A Novel Sketch Attack for H.264/AVC Format-Compliant Encrypted Video

Kazuki Minemura, *Student Member, IEEE*, KokSheik Wong, *Senior Member, IEEE*, Raphael C.-W. Phan, *Member, IEEE*, and Kiyoshi Tanaka, *Member, IEEE*

***Abstract*— In this paper, we propose a novel sketch attack for H.264 advanced video coding (H.264/AVC) format-compliant encrypted video. We briefly describe the notion of sketch attack, review the conventional sketch attacks designed for discrete cosine transform (DCT)-based compressed image, and identify their shortcomings when applied to attack compressed video. Specifically, the conventional DCT-based sketch attacks are incapable in sketching outlines for inter frame, which is deployed to significantly reduce temporal redundancy in video compression. To sketch directly from inter frame, we put forward a sketch attack by considering the partially decoded information of the H.264/AVC compressed video, namely, the number of bits spent on coding a macroblock. To evaluate the sketch image, we consider the Canny edge map as the ideal outline image. Experiments are conducted to verify the performance of the proposed sketch attack using ICADR2013, High Efficiency Video Coding dash, and Xiph video data sets. Results suggest that the proposed sketch attack can generate the outline image of the original frame for not only intra frame but also inter frame.**

***Index Terms*— Format-compliant encryption, H.264 advanced video coding (H.264/AVC), integer transform, macroblock (MB), sketch attack.**

I. Introduction

With advances in multimedia signal processing, the state-of-the-art technology enables us to handle high-definition (HD) video across devices of different computational capabilities. Notably, video-based services, such as video on demand (VOD), video sharing (VS), video conferencing (VC), advertisement (AD), and surveillance system (SS), have enriched and revolutionized our daily life. One of the main motivations that enables the aforementioned applications is the advent of highly efficient video compression standards. Among the compression standards, H.264 advanced video coding (H.264/AVC) [1] is one of the most commonly utilized standards for the recording, compression, and distribution of HD video content.

As the demand for HD video increases, practical security tools for H.264/AVC compressed video also become increasingly important. Stutz and Uhl [2] defined transparent encryption, where one of the requirements is the ability to control the perceptual video quality. Perceptual quality control for entertainment is of high significance for the aforementioned applications (e.g., VOD, VS, VC, AD, and SS), because the content provider can flexibly generate a video of the desired level of distortion prior to transmission to attract potential purchasers while not revealing the high-quality content. Furthermore, Stutz and Uhl [2] categorized the conventional encryption methods into three classes, namely: 1) encryption before compression; 2) encryption during compression; and 3) encryption on bitstream. Specifically, Class 1) encrypts the video prior to compression, but it compromises significantly on the video compression performance, because encryption by its nature mixes up any redundancies. On the other hand, Class 2) encrypts the components in the H.264/AVC compression standard, such as integer transformed coefficients, scanning order, and sign information. Finally, Class 3) encrypts the encoded bitstream while ensuring format compliance. This paper focuses on attacking video encryption methods of Class 2).

Most conventional methods are designed to meet a certain security requirements against cryptographic attacks, such as known/chosen-plaintext attacks. However, a format-compliant transparent/perceptual video encryption may not be completely secure from other forms of attack. Specifically, in the literature, there are two kinds of perceptual attacks: 1) error-concealment-based attack, which treats the encrypted parts as the lost packets and reconstructs the video using error-concealment techniques [3], and; 2) replacement attack that replaces the encrypted parts with arbitrary data [4], [5]. These perceptual attacks failed to reconstruct low-quality images when discrete cosine transform (DCT) coefficients are shuffled within a block.

While the conventional perceptual encryption methods are argued to be cryptographically secure, Li and Yuan [6] proposed a primitive signal processing operation to sketch the outline of the original (i.e., plaintext) JPEG image directly from its encrypted counterpart. In particular, a simple feature extraction is performed on the encrypted image, which is degraded by transparent/perceptual encryption [2], [7], [8]. Li and Yuan's [6] method [nonzero coefficient count attack (NZCA)] relies on the number of nonzero DCT

Manuscript received June 28, 2015; revised April 26, 2016; accepted July 2, 2016. Date of publication July 11, 2016; date of current version November 8, 2017. This work was supported by the University of Malaya–High Impact Research Central through the Project entitled Unified Scalable Information Hiding under Grant UM.C/625/1/HIR/MOHE/FCSIT/01 and Grant B000001-22001. The work of R. C.-W. Phan was supported by Telekom Malaysia through the Project 2beAware under Grant RDTC/150879. This paper was recommended by Associate Editor P. Salama.

K. Minemura and K. Wong are with the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur 50603, Malaysia (e-mail: kazuki.minemura@siswa.um.edu.my; koksheik@um.edu.my).

R. C.-W. Phan is with the Faculty of Engineering, Multimedia University, Cyberjaya 63000, Malaysia (e-mail: raphael@mmu.edu.my).

K. Tanaka is with the Academic Assembly, Institute of Engineering, Shinshu University, Matsumoto 390-0802, Japan (e-mail: ktanaka@shinshu-u.ac.jp).

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/TCSVT.2016.2589742

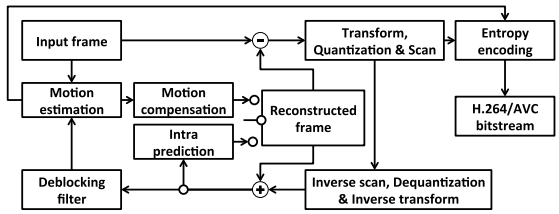


Fig. 1. H.264/AVC encoder [1].

coefficients in each block to generate a binary sketch image. However, NZCA requires manual thresholding to generate a clear outline image. To solve the manual thresholding problem, Minemura *et al.* [9], [10] proposed three sketching methods, namely, improved nonzero coefficient count (INCC), position of the last nonzero coefficient (PLZ), and energy of ac coefficients, to generate the outlines of higher fidelity without the need to tune any threshold value. To distinguish these attacks from cryptanalysis, we name them as sketch attack.

Although Minemura *et al.*'s [9], [10] method achieved promising results, the features are limited to DCT coefficients in JPEG. However, the structure of H.264/AVC differs significantly from that of JPEG, including transformation block size and intra prediction. In addition, inter frame is deployed to significantly reduce temporal redundancy in video compression by performing motion compensation, where its coefficients (holding the residual information) are small in magnitude or of zero value. Therefore, the features considered in the conventional sketch attacks may not be directly applicable to handle encrypted H.264/AVC compressed video, especially for the case of inter frame. Furthermore, the components in H.264/AVC compressed video, including integer transformed coefficients, quantization parameter, intra-prediction mode, motion vector, and so on, which collectively affect the number of bits required to encode a macroblock (MB), remain unexplored for the purpose of sketch attack.

In this paper, a new sketch attack based on macroblock bitstream size (MBS) is proposed to generate the outline of the original video directly from its encrypted counterpart. The rest of this paper is organized as follows. Section II briefly reviews the H.264/AVC coding standard, summarizes the conventional sketch attack methods, and dissects the conventional format-compliant video encryption methods. In Section III, a novel sketch attack is proposed to generate the outline of H.264/AVC compressed video using the MBS, followed by the extension of an existing evaluation method to evaluate the fidelity of sketch video. Section IV presents the experiment results and further discusses about the proposed MBS sketch attack. Finally, Section V summarizes this paper and comments on our future directions for this research.

II. Related Works

In this section, we briefly review the H.264/AVC coding standard, summarize the conventional sketch attacks, and dissect the existing format-compliant video encryption methods.

A. H.264 Advance Video Coding

Fig. 1 shows the H.264/AVC [1] encoder. Each video consists of a sequence of frames and each frame is classified as intra frame (I-frame), predictive frame (P-frame), or bipredictive frame (B-frame), depending on the order in which it appears. Each frame is further partitioned into nonoverlapping 16×16 pixels MB for further processing. Each I-frame is coded independently, while P-frame can make the decision whether to code an MB independently or to refer an MB in the previously coded I- or P-frame. B-frame is similar to P-frame, but it can also refer to an MB in the future I- or P-frame.

In I-frame, each MB is further divided into nonoverlapping square blocks of various sizes (i.e., one 16×16, four 8×8, or sixteen 4×4 pixel blocks) depending on its spatial activity. The pixel values of each square block are then predicted based on the pixel values from the neighboring blocks (i.e., northwest, north, north-east, and west neighbors) in the same frame. Next, integer DCT (IntDCT) and quantization are performed on each predicted MB. The output is entropy coded using either context adaptive variable length coding (CAVLC) or context adaptive binary arithmetic coding (CABAC). On the other hand, in P- or B-frame, each MB is divided into blocks of various sizes (i.e., 16×16, 16×8, 8×16, 8×8, 8×4, or 4×4) depending on the quarter-pixel motion vector estimation results using single/multiple reference frames. Next, each MB is processed in the same manner as in the I-frame. The output slice is preceded or followed by various markers to form the complete bitstream following the H.264/AVC syntax.

B. Conventional Sketch Attacks

To the best of our knowledge, there are only three publications [6], [10], [11] related to sketch attack on block-transform-compressed images. Specifically, these methods aim to generate the outline of an image compressed by the JPEG standard. In particular, the features of dc and ac coefficients are exploited. These methods are detailed in Sections II-B-1–4.

