Moon *et al. EURASIP Journal on Image and Video Processing* (2019) 2019:12
https://doi.org/10.1186/s13640-018-0389-0

EURASIP Journal on Image and Video Processing

RESEARCH Open Access

Integrated visual security management for optimal condition explorations in resource-constrained systems

Junhyung Moon[1*], Heemin Park[2†] and Kyoungwoo Lee[1]

Abstract

As more and more mobile video applications contain important and private video data, various video encryption techniques have been proposed to protect them. Many of those applications typically utilize the standard video compression while running on mobile platforms which have limited amount of resources. Therefore, we need to consider not only the protection but also the compression and the energy in the mobile video applications. Accordingly, the selective video encryption which protects partial data within the standard video compression has been challenging thanks to format compliance and low computational complexity. However, various parameters of the video compression and the video encryption result in different amount of visual security, compression efficiency, and energy efficiency. Further, it is difficult to find the one solution to maximize all those performance indices at once since there exist tradeoff relationships among them. Therefore, based on the tradeoff relationships, design space exploration should be required to find the interesting parameter set under the given requirements. For the efficient design space exploration, we propose the BEVS (bitrate and energy-bound visual security) to examine each Joint Video Compression and Encryption scheme in terms of the visual security, the compression efficiency, and the energy efficiency. Utilizing the proposed BEVS in our experimental design space, we achieve up to 36.7% visual security improvement under the empirical budgets.

Keywords: Video compression, Video encryption, Energy efficiency, Compression efficiency, Visual security, Tradeoff, Limited resource, Design space exploration, Optimal condition

1 Introduction

1.1 Background

Thanks to the technology advances, the embedded systems have been widely utilized in human lives such as households, industries, stock markets, and hospitals. Based on the embedded systems, lots of mobile devices and applications have been designed to conveniently control and activate physical and cyber functionalities of various services. Among them, mobile video applications support various services exploiting visual data collected from the image sensors. According to the burst of the IoT (Internet of Things) devices and the growth of the image sensor markets, the number of the mobile video applications keeps increasing.

As more and more mobile video applications have been employed for various services, some of them become responsible for critical tasks. They contain important information such as finance, confidentiality, security, and privacy [1–5]. For example, the video surveillance systems installed for the individual's safety have been ironically argued that they invasively record the individual's privacy. In addition, the confidential information such as important contract is discussed between the companies and the governments through the video conferencing. Further, the medical video applications for the telemedicine and the surgery assistance handle the private information and possibly the life-critical data of the patients. Military also have employed several video applications such as the reconnaissance systems using the unmanned aerial vehicles to monitor the important information. If those data

*Correspondence: jh.moon.cs@yonsei.ac.kr
†Heemin Park contributed equally to this work.
[1]Department of Computer Science, Yonsei University, Yonsei-ro 50, Seoul, Republic of Korea
Full list of author information is available at the end of the article

are abused by anyone who has vicious purposes, it must be definitely dangerous threats and harms.

In order to protect the private and important video information, various video encryption techniques have been designed. Among them, the selective video encryption which typically embeds lightweight encryption modules into the standard video encryption process has been popular since it reduces computational complexity and keeps format compliance of the standard video codec as compared to the full video encryption that usually protects the video data separately with the codec. As an example, a simple operation like the bit-wise flipping partially protects the quantized coefficients within the standard video compression process, incurring negligible overheads without any impact on the format of the codec. Reducing computations and preserving the format compliance of the video compression are significantly important since many of the video communication services require low latency and utilize the standard video codec. Accordingly, various selective encryption methods [6–18] have been designed to protect the video regarding the video compression process through simple operations.

Besides the protection of the important videos, we also need to care the energy consumption and the compression bitrate since the amount of resources is constrained in the mobile platforms. The video compression has been well-known to require lots of computations, thereby consuming huge energy [19–21]. Further, the video encryption techniques additionally incur the extra overheads. Moreover, the video compression and the video encryption influence each other in terms of the energy consumption and the compression bitrate. On the one hand, the overheads of several encryption methods differ according to the configurations of the video compression parameters to be utilized. For example, the amount of data to be protected by the motion vector encryption techniques depends on the GOP (Group Of Picture) size of the video compression. On the other hand, several video encryption methods possibly break the redundancy of the video data increasing the size of the compressed output. Thus, we need to consider both the energy consumption and the compression bitrate when conducting the combination of the video compression and the video encryption. In our study, we use the term *JVCE (Joint Video Compression and Encryption)* to represent the combination of the video compression and the video encryption.

1.2 Problem definition

By considering the protection of the video, the energy consumption, and the compression bitrate, we like to determine the optimal solution among various parameter sets of the JVCE to satisfy the requirements of the target video service. In order to determine which parameter set is better than others, we need to compare the candidates each other in the quantitative manner. However, as compared to the objective degrees of the energy consumption and the compression bitrate, the human eye's perception about the encrypted videos is definitely subjective.

Accordingly, various evaluation metrics have been proposed to represent the visual distortion in the objective manner such as the numbered levels. They estimate the visual degradation introduced by the video compression, the error-prone transmission, and the video encryption, as such. However, conventional metrics show several inaccurate estimations of the human eye's perception due to the lack of considering the temporal distortion of the encrypted videos. Therefore, we proposed the STVSM (spatio-temporal visual security metric) to capture both the spatial distortion within each encrypted frame and the temporal distortion in the consecutive encrypted frames [22].

By using the STVSM and estimating the energy consumption and the compression bitrate, we can compare various parameter sets of the JVCE each other in terms of the visual security, the compression efficiency, and the energy efficiency. We refer the perceptual distortion introduced by the video encryption techniques as the visual security in this work. Further, we exploit the terms energy efficiency and compression efficiency to claim that the larger, the better. For example, if a parameter set A of the JVCE consumes 30 joules and another one B consumes 20 joules, we say the B has higher energy efficiency than the A. As for the compression efficiency, same implication works.

Depending on the set of parameters such as the GOP size and the kind of the protection scheme, the JVCE results in different amount of the visual security, the compression efficiency, and the energy efficiency. The visual security could show tradeoff relationships against the energy efficiency and the compression efficiency. For example, a video encryption technique achieves higher visual security by significantly modifying the magnitudes of the coefficients. However, in this case, the compression efficiency could be degraded since the local redundancy of the consecutive coefficients is messed up. As an another example, if we utilize several encryption techniques rather than just one method for higher visual security, the energy efficiency would decrease due to more computations.

The amount of the improvement of the visual security and the amount of loss of the energy efficiency and the compression efficiency in the tradeoffs are not linear. In the tradeoff space, the JVCE with one set of parameters can significantly enhance the visual security with a little degradation in the compression efficiency and the energy efficiency. In contrast, the JVCE with another set could improve the visual security at the huge cost of both efficiencies. Further, various kinds of parameters of the JVCE make the tradeoff space complex enough to be examined.

For example, we investigate 1629 sets of 7 parameters of the JVCE in this study. Moreover, the tradeoff space differs according to characteristics of each video such as the movements even if we utilize same JVCE schemes.

Therefore, it is difficult to find the optimal condition to enhance the visual security that is the visual quality, the compression efficiency, and the energy efficiency to the best in the complex tradeoff space. Alternatively, what we expect to do is to find the interesting condition in the tradeoff space to maximize the visual security while just managing the compression efficiency and the energy efficiency not to exceed the certain requirements. Several design space studies have been conducted for the video compression and decompression to optimize the multimedia systems and services based on the tradeoff among the energy efficiency, the QoS (quality of service), and the latency [23–28]. However, there have been few works to examine the design space of the combination of the video compression and the video encryption techniques that is the JVCE in this paper. The JVCE incurs more overheads and has more sophisticated tradeoff relationships among the outcomes as compared to the video compression without the protection. Therefore, the design space of the JVCE should be comprehensively studied in detailed manner. This design space study helps us meet the requirements according to target services while satisfying the constraints. For example, for a video service not containing critical information, we can save resources to be utilized by supporting the least protection. Further, in case that we should thoroughly protect the confidential information, we could minimize the overhead introduced by heavy encryption.

1.3 Our contributions

Based on our previous work [29], we study the design space of the JVCE in terms of the energy efficiency and the visual security while additionally considering the video compression efficiency together. The compression efficiency is one of the most important things as well as the energy efficiency and the visual security. Even though the JVCE with a specific set of parameters properly protects the video under the given energy budget, that could not become the efficient solution in case that it significantly degrades the compression efficiency. Table 1 shows the example results after performing the JVCE for the Foreman QCIF (Quarter Common Intermediate Format) video with two different parameter sets. Two different JVCE schemes are denoted as A and B in Table 1. Details about the parameters and the estimation setup are introduced in the following sections. In terms of the energy efficiency and the visual security where higher value indicates more perceptual distortion, the A guarantees extremely higher visual security as compared to the B while losing only about 3.8% energy efficiency which can be negligible. However, when it comes to the compression efficiency, the A entails the significant degradation of the compression efficiency as about 175% as compared to the B. Therefore, we need to consider the visual security, the compression efficiency, and the energy efficiency at once in the design space of the JVCE.

Table 1 Example tradeoff in terms of visual security, energy consumption, and compression bitrate

Encryption	Visual security (STVSM)	Energy consumption (J)	Compression bitrate (Mbps)
A	0.63	34.58	2.48
B	0.12	33.33	0.90

Accordingly, we propose the BEVS (bitrate and energy-bound visual security) to investigate each set of parameters in the design space of the JVCE under several requirements in terms of the visual security, the compression efficiency, and the energy efficiency. Thanks to the BEVS, we improve up to 36.7%, 30.5%, and 26.6% visual security between the minimum and the maximum of the BEVS for three QCIF videos, respectively in our experiments. Finally, we summarize the contributions of this paper as follows.

