. Specifically, the solution we are proposing works as follows:

- Extract the footprint $\mathcal{F}_V$ from V;
- Re-encode V with AVC with increasing $QP_{avc}^{(id)} = \{QP_{hevc} + 1, \dots, 51\}$ followed by HEVC with QP_{hevc} to obtain a sequence of videos V_i', $i = 1, \dots, length(QP_{avc}^{(id)})$;
- Extract the footprint $\mathcal{F}_{V_i'}$ from each V_i';
- Measure the distances $d(\mathcal{F}_V, \mathcal{F}_{V_i'}) = |\mathcal{F}_V - \mathcal{F}_{V_i'}|\ \forall i$, where each difference is element-wise.
- Use the distances computed at the previous point to detect AVC/HEVC re-encoded video.

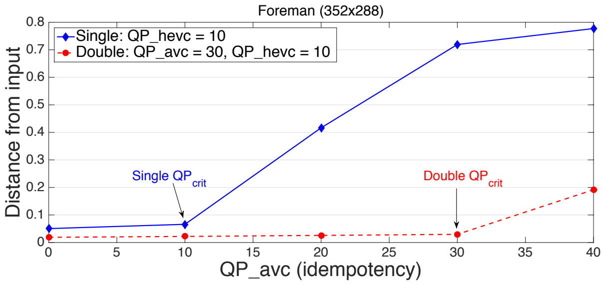

Fig. 2. Example of distances following idempotency analysis.

With regard to the last point, if V was compressed just once with HEVC, we expect $d(\mathcal{F}_V, \mathcal{F}_{V_i'})$ to be large for all $QP_{avc}^{(id)}$. This is due to the fact that the frequencies of motion prediction modes in all the sequences V_i' have been altered with respect to the original single compressed HEVC video as in Fig. 1. Conversely, if V was compressed twice, we expect $d(\mathcal{F}_V, \mathcal{F}_{V'})$ to be minimum in correspondence of the $QP_{avc}^{(id)}$ that coincides with the original QP_{avc} used to encode V. To show that this is indeed the case, let us consider the example in Fig. 2 (a more detailed analysis will be given in in Sec. IV-B): the solid (resp. dashed) line plots the set of distances for the single (resp. double) compressed `Foreman` sequence ($QP_{avc} = 30$, $QP_{hevc} = 10$). For both the single and the double-compressed sequence there exists a critical $QP_{avc}^{(id)}$ past which the distance between the footprints starts increasing; in the case of single compression, this happens in correspondence of the QP_{hevc} declared in the bit-stream information of the video under scrutiny; in the case of double compression, in correspondence of the original QP_{avc}. Let QP_{crit} be such critical value, adopt the following decision rule to classify V:

$$V \in \begin{cases} \textit{HEVC} & \text{if } QP_{crit} = QP_{hevc} \\ \textit{AVC/HEVC} & \text{if } QP_{crit} > QP_{hevc}. \end{cases} \tag{1}$$

In the latter case, the quantization parameter of the old AVC compression coincides with QP_{crit}.

IV. EXPERIMENTAL VALIDATION

In this section, we first provide a qualitative analysis of the proposed footprint on two video sequences, then we evaluate the performance of the method in terms of detection accuracy on a data set of 20 YUV test sequences[4]. Unless specified otherwise, all sequences have resolution 532×288 and their frame count ranges from 150 to 2100. Similarly to other works, we focused on low resolution videos to keep the time complexity under control. We compressed the videos by means of the open-source *libx264* (AVC) and *libx265* (HEVC) codecs included in the well established *FFmpeg* multimedia framework[5]. We used a commercial HEVC bit-stream analyser and a in-house C++ parser to extract all the information. We made two assumptions regarding the quantization parameters:

$$QP_{avc} \in \{0, 1, 2, \dots, 40\} \tag{2}$$

$$QP_{avc} > QP_{hevc}. \tag{3}$$

Eq. (2) restricts QP_{avc} to the range of compressions that do not impact too much on the visual quality (otherwise, the to-be-considered HEVC video would hardly fool anyone), while Eq. (3) is in line with the scenario of Sec. III.