1) DC Error Category [11]: The array of dc coefficients is essentially the original image, but at the reduced resolution (i.e., 1/8) of the original image in the case of JPEG. Since the dc prediction process in JPEG considers a function similar to that of edge detection, the prediction errors carry some form of edge information. For coding purposes, the prediction errors are grouped into categories depending on their magnitudes; therefore, a sketch ϕ_D can be generated by considering a representative value for each category expressed as follows:

$$\phi_D(i,j) \leftarrow \text{round}\left(255 \times \frac{d(i,j)}{\max\{d(i,j)\}}\right) \tag{1}$$

$$d(i,j) \leftarrow \begin{cases} 2^{\log_2 |r(i,j)|} & \text{if } |r(i,j)| > 0 \\ 0 & \text{otherwise} \end{cases} \tag{2}$$

where $r(i,j)$ is the residue dc value (i.e., prediction error) at the (i,j)th 8×8 block.

2) Improved Nonzero Coefficient Count [10]: The number of nonzero coefficients in each block is exploited to generate

a sketch ϕ_N as follows:

$$\phi_N(i,j) \leftarrow \text{round}\left(255 \times \frac{c(i,j)}{\max\{c(i,j)\}}\right) \quad (3)$$

where $c(i,j)$ denotes the number of nonzero ac coefficients in the (i,j)th 8×8 block. Here, the number of ac coefficients infers the complexity of a block. A larger INCC implies higher spatial activity within the block, and vice versa.

3) Position of Last Nonzero Coefficient [9]: The position (i.e., index with respect to zigzag order) of the last nonzero coefficient is exploited to generate a sketch ϕ_P. Specifically

$$\phi_P(i,j) \leftarrow \text{round}\left(255 \times \frac{p(i,j)}{\max\{p(i,j)\}}\right) \quad (4)$$

where $p(i,j)$ denotes the PLZ ac coefficients in the (i,j)th 8×8 block. Larger PLZ implies more complex pattern (i.e., 2D DCT basis vector) in the block, and vice versa.

4) Sum of Absolute AC Coefficients [10]: The ac coefficients are exploited to generate a sketch ϕ_S as follows:

$$\phi_S(i,j) \leftarrow \text{round}\left(255 \times \frac{s(i,j)}{\bar{s}}\right) \quad (5)$$

where $s(i,j)$ denotes the sum of magnitude of ac coefficients in the (i,j)th block and

$$\bar{s} \leftarrow \frac{1}{64}\sum_{i=1}^{8}\sum_{j=1}^{8} s(i,j). \quad (6)$$

Informally, $s(i,j)$ indicates the strength of the textures/edges, if any, in the (i,j)th block. Therefore, larger $s(i,j)$ suggests stronger edges, and vice versa.

Although the aforementioned sketch attacks are able to generate outline images, they are limited to JPEG compressed image or I-frame of an MPEG-1/2 compressed video. Specifically, the conventional sketch attacks are not directly applicable to H.264/AVC compressed videos due to the intra-prediction process [1], where the quantized prediction errors are encoded instead of the coefficient values themselves as considered in MPEG-1/2. Furthermore, even when the conventional sketch attacks can reveal some outline of the I-frame, they are not viable in handling inter frame (i.e., P- and B-frames), because the residual information (i.e., prediction error during motion compensation) is limited and does not correlate to the outline of the original frame directly. Moreover, sketching only the I-frame is insufficient, especially when the group of picture size of the video is large or when there are significant scene changes within the inter frames that end before encountering the next I-frame. Hence, in this paper, we propose a new sketch attack that is capable in sketching the outline of an H.264/AVC compressed video regardless of the frame type and its form (i.e., plaintext or encrypted video).

C. Conventional Format-Compliant Video Encryption Methods

To achieve format-compliant video encryption, nine elementary operations (hereinafter referred to as modules) are commonly considered to handle the components in H.264/AVC compressed video [2]. Conventionally, these elementary operations are selectively considered to achieve the desired effects and properties. Sections II-C1)–II-C9) briefly describe each of these elementary operations.

TABLE I
FORMAT-COMPLIANT ENCRYPTION MODULES FOR H.264/AVC VIDEO

	S	L	ST	Q	DF	MVD	SSO	Inter	Intra
Ref [12]	×	×	×	×	×	✓	×	✓	✓
Ref [13]	✓	✓	×	×	×	✓	×	×	✓
Ref [14]	×	×	×	×	×	✓	✓	×	✓
Ref [15]	×	✓	×	×	×	✓	×	×	×
Ref [16]	×	×	✓	×	×	×	×	×	×
Ref [17]	×	×	×	✓	✓	×	×	×	✓
Ref [18]	✓	×	×	×	×	✓	×	×	✓
Ref [19]	×	×	✓	×	×	×	×	×	×

1) IntDCT Coefficient Sign Randomization (S) [13]: This approach randomly flips the sign of each coefficient. It maintains the bitstream size of the original input video when CAVLC is in use.

2) IntDCT Coefficient Level Modification (L) [15]: This approach modifies the nonzero coefficients prior to entropy coding. In particular, the magnitude of the coefficients is modified.

3) Secret Transformation [16]: This approach randomly selects either IntDCT or a different 4×4 integer transformation function with similar property as that of the 4×4 IntDCT.

4) Quantization Parameter Manipulation (Q) [17]: Quantization parameter that defines the quantization step size is modified.

5) Deblocking Filter Value Modification [16]: The deblocking filter (DF) is utilized prior to motion estimation and compensation. The global DF value is modified to severely distort the quality of the H.264/AVC compressed video frame.

6) MV Difference Modification [14]: The motion vectors and their differences are modified.

7) Secret Scan Order [14]: Instead of using the standard zigzag order designed for 4×4 block coefficients, a secret scan order is considered.

8) Inter-Prediction Mode Modification (Inter) [12]: The set of inter-frame MBs with the same structure is permuted. Here, structure refers to the subdivisions/sub-MB (e.g., number of 8×8, 8×16, 16×8, and 16×16 blocks) and the number of motion vectors within an MB.

9) Intra-Prediction Mode Modification (Intra) [12]: Intra prediction is performed prior to transformation, where the prediction mode can be modified accordingly to distort the video.

Table I marks the modules involved in the construction of the conventional format-compliant video encryption methods. Here, if row α and column β are marked with ✓, it signifies that method α performs module β as part of its encryption process, and × signifies otherwise. For the proof of concept, in this paper, we consider two recently published encryption methods for H.264/AVC video, namely, [18] and [19], because they had been shown to be secure against cryptanalytic attacks while achieving minimal bitstream size expansion.

III. PROPOSED SKETCH ATTACK

This section first identifies the shortcomings of the conventional sketch attack methods, then puts forward a new method

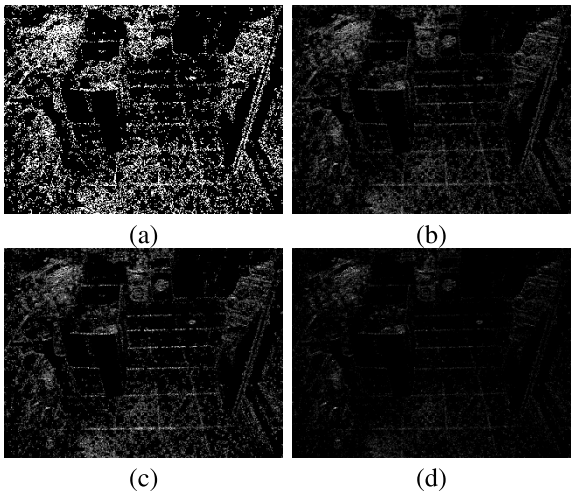


Fig. 2. Outline of the original I-frame (in Video 17 from ICDAR2013) sketched by using (a) DCEC, (b) INCC, (c) PLZ, and (d) SAC.

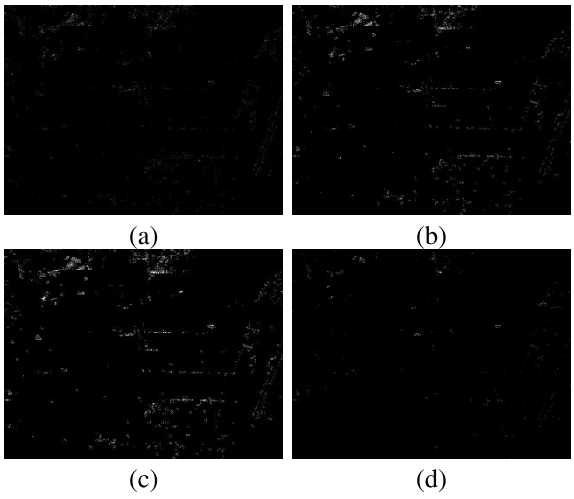


Fig. 3. Outline of the original P-frame (in Video 17 from ICDAR2013) sketched by using (a) DCEC, (b) INCC, (c) PLZ, and (d) SAC.

to sketch the outline of the original (plaintext) video directly from the encrypted (ciphertext) video based on the bitstream size of MB. Next, it formulates an evaluation criteria to assess the effectiveness of the proposed sketch attack.

A. Limitation of the Conventional Sketch Attacks

In order to design an effective sketch attack, first, we sketch directly from an H.264/AVC compressed video instead of an encrypted video. Specifically, we applied the conventional sketch attacks to handle H.264/AVC compressed video. Fig. 2 shows the sketches obtained by applying the conventional methods, namely: dc error category (DCEC), INCC, PLZ, and sum of absolute ac coefficients (SACs),[1] on the first frame of an H.264/AVC compressed video. For comparison purpose, the original frame is shown in Fig. 4(a). All four conventional methods can generate an outline image for the first frame (i.e., I-frame). On the other hand, Fig. 3 shows the

[1]We modified (5) by replacing $\bar{s}$ in the denominator by $\max\{s(i, j)\}$, because it is empirically found to be generating better sketch image.