- We have investigated the problem of the video encryption regarding the video compression in the resource-constrained systems. In this problem, we examine the design space of the JVCE consisting of detailed and specific parameters.
- In order to objectively examine the design space in terms of the visual security, we thoroughly demonstrate the visual security evaluation of several metrics including our previous work [22] for various encrypted videos.
- Based on the proposed BEVS and customDSE, we efficiently explore the design space of the JVCE to improve the visual security while satisfying the requirements of the energy efficiency and the compression efficiency.

2 Joint Video Compression and Encryption

In order to compare each set of parameters of the JVCE, it is necessary to adjust those parameters and to analyze the results after conducting the JVCE. Accordingly, we implement the framework to support the configurations of the video compression and several video encryption techniques during the compression.

2.1 Configuration framework to support the JVCE

For the video compression, we utilize the JM H.264 reference software (version 18.4) (http://iphome.hhi.de/suehring/tml). Further, we need to select and implement several video encryption techniques among various ones.

The video encryption methods of the early stage were designed to protect the whole video data by using the conventional encryption algorithms such as the DES (Data Encryption Standard) [30] and the AES (Advanced Encryption Standard) [31]. Those techniques have high computational complexity [32–34]. Moreover, they do not consider the format of the compressed bitstream defined in the standard video codecs. If the format is broken, direct manipulations on the bitstream such as parsing, transcoding, and decoding cannot be supported [12, 35, 36]. This is not appropriate for various video services such as the subscription-based television, the video-on-demand, and the video surveillance systems. In order to keep the format compliance while reducing the computational complexity of the algorithm, several video encryption techniques are introduced considering the characteristics of the video compression process and the video data [6–15]. They protect partial data within the compression process through just simple operations. Among them, we implement several methods [6–8] in the JM software as presented in Fig. 1.

In more detail, the utilized encryption methods protect the coefficients and the motion vectors that are spatial and temporal information, respectively. They pseudo-randomly flip sign bits of those data [6, 8], xor them with pseudo-random numbers [7, 8], pseudo-randomly shuffle them with each other [7], and apply possible combinations of those three methods. We do not shuffle the coefficients since we need to re-calculate the number of zero coefficients prior to each non-zero coefficient, incurring the extra computations. If we do not re-calculate that numbers after shuffling the coefficients, the format of the bitstream could be messed up. In case of the coefficient protection, the encryption methods are applied to the DC (direct current) coefficient only, the AC (alternating current) coefficient only, and both of them in the luminance channel. Further, we change possible range of pseudo-random numbers for the XOR method from 0 to small and large maximum in case of both coefficients and motion vectors. When using the pseudo-random numbers from the larger range, the video is distorted by the XOR method more heavily than the case of using the smaller range. This can increase the size of the compressed bitstream so that it should be discussed in the design space exploration in our results.

In the rest of this paper, we utilize the annotations in Table 2 to clearly represent both the data to be encrypted and the utilized encryption techniques. Table 2 presents several video encryption techniques as the functions requiring input data. Further, the input data to be encrypted by the functions are also listed in Table 2. For example, sign_enc{xor_enc(*large_r*){DC}}no_enc{MV} xors DC coefficients with pseudo-random numbers from the large range and then pseudo-randomly flip sign bits of the xored DC coefficients. It does not encrypt motion vectors at all. As an another example, sign_enc{DCAC}shuffle_enc{xor_enc(*small_r*){MV}} pseudo-randomly flips sign bits of both DC and AC coefficients. Further, it xors x- and y-coordinates of motion vectors with pseudo-random numbers from small range and then shuffles them each other. The order of applying different techniques to the same data could result in different outcomes including the visual security of the encrypted video. However, the order of applying the encryption methods makes the design space exploration not that interesting but too much complex as compared to other parameters examined in this paper. Further, the encryption key of the utilized encryption algorithms in this paper is the random seed for pseudo-random number generator. Although changing the key results in different encryption results, that shows the unclear differences in terms of the visual quality and the energy consumption as compared to changing other parameters utilized in Section 5 .

2.2 Design space exploration framework of the JVCE

As shown in Fig. 2, we complete the experimental framework for the design space exploration of the JVCE using the raspberry pi 3 model B board (https://www.raspberrypi.org/products/raspberry-pi-3-model-b/), the

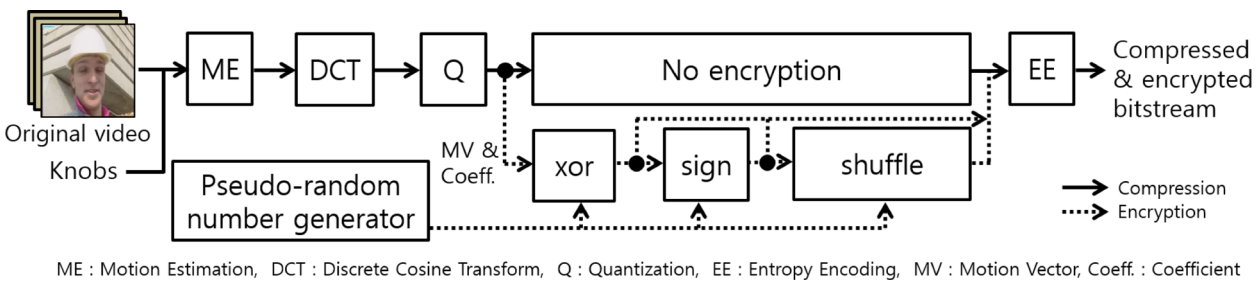


Fig. 1 Overall process of Joint Video Compression and Encryption (JVCE) exploiting the standard video codec and several encryption modules. This shows the overview of the joint compression and encryption process that has been implemented and utilized through the entire paper. Several encryption methods from previous works can be applied to the intermediate data such as coefficients during the standard video compression process

Table 2 Annotations to represent the video encryption functions and the data to be protected by the video encryption functions

Annotation	Type	Meaning
no_enc{data}	Function	Not encrypt data
sign_enc{data}	Function	Pseudo-randomly flip sign bit of data
shuffle_enc{data}	Function	Pseudo-randomly shuffle data each other
xor_enc(*large_r*){data}	Function	XOR data with pseudo-random number from *large* range
xor_enc(*small_r*){data}	Function	XOR data with pseudo-random number from *small* range
DC	Input data	DC coefficients
AC	Input data	AC coefficients
DCAC	Input data	both DC and AC coefficients
MV	Input data	x and y coordinates of motion vectors

Monsoon power monitor (http://www.msoon.com/LabEquipment/PowerMonitor), and the desktop. The Monsoon power monitor accurately estimates the energy consumption of the JVCE running in the Ubuntu 16.04 LTS (https://wiki.ubuntu.com/ARM/RaspberryPi) which is specified for the raspberry board. After conducting the JVCE on the board, we can get the compression efficiency returned from the H.264 codec software. Then, we decompress the encrypted bitstream without the video decryption in the desktop to obtain the encrypted videos. Further, the encrypted videos are estimated by the implemented visual security metrics to represent the visual security. For the visual security evaluation, we exploit (https://ece.uwaterloo.ca/texttildelowz70wang/research/ssim/) [37] and (http://live.ece.utexas.edu/research/Quality) for the SSIM (Structural SIMilarity) [38], the LEG (Local Edge Gradients) [39], and the VIF (Visual Information Fidelity) [40], respectively, and implement other evaluation methods presented in the following section. Finally, we can examine the design space consisting of every possible set of parameters of the JVCE utilized in this study in terms of the visual security, the compression efficiency, and the energy efficiency.

For the design space exploration, we adjust various parameters of the video encryption explained in previous subsection such as the kind of data, the type of the coefficient, and the number of the applied encryption techniques. Additionally, we also adjust the video compression parameters such as the GOP size and the QP (quantization parameter) level. The number of motion vectors to be protected by the encryption methods depends on the GOP size leading to different amount of the overheads. In our experiments, we vary the GOP size as 1, 5, 10, 30, and 300. In addition, we examine the trend while changing the QP levels as 15, 20, and 25, since the QP significantly incurs the impact on the compression efficiency and the energy efficiency. Finally, we protect the Akiyo, Coastguard, and Foreman QCIF videos of 300 frames while varying all the parameters in the JVCE as described so far.

3 Visual security evaluation

In order to explore the interesting design space of the JVCE, we need to objectively compare every set of parameters in the space in terms of the visual security, the compression efficiency, and the energy efficiency. Accordingly, we need to quantify those three indices in numbered levels. The energy efficiency and the compression efficiency are quantified by estimating the energy consumption and the compression bitrate. However, the visual security that is the degree of the visual distortion caused by the encryption requires an objective metric which represents human eye's perception. Therefore, this section thoroughly studies the visual security evaluation in the objective manner.

