IET Information Security

Research Article

Design of new scan orders for perceptual encryption of H.264/AVC videos

ISSN 1751-8709
Received on 15th November 2015
Revised 21st February 2016
Accepted on 4th March 2016
E-First on 12th September 2016
doi: 10.1049/iet-ifs.2015.0492
www.ietdl.org

Xiangling Ding[1,2], Yingzhuo Deng[1], Gaobo Yang[1] ✉, Yun Song[1], Dajiang He[2], Xingming Sun[3]

[1]School of Information Science and Engineering, Hunan University, Changsha 410082, People's Republic of China

[2]Key Laboratory of Intelligent Control Technology for Ecological Agriculture in Wuling Mountain Areas of Hunan Province, Huaihua University, Huaihua 418008, People's Republic of China

[3]School of Computer and Software, Nanjing University of Information Science and Technology, Nangjing 210044, People's Republic of China

✉ *E-mail: yanggaobo@hnu.edu.cn*

Abstract: In this study, a perceptual encryption algorithm is proposed for H.264/AVC video to enhance the scrambling effect and encryption space. Six new scan orders are designed for H.264/AVC encoder by analysing the energy distribution of discrete cosine transform coefficients. They are proven to have similar performance as the conventional zigzag scan order and its symmetrical scan order. These six new scan orders are combined with two existing scan orders to design a scan-order based perceptual encryption algorithm. Specifically, video encryption is achieved more specifically by randomly selecting one scan order from the eight scan orders with a security key, and the sign bit flipping of DC coefficients is also incorporated to further increase the encryption space. Experimental results show that the proposed approach has the advantages of both low bitrate increase and low computational cost. Furthermore, it is more flexible and has stronger security than the existing scan-order based video encryption schemes.

1 Introduction

With the advancements of video compression and transmission techniques, it has become much easier than ever to exchange video information with any person at any time. To prevent the illegal access to video information, video encryption is one of the most important research topics in the field of information security [1, 2]. Most existing video encryption approaches are complete encryption, leaving only a very small portion for the perceptual encryption. For complete encryption approaches, video sequences are fully encrypted and no meaningful visual information can be reconstructed by an un-authorised user. These complete video encryption approaches have high securities, which mainly benefit from traditional ciphers. However, they usually have disadvantages such as high complexity and format in-compatibility. Thus, they are more suitable for strict security applications such as secure video storage. For more popular entertainment applications such as video-on-demand (VoD), pay-TV and live video broadcasting, their encryptions have some special requirements. First, the encrypted video should only degrade its visual quality, but still retains some video information as a preview to attract interested users. Second, the encrypted video bit-streams should be decoded by any normal video decoder, even without knowing the encryption key. Third, the encryption scheme should be efficient and simple, which implies that it cannot be broken easily but may not necessarily be immune to some complicated attacks. This leads to the so-called perceptual encryption, in which a user can still obtain some visible video contents (but at an annoying quality) even without knowing the encryption key. There are also some other video encryption approaches which are referred as partial encryption or selective encryption [3], which have many similarities with perceptual encryption but put emphasis on the encryption of partial regions such as region of interest. Please note that since partial encryption or selective encryption share similar application scenarios and encryption purposes with the perceptual encryption, the terminology of perceptual encryption is used in the rest paper for simplicity.