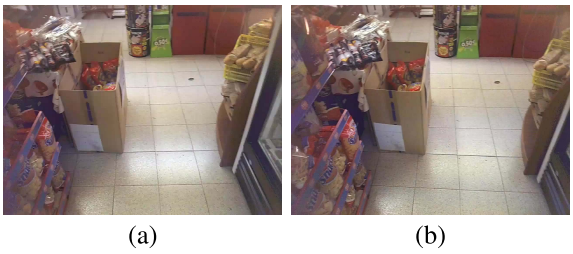


Fig. 4. Original first (I-) and third (P-) frames in Video 17 from ICDAR2013. (a) First frame. (b) Third frame.

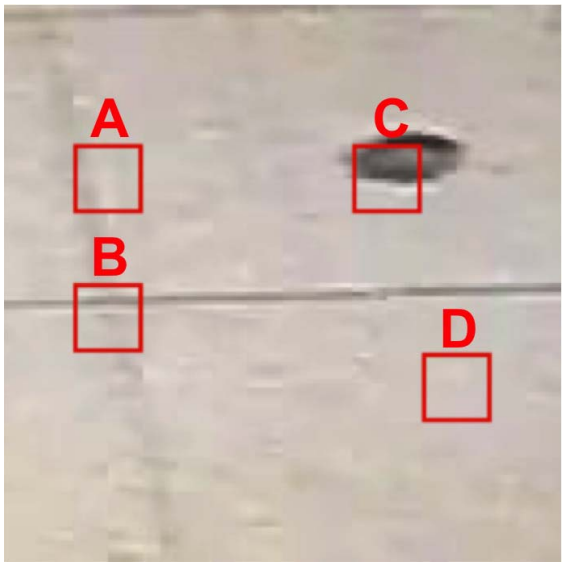


Fig. 5. Four MBs (regions) with different spatial activities in a video frame compressed by using H.264/AVC.

sketches generated by applying the same conventional methods to the third frame (i.e., P-frame) from the same video. For comparison purpose, the original frame is shown in Fig. 4(b). Since these methods are designed to attack still images, as expected, they fail to generate a perceptually meaningful sketch image when handling an inter frame, such as P- or B-frame. It is because the I-frame is encoded independently of the rest of the frames (similar to the scenario of still image), while the P- or B-frame is encoded based on the differences between the current and previous/future frames through motion estimation. In addition, the conventional methods are less effective in sketching due to the introduction of various components in the H.264/AVC compression standard (e.g., IntDCT for H.264/AVC, various MB types, and quarter-pixel motion compensation) when compared with JPEG [20].

B. Sketch Attack Using Macroblock Bitstream Size

In this section, we propose a sketch attack to generate the outline of H.264/AVC compressed video. To facilitate the discussion and without loss of generality, we consider H.264/AVC compressed video in level 5.1 and baseline profile (BP), which utilizes only one slice within a frame, 4×4 IntDCT, and seven block sizes, namely, 16×16, 16×8, 8×16, 8×8, 8×4, 4×8, and 4×4.

H.264/AVC codes the coefficients in an MB by using either CAVLC or CABAC. It is observed that an MB with higher

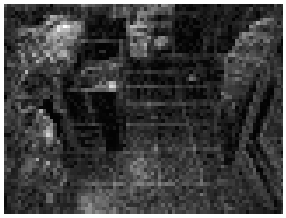


Fig. 6. Outline of the original P-frame sketched by using MBS (#1).

spatial activity requires larger number of bits for encoding purposes, and vice versa. In other words, the bitstream size of an MB infers its complexity, and it is an intrinsic property of H.264/AVC compressed video that varies according to the video content. For example, Fig. 5 shows four MBs (labeled as A, B, C, and D) in a 128×128 pixels region of an H.264/AVC compressed video frame. Each marked MB shows different spatial activities. Specifically, region A consists of a weak diagonal edge and is allocated 80 bits. On the other hand, region B consists of borders of the floor tiles (i.e., edges) and is allocated 208 bits. Meanwhile, region C contains two areas of different colors along with an edge separating them, and is allocated 146 bits. Finally, region D is smooth, where it has no edges and is allocated merely 24 bits. Based on these observations, we propose a sketch attack as follows:

$$\phi_B(i, j) \leftarrow \text{round}\left(255 \times \frac{b(i, j)}{\max\{b(i, j)\}}\right) \tag{7}$$

where $b(i, j)$ denotes the number of bits spent on coding the (i, j)th MB. Note that MBS considers the ratio of $b(i, j)$ over $\max\{b(i, j)\}$ regardless of the frame type. Fig. 6 shows the sketch generated by (7), where the resolution is $1/16$ of its original counterpart.

C. Evaluation of Sketch Images

Li and Yuan [6] proposed a method to evaluate the output generated by their sketch attack by using the Sobel edge image [21] of the original image as the ideal outline image. Specifically, Otsu thresholding [22] is deployed to obtain the binary edge image. Since Li and Yuan's [6] evaluation method is designed to map one output value for every 8×8 pixel block and therefore is not relevant for H.264/AVC, we extend their evaluation framework to cater for different block sizes by considering two steps: 1) binary edge image definition and 2) edge similarity evaluation.

1) Definition of Reference Binary Edge Image: There are many edge detectors in the literature, including Canny [23], Sobel [21], and Laplacian filter [21]. Here, we adopt the Canny operator as the edge detector (instead of Sobel operator used by Li and Yuan [6]), because it employs hysteresis thresholding, which assumes an edge as continuous curve and it can detect some weak parts of an edge. We utilize the output B generated by the Canny operator as the reference edge image for evaluation. Since a sketch ϕ is of size $M/k \times N/k$ for k being the size of a block or MB, ϕ is smaller than the Canny edge map B, which is of size $M \times N$. As such, the Canny binary edge map needs to be resized to match the dimension of the sketch image. To generate the downscaled Canny binary edge map ϵ^k, we first divide the Canny binary edge map into nonoverlapping blocks each of $k \times k$ pixels. Next, the number of edge candidates for each block is calculated as follows:

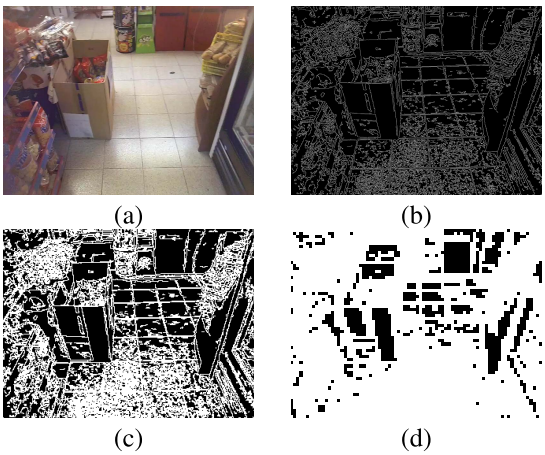


Fig. 7. Illustration of Canny edge map and downscaled Canny edge maps. (a) Original. (b) Canny. (c) Downscaled ($k = 4$). (d) Downscaled ($k = 16$).

$$c(i, j) \leftarrow \sum_{u=1}^{k} \sum_{v=1}^{k} B_{u,v}^{k}(i, j) \tag{8}$$

where $B_{u,v}^{k}(i, j)$ denotes the (u, v)th value in the (i, j)th $k \times k$ block, where $1 \le i \le M/k$, $1 \le j \le N/k$, and $1 \le u, v \le k$.

To generate the downscaled Canny binary edge map $\epsilon^k(i, j)$, the number of edge candidates for each block is classified as follows:

$$\epsilon^k(i, j) \leftarrow \begin{cases} 1 & \text{if } c(i, j) > 0 \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

The original image, Canny edge map, and downscaled Canny edge maps are shown in Fig. 7.

2) Edge Similarity Score: Li and Yuan [6] proposed an evaluation method for binary sketch images defined as follows:

$$\zeta_0 \leftarrow \frac{n_1}{2B} + \frac{n_2}{2W} - \frac{n_3}{2B} - \frac{n_4}{2W} \tag{10}$$

where B and W denote the number of 0's and 1's in the sketch image ϵ^k, respectively. In addition

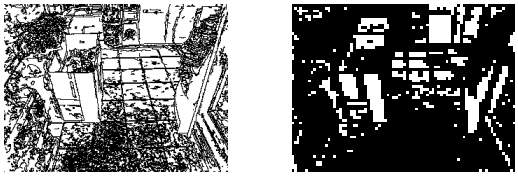
$$n_1 \leftarrow |\{(i, j) : \epsilon^k(i, j) = 0 \ \& \ \phi(i, j) = 0\}| \tag{11}$$

$$n_2 \leftarrow |\{(i, j) : \epsilon^k(i, j) = 1 \ \& \ \phi(i, j) = 1\}| \tag{12}$$

$$n_3 \leftarrow |\{(i, j) : \epsilon^k(i, j) = 0 \ \& \ \phi(i, j) = 1\}| \tag{13}$$

$$n_4 \leftarrow |\{(i, j) : \epsilon^k(i, j) = 1 \ \& \ \phi(i, j) = 0\}|. \tag{14}$$

In particular, Li and Yuan's [6] method behaves as follows: 1) $\zeta = 0$ when the sketch image is entirely black (0s) or white (i.e., 1s); 2) $\zeta = 0$ when the sketch image is random; 3) $\zeta = 1$ when the sketch image is exactly the same as the downscaled reference edge (binary) image (in our case, the reference binary image is a Canny edge map); and 4) $\zeta = -1$ when the sketch image is exactly the negated downscaled reference edge image.



(a) (b)

Fig. 8. Illustration of negated downscaled Canny edge maps. (a) Negated downscaled ($k = 4$). (b) Negated downscaled ($k = 16$).