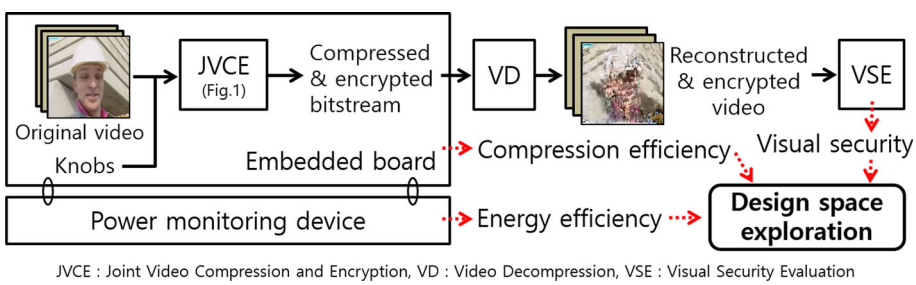


Fig. 2 Design space exploration framework of the JVCE regarding visual security, compression efficiency, and energy efficiency. This figure shows the experimental framework that we construct for the design space exploration of the JVCE. The framework consists of the embedded board for running the JVCE with various parameters, the power monitoring device for estimating the energy consumption of the JVCE, and the desktop for estimating the visual security of the encrypted videos after the JVCE

3.1 Objective visual quality/security evaluation metrics

In the image processing studies, many wonder how much an image is perceptually changed after the compression, the encryption, and the transmission on the error-prone network, as such. The perceptual change typically represents the quality degradation or the protection power. In this case, we can compare the processed image with the reference one in terms of the human eye's perception such that *"the processed image is heavily distorted, slightly twisted, or quite unseen"*. Based on those subjective expressions, it is difficult to objectively determine whether one image is more perceptually distorted than other ones in the way that many people agree with. Thus, for the clear comparison, we need to quantify the degree of the visual difference between the target images as close as possible to the human eye's perception and also the way many people agree with. Accordingly, various evaluation metrics have been proposed to estimate how much the target images perceptually differ as the numerical levels. These metrics are typically called the visual quality metrics. Further, the term visual security metric is often used to indicate the evaluation method to show the amount of the perceptual distortion introduced by the video encryption techniques.

The PSNR (peak signal-to-noise ratio) is one of the most popular metrics in the signal processing to represent the difference between two images. However, the PSNR has often been argued that it is inaccurate to show the evaluation results similar to the ones from the human eye's perception [41–43]. For the better representation of the visual difference between the images than the PSNR, there have been several approaches including the SSIM and the LFBVS (local feature-based visual security) [44]. The SSIM captures the visual difference between the images better than the PSNR [45]. However, it has been turned out that it shows inaccurate evaluation for the heavily blurred images [46]. The LFBVS [44] has been designed to more effectively represent the visual difference between the original video and the encrypted video as compared to the conventional metrics including the PSNR and the SSIM. The LFBVS appropriately estimates the visual distortion introduced by the video encryption techniques. However, it shows the inaccurate evaluation on several encrypted videos since it does not consider the distortion of the temporal information such as the movements of the objects in the consecutive frames [22]. The LEG appears a good correlation with human eye's perception, and further, the VIF outperforms various metrics mentioned above including LEG [47]. However, they still lack in considering temporal information of consecutive images in the same video.

In order to compensate the inaccurate estimation cases of the conventional metrics, we designed the STVSM (spatio-temporal visual security metric) in our previous research [22]. The STVSM considers the distortions of both the spatial and the temporal information such as the luminance coefficient and the motion vector, respectively. Finally, the STVSM effectively captures not only the spatial distortion of each frame but also the temporal distortion of the consecutive frames of the encrypted video [22]. Accordingly, we employ the STVSM for the design space exploration of the JVCE regarding the visual security. Before the exploration, we comprehensively demonstrate the STVSM's evaluation as compared to several conventional metrics in the following subsections.

3.2 Visual security evaluation setup

By utilizing the JVCE framework, we validate the visual security evaluation of the objective metrics for various encrypted videos. However, if all the parameter of the JVCE are arbitrarily adjusted at once, it could be too much complex to understand the results. Accordingly, we consistently fix the video compression parameters to clearly show the visual distortion introduced by the video encryption techniques. In addition, we construct four scenarios varying the parameters of the coefficient encryption as shown in Table 3. For each scenario from 1 to 3, different types of the coefficients are encrypted by different methods while all possible motion vector encryption schemes are examined. Further, we configure the special option only for the xor encryption which determines the possible range of pseudo-random numbers in the scenario Extended-1.

For the encrypted videos by using those four scenarios, we clearly demonstrate how much the objective evaluation correlates with the subjective evaluation that is the

Table 3 Experimental scenarios of the JVCE under consistent set of parameters of video compression while varying several parameters of video encryption (Foreman video)

Scenario	1	2	3	Extended-1
Coefficient type	AC	DC	DCAC	AC
Coefficient encryption	xor_enc	sign_enc & xor_enc	sign_enc	xor_enc
Motion vector encryption		One of all possible combinations		
Option for xor range		Small		Small, large
Input video	Foreman video (QCIF format, 300 frames)			
Compression parameters	GOP : 30, QP : 25, Others : default			

human eye's perception. To obtain the subjective evaluation results as the reference, we conduct a survey for 66 university students. In the survey, the participants watch both the original QCIF videos and the distorted videos by the encryption schemes defined in Table 3. To control the lighting condition which could affect the visual assessments of the participants, we use the same laptop under the same contrast and brightness for all the survey. For each encrypted video, the participants score the degree of how much the video is perceptually distorted as compared to the reference one from 1 to 5. The scores indicate "Highly recognized (1)", "Partially recognized (2)", "Medium (3)", "Slightly distorted (4)", and "Totally distorted (5)", respectively. The higher score indicates the higher visual security. Based on the survey, the MOS (mean opinion score) is calculated by averaging the subjective scores for each encrypted video. We do not consider the outlier detection when averaging the MOS, since most of the scores typically appear in the similar range. The averaged MOS scores are compared with the objective evaluation results.

For the objective evaluation, we exploit the luminance channel data since the change of the luminance information is more sensitive to human eyes than that of the chrominance information [48]. For the SSIM, we use the default setting in the [38] so that the MSSIM (mean SSIM) is estimated per each frame by averaging the SSIM of the local windows within the frame. The PSNR, the MSSIM, the LEG, and the VIF represent the visual difference of each image in the video so that we utilize the average of each metric score of total frames in the encrypted video. From 0 to 1, lower MSSIM, higher LFBVS, lower LEG, lower VIF, and higher STVSM of the encrypted video indicate that the video is more visually protected than others. Further, the PSNR is ranged from 0 to infinite where closer value to 0 implies that the encrypted video is more different from the reference one than others. In order to claim that higher values of the estimation results from four metrics indicate higher visual security, we examine the $1/PSNR$, 1-MSSIM, 1-LEG, and 1-VIF instead of PSNR, MSSIM, LEG, and VIF, respectively.

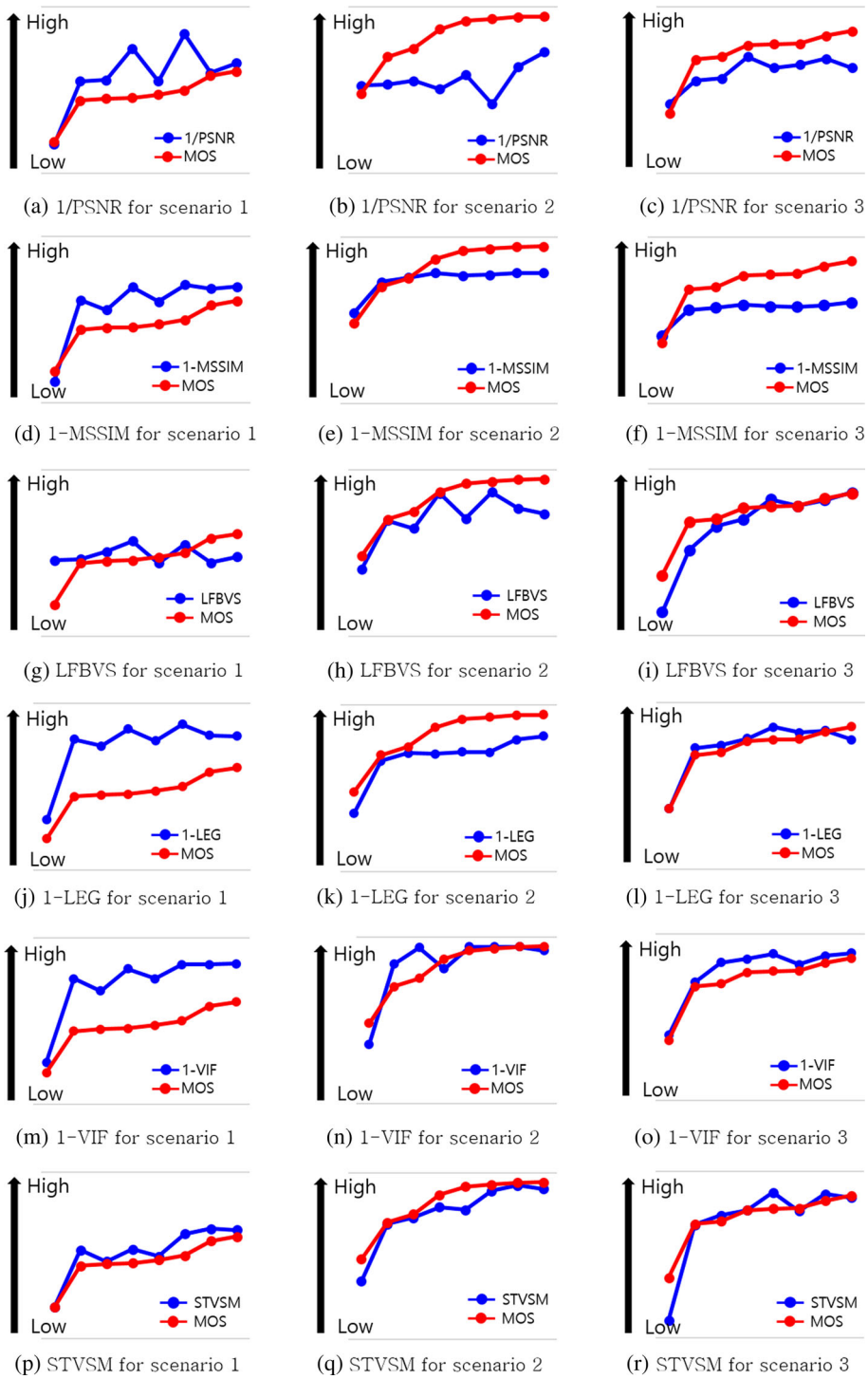
3.3 Objective visual security evaluation results as compared to subjective ones