To determine QP_{crit}, we computed the first order forward difference of the distance vector, i.e. $f(k) = d(k+1) - d(k)$, $k \in [0, length(d) - 1]$; then, we selected the $QP_{avc}^{(id)}$ corresponding to the first k such that $f(k) > T$, where T is a fixed threshold experimentally derived in Sec. IV-B.

---

[4] Namely: *Akiyo*, *BasketBallDrill* (qualitative analysis only), *BridgeClose*, *BridgeFar*, *Bus*, *Carphone*, *Coastguard*, *Container*, *Flower*, *Football* (qualitative analysis only), *Foreman*, *Ice*, *MissAmerica*, *Mobile*, *MotherDaughter*, *News*, *Paris*, *Silent*, *Soccer*, *Stefan*, *Tempete*, *Tennis* (Sec. III only), *Waterfall*. Download: http://videocoders.com/yuv.html and http://trace.eas.asu.edu/yuv/

[5] Software downloadable from: https://www.ffmpeg.org (*FFmpeg*); http://www.videolan.org/developers/x264.html (*libx264*); http://x265.org (*libx265*).

A. *Qualitative analysis of double AVC/HEVC detection*

In this experiment, we compressed the `Football` sequence with HEVC with $QP_{hevc} = \{0, 10, 20, 30, 40\}$ to obtain five single compressed videos. Then, we re-encoded each video with AVC with increasing $QP_{avc}^{(id)}$ (step 5 to reduce computational complexity) followed by HEVC with the original QP_{hevc}. Finally, we computed the distance between the input footprint and the footprint of the re-encoded videos. The results we obtained are shown in Fig. 3, where each curve corresponds to a sequence[6].

Expectedly, when $QP_{avc}^{(id)} \leq QP_{hevc}$ no valuable information can be obtained, since double encoding can not be revealed when the first AVC compression is not strong enough to alter HEVC motion prediction. Conversely, when $QP_{avc}^{(id)} > QP_{hevc}$, the distance between footprints rapidly increases.

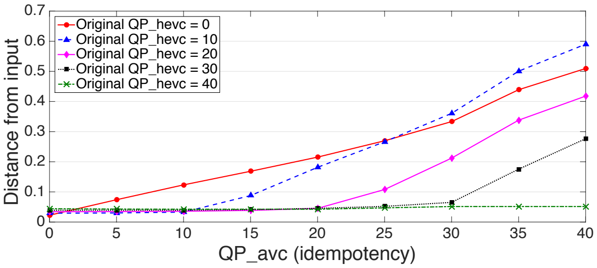

Fig. 3. Distance between the single compressed input sequence V and idempotency's outputs V'. Each curve corresponds to a different input sequence obtained by varying the original HEVC quantization parameter.

In the second experiment, we compressed the `Football` sequence twice, with AVC with $QP_{avc} = \{15, 20, 25, 30, 35\}$ followed by HEVC with fixed $QP_{hevc} = 10$. Then, we re-encoded each of the five double compressed sequences with AVC with increasing $QP_{avc}^{(id)}$ followed by HEVC with $QP_{hevc} = 10$. By observing the results shown in Fig. 4, we notice that the distance between footprints is null until the $QP_{avc}^{(id)}$ used for re-encoding is lower than the original QP_{avc}. Again, this happens because re-compression is higher quality than the original compression, thus producing a sequence that is extremely similar to the input sequence. When $QP_{avc}^{(id)}$ is greater or equal than the real QP, the distance starts growing.

One advantage of the proposed method is that it does not require many frames to obtain a reliable footprint. In Fig. 5, we can observe that the shape of $d(\mathcal{F}_V, \mathcal{F}_{V'})$ does not depend significantly on the number of frames, either for single (left) or double (right) compressed sequences (`BasketBallDrill`, 832×480, $QP_{avc} = 30$, $QP_{hevc} = 10$). Recall that we resort only to B- frames, which, according to our analysis over all the compressed videos, represent on average about the 70% of the totality when *x265* default options are used [19]. Then, we do not need more than a few seconds of content to classify a sequence as single or double compressed.