Fig. 8(a) and (b) shows the negated downscaled edge images shown in Fig. 7(c) and (d), respectively. It is observed that both outlines are similar. However, since we are only interested in the edges regardless of whether the sketch is positively or negatively correlated, the absolute value of (15) is considered in this paper. That is, we consider the edge similarity score (ESS) ζ as follows:

$$\zeta = |\zeta_0| = \left| \frac{n_1}{2B} + \frac{n_2}{2W} - \frac{n_3}{2B} - \frac{n_4}{2W} \right|. \tag{15}$$

Note that $\zeta \in [0, 1]$, where any value nearer to unity indicates a better match between the sketch image and the downscaled Canny edge map, and vice versa. Therefore, with the assumption that the Canny edge map is the ideal outline, a larger value implies better sketch quality, and vice versa. Fig. 9 shows the MBS sketch images and their corresponding ESS. It is observed that MBS can sketch the outline of both steady (upper row, i.e., Video 17 in [24]—browsing products in a super market) and nonsteady (lower row, i.e., Video 20 in [24]—captured from driver's seat) videos. The nonsteady video yields higher ESS most of the time when compared with the steady video, because there are more scene changes in the nonsteady video, and hence, more bits are spent on coding the MBs, which can be exploited by MBS to generate better outline. Fig. 9 also suggests that blurred sketches yield lower ESS scores, and vice versa.

IV. Experiments and Discussion

To verify the viability of our proposed sketch attack, six video sequences, namely: two video sequences from the ICDAR2013 video data set [24], two video sequences from ultra-HD High Efficiency Video Coding (UHD HEVC) dash data set [25], and two video sequences from Xiph.org video test media (HD content and above) [26], are considered. The resolution of the ICDAR2013 videos is 1280×960 pixels, while the resolution of the HEVC dash and Xiph.org video test media videos ranges from 1920×1080 to 3840×2160 pixels. The first 30 frames of each test video are encoded into H.264/AVC format using the BP with level 5.1 and then partially decoded for sketching the outline. Unless specified otherwise, the initial quality parameter (QP) in H.264/AVC is set at 30. In particular, the proposed and conventional methods are implemented on top of the H.264/AVC reference softwares [27] for partial decoding purpose, while the MATLAB is utilized to generate the sketch images. Here, when we encounter an MB of type PCM, the value of the second largest MBS within the same frame will be an output. To compute ESS, Otsu thresholding [22] is applied to the obtained sketch image for generating the final binary outline image.

TABLE II

ESS of Sketched Frames for H.264/AVC I-Frame With Initial QP = 20, 30, and 40

Dataset	Sequenc	QP	DCEC	INCC	PLZ	SAC	MBS
ICDAR 2013	Video 17 (1280×960)	20	0.1884	0.4062	0.3595	0.2140	**0.4595**
		30	0.2194	0.1531	0.2664	0.1684	**0.3406**
		40	0.0436	0.0345	0.0345	0.0345	**0.3162**
	Video 20 (1280×960)	20	0.1851	0.5795	**0.6194**	0.2707	0.5210
		30	0.1665	0.3527	0.3686	0.2193	**0.4352**
		40	0.1186	0.1233	0.1233	0.1233	**0.4524**
HEVC dash	v5 (1920 × 1080)	20	0.0920	0.4653	0.4881	0.2092	**0.7494**
		30	0.1340	0.2426	0.2946	0.1491	**0.5001**
		40	0.0947	0.0394	0.0750	0.0409	**0.4549**
	v9 (3840 × 2160)	20	0.0901	0.4912	0.5660	0.1559	**0.8062**
		30	0.1032	0.2346	0.2664	0.1033	**0.4209**
		40	0.0650	0.0608	0.0464	0.0608	**0.3780**
Xiph	Old town cross (3840×2160)	20	0.2503	0.3228	0.2748	**0.3798**	0.3299
		30	0.2910	0.3576	0.2910	0.1490	**0.6857**
		40	0.0421	0.0204	0.0531	0.0204	**0.3989**
	Rush hour (1920×1080)	20	0.1413	0.1772	0.1090	0.1076	**0.3853**
		30	0.1235	0.0606	0.0606	0.0606	**0.4914**
		40	0.0162	0.0035	0.0169	0.0035	**0.1876**

A. Nonencrypted Video

Since an H.264/AVC compressed video consists of two types of frame (i.e., I-frame and inter frame), we consider them separately. Here, the nonencrypted video (i.e., the best case scenario) is considered to show the performance of the proposed sketch attack. This test case is crucial, because if the proposed sketch attack cannot even sketch the outline from the plaintext video, it will not be able to sketch from the encrypted video. More importantly, since sketch attack exploits the features that stay intact before and after encryption (i.e., complexity of each MB), a sketch obtained from the plaintext video will not differ significantly from that obtained from the encrypted video if the complexity features are similar in both videos.

1) Intra Frame: Table II records the ESS values when applying five sketch attacks on six videos (i.e., two videos from every data set, three data sets in total) for the case of I-frame. Here, the best ESS result among five sketch attacks is bolded for quick lookup purpose. Results suggest that, when handling I-frame, the ESS value for all sketch attacks decreases when QP increases regardless of the video in question, except for old town cross and rush hour. In addition, MBS yields the best results for QP $\geq$ 30. Notably, the ESS values attained by MBS are particularly high when compared with all considered conventional methods for QP = 40 (i.e., the low bitrate case). It is because most coefficients (both ac and dc) are quantized to zeros, and hence, DCEC, INCC, PLZ, and SAC are ineffective in extracting information from the compressed video. On the other hand, MBS appears to be robust against heavy compression (i.e., using large QP value) as suggested by the higher ESS scores in Table II. Therefore, we conclude that, when handling I-frame, MBS is able to sketch at a quality comparable with that of the conventional sketch methods for smaller QPs, and further outperforms the rest for QP $\geq$ 30.

In addition, the quality of the sketch images improves when the resolution increases. The trend is confirmed by considering the same video frame at different resolutions, i.e., 1920×1080 and 3840×2160 pixels [see Fig. 10(a) and (b)]. It is an expected result, because more MBs are needed to code the

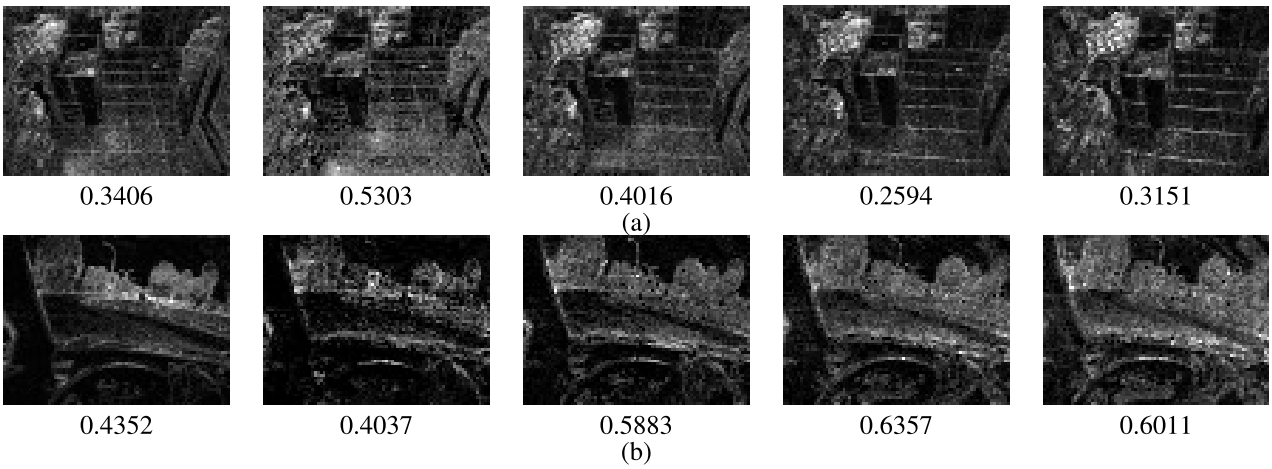


Fig. 9. Sketched outline using MBS for frames #1 to #5 and the corresponding ESS. (a) Video 17 in ICDAR2013. (b) Video 20 in ICDAR2013.

TABLE III
AVERAGE ESS OF SKETCHED FRAMES FOR H.264/AVC INTER FRAME WITH INITIAL QP = 20, 30, AND 40

Dataset	Sequenc	QP	DCEC	INCC	PLZ	SAC	MBS
ICDAR 2013	Video 17 (1280×960)	20	0.0420	0.1987	0.2044	0.0886	**0.4220**
		30	0.0432	0.1779	0.1871	0.0808	**0.4115**
		40	0.0376	0.1372	0.1442	0.0684	**0.3841**
	Video 20 (1280×960)	20	0.0235	0.0330	0.0177	0.0444	**0.5490**
		30	0.0219	0.0317	0.0162	0.0421	**0.5469**
		40	0.0206	0.0302	0.0178	0.0396	**0.5298**
HEVC dash	v5 (1920×1080)	20	0.0001	0.0005	0.0004	0.0002	**0.2636**
		30	0.0003	0.0006	0.0006	0.0003	**0.3061**
		40	0.0012	0.0018	0.0016	0.0009	**0.3357**
	v9 (3840×2160)	20	0.0002	0.0006	0.0004	0.0004	**0.5671**
		30	0.0003	0.0007	0.0006	0.0004	**0.5668**
		40	0.0019	0.0027	0.0027	0.0025	**0.5496**
Xiph	Old town cross (3840×2160)	20	0.0293	0.0288	0.0139	0.0306	**0.4407**
		30	0.0205	0.0210	0.0136	0.0234	**0.4059**
		40	0.0202	0.0267	0.0202	0.0261	**0.3940**
	Rush hour (1920×1080)	20	0.1156	0.0876	0.1223	0.0711	**0.3337**
		30	0.1052	0.0834	0.1163	0.0623	**0.3321**
		40	0.0975	0.0787	0.1051	0.0566	**0.3366**

same semantic of the frame, hence generating a sketch image of higher resolution. When scaled to the same size for display, the sketch of higher resolution will show greater detail, and vice versa.