Figures 3 and 4 show the objective scores and the subjective scores for all the scenarios in Table 3. For all the graphs in both Figs. 3 and 4, the evaluation scores are represented in the abstract manner such as from low to high since the range of the scores differ based on the type of metric, target video, and parameters. Moreover, the examined range of the objective scores is appropriately scaled to clearly show the trend as compared to that of the subjective scores. Note that even if several objective scores present the trend not corresponding to that of the subjective scores such as the zigzag pattern in Fig. 3 (g), they reveal several inaccurate cases, not total failure. From left to right on the x-axes of all the graphs, the evaluation results in one scenario are arranged in increasing order of MOS.

In the results, the evaluation scores by using the STVSM show more similar trend to that of the subjective score MOS as presented in Fig. 3(p), (q), and (r) as compared to other metrics. Moreover, we examine the visual security evaluation results additionally adjusting the range option for the xor encryption operation in Fig. 4. Similarly, it is demonstrated that the STVSM estimates the perceptual distortion more close to the subjective score than other metrics. Figure 5 shows the sample frames of each encrypted video corresponding to the same order of the x-axes of the graphs in Fig. 4. From Fig. 5(a) to (p), it is demonstrated that the visual contents of the video become more noisy and blurred.

Further, we estimate the correlation between the evaluation results of the objective metrics and the subjective scores, that are MOS, by using two correlation coefficients: the PPMCC (Pearson product-moment correlation coefficient) [49] and the SRCC (Spearman's rank correlation coefficient) [50]. The PPMCC is a measure of how much two variables are linearly dependent, and the SRCC shows the degree of the dependence between the ranking of two variables in statistical manner. These coefficients range from -1 to 1 where higher magnitude indicates two input data have higher correlation and the sign represents whether the correlation is positive or negative. Table 4 presents those correlation coefficients between the objective scores and the subjective ones. Among the objective metrics, the STVSM represents the highest correlation against the subjective score in terms of both correlation coefficients for the scenario Extended-1. For the Akiyo and the Coastguard videos, the STVSM also shows the highest correlation with the subjective score under the same encryption scenario.

In the scenario Extended-1, the visual distortion of the encrypted videos differs highly depending on whether the motion vectors are protected or not and which encryption techniques are applied to motion vectors. For example, we compare two encrypted videos by encryption schemes P and Q which have different motion encryption methods in the scenario Extended-1: xor_enc(large_r){AC}shuffle_enc{MV} and xor_enc (large_r){AC}shuffle_enc{sign_enc{xor_enc(large_r) {MV}}}, respectively. Figure 6 presents four sample frames from the original Foreman video and two encrypted versions by P and Q. Both P and Q distort the original video in noisy manner, but we can see the perceptual difference between the encrypted videos due to the different motion vector encryption techniques. Note that not all the frames of the encrypted video by P leak the visual information more



(a) 1/PSNR for scenario 1 (b) 1/PSNR for scenario 2 (c) 1/PSNR for scenario 3

(d) 1−MSSIM for scenario 1 (e) 1−MSSIM for scenario 2 (f) 1−MSSIM for scenario 3

(g) LFBVS for scenario 1 (h) LFBVS for scenario 2 (i) LFBVS for scenario 3

(j) 1−LEG for scenario 1 (k) 1−LEG for scenario 2 (l) 1−LEG for scenario 3

(m) 1−VIF for scenario 1 (n) 1−VIF for scenario 2 (o) 1−VIF for scenario 3

(p) STVSM for scenario 1 (q) STVSM for scenario 2 (r) STVSM for scenario 3

Fig. 3 Visual security evaluation results of objective (blue) and subjective (red) metrics for scenarios 1, 2, and 3 (from the subfigure (**a**) to (**r**) in case of Foreman video, *x*-axis represents different encryptions which result in different MOS in the increasing order and *y*-axis represents visual security). This figure shows the evaluation results from the objective metrics (PSNR, MSSIM, LFBVS, LEG, VIF, and STVSM) and the subjective scores of 66 subjects for various encrypted videos. From the subfigure (**a**) to (**r**) in case of Foreman video, *x*-axis represents different encryptions which result in different MOS in the increasing order and *y*-axis represents visual security. Figure 3 aims to present how each objective metric correlates with the subjective scores from the human eye's perception

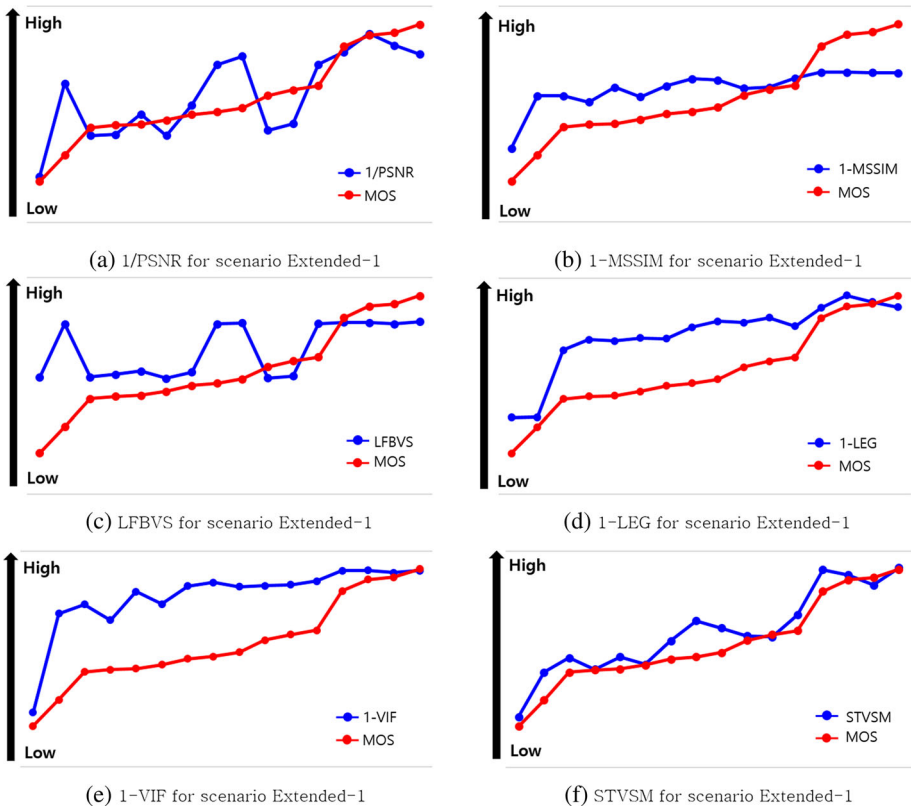


Fig. 4 Visual security evaluation results of objective (blue) and subjective (red) methods for scenario Extended-1 (from the subfigure (**a**) to (**f**) in case of Foreman video, *x*-axis represents different encryptions which result in different MOS in the increasing order and *y*-axis represents visual security). This shows the evaluation results from the objective metrics (PSNR, MSSIM, LFBVS, LEG, VIF, and STVSM) and the subjective scores of 66 subjects for various encrypted videos additionally considering a special option of the implemented encryption method. By considering the special option, we have 16 encrypted videos which enables us to study the visual security evaluation more detailed as compared to Fig. 3. From the subfigure (**a**) to (**f**) in case of Foreman video, *x*-axis represents different encryptions which result in different MOS in the increasing order and *y*-axis represents visual security. Similar to Fig. 3, Fig. 4 aims to present how each objective metric correlates with the subjective scores from the human eye's perception

than those by Q. However, it is true that the encrypted video by P has more frames to leak perceptual clues to guess the original contents shown in Fig. 6(e), (f), (g), and (h). The subjective scores which are MOS scores in Table 5 also demonstrate the clear difference between two encrypted videos. However, the $1/PSNR$, the LFBVS, and the 1-LEG cannot determine that which one of P and Q perceptually distorts the video more than the other. Although the scores of the 1-MSSIM and the 1-VIF show the perceptual difference of both videos, the difference is only 3% which is ambiguous and even negligible as compared to the clear difference as 12% in case of the STVSM.

One of the main reasons of the inaccurate evaluation cases of the conventional visual quality metrics is lacking in the consideration of the temporal distortion of the encrypted videos. The temporal information such as the trajectories of the moving objects can be leaked when playing the consecutive frames even if each frame is well encrypted. Unfortunately, the conventional visual quality metrics estimate the visual security of the encrypted video only in the spatial manner which considers the visual distortion between the target image and the reference one. On the other hand, the STVSM compares the degree of the visual distortion not only between the encrypted frame and the reference one but also between the consecutive frames in the encrypted video and the corresponding ones in the reference video. Based on thorough investigation in this section, the STVSM works as a good measure to objectively represent the perceptual distortion of the encrypted videos as compared to several conventional metrics. Therefore, we utilize the STVSM to examine the design space in terms of the visual security in the following section.