The existing perceptual encryption approaches can be divided into two categories, either independent of video coding or combined with a video encoder. In general, video encryption combined with an encoder can achieve better encryption performance by fully exploiting the flexibility of video encoder, which makes it the mainstream of current perceptual video encryption. Therefore, this category can be incorporated into different stages of video encoding such as transform [3–6], discrete cosine transform (DCT) coefficients post-processing [7, 8], entropy coding [9–15], and zigzag scanning [13, 14] and so on. Transform-based video encryption is to design new transforms, which are alternately employed in video encoder by a security key. Yeung *et al.* [3] proposed to encrypt video by alternately using four new unitary transforms for 4 × 4 blocks, which is controlled by a pre-designed security key. Later, they extended this encryption method to 8 × 8 block and designed a new one-dimensional 16 point (4 × 4 block) DCT [4, 5]. In addition, Bing *et al.* [6] proposed a perceptual encryption approach for H.264/AVC videos by randomly embedding sign-flips into the butterfly structure of integer-based transform, which further increases the key space. The new alternative transforms are as efficient as DCT in these transform-based encryption approaches. However, they are actually obtained by an angle rotation, which significantly increases the computational complexity of transform coding. For the DCT coefficients post-processing, Li *et al.* [7] and Magli *et al.* [8] proposed two perceptual encryption approaches by flipping the sign bit of DCT coefficients or directly encrypting the DCT coefficients with a linear transformation. Since each DCT coefficient is processed individually, which greatly increases the computational complexity as well. Moreover, the statistical distribution of video data might be changed, which degrades the subsequent entropy coding and increases the final bitrate. Entropy coding is the final step for video encoder, which is lossless. It is not difficult to flip the sign bits of motion vector residuals [9], intra prediction modes [10], encrypt some bits in intra-coded/inter-coded macroblock [11, 12] and other significant codewords (i.e. trailing coefficients etc.) [9, 13–15] in this stage, which meets the real time requirements. However, the video encryption schemes in the stage of entropy coding have a common drawback that if the encrypted codewords are the fixed-length parts, they can be independently extracted from video bitstream, thereby having side effects on the security of encryption. Up to now, there are relatively few video encryption approaches in the coefficient scanning stage. The most

© The Institution of Engineering and Technology 2016

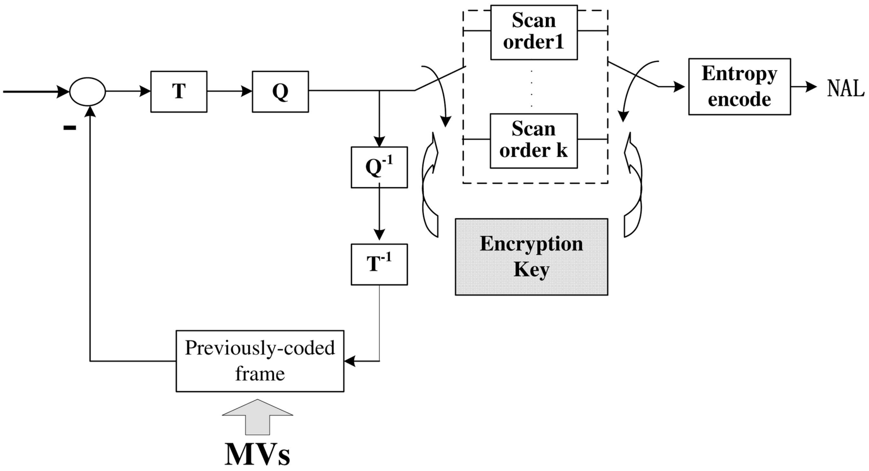


Fig. 1 *Block diagram of the proposed scan-order based perceptual video encryption*