2) Inter Frame: Table III records the average ESS values for the case of inter frame. Results suggest that MBS always yields, by far, the highest ESS regardless of the video and QP value under consideration, which show the superiority and capability of MBS in extracting information from inter frame, where much of the redundancy is removed. Hence, the performance of the proposed MBS sketch attack is consistent and more robust to frame type and QP parameter when compared with those of the conventional sketch attacks.

B. Format-Compliant Encrypted Video

In this section, we discuss the viability of the proposed sketch attack for two recently published format-compliant encryption methods [18], [19]. Due to the low performance of DCEC, as suggested in Section IV-A, only four sketch attacks are, henceforth, considered, namely: INCC, PLZ, SAC, and MBS.

Specifically, 30 frames from the videos in the ICDAR2013 data set are utilized as the representative test video sequences and they are encrypted by using [18] and [19]. Figs. 11–14 show the graph of ESS for INCC, PLZ, SAC, and MBS, respectively.

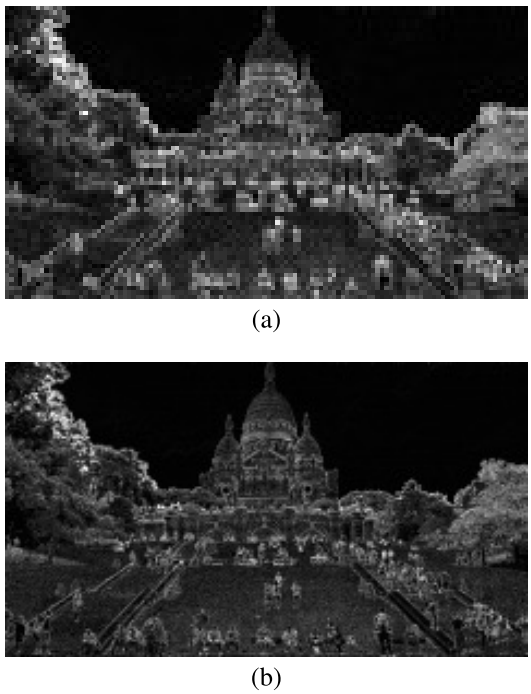


Fig. 10. Outline frames sketched by MBS for the same video (UHD HEVC dash data set) with two different resolutions. (a) Sketch generated from original video frame with a resolution of 1920 × 1080 pixels. (b) Sketch generated from original video frame with a resolution of 3840 × 2160 pixels.

Here, the not encrypted curve refers to the ESS score attained when comparing the ground truth (i.e., Canny binary edge map) with the outline generated by the respective sketch attacks on the plaintext (i.e., not encrypted) frames. Based on these graphs, the ESS value for Wang *et al.*'s [18] encryption method is almost identical to that of the original plaintext video. This trend suggests that, Wang *et al.* [18] fail to mask the complexity of each MB, which could be exploited to generate sketches of the original video. On the other hand, for all sketch attacks considered, Zeng *et al.*'s [19] method yields

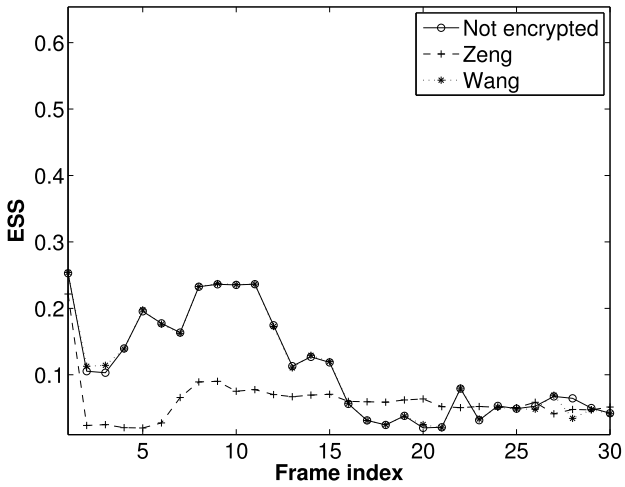


Fig. 11. ESS of INCC for various encryption methods.

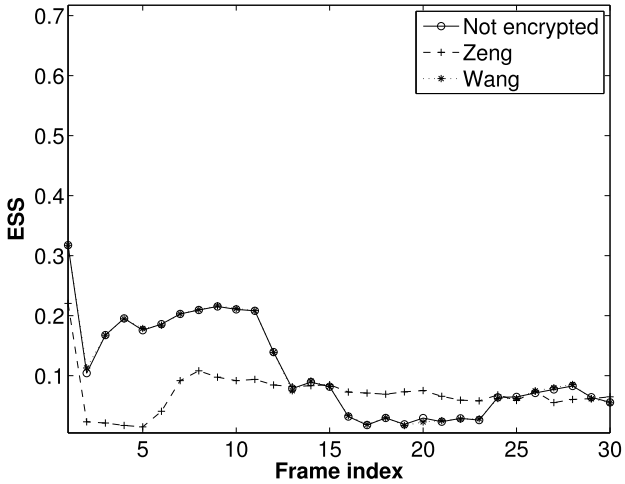


Fig. 12. ESS of PLZ for various encryption methods.

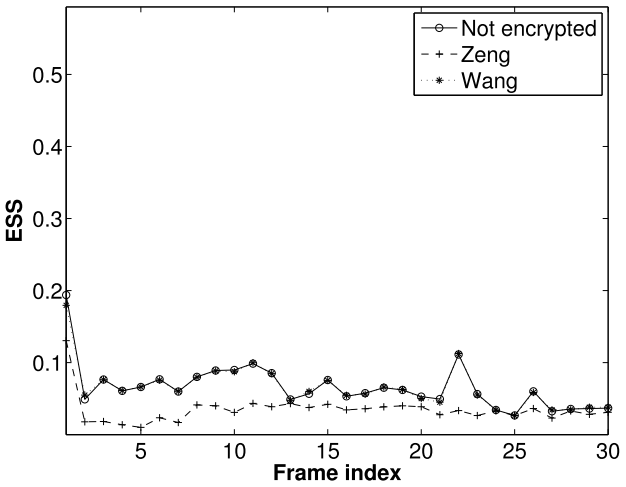


Fig. 13. ESS of SAC for various encryption methods.

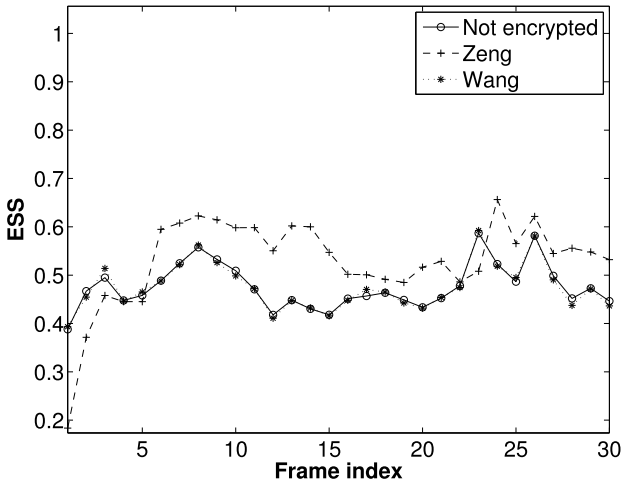


Fig. 14. ESS of MBS for various encryption methods.

different trends, including both higher and lower ESS values when compared with those of the original plaintext videos. A possible reason to these trends is that Zeng *et al.*'s [19] encryption method changes the number of coefficients, level of coefficients, and position of coefficients and hence the MBS. More importantly, it is observed that the ESS values vary in a similar manner for all sketch attacks considered, but nonetheless, the proposed MBS yields the highest ESS scores.

Next, the first five frames of a representative steady video and a representative nonsteady video (i.e., Video 17 and Video 20 from ICDAR2013, respectively) are encrypted by using Zeng *et al.*'s [19] method, and the corresponding sketch images are shown in Figs. 15 and 19. It is observed that all considered sketch attacks are able to sketch the first (i.e., I-frame) frame, but only MBS is able to sketch the inter frames. Moreover, the results shown in Fig. 14 suggest that the MBS yields stable scores across the inter frames.

In addition, Figs. 16 and 17 show the graph of ESS against QP for all sketch attacks when considering Video 17 from ICDAR2013 [24] encrypted by Zeng *et al.*'s [19] method for the I-frame and inter frame, respectively. In the case of I-frame, the ESS of MBS is always higher than other sketch attacks for all QPs considered, although the differences are smaller for QP between 25 and 35. On the other hand, for inter frame, the ESS of INCC, PLZ, and SAC dropped drastically due to significant reduction in information during video encoding, but MBS still yields stable performance with high ESS values. Similar results are achieved when considering other test videos. Therefore, it is concluded that the MBS can be considered as a practical sketch attack method regardless of frame type and QP.