4 Proposed methods—design space exploration of the JVCE

In this paper, we comprehensively examine the design space of the JVCE to find the interesting set of parameters in terms of the visual security, the compression

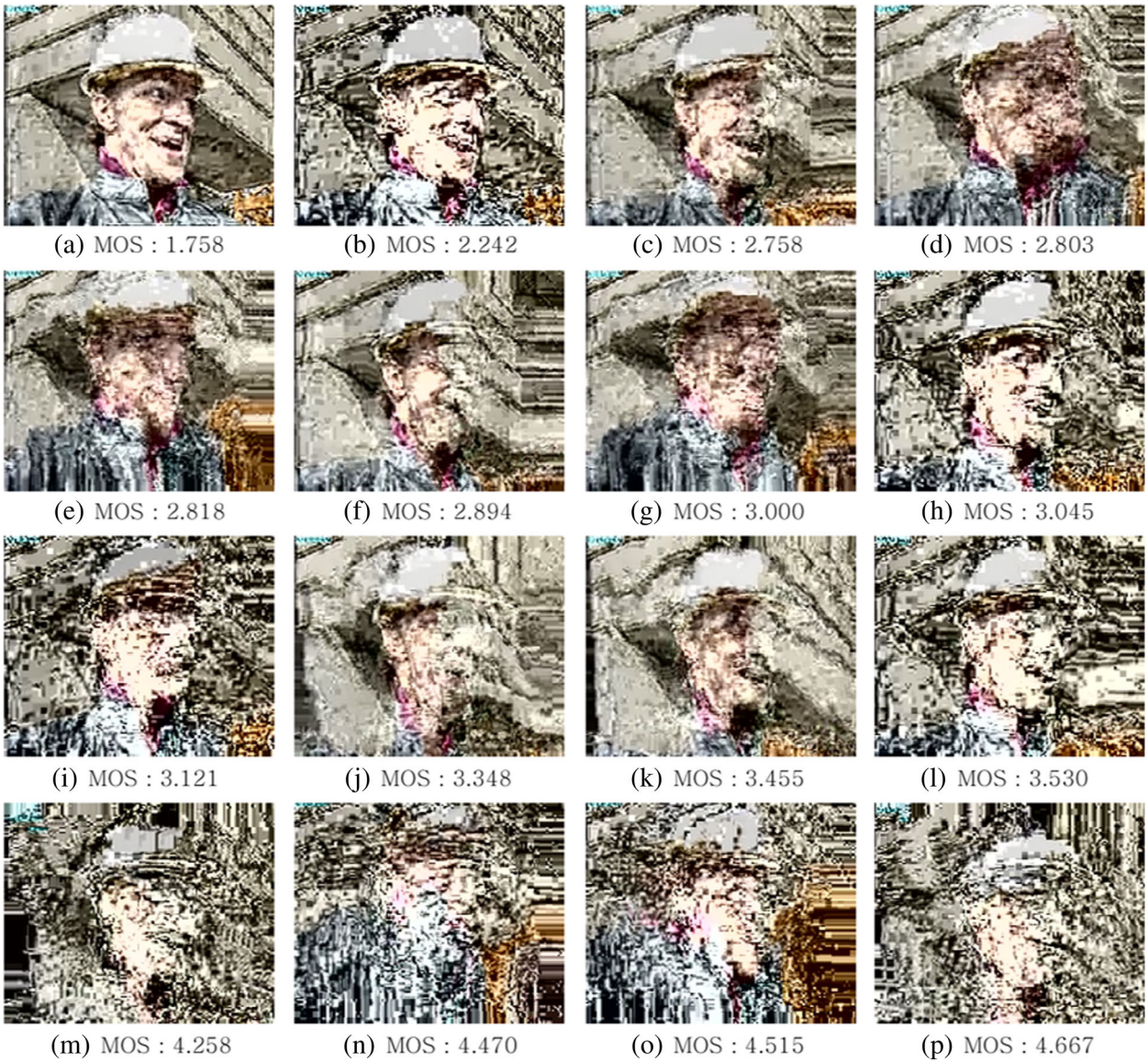


Fig. 5 a–p Each sample frame (131st) corresponding to each dot on the *x*-axis in Fig. 4. These figures show sample frames (131st frames) and their scores from 66 subjects of all the videos evaluated in Fig. 4. Figure 5 presents the sample frames of different videos in the order of increasing MOS that is the subjective score, corresponding to the order of dots in terms of *x*-axis in Fig. 4. Accordingly, we like to claim that the STVSM which shows the most similar trend to increasing MOS appropriately captures the human eye's perception for the encrypted videos

Table 4 Correlation coefficients of objective scores against subjective score (MOS) for scenario Extended-1

Video	Coefficient	1/*PSNR*	1-MSSIM	LFBVS	1-LEG	1-VIF	STVSM
Akiyo	PPMCC	0.753	0.754	0.300	0.640	0.709	0.976
	SRCC	0.547	0.550	0.341	0.182	0.218	0.671
Coastguard	PPMCC	0.736	0.911	0.523	0.540	0.254	0.936
	SRCC	0.609	0.650	0.338	0.503	-0.014	0.835
Foreman	PPMCC	0.713	0.763	0.562	0.753	0.728	0.869
	SRCC	0.715	0.815	0.649	0.744	0.826	0.894

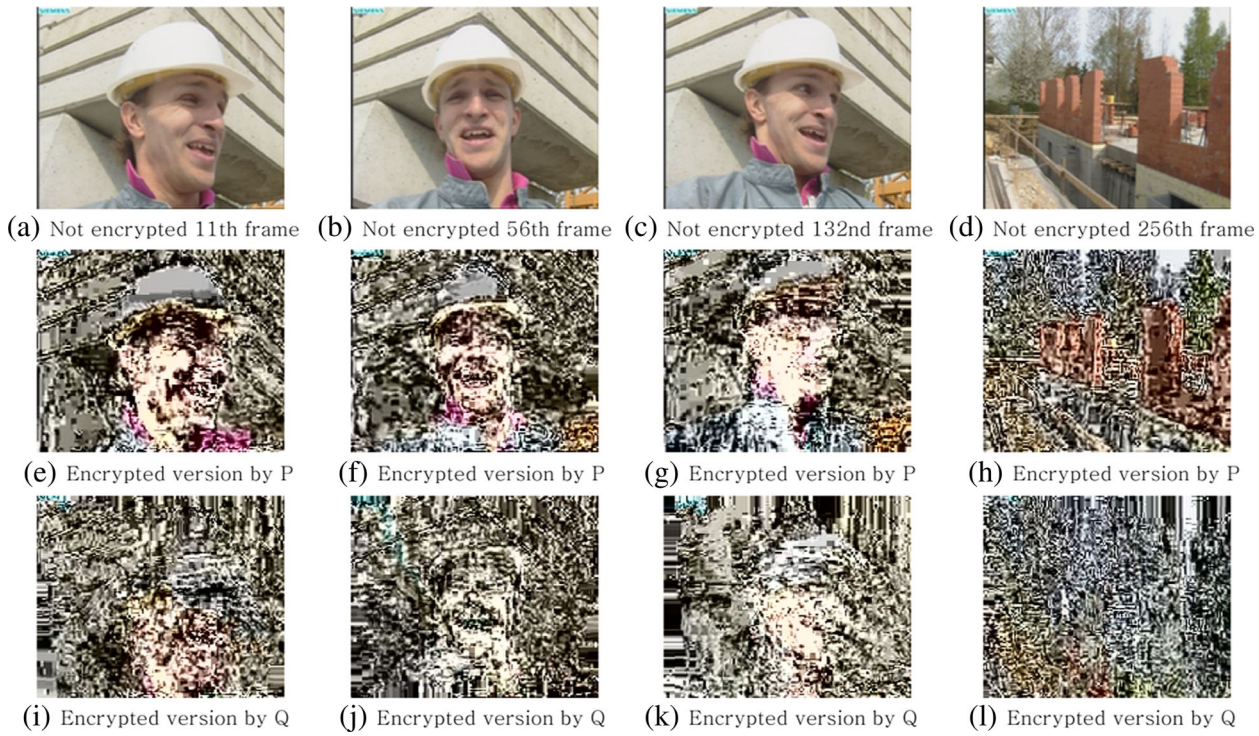


Fig. 6 a–l Sample frames of original and two encrypted videos by using different encryption techniques in scenario Extended-1. These figures show sample frames (11st, 58th, 132nd, and 256th) of original Foreman video and its encrypted versions using two JVCE schemes P and Q. P and Q differ only in terms of how they protect the temporal information of the video such as the motion vector. P and Q result in visually different encrypted videos. Regarding Table 5 together, Fig. 6 aims to present that the STVSM appropriately captures the difference of the protection on the temporal information in perceptual manner since it considers both the spatial and temporal information as compared to other metrics

efficiency, and the energy efficiency satisfying the given requirements and thresholds.