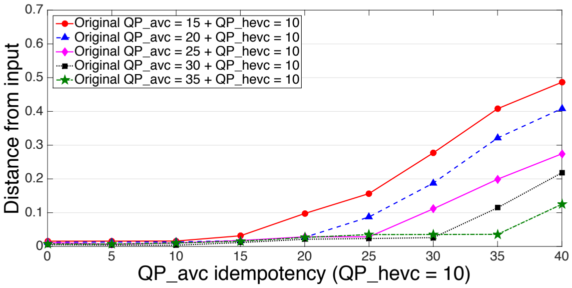

Fig. 4. Distance between the double compressed input sequence V and V'. Each curve corresponds to a different input sequence obtained by varying the first AVC quantization parameter.

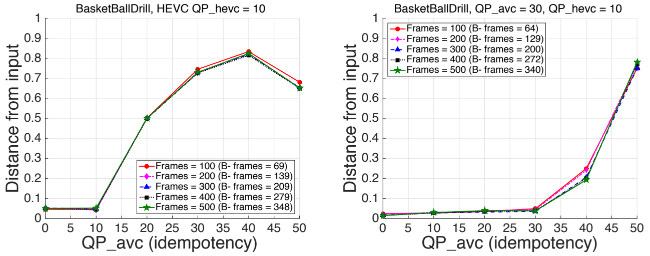

Fig. 5. Distance between footprints depending on the number of analysed frames (`BasketBallDrill`). Left: HEVC; right: AVC/HEVC.

B. *Detection accuracy of double AVC/HEVC detection*

We evaluated the detection accuracy of the method in terms of ROC (Receiver Operating Characteristic) curves as follows. We encoded the 20 YUV test sequences to create 20 single compressed and 20 double compressed versions; we computed the distance as in Sec. III-A; we varied the decision threshold T in the interval $[0, 1]$ with step 0.05; for each T, we carried out the classification task as in Eq. (1). Results are shown in Fig. 6, where we have repeated the experiment twice for $(QP_{avc}, QP_{hevc}) = (30, 10)$ and $(20, 10)$ (solid/dashed line). Expectedly, the closer the QPs, the harder the classification. Nevertheless, results are indeed satisfactory, considering that the AuC (Area Under Curve) is 0.973 and 0.926 respectively.

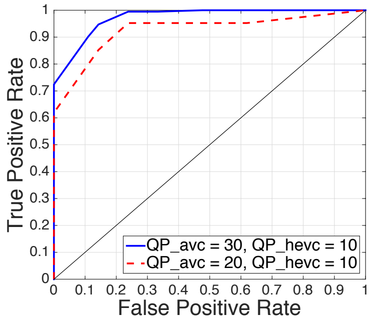

Fig. 6. Accuracy of single vs. double classification. ROC curves obtained on the data set of 40 video sequences by varying the decision threshold of Eq. (1), with fixed $QP_{avc} = 30$ (solid line) / 20 (dashed line) and $QP_{hevc} = 10$.

Eventually, we evaluated the accuracy of the estimation of the old AVC quantization parameter. To do so, we fixed

---

[6] For example, the dashed blue line with triangle markers (Fig. 3) is the distance between the footprint of the input sequence, originally compressed once with HEVC $QP_{hevc} = 10$, from all the sequences that are re-encoded with increasing QP_{avc} followed by $QP_{hevc} = 10$.

$T = 0.35$, corresponding to a false alarm rate of 0.1, and we focused on those videos that were correctly classified as double compressed (only in the case $QP_{avc} = 30, QP_{hevc} = 10$), i.e. 18 out of 20 sequences. We found out that for 16 of them, QP_{crit} coincides with the true value 30, whereas for the remaining two, we got 20. Interestingly, the outcome of QP estimation does not change as T increases: instead, it seems related to the smooth content and very small motion of such two videos. Arguably, a correct estimation would require more B- frames than those among the 100 frames under analysis.

V. CONCLUSION

We have proposed a new forensic technique to discriminate between uncompressed video sequences encoded once with HEVC and lower-quality AVC sequences re-encoded with HEVC. Discrimination is based on the presence of alterations introduced by the initial AVC compression to the decision strategy of the subsequent HEVC encoding regarding the motion prediction modes. We showed that, in the case of double encoding, we can also estimate the former AVC quantization parameter. We plan to devote further research to assessing the performance of our method on High Definition content; determining the relationship between the amount of video motion and the number of frames for fingerprint stability; considering different codecs for the first compression. Finally, we will adapt our method locally to the detection of splicing of HEVC and AVC sequences.