TABLE IV
SPECIFIC FEATURES EXPLOITED BY EACH SKETCH ATTACK

	Level_{DC}	Level_{AC}	Coef. number	Coef. position	MB bits
DCEC [11]	✓	×	×	×	×
INCC [10]	×	×	✓	×	×
PLZ [10]	×	×	×	✓	×
SAC	×	✓	×	×	×
MBS	×	×	×	×	✓

C. Analysis

To summarize the sketch attacks, the features exploited by each sketch attack are listed in Table IV. In particular, Level_{dc} denotes the residual dc values, Level_{ac} refers to the quantized

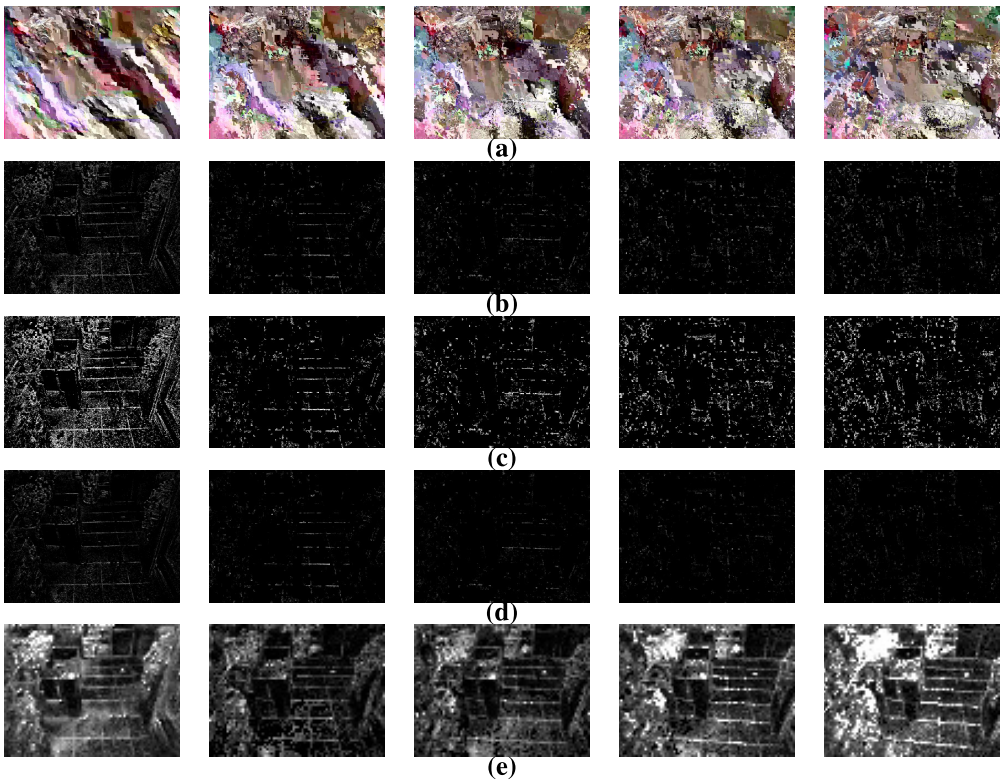


Fig. 15. Video 17 in ICDAR2013 (first five frames) encrypted by using [19] and the corresponding sketched frames (second to fifth rows). (a) Encrypted. (b) INCC. (c) PLZ. (d) SAC. (e) MBS.

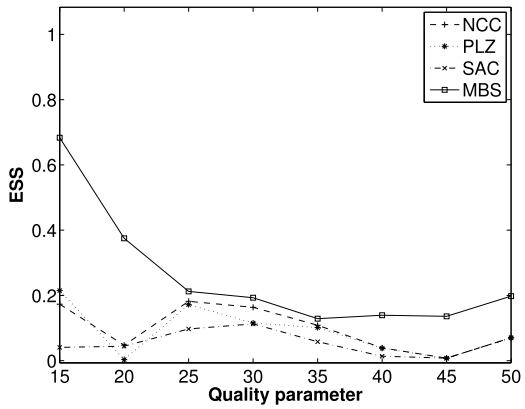


Fig. 16. Graph of ESS versus QP for various sketch attacks in the I-frame.

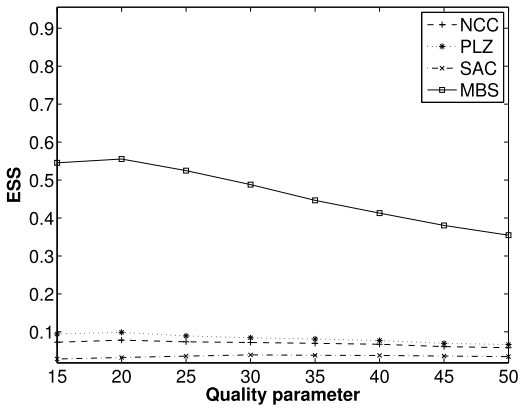


Fig. 17. Graph of ESS versus QP for various sketch attacks in the inter frame.

ac coefficient values, Coef. number indicates the number of nonzero ac coefficients of each block, Coef. position is the PLZ ac coefficient in the order of zigzag in each block, and MB bits denotes the number of bits allocated to each MB. It is observed that the coefficient-based sketch attacks, i.e., INCC, PLZ, and SAC, fail to sketch the outline of the frame when the coefficients are modified. On the other hand, MBS is still viable in sketching the outline of the frame even when the coefficients are modified.

Table V records the viability of each sketch attack in generating the outline directly from the encrypted H.264/AVC video. Here, the first row is interpreted in the same manner as in Table I. If row α and column β are marked with ✓ (×), it means that the sketch attack method α can (cannot) generate the outline of the video when module β is implemented into the video encryption method. Results suggest that the MBS can generate the outline images directly from the encrypted video except when IntDCT coefficient

TABLE V

VIABILITY OF SKETCH ATTACKS FOR VARIOUS H.264/AVC FORMAT-COMPLIANT VIDEO ENCRYPTION MODULES

	S	L	ST	Q	DF	MVD	SSO	Inter	Intra
DCEC [11]	✓	×	×	✓	✓	✓	✓	×	✓
INCC [10]	✓	✓	×	✓	✓	✓	✓	×	✓
PLZ [10]	✓	✓	×	✓	✓	✓	×	×	✓
SAC	✓	×	×	✓	✓	✓	✓	×	✓
MBS	✓	×	✓	✓	✓	✓	✓	✓	✓

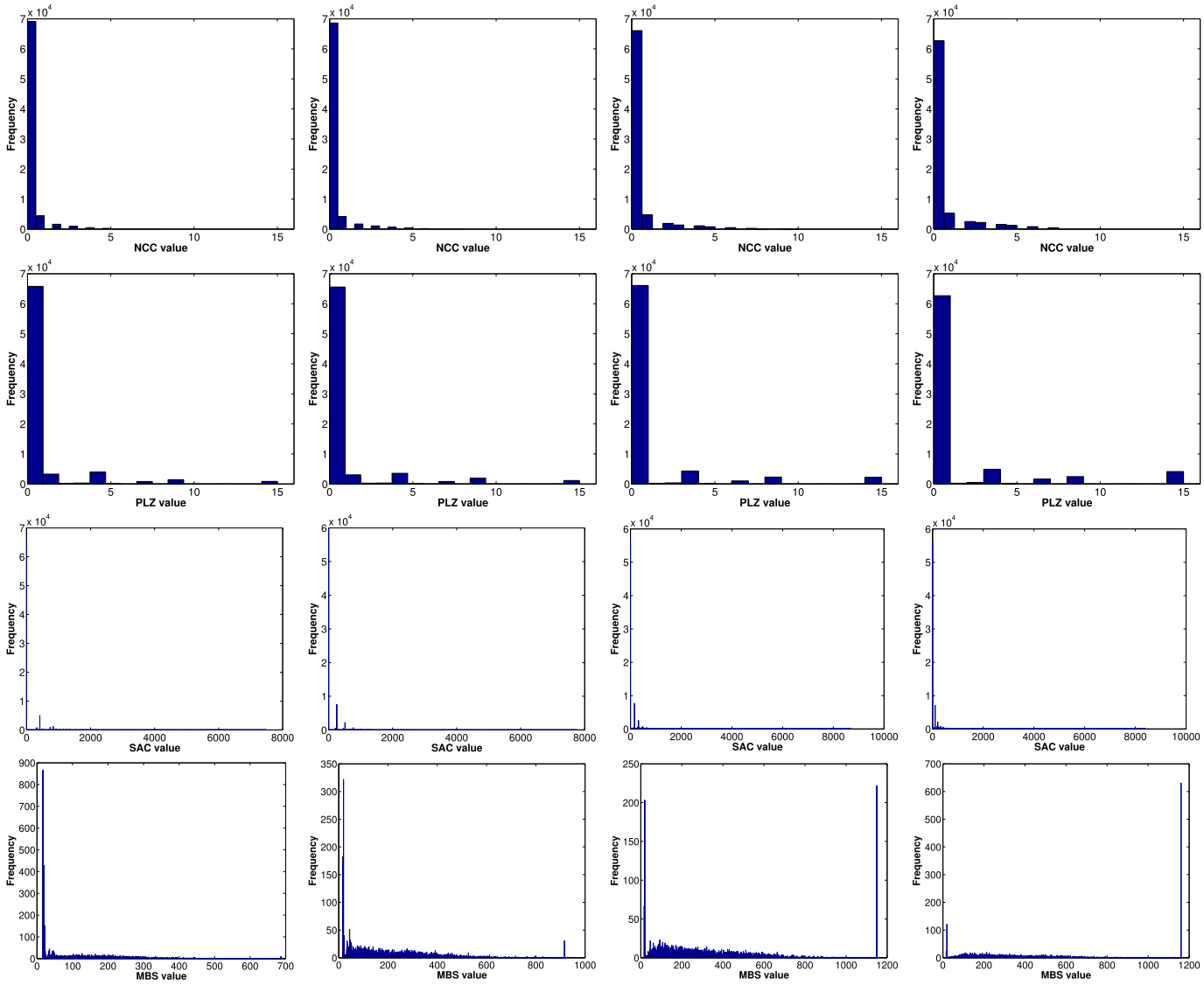


Fig. 18. Histogram of each sketch attack method before applying Otsu's binarization for frames #2 to #5 shown in Fig. 15.

level modification is involved. However, INCC and PLZ are robust to IntDCT coefficient level modification. Therefore, when some information about the encryption method in use is available, we can launch appropriate attacks to obtain better sketch images instead of using one attack. For example, the encryption modules in use can be determined (i.e., in use or not) directly by analyzing the encrypted video. Statistical analysis can be performed on the distribution of coefficient values, motion vectors, scanning orders, and so on, and then launch the appropriate attack, but it is beyond the scope of this paper and we shall pursue in this direction as our future work.