4.1 Bitrate and energy-bound visual security

In the design space of the JVCE, exploiting various sets of parameters result in different amount of the visual security, the compression efficiency, and the energy efficiency. The values of three indices have been turned out that they have tradeoff relationships between each other. Accordingly, it is difficult to find one optimal set of parameters among various ones to achieve all those indices to the best. Therefore, we like to find the interesting solution in the design space to maximize the visual security while satisfying the energy budget, the visual security threshold, and the compression bitrate threshold. In order to achieve that, we extend the design space exploration of the JVCE in terms of the visual security and the energy efficiency [29] by additionally considering the compression efficiency under several budgets. The compression efficiency should be considered since reducing the size of the video is the main goal of the video compression. Although the JVCE with one specific set of parameters achieves the maximum visual security under the budgets, the set could significantly degrade the compression efficiency.

Table 5 Objective and subjective evaluation results for encryption cases in Fig. 6

Encryption	1/*PSNR*	1-MSSIM	LFBVS	1-LEG	1-VIF	STVSM	MOS
P	0.122	0.463	0.666	0.880	0.934	0.561	3.530
Q	0.122	0.476	0.668	0.886	0.963	0.631	4.667

In order to examine all sets of parameters in the design space of the JVCE, we propose the BEVS (bitrate and energy-bound visual security) regarding the visual security, the compression efficiency, and the energy efficiency. The BEVS estimates the results of the JVCE with a set of parameters as the normalized level of the visual security while considering the given requirements. To calculate the BEVS, the energy consumption E, the compression bitrate B, and the visual security VS should be estimated after performing the JVCE for the input video by using a set of parameters as shown in the Eq. 1.

$$[E, B, VS] = JVCE(video,\ a\ set\ of\ parameters) \quad (1)$$

In our study, we utilize the STVSM to represent VS. Based on E, B, and VS, the BEVS in the Eq. 2 examines the JVCE with a set of parameters while considering the

compression bitrate threshold, the energy budget, and the visual security threshold denoted as $B_{threshold}$, E_{budget}, and $VS_{threshold}$, respectively. If the JVCE with a set of parameters satisfies all the requirements, the BEVS returns the normalized visual security ranged from 0 to 1. Otherwise, the BEVS is set to zero. By using the BEVS, we can find the interesting set of parameters which achieves the maximum visual security under the given requirements. Besides the BEVS, we can easily think the BVSE (bitrate and visual security-bound energy) and the EVSB (energy and visual security-bound bitrate) to find the maximum energy efficiency and the maximum compression efficiency, respectively under the same requirements. In the following subsections, we exploit all three proposed metrics for the design space exploration.

$$\begin{aligned} \text{BEVS} &= f(B, B_{\text{threshold}}, E, E_{\text{budget}}, VS, VS_{\text{threshold}}) \\ &= \begin{cases} \frac{VS}{VS_{\max}} & \text{iff } E \le E_{\text{budget}} \\ & \quad \& \; B \le B_{\text{threshold}} \\ & \quad \& \; VS \ge VS_{\text{threshold}} \\ 0 & \text{otherwise} \end{cases} \end{aligned} \tag{2}$$

4.2 Heuristic exploration algorithm for the design space of the JVCE

According to various parameters of the JVCE such as the GOP size and the number of utilized encryption methods, the design space of the JVCE contains lots of parameter sets to be examined. Therefore, an efficient exploration algorithm for the design space of the JVCE is required. In this work, we design a simple but effective heuristic algorithm to relieve the exploration overheads. Based on the experimental results of the JVCE in this work, we empirically design the exploration algorithm DSE in Algorithm 1. The DSE finds the interesting set of parameters among various ones to obtain the maximum BEVS while satisfying the compression bitrate threshold B_{thres}, the energy budget E_{bud}, and the visual security threshold VS_{thres}. The key idea of the DSE is that if we can classify the sets of parameters into groups resulting in different range of the visual security without overlapping, it is possible to save the exploration overheads by examining the group of higher visual security range prior to other groups. By doing so, the DSE does not examine all the sets at one time but classifies the total sets into several groups based on the visual security (line 2). The criteria to determine the number of groups and the branch points of those groups without overlapping should be specific to target applications. Further, the DSE searches for one set providing the maximum visual security group by group under the given requirements (lines 3–6). In this case, we examine the group of the range of higher visual security prior to the groups of the range of lower visual security. Therefore, if we find one set to obtain the maximum visual security within one group under the requirements, the DSE immediately finishes the exploration returning it as the solution (line 7).

Algorithm 1 Design space exploration (DSE)

```
1: procedure DSE(E_bud, B_thres, VS_thres)
2:     GROUPS ← groups of sets of parameters in terms of VS
3:     COND ← (E_bud, B_thres, VS_thres)
4:     for group ← GROUPS do
5:         for set ← group do
6:             if BEVS(set, COND) is MAX then
7:                 return set
```

Algorithm 2 represents the example of the customized exploration algorithm customDSE which is designed for the design space of the JVCE in this paper. In Algorithm 2, we exploit three observations from our experimental results; (1) the trend of the visual security differs according to the GOP size and the QP level, (2) the xor encryption significantly enhances the visual security, and (3) the groups determined by the options for the xor encryption and the range of the pseudo-random numbers have different ranges of the visual security which are not overlapped. Based on these observations, we decide to show the heuristic exploration algorithm under the fixed QP level since considering the combination of the GOP size and the QP level at once makes the algorithm too much complex and long to represent in this paper. Moreover, the GOP size is more related with the amount of the data to be encrypted by the schemes that we implement as compared to the QP level. Therefore, we conduct the design space exploration under the fixed QP level 25 while adjusting the GOP size utilized in this study (lines 4–34). Initially, we examine the sets of parameters under the GOP size 1 (lines 5–17). Under the GOP size 1, we classify the sets of parameters into three groups based on the xor encryption and the range of the pseudo-random numbers (lines 6–8). The *group1*, *group2*, and *group3* are in the decreasing order of the visual security and not overlapped. From *group1* to *group3*, if we find a set of parameters to satisfy all the requirements COND consisting of E_{bud}, B_{thres}, and VS_{thres}, the set is assigned to $set_{gop==1}$ (lines 9–17). For the GOP size 5, 10, and 30, we commonly classify the sets of parameters into two groups based on the option for the xor encryption (lines 19–21). And then the similar exploration of the GOP size 1 is conducted (lines 22–27). In case of the GOP size 300, there are no intuitive ways to group the sets of parameters so that we use the exhaustive search (lines 28–32). After the exploration, we have the candidate sets of parameters (line 35). Finally, we find the interesting set of parameters denoted as set_{result} to have the maximum BEVS among the *candidates* of five GOP sizes and the customDSE returns it (line 36 and 37).

Algorithm 2 Example of customized design space exploration (customDSE)

```
procedure CUSTOMDSE(E_bud, B_thres, VS_thres)
    GOP ← [1, 5, 10, 30, 300]
    COND ← (E_bud, B_thres, VS_thres)
    for gop ← GOP do
        if gop==1 then
         group1 ← sets w/ xor(large_r) for coefficients
         group2 ← sets w/ xor for coefficients except
                  group1
        group3 ← all sets except group1 and group2
            set ← MaxBEVS(group1, COND)
            if set != NULL then
                set_gop ← set; continue;
            set ← MaxBEVS(group2, COND)
            if set != NULL then
                set_gop ← set; continue;
            set ← MaxBEVS(group3, COND)
            if set != NULL then
                set_gop ← set; continue;
        else if gop==5 || gop==10 || gop==30 then
            group1 ← sets w/ xor(large_r) for
                     coefficients and motion vectors
            group2 ← all sets except group1
            set ← MaxBEVS(group1, COND)
            if set != NULL then
                set_gop ← set; continue;
            set ← MaxBEVS(group2, COND)
            if set != NULL then
                set_gop ← set; continue;
        else if gop==300 then
            group ← all parameter sets
            set ← MaxBEVS(group, COND)
            if set != NULL then
                set_gop ← set; continue;
        else
            Wrong GOP size!
    candidates ←[ set_1, set_5, set_10, set_30, set_300]
    set_result ← MaxBEVS(candidates, COND)
    return set_result
```

5 Results and discussion

Based on the BEVS, we explore the design space consisting of various sets of parameters after performing the JVCE for three QCIF video. Details of the experimental setup are described in the Joint video compression and encryption section. For the clear understanding, all the experimental results for the Foreman video are examined in details while the results for the Akiyo and the Coastguard videos are summarized. The JVCE with all possible sets of parameters for the Foreman video results in huge space between the minimum and the maximum in terms of the visual security, the compression efficiency, and the energy efficiency as shown in Table 6. In this paper, we empirically decide the budgets for the visual security, the energy consumption, and the compression bitrate as 0.50, 29.84 joule, and 0.5 Mbps, respectively by averaging all the results in our experiment. The budgets can be practically determined by the requirements of the mobile video applications in real cases.