representative works are summarised as follows: Tang [16] first proposed a zigzag-permutation based encryption algorithm for MPEG video. The basic idea is that the 64 quantised DCT coefficients are scanned by a random permutation list instead of the conventional zigzag scan order for each 8 × 8 block. Although it can achieve efficient encryption, it does not provide enough security against known-plaintext and ciphertext-only attacks [17]. Moreover, its compression performance serious decreases up to 45% for MPEG, simply because the conventional zigzag scan is replaced by a random scan, whereas the conventional zigzag scan is an optimised design for effective compression in the subsequent entropy coding. It cannot achieve perceptual video encryption as well. To maintain the compression efficiency and achieve perceptual encryption, Wang *et al.* [18] presented a novel perceptual encryption approach by improved randomised zigzag scanning (IRZZ) for H.264/AVC video. Its basic idea is that the alternative scan order in the symmetrical direction can produce a DCT coefficient sequence similar to the conventional zigzag scan order, because it is also symmetrical along the main diagonal line. Thus, the alternative symmetrical scan order has no side effects on the coding efficiency of subsequent entropy coding. However, when the conventional zigzag scan order is used in the decoding procedure, a macroblock scanned with the symmetrical scan order can be decoded as the transpose of the block which is produced by the conventional zigzag scan order. Thus, the approach has limited scrambling effect. Moreover, it has small encryption space since the IRZZ approach is achieved by randomly selecting one from two scan orders. Therefore, the design of new scan orders and incorporating them into the SCAN pattern-based perceptual video encryption is worth further investigation, which is anticipated to achieve bigger encryption space and a more desirable scrambling effect.

In this paper, a real-time perceptual video encryption approach is proposed for H.264/AVC video to achieve more desirable scrambling effects. Motivated by the IRZZ approach [18], a set of new scan orders are designed by analysing the energy distribution of DCT coefficients within a block. These new scan orders have similar performance with conventional zigzag scan order adopted by H.264/AVC encoder. Then, an improved perceptual video encryption approach is presented by introducing the new scan orders into video encryption. The contributions of the proposed approach are four-folds. First, a set of new scan orders are proposed by considering the energy distribution of DCT coefficients, which can meet the requirement of successive entropy coding. It is proved by theoretical analysis and experimentally verified that they have similar performance with conventional zigzag scan order. Second, the proposed new set of scan orders is combined with the existing two scan orders for perceptual video encryption. A security key is exploited to randomly select one scan order from a scan pattern set with more scan orders. Thus, the proposed approach has much bigger encryption space than the existing IRZZ approach. Meanwhile, the flipping of sign bits is also exploited for DC coefficients to further increase the encryption space and the security. Third, compared with the IRZZ based encryption which exploits the symmetrical scan order, the decoded blocks will not be simply the transpose of the original residual block. Therefore, the proposed approach achieves much better scrambling effect. Forth, the proposed approach still keeps real-time performance and low increase of bitrates, even it introduces some new scan orders.

The rest of the paper is organised as follows: Section 2 presents the new scan orders for video encoder, and both theoretical analysis and experimental results are provided to compare their performances with the conventional zigzag scan order. Section 3 presents the improved perceptual video encryption approach by using randomised scan orders. Section 4 reports the experimental results and security analysis. We conclude this paper in Section 5.

2 New scan orders and their performance analysis

Fig. 1 shows the block diagram of the proposed perceptual encryption approach, which is integrated with the H.264/AVC video encoder. We choose H.264/AVC video coding standard because of its excellent coding performance and wide applications. The proposed approach is also based on randomised scan orders, but the set of scan orders is enlarged with new scan orders. Because the scan process is an integral step for video encoder, the proposed approach does not lead to too much extra computations. However, since the change of scan order might have influence on subsequent entropy coding, which might further increase the bitrates of both encoded and encrypted video. Therefore, the design of new scan order must consider the requirement of entropy coding, and tradeoff should be made between the increase of encryption space and the increase of final bitrate after encryption. In the following, we first propose a set of new scan orders, and then prove their similar performances with zigzag scan order by both theoretical analysis and experimental results. Then, the proposed video encryption approach is discussed in detail.

2.1 Design of new scan orders

For H.264/AVC, the quantised DCT coefficients are first scanned for subsequent entropy coding. After scanning stage, a one-dimensional array of DCT coefficients is obtained. The purpose of the scanning stage is to gather the non-zero quantised coefficients with large amounts of energy and leave zero or near-zero coefficients in the tail, which will facilitate subsequent entropy coding to improve the coding efficiency. When all quantised DCT coefficients or all AC components in the quantised DCT coefficients are scanned in a random permutation pattern, it will lead to improper distribution of non-zero DCT coefficients after scanning. That is, though this kind of random permutation scan pattern increases the key space, but does not meet the requirements of subsequent entropy coding, which will inevitably lead to the increase of final bitrate after video compression and encryption.