On the other hand, since the proposed and conventional sketch attacks exploit the complexity of MBs to sketch the outline of the original video, each method can be treated as a specific approach to approximate complexity by considering the specific entity of the H.264/AVC compressed video in question. To clearly differentiate the information/statistics captured by each sketch attack, we consider the histogram of the captured values. Fig. 18 shows the histograms of complexity feature values collected by using INCC, PLZ, SAC, and MBS for the plaintext video—Video 17 from ICDAR2013 for frames #2 to #5. Note that frame #1 (i.e., I-frame) is not considered, because INCC, PLZ, SAC, and MBS yield similar ESS values when handling I-frame albeit MBS consistently achieves the best result for QP $\geq$ 30.

When considering INCC and PLZ, the theoretical range of bin values is $[0, 16]$, and the actual observed ranges are smaller, e.g., $[0, 15]$ in the case of PLZ for frame #5. This small range of values has similar effect of displaying a frame/image at reduced bit depth, while the range for SAC bins is significantly wider. However, the values of SAC are mostly concentrated at the zeroth bin (i.e., low contrast) due to the high compression efficiency of H.264/AVC.

On the other hand, among all the histograms considered, only the distribution of MBS histogram is wide and spreads across a large number of bins (i.e., higher contrast). These phenomena suggest that the MBS can generate the outlines of higher entropy (i.e., greater detail) when compared with the conventional methods.

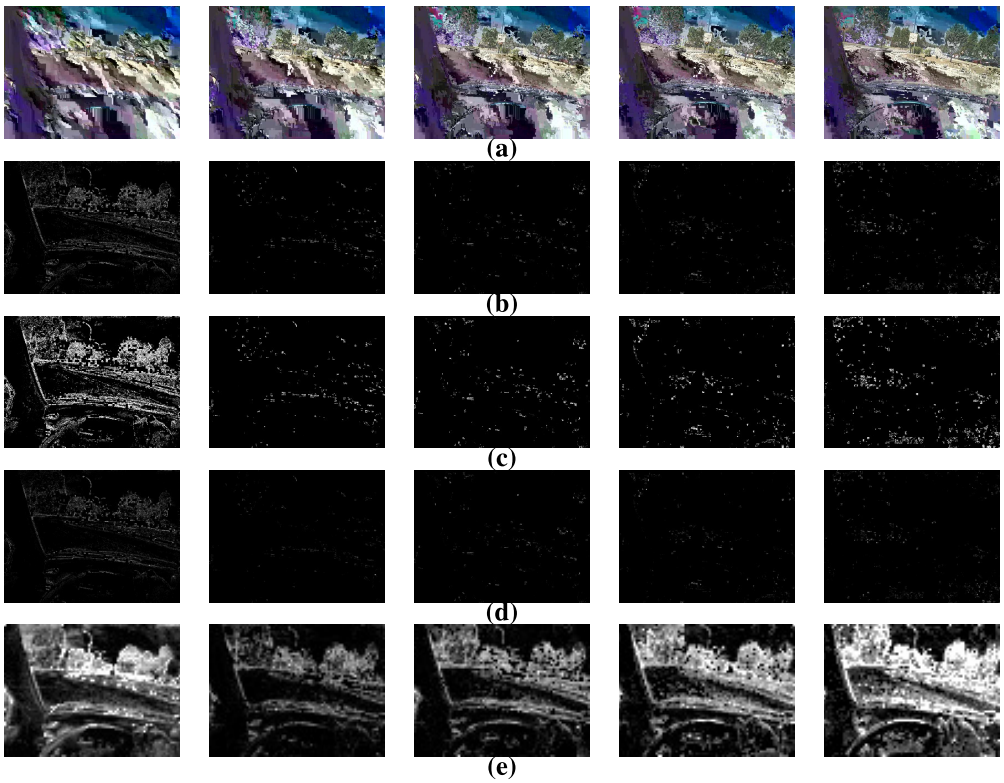


Fig. 19. Video 20 in ICDAR2013 (first five frames) encrypted by using [19] and the corresponding sketched frames (second to fifth rows). (a) Encrypted. (b) INCC. (c) PLZ. (d) SAC. (e) MBS.

Moreover, recall that, to compute ESS, the INCC, PLZ, SAC, and MBS output images are further processed by Otsu's binarization method, which assumes that an image contains two classes of pixels following bimodal histogram to find the optimum threshold value. Since the distribution of INCC, PLZ, and SAC is mostly concentrated at the zeroth bin, most pixel values will be treated as background, i.e., intensity of zero. These observations are supported by the output shown in Figs. 15 and 19.

D. Is MBS Viable?

In Section IV-C, MBS was empirically found to be effective in sketching the outline of both I-frame and inter frame encrypted by the conventional H.264/AVC format-compliant encryption modules. However, since the proposed MBS sketch attack relies on the local complexity information derived from each MB (viz., nonoverlapping the 16×16 pixels block before compression), a possible approach to resist the proposed MBS sketch attack is to permute the MBs while ensuring format compliance. It is because each MB corresponds to a pixel in the sketch generated by MBS, and when the MBs are permuted, the pixels in the output sketch will also be permuted. Hence, the information on local complexity of the corresponding plaintext video will not be leaked. However, in the case of permuting the 16×16 pixel blocks before or during video compression, researchers refrained from deploying the permutation operation, because the correlations among MBs become weak after permutation, which leads to significant bitstream size increment. In other words, the compression efficiency is compromised. Note that additional operations dealing with the pixels values (after decoding the encrypted video) must be included to complete the decryption process, hence making this encryption approach less practical. On the other hand, to permute MBs during or after video compression, the flexible macroblock ordering (FMO), which is included in the H.264/AVC standard as an error resilient tool, can be exploited. Specifically, FMO allows a video frame to be arranged in different scan patterns of MBs, where six built-in scan patterns as well as one option to include an entire explicitly user-defined pattern (assigned through the parameter MBAmap) are included. By means of FMO, the MBs can be permuted (e.g., by masking the scan pattern type information in MBAmap) to resist the proposed MBS sketch attack. However, the deployment of FMO is still rare as mentioned in [28], and FMO is only supported in the BP and extended profile of H.264/AVC [1].

Another possible approach for resisting the proposed MBS sketch attack is to diffuse (or distribute) the visual information of a region (e.g., MB) into other regions (e.g., MBs). In the case of diffusion before or during video compression, the compression efficiency is compromised, because the spatial correlation among the pixels is destroyed, i.e., little to no redundancy is available to exploit for compression purpose. In the case of diffusion during or after compression (i.e., manipulation in the compressed form), this approach is theoretically possible, but it is not straightforward to be implemented while maintaining format compliance due to the nature of context adaptive entropy coding (i.e., CAVLC and CABAC) in the H.264/AVC standard. Although the

decode-encrypt-encode route is always available to encrypt a compressed video, the cost in terms of time and space complexities is too high, especially for smart devices with limited battery power. It should also be noted that the bitstream size will increase significantly by following this route.

All in all, based on the aforementioned discussions, given the current deployment trend of H.264/AVC (i.e., less popularity in implementing FMO feature) as well as the needs for format compliance and suppression of bitstream size increment, the proposed MBS sketch attack is viable and effective in extracting the outline information of the original frame directly from the encrypted video. Specifically, if a video encryption scheme wants to maintain the bitstream size of the original input video as well as format compliance, it will leak the information on local complexity, and thus vulnerable to the proposed MBS based attack. However, if security is the uppermost important issue for a particular video, then the user should consider the permutation, FMO, or decode-encrypt-encode route to resist the proposed MBS at the expense of bitstream size increment or decoding issue. In other words, there is a tradeoff among security, bitstream size increment, and format compliance.

E. Contributions

This paper makes the following three contributions. First, the limitations of the conventional sketch attacks were demonstrated. Specifically, the conventional sketch attacks considered in this paper (i.e., four in total) were verified to be somewhat viable in extracting the perceptual information from the encrypted I-frame, but they failed to extract any information from the encrypted inter frame under the H.264/AVC standard. Second, a novel sketch attack called MBS was proposed, and it was found to be effective in sketching both I-frame and inter frame. Notably, unlike the conventional plaintext-only, chosen-plaintext, and error-concealment attacks, the proposed MBS sketch attack considers the number of bits allocated to encode an MB. Therefore, an encryption scheme that was analyzed to be secure against conventional cryptanalysis might be vulnerable to the proposed MBS sketch attack. Third, the performance of the proposed sketch attack was benchmarked using three video data sets encoded with various parameter settings and compared with those of the conventional sketch attacks. It was found that only the proposed MBS sketch attack could extract the outline information from the inter frame, which outnumbers I-frame by far in any video encoding standard including H.264/AVC.

V. Conclusion

In this paper, we proposed a novel sketch attack for H.264/AVC format-compliant encrypted video. Specifically, the MBS was exploited to sketch the outline of the original video frame directly from the encrypted video. In addition, the Canny edge map was considered as the ideal outline image, and an edge similar score was modified for performance evaluation purposes. Experimental results suggest that the proposed sketch attack can extract visual information directly from the format-compliant encrypted video. Although the proposed and conventional methods can sketch the outline from the encrypted I-frame, only the proposed MBS sketch attack method can sketch the outline from the encrypted inter frame, which outnumbers I-frame, by far, in compressed video. Moreover, the proposed MBS sketch attack is verified to be more robust against compression when compared with the conventional sketch attacks.