Figure 7 shows the interesting area within the whole design space for the Foreman video consisting of 1629 sets of 7 parameters of the JVCE. In this area, there exist the sets of parameters to obtain the maximum and the minimum BEVS under the budgets. Among the results, the sets to satisfy the visual security threshold and the energy budget locate in the left-top region among four regions in the graph divided by vertical and horizontal dotted lines which represent the budgets. Based on the EVS (energy-bound visual security) [29] which does not consider the compression efficiency, all sets in this region can be the candidate for the efficient solution. However, the mobile video applications should significantly consider the compression efficiency because of the huge size of the video data. By additionally considering the compression efficiency using the BEVS, we find the sets denoted as dots in this interesting region in Fig. 7. The sets represented as triangles in this region fail to be determined as the efficient solutions when considering the visual security, the compression efficiency, and the energy efficiency all together. Finally, we can achieve up to 26.6% visual security enhancement in this experiment by comparing the minimum and the maximum of BEVS while satisfying the budgets of three indices as shown in Table 7. The interesting set of the maximum BEVS is xor_enc(large_r){DCAC}shuffle_enc{sign_enc{xor_enc (large_r){MV}}} under the GOP size 10 and the QP level 25. Using this design space exploration, we can suggest the parameter set of the JVCE to satisfy each requirement. For example, applying xor_enc to both DC and AC coefficients and all of sign_enc, xor_enc, and shuffle_enc to motion vectors can guarantee high confidentiality for important medical images. Its perceptual protection result is shown in Fig. 5(p). Further, we can reduce the bitrate about 29% without increasing energy consumption by replacing the encryption scheme for the

Table 6 Size of the design space between the minimum and the maximum in terms of the visual security, the compression efficiency, and the energy efficiency (Foreman video)

	Min	Max	Size (%)
Visual security (STVSM)	0.119	0.634	430.8
Energy efficiency (joule)	14.839	34.799	134.5
Compression efficiency (Mbps)	0.243	6.790	2692.8

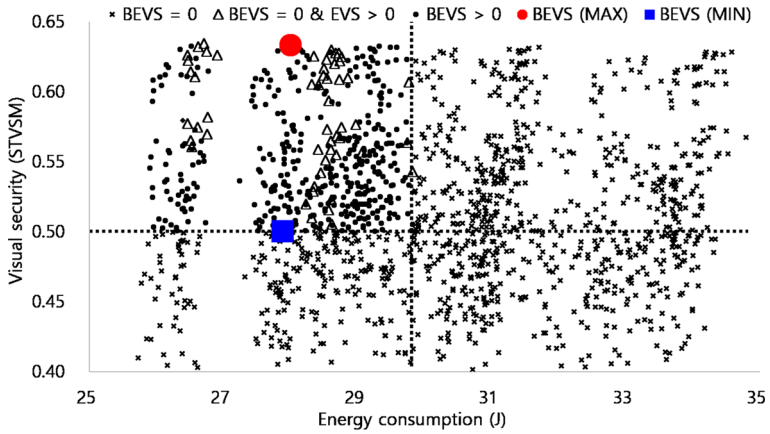


Fig. 7 Visual security enhancement through the BEVS-based design space exploration of the JVCE (Foreman video). This figure shows the design space of the JVCE which has been studied in this paper. Each dot represents the result after conducting the JVCE with one set of values of parameters. Thanks to the BEVS, we can maximize the visual security by selecting the optimal condition of the maximum BEVS instead of the one of the minimum BEVS while satisfying the requirements of the visual security, the compression bitrate, and the energy consumption

motion vectors with sign_enc only. This should be effective for the video surveillance with limited bandwidth. In this case, the perceptual security looks like Fig. 5(i). Although these two example JVCEs are conducted under the same compression parameter of the GOP size 30 and the QP level 25, varying those parameters, of course, changes the visual security, the energy efficiency, and the compression efficiency.

For the Akiyo and the Coastguard videos, similar design space explorations using the BEVS are conducted. In the experimental results under the empirical budgets of the visual security, the compression bitrate, and the energy consumption, we obtain up to 36.7% and 30.5% visual security improvement for the Akiyo and the Coastguard, respectively. Consequently, thanks to the design exploration with the BEVS, we enhance the visual security under the budgets in this study by selecting the set of parameters of the maximum BEVS instead of that of the minimum BEVS.

Moreover, we study the design space of the JVCE in terms of the EVSB and the BVSE to find the interesting set of parameters to achieve the maximum compression efficiency and the energy efficiency, respectively, under the same budgets. For the Foreman video, we obtain up to 267.9% compression efficiency improvement and the 15.1% energy efficiency enhancement by comparing the minimum and the maximum results. Similarly, it is demonstrated that the compression efficiency and the energy efficiency are improved as 25.2% and 7.4%, respectively, for the Akiyo and 245.6% and 16.7%, respectively, for the Coastguard.

Further, we exploit the customDSE to save the exploration overhead for the design space of the JVCE. Thanks to the customDSE, the exploration for the GOP size 10 ends up with the group1 prior to investigating the group2 in Algorithm 2. Accordingly, we reduce the number of the sets of parameters to be investigated as about 19.9% for the Foreman video. Among the sets of the local maximum BEVS for each GOP size, we conclude the interesting set to have the global maximum BEVS. The amount of the reduction in the exploration overheads differ based on the utilized encryption techniques and the type of the video. Moreover, we can decrease the amount of the exploration overhead through more sophisticated customization for the target design space.

Note that as more performance indices and parameters should be considered, the complexity of the design space of the JVCE will exponentially increase. The proposed BEVS is flexible enough to include additional performance indices such as QoS of the reconstructed video. In addition, the proposed DSE algorithm can be configurable according to other criteria such as size of the compressed output instead of the visual security. Even for the larger and more complex design space of the JVCE, our work will contribute to maximize the targeted performance indices while properly satisfying the system and user requirements.

Table 7 Visual security improvement through the design space exploration based on the BEVS for three QCIF videos

Video	Min BEVS	Max BEVS	Improvement
Akiyo	0.474	0.648	36.7%
Coastguard	0.465	0.607	30.5%
Foreman	0.500	0.633	26.6%

6 Conclusion

In this paper, we thoroughly investigate the interesting design space of the JVCE consisting of various sets of parameters in terms of the visual security, the compression efficiency, and the energy efficiency. For the design space exploration, we quantify those three indices in the objective manner. In order to quantify the visual security which is the visual distortion from the human eye's perception, we conduct in-depth study on the visual security evaluation in terms of various metrics. As compared to the conventional metrics, the STVSM which is our previous work accurately captures the human eye's perception. Further, we propose the BEVS to estimate every set of parameters as the normalized visual security while investigating the results under the given requirements. In addition, we design a heuristic exploration algorithm customDSE to reduce the exploration overhead of the design space. Finally, thanks to the STVSM and the BEVS, we improve the visual security up to 36.7% under the visual security threshold, the energy budget, and the compression bitrate threshold in this study. Further, the customDSE show about 19.9% reduction in the exploration overhead of the design space.

In future works, we examine the JVCE for the videos of higher resolutions by exploiting more parameters such as the kind of the motion estimation algorithm and other video encryption techniques. Further, we study the design space of the JVDD (Joint Video Decompression and Decryption) while additionally investigating the QoS of the reconstructed video. Finally, we would like to explore the total design space of the combination of the JVCE and the JVDD in order to find the optimal condition for the mobile video communication services.

Abbreviations

AC: Alternating current; AES: Advanced encryption standard; BEVS: Bitrate and energy-bound visual security; BVSE: Bitrate and visual security-bound energy; Coeff.: Coefficient; DC: Direct current; DCT: Discrete cosine transform; DES: Data encryption standard; EE: Entropy encoding; EVSB: Energy and visual security-bound bitrate; EVS: Energy-bound visual security; GOP: Group of picture; IoT: Internet of things; JVCE: Joint video compression and encryption; JVDD: Joint video decompression and decryption; LEG: Local edge gradient; LFBVS: Local feature-based visual security; ME: Motion estimation; MOS: Mean opinion score; MSSIM: Mean SSIM; MV: Motion vector; PPMCC: Pearson product-moment correlation coefficient; PSNR: Peak signal-to-noise ratio; Q: Quantization; QCIF: Quarter common intermediate format; QoS: Quality of service; QP: Quantization parameter; SRCC: Spearman's rank correlation coefficient; SSIM: Structural SIMilarity; STVSM: Spatio-temporal visual security metric; VD: Video decompression; VIF: Visual information fidelity; VSE: Visual security evaluation

Acknowledgements

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2016-0-00562, Emotional Intelligence Technology to Infer Human Emotion and Carry on Dialogue Accordingly).

Funding

This work was supported by Institute for Information & Communications Technology Promotion (IITP) grant funded by the Korea government (MSIT) (2016-0-00562, Emotional Intelligence Technology to Infer Human Emotion and Carry on Dialogue Accordingly).

Availability of data and materials

The dataset(s) supporting the conclusions of this article is(are) available in the github repository, https://github.com/junhyungmoon/jvce.

Authors' contributions

All the authors contributed to the original research in the paper. All of them read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Author details

[1]Department of Computer Science, Yonsei University, Yonsei-ro 50, Seoul, Republic of Korea. [2]Department of Software, Sangmyung University, Sangmyeongdae-gil 31, Cheonan-si, Republic of Korea.