REFERENCES

[1] G. J. Sullivan, J.-R. Ohm, W.-J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 22, no. 12, pp. 1649–1668, 2012.

[2] V. Sze, M. Budagavi, and G. J. Sullivan, *High Efficiency Video Coding (HEVC)*. Springer, 2014.

[3] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H. 264/AVC video coding standard," *Circuits and Systems for Video Technology, IEEE Transactions on*, vol. 13, no. 7, pp. 560–576, 2003.

[4] S. Milani, M. Fontani, P. Bestagini, M. Barni, A. Piva, M. Tagliasacchi, and S. Tubaro, "An overview on video forensics," *APSIPA Transactions on Signal and Information Processing*, vol. 1, p. e2, 2012.

[5] M. Chen, J. Fridrich, M. Goljan, and J. Lukáš, "Source digital camcorder identification using sensor photo response non-uniformity," in *Electronic Imaging 2007*. International Society for Optics and Photonics, 2007, pp. 65 051G–65 051G.

[6] S. Bayram, H. T. Sencar, and N. Memon, "Efficient techniques for sensor fingerprint matching in large image and video databases," in *IS&T/SPIE Electronic Imaging*. International Society for Optics and Photonics, 2010, pp. 754 109–754 109.

[7] M. Visentini-Scarzanella and P. L. Dragotti, "Video jitter analysis for automatic bootleg detection," in *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*. IEEE, 2012, pp. 101–106.

[8] S. Tjoa, W. S. Lin, and K. Liu, "Transform coder classification for digital image forensics," in *Image Processing, 2007. ICIP 2007. IEEE International Conference on*, vol. 6. IEEE, 2007, pp. VI–105.

[9] W. Luo, Y. Wang, and J. Huang, "Detection of quantization artifacts and its applications to transform encoder identification," *Information Forensics and Security, IEEE Transactions on*, vol. 5, no. 4, pp. 810–815, 2010.

[10] G. Valenzise, M. Tagliasacchi, and S. Tubaro, "Estimating QP and motion vectors in H. 264/AVC video from decoded pixels," in *Proceedings of the 2nd ACM workshop on Multimedia in forensics, security and intelligence*. ACM, 2010, pp. 89–92.

[11] M. Tagliasacchi and S. Tubaro, "Blind estimation of the QP parameter in H. 264/AVC decoded video," in *Image Analysis for Multimedia Interactive Services (WIAMIS), 2010 11th International Workshop on*. IEEE, 2010, pp. 1–4.

[12] P. Bestagini, A. Allam, S. Milani, M. Tagliasacchi, and S. Tubaro, "Video codec identification," in *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*. IEEE, 2012, pp. 2257–2260.

[13] T. Gloe, A. Fischer, and M. Kirchner, "Forensic analysis of video file formats," *Digital Investigation*, vol. 11, pp. S68–S76, 2014.

[14] S. Milani, P. Bestagini, M. Tagliasacchi, and S. Tubaro, "Multiple compression detection for video sequences," in *Multimedia Signal Processing (MMSP), 2012 IEEE 14th International Workshop on*. IEEE, 2012, pp. 112–117.

[15] D. Vazquez-Padin, M. Fontani, T. Bianchi, P. Comesaña, A. Piva, and M. Barni, "Detection of video double encoding with GOP size estimation," in *Information Forensics and Security (WIFS), 2012 IEEE International Workshop on*. IEEE, 2012, pp. 151–156.

[16] W. Wang and H. Farid, "Exposing digital forgeries in video by detecting duplication," in *Proceedings of the 9th workshop on Multimedia & security*. ACM, 2007, pp. 35–42.

[17] J. Zhang, Y. Su, and M. Zhang, "Exposing digital video forgery by ghost shadow artifact," in *Proceedings of the First ACM workshop on Multimedia in forensics*. ACM, 2009, pp. 49–54.

[18] P. Bestagini, S. Milani, M. Tagliasacchi, and S. Tubaro, "Video codec identification extending the idempotency property," in *Visual Information Processing (EUVIP), 2013 4th European Workshop on*. IEEE, 2013, pp. 220–225.

[19] "MulticoreWare x265 documentation," Tech. Rep. MulticoreWare Inc. Revision e951ab673b1c+. [Online]. Available: https://x265.readthedocs.org/en/default/