In view of this proposed sketch attack framework, we suggest that this framework should be considered for format-compliant video encryption security analysis. In addition to determining the encryption modules in use by analyzing the encrypted video, we would like to extend this sketch attack framework to handle different video standards, such as HEVC, Audio Video Standard, and Google VP9 as our future work.

References

[1] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 13, no. 7, pp. 560–576, Jul. 2003.

[2] T. Stutz and A. Uhl, "A survey of H.264 AVC/SVC encryption," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 3, pp. 325–339, Mar. 2012.

[3] J. Wen, M. Severa, W. Zeng, M. H. Luttrell, and W. Jin, "A format-compliant configurable encryption framework for access control of video," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 12, no. 6, pp. 545–557, Jun. 2002.

[4] C.-P. Wu and C.-C. J. Kuo, "Design of integrated multimedia compression and encryption systems," *IEEE Trans. Multimedia*, vol. 7, no. 5, pp. 828–839, Oct. 2005.

[5] M. Podesser, H.-P. Schmidt, and A. Uhl, "Selective bitplane encryption for secure transmission of image data in mobile environments," in *Proc. 5th IEEE Nordic Signal Process. Symp.*, Oct. 2002, pp. 4–6.

[6] W. Li and Y. Yuan, "A leak and its remedy in JPEG image encryption," *Int. J. Comput. Math.*, vol. 84, no. 9, pp. 1367–1378, 2007.

[7] S. Li, G. Chen, A. Cheung, B. Bhargava, and K.-T. Lo, "On the design of perceptual MPEG-video encryption algorithms," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 17, no. 2, pp. 214–223, Feb. 2007.

[8] F. Liu and H. Koenig, "A survey of video encryption algorithms," *Comput. Secur.*, vol. 29, no. 1, pp. 3–15, 2010.

[9] K. Minemura, Z. Moayed, K. Wong, X. Qi, and K. Tanaka, "JPEG image scrambling without expansion in bitstream size," in *Proc. IEEE Int. Conf. Image Process.*, Sep./Oct. 2012, pp. 261–264.

[10] K. Minemura, K. Wong, X. Qi, and K. Tanaka, "A scrambling framework for block transform compressed image," *Multimedia Tools Appl.*, vol. 75, pp. 1–21, Feb. 2016.

[11] S. Y. Ong, K. Minemura, and K. S. Wong, "Progressive quality degradation in JPEG compressed image using DC block orientation with rewritable data embedding functionality," in *Proc. IEEE Int. Conf. Image Process.*, Sep. 2013, pp. 4574–4578.

[12] Y. Li, L. Liang, Z. Su, and J. Jiang, "A new video encryption algorithm for H.264," in *Proc. Int. Conf. Inf., Commun. Signal Process.*, 2005, pp. 1121–1124.

[13] S. Lian, Z. Liu, Z. Ren, and H. Wang, "Secure advanced video coding based on selective encryption algorithms," *IEEE Trans. Consum. Electron.*, vol. 52, no. 2, pp. 621–629, May 2006.

[14] P.-C. Su, C.-W. Hsu, and C.-Y. Wu, "A practical design of content protection for H.264/AVC compressed videos by selective encryption and fingerprinting," *Multimedia Tools Appl.*, vol. 52, nos. 2–3, pp. 529–549, 2011.

[15] E. Magli, M. Grangetto, and G. Olmo, "Conditional access to H.264/AVC video with drift control," in *Proc. IEEE Int. Conf. Multimedia Expo*, Jul. 2006, pp. 1353–1356.

[16] S.-K. A. Yeung, S. Zhu, and B. Zeng, "Partial video encryption based on alternating transforms," *IEEE Signal Process. Lett.*, vol. 16, no. 10, pp. 893–896, Oct. 2009.

[17] S. Spinsante, F. Chiaraluce, and E. Gambi, "Masking video information by partial encryption of H.264/AVC coding parameters," in *Proc. Eur. Signal Process. Conf.*, Sep. 2005, pp. 1–4.

[18] Y. Wang, M. O'Neill, and F. Kurugollu, "A tunable encryption scheme and analysis of fast selective encryption for CAVLC and CABAC in H.264/AVC," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 23, no. 9, pp. 1476–1490, Sep. 2013.

[19] B. Zeng, S.-K. A. Yeung, S. Zhu, and M. Gabbouj, "Perceptual encryption of H.264 videos: Embedding sign-flips into the integer-based transforms," *IEEE Trans. Inf. Forensics Security*, vol. 9, no. 2, pp. 309–320, Feb. 2014.
[20] W. B. Pennebaker and J. L. Mitchell, *JPEG: Still Image Data Compression Standard*. New York, NY, USA: Van Nostrand, 1992.
[21] R. C. Gonzalez and R. E. Woods, *Digital Image Processing*, 3rd ed. Upper Saddle River, NJ, USA: Prentice-Hall, 2006.
[22] N. Otsu, "A threshold selection method from gray-level histograms," *IEEE Trans. Syst., Man, Cybern.*, vol. 9, no. 1, pp. 62–66, Jan. 1979.
[23] J. Canny, "A computational approach to edge detection," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 8, no. 6, pp. 679–698, Nov. 1986.
[24] D. Karatzas *et al.*, "ICDAR 2013 robust reading competition," in *Proc. Int. Conf. Document Anal. Recognit.*, 2013, pp. 1484–1493.
[25] J. L. Feuvre, J. Thiesse, M. Parmentier, M. Raulet, and C. Daguet, "Ultra high definition HEVC DASH data set," in *Proc. Multimedia Syst. Conf.*, 2014, pp. 7–12.
[26] *Xiph.org Video Test Media*. [Online]. Available: http://media.xiph.org/video/derf/
[27] *H.264/AVC Reference Software*. [Online]. Available: http://iphome.hhi.de/suehring/tml/
[28] T. Schierl, M. M. Hannuksela, Y.-K. Wang, and S. Wenger, "System layer integration of High Efficiency Video Coding," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 22, no. 12, pp. 1871–1884, Dec. 2012.

Kazuki Minemura (S'13) received the B.S. and M.S. degrees in electrical and electronics engineering from Shinshu University, Matsumoto, Japan, in 2010 and 2012, respectively. He is currently pursuing the Ph.D. degree with the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia.

He is a Research Assistant with the Faculty of Computer Science and Information Technology, University of Malaya. His current research interests include information hiding, multimedia encryption, and attacking.

KokSheik Wong (S'06–M'10–SM'15) received the B.S. and M.S. degrees in computer science and mathematics from Utah State University, Logan, UT, USA, in 2002 and 2005, respectively, and the D.Eng. degree from Shinshu University, Matsumoto, Japan, in 2009, with the Monbukagakusho Scholarship.

He joined the Faculty of Computer Science and Information Technology, University of Malaya, Kuala Lumpur, Malaysia, in 2010, where he is currently a Senior Lecturer. He is currently a member of the Centre for Image and Signal Processing with the University of Malaya, where he leads the Multimedia Signal Processing and Information Hiding Group. His current research interests include information hiding, steganography, watermarking, multimedia perceptual encryption, multimedia signal processing, and their applications.

Dr. Wong is a member of the Asia Pacific Signal and Information Processing Association.

Raphael C.-W. Phan (M'03) received the Ph.D. (Eng.) degree in security from Multimedia University, Cyberjaya, Malaysia.

He held academic positions with Australian, Swiss, and British universities before taking up his current Chair position. His current research interests include diverse areas of security and privacy with a focus on privacy preservation and processing of data in the encrypted domain.

Dr. Phan was/is the General Chair of Mycrypt 05 and Asiacrypt 07, and the Program Chair of ISH 05 and Mycrypt 16. He has served on the technical program committees of international conferences since 2005. He is a Co-Designer of BLAKE, one of the five hash function finalists of the NIST SHA-3 competition. He has an Erdős number of 2.

Kiyoshi Tanaka (M'95) received the B.S. and M.S. degrees in electrical engineering and operations research from the National Defense Academy, Yokosuka, Japan, in 1984 and 1989, respectively, and the D.Eng. degree from Keio University, Tokyo, Japan, in 1992.

He joined the Department of Electrical and Electronic Engineering, Faculty of Engineering, Shinshu University, Matsumoto, Japan, in 1995, where he is currently a Full Professor with the Academic Assembly, Institute of Engineering. He is the Vice President of Shinshu University, where he is the Director of the Global Education Center. He has been a Project Leader of the JSPS Strategic Young Researcher Overseas Visits Program for Accelerating Brain Circulation entitled Global Research on the Framework of Evolutionary Solution Search to Accelerate Innovation since 2013. His current research interests include image and video processing, information hiding, human visual perception, 3D point cloud processing, evolutionary computation, multiobjective optimization, smart grid, and their applications.

Dr. Tanaka is a member of the Institute of Electronics, Information and Communication Engineers, the Information Processing Society of Japan, and JSEC. He is a fellow of the Institute of Image Electronics Engineers of Japan (IIEEJ). He received the IEVC2010 Best Paper Award from IIEEJ, the iFAN2010 Best Paper Award from SICE, the GECCO2011 Best Paper Award, the GECCO2015 Best Paper Award from ACM-SIGEVO, the ISPACS2011 Best Paper Award from the IEEE, the Excellent Journal Paper Award from IIEEJ two times, in 2012 and 2014, and the Best Journal Paper Award from JSEC in 2012. He is the Editor-in-Chief of the *Journal of the Institute of Image Electronics Engineers Japan* and the *IIEEJ Transactions on Image Electronics and Visual Computing*.