© The Institution of Engineering and Technology 2016

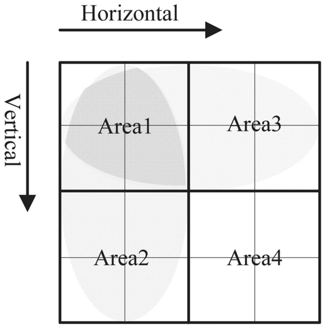


Fig. 2 *Energy distribution of four areas in a 4 × 4 block*

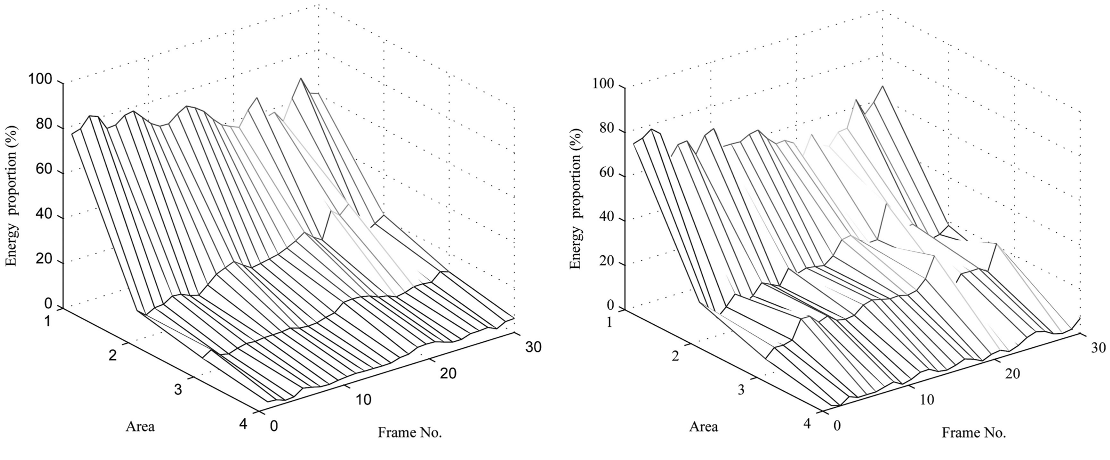


Fig. 3 *Energy distribution of quantised DCT coefficients in four areas*
a Luminance component
b Chrominance component

Actually, this has been verified in the existing works [17]. Therefore, the design of new scan orders should fully consider the energy distribution of all DCT coefficients in a 4 × 4 block to tradeoff among the encryption space, bitrate and coding efficiency.

Instead of the 8 × 8 block for transform coding in previous video coding standards, the H.264/AVC encoder adopts a novel 4 × 4 integer transform [2]. The 4 × 4 DCT coefficients can be divided into four areas, as shown in Fig. 2, to describe the energy distribution of these DCT coefficients. Area 1 contains four coefficients in the upper left-hand side corner of the 4 × 4 DCT coefficients. Area 2 and area 3 contain the frequency components in vertical and horizontal directions except area 1, respectively. Area 4 contains four coefficients in the bottom right-hand side corner of the 4 × 4 coefficient block. To model the energy distribution of the 4 × 4 coefficient block, the concept of energy packing efficiency [3] is introduced as follows:

$$EPE = \sum_{t=0}^{M_0 - 1} \frac{E\{S_t^2\}}{\sum_{t=0}^{15} E\{S_t^2\}} \tag{1}$$

It reflects the percentage of energy for the former M_0 coefficients (after zigzag scanning) to the entire block energy, where S_t refers to the tth $(0 < t < 15)$ coefficient in a block, and $E\{S_t^2\}$ is the energy of the tth coefficient. Thus, the energy percentage can be calculated for each DCT coefficient in this block by simply setting different M_0.