Received: 22 May 2018 Accepted: 3 December 2018
Published online: 14 January 2019

References

1. A. Massoudi, F. Lefebvre, C. De Vleeschouwer, B. Macq, J.-J. Quisquater, Overview on selective encryption of image and video: challenges and perspectives. EURASIP J. Inf. Secur. **2008**, 5 (2008)
2. F. Liu, H. Koenig, A survey of video encryption algorithms. Comput. Secur. **29**(1), 3–15 (2010)
3. S. Jana, A. Pande, A. Chan, P. Mohapatra, Mobile video chat: issues and challenges. IEEE Commun. Mag. **51**(6), 144–151 (2013)
4. C. Xiao, T. Li, N. Yang, L. Wang, S. Ma, in *Computing, Communications and IT Applications Conference (ComComAp), 2014 IEEE*. A speed adjustable scheme based selective encryption control model for large-scale multimedia sensing system under resources constraints (IEEE, 2014), pp. 104–109
5. P. Deshmukh, V. Kolhe, in *Information Communication and Embedded Systems (ICICES), 2014 International Conference On*. Modified aes based algorithm for mpeg video encryption (IEEE, 2014), pp. 1–5
6. W. Zeng, S. Lei, in *Proceedings of the Seventh ACM International Conference on Multimedia (Part 1)*. Efficient frequency domain video scrambling for content access control (ACM, 1999), pp. 285–294
7. Z. Liu, X. Li, in *Multimedia Modelling Conference, 2004. Proceedings. 10th International*. Motion vector encryption in multimedia streaming (IEEE, 2004), pp. 64–71
8. Y. Li, L. Liang, Z. Su, J. Jiang, in *2005 5th International Conference on Information Communications & Signal Processing*. A new video encryption algorithm for h. 264 (IEEE, 2005), pp. 1121–1124
9. Y. Wang, M. O'Neill, F. Kurugollu, in *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*. The improved sign bit encryption of motion vectors for h. 264/avc (IEEE, 2012), pp. 1752–1756
10. M. Pazarci, V. Dipcin, A mpeg2-transparent scrambling technique. IEEE Trans. Consum. Electron. **48**(2), 345–355 (2002)
11. M. Grangetto, E. Magli, G. Olmo, Multimedia selective encryption by means of randomized arithmetic coding. IEEE Trans. Multimed. **8**(5), 905–917 (2006)
12. S. Lian, J. Sun, G. Liu, Z. Wang, Efficient video encryption scheme based on advanced video coding. Multimedia Tools Appl. **38**(1), 75–89 (2008)
13. Z. Shahid, M. Chaumont, W. Puech, Fast protection of h. 264/avc by selective encryption of cavlc and cabac for i and p frames. IEEE Trans. Circ. Syst. Video Technol. **21**(5), 565–576 (2011)
14. M. Roy, C. Pradhan, in *Electronics Computer Technology (ICECT), 2011 3rd International Conference On, vol. 2*. Secured selective encryption algorithm for mpeg-2 video (IEEE, 2011), pp. 420–423
15. C. Mansour, D. Chasaki, in *2014 International Wireless Communications and Mobile Computing Conference (IWCMC)*. Power-aware selective encryption of video transmissions in smart devices (IEEE, 2014), pp. 967–972

16. J. Moon, H. So, K. Lee, in *Systems, Man, and Cybernetics (SMC), 2016 IEEE International Conference On*. Configurable privacy management for secure video surveillance in energy-constrained systems (IEEE, 2016), pp. 003800–003805
17. X. Di, Y. Wang, J. Li, L. Cong, H. Qi, Y. Zhang, in *Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), 2017 10th International Congress On*. An optimized video selective encryption algorithm (IEEE, 2017), pp. 1–5
18. B. Boyadjis, C. Bergeron, B. Pesquet-Popescu, F. Dufaux, Extended selective encryption of h. 264/avc (cabac)-and hevc-encoded video streams. IEEE Trans. Circ. Syst. Video Technol. **27**(4), 892–906 (2017)
19. E. Magli, M. Mancin, L. Merello, in *Multimedia and Expo, 2003. ICME'03. Proceedings. 2003 International Conference On, vol. 3*. Low-complexity video compression for wireless sensor networks (IEEE, 2003), p. 585
20. S. Mohapatra, R. Cornea, N. Dutt, A. Nicolau, N. Venkatasubramanian, in *Proceedings of the Eleventh ACM International Conference on Multimedia*. Integrated power management for video streaming to mobile handheld devices (ACM, 2003), pp. 582–591
21. S. Mohapatra, R. Cornea, H. Oh, K. Lee, M. Kim, N. Dutt, R. Gupta, A. Nicolau, S. Shukla, N. Venkatasubramanian, in *19th IEEE International Parallel and Distributed Processing Symposium*. A cross-layer approach for power-performance optimization in distributed mobile systems (IEEE, 2005), p. 8
22. J. Moon, K. Lee, in *Proceedings of the 7th ACM International Workshop on Mobile Video*. Spatio-temporal visual security metric for secure mobile video applications (ACM, 2015), pp. 9–14
23. C. N. Taylor, S. Dey, D. Panigrahi, in *Software Radio*. Energy/latency/image quality tradeoffs in enabling mobile multimedia communication (Springer, London, 2001), pp. 55–66
24. C.-F. Chiasserini, E. Magli, in *Personal, Indoor and Mobile Radio Communications, 2002. The 13th IEEE International Symposium On, vol. 5*. Energy consumption and image quality in wireless video-surveillance networks (IEEE, 2002), pp. 2357–2361
25. T.-H. Lan, A. H. Tewfik, A resource management strategy in wireless multimedia communications-total power saving in mobile terminals with a guaranteed qos. IEEE Trans. Multimed. **5**(2), 267–281 (2003)
26. S. Mohapatra, N. Dutt, A. Nicolau, N. Venkatasubramanian, Dynamo: A cross-layer framework for end-to-end qos and energy optimization in mobile handheld devices. IEEE J. Sel. Areas Commun. **25**(4), 722–737 (2007)
27. K. Lee, M. Kim, N. Dutt, N. Venkatasubramanian, in *Distributed Embedded Systems: Design, Middleware and Resources*. Error-exploiting video encoder to extend energy/qos tradeoffs for mobile embedded systems (Springer, Boston, 2008), pp. 23–34
28. C. Herglotz, R. Rosales, M. Glaß, J. Teich, A. Kaup, in *Picture Coding Symposium (PCS), 2016*. Multi-objective design space exploration for the optimization of the hevc mode decision process (IEEE, 2016), pp. 1–5
29. J. Moon, K. Lee, in *Embedded Systems For Real-time Multimedia (ESTIMedia), 2015 13th IEEE Symposium On*. Integrated visual security management for video encryption in limited battery devices (IEEE, 2015), pp. 1–8
30. P. FIPS, 46-3: Data encryption standard (des). Natl Inst. Stand. Technol. **25**(10), 1–22 (1999)
31. N. F. Pub, 197: Advanced encryption standard (aes). Fed. Inf. Process. Stand. Publ. **197**, 441–0311 (2001)
32. B. Bhargava, C. Shi, S.-Y. Wang, Mpeg video encryption algorithms. Multimedia Tools Appl. **24**(1), 57–79 (2004)
33. L. Qiao, K. Nahrstedt, Comparison of mpeg encryption algorithms. Comput. Graph. **22**(4), 437–448 (1998)
34. C. N. Raju, G. Umadevi, K. Srinathan, C. Jawahar, in *Computer Vision, Graphics & Image Processing, 2008. ICVGIP'08. Sixth Indian Conference On*. Fast and secure real-time video encryption (IEEE, 2008), pp. 257–264
35. S. Lian, Z. Liu, Z. Ren, H. Wang, Secure advanced video coding based on selective encryption algorithms. IEEE Trans. Consum. Electron. **52**(2), 621–629 (2006)
36. D. Socek, H. Kalva, S. S. Magliveras, O. Marques, D. Culibrk, B. Furht, New approaches to encryption and steganography for digital videos. Multimedia Systems. **13**(3), 191–204 (2007)
37. University of Salzburg, VQI – Visual Quality Iindex Implementations (2018). http://www.wavelab.at/sources/VQI. version 1.x
38. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity. IEEE Trans. Image Process. **13**(4), 600–612 (2004)
39. H. Hofbauer, A. Uhl, in *EUVIP*. An effective and efficient visual quality index based on local edge gradients (Citeseer, 2011), pp. 162–167
40. H. R. Sheikh, A. C. Bovik, in *Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference On, vol. 3*. Image information and visual quality (IEEE, 2004), p. 709
41. P. C. Teo, D. J. Heeger, in *IS&T/SPIE 1994 International Symposium on Electronic Imaging: Science and Technology*. Perceptual image distortion (International Society for Optics and Photonics, 1994), pp. 127–141
42. M. P. Eckert, A. P. Bradley, Perceptual quality metrics applied to still image compression. Signal Process. **70**(3), 177–200 (1998)
43. S. Winkler, in *Electronic Imaging'99*. Perceptual distortion metric for digital color video (International Society for Optics and Photonics, 1999), pp. 175–184
44. L. Tong, F. Dai, Y. Zhang, J. Li, in *Proceedings of the 18th ACM International Conference on Multimedia*. Visual security evaluation for video encryption (ACM, 2010), pp. 835–838
45. W. Lin, C.-C. J. Kuo, Perceptual visual quality metrics: A survey. J. Vis. Commun. Image Represent. (2011)
46. G.-H. Chen, C.-L. Yang, S.-L. Xie, in *2006 International Conference on Image Processing*. Gradient-based structural similarity for image quality assessment (IEEE, 2006), pp. 2929–2932
47. H. Hofbauer, A. Uhl, Identifying deficits of visual security metrics for images. Signal Process. Image Commun. **46**, 60–75 (2016)
48. J. Klaue, B. Rathke, A. Wolisz, in *International Conference on Modelling Techniques and Tools for Computer Performance Evaluation*. Evalvid–a framework for video transmission and quality evaluation (Springer, 2003), pp. 255–272
49. K. Pearson, Note on regression and inheritance in the case of two parents. Proc. R. Soc. Lond. **58**, 240–242 (1895)
50. C. Spearman, The proof and measurement of association between two things. Am. J. Psychol. **15**(1), 72–101 (1904)