Some experiments are performed on *Foreman* sequence to verify the energy distribution of 4 × 4 coefficient block. The original *Foreman* sequence (QCIF, 4:2:0, 30 frames) is encoded with the reference software of H.264/AVC [19] (JM10.2, baseline profile, IPPPP, QP value 18). Fig. 3 shows the energy distribution of luminance and chrominance components, respectively. Apparently, area 1 contains about 60–80% energy for its luminance component, and more than 60% energy for its chrominance component. That is, area 1 occupies most energies of the entire block, which means that the block energy is gathered by these four low-frequency coefficients in the upper left-hand side corner. Thus, if the scan orders of the coefficients in area 1 are adjusted, it can still retain the fact that the four coefficients with relatively bigger energies stay ahead after scanning. Since the DC coefficient is involved in the Hadamard transformation [20], only the order of the rest three AC coefficients in area 1 should be adjusted to design new scan orders.

There are eight candidate scan orders for these three coefficients including the conventional zigzag scan order and the symmetrical scan order. They are shown in Fig. 4, in which the left-hand side column shows the conventional zigzag scan order and the symmetrical scan order and the rest three columns are the proposed six scan orders (S1–S6). For the coefficients of the rest three areas, they may keep the order of either the conventional zigzag scan or its symmetrical order. However, since the encryption performance by changing the scan direction of DCT coefficients in these three areas has been analysed in literature [18], similar analysis is not provided here to avoid repetition. In the following, the performance of new scan orders by changing the first three AC coefficients are discussed. Moreover, they can provide sufficient encryption performance, which is discussed in Section 3.

2.2 Experimental analysis of the new scan orders

Bitrate is an important performance evaluation metric for video encryption approaches. Since the proposed video encryption approach is combined with video encoder, the proposed scan orders are tested with the reference software of H.264/AVC (baseline profile, IPPPP, QP value 18) to evaluate their potential influences towards the bitrate of encoded video. The first 30 frames are tested for three typical video sequences including *Foreman*, *Mobile* and *Coastguard*. Comparisons are made among the conventional zigzag scan order, symmetrical scan order [18] and the proposed new scan orders (S1–S6) when they are used for the scanning of quantised DCT coefficients, respectively.

Table 1 summarises the results of bitrate comparison, where bitrate difference refers to the percentage of bitrate change. The proposed scan orders (S1–S6) and the symmetrical scan order are used in H.264/AVC encoder to compare with the ground-truth bitrate by the conventional zigzag scan order. For the scan orders (S1–S6), their biggest values of average bitrate increases under different QPs are 0.49, 1.61, 0.96, 0.40, 1.41 and 1.84%, respectively. Apparently, there are less bitrate decreases for S1, S3 and S4 when they are compared with the symmetrical scan order. For the rest scan orders S2, S5 and S6, though their bitrate increases are bigger than the symmetrical scan order, but they are still acceptable. Meanwhile, compared with the symmetrical scan

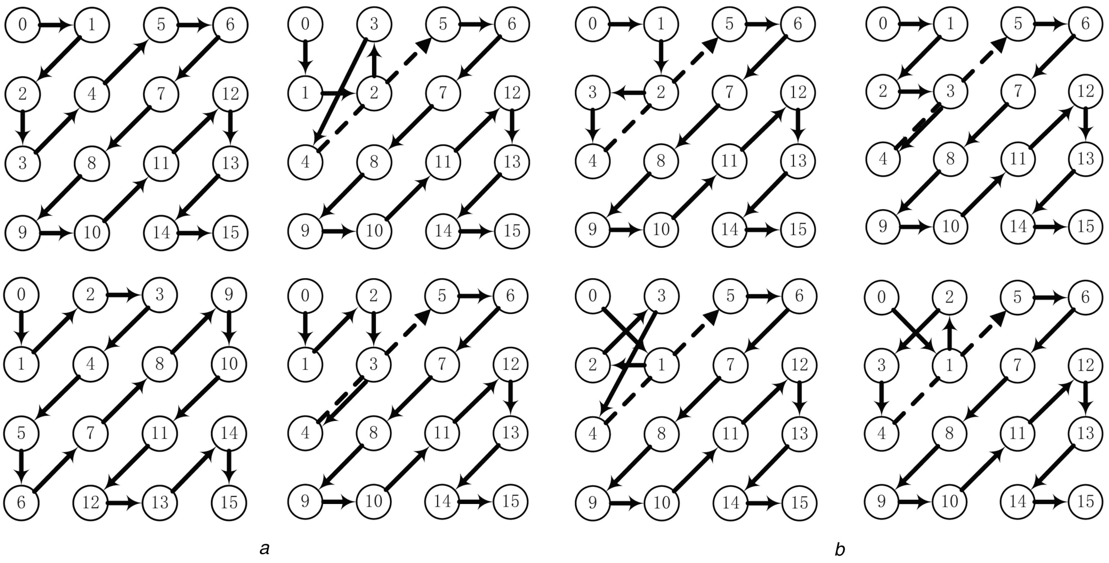


Fig. 4 *Scan orders*
a Two existing scan orders
b Proposed six scan orders (S1–S6)

order, the average differences of bitrate increase by these six scan orders under all QP values are −0.15, 0.03, −0.18, −0.25, 0.07, and 0.18%, respectively. We can conclude that these six new scan orders (S1–S6) by random permutation of the three AC coefficients in area 1 can achieve similar compression efficiency as the zigzag scan order and the symmetrical scan order.

2.3 Theoretical analysis of the proposed scan orders

Let $\boldsymbol{X}$ be the residual matrix after Inter/Intra prediction, the complete forward integer DCT transform and quantisation process can be expressed as:

$$\boldsymbol{Y} = \mathrm{round}(\boldsymbol{C} \times \boldsymbol{X} \times \boldsymbol{C}^{\mathrm{T}} \cdot \boldsymbol{M}) \tag{2}$$

where $\boldsymbol{C}$ is a constant matrix, $\boldsymbol{M}$ is a symmetric matrix which is dependant on the quantisation parameter. The elements of matrix $\boldsymbol{M}$ are none-zero, which can be obtained by a lookup-table [20]. The operators × and · are matrix multiplication and element-by-element multiplication, respectively. Let the quantised DCT coefficient matrix $\boldsymbol{Y}$ be:

$$\boldsymbol{Y} = \begin{pmatrix} a_{00} & a_{01} & a_{02} & a_{03} \\ a_{10} & a_{11} & a_{12} & a_{13} \\ a_{20} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{3}$$

Correspondingly, the inverse quantisation and inverse DCT transform process are expressed as follows:

$$\boldsymbol{Z} = \mathrm{round}(\boldsymbol{C}_i^{\mathrm{T}} \times (\boldsymbol{Y} \cdot \boldsymbol{V}) \times \boldsymbol{C}_i) \tag{4}$$

where $\boldsymbol{C}_i$ is a constant matrix, $\boldsymbol{V}$ is a matrix which can be computed from $\boldsymbol{V} \cdot \boldsymbol{M} = \boldsymbol{S}$. Since $\boldsymbol{S}$ is a symmetric constant matrix, so $\boldsymbol{V} = \boldsymbol{V}^T$. $\boldsymbol{Z}$ is the reconstructed residual block of $\boldsymbol{X}$, which can be represented as follows.

$$\boldsymbol{Z} = \begin{pmatrix} z_{00} & z_{01} & z_{02} & z_{03} \\ z_{10} & z_{11} & z_{12} & z_{13} \\ z_{20} & z_{21} & z_{22} & z_{23} \\ z_{30} & z_{31} & z_{32} & z_{33} \end{pmatrix} \tag{5}$$

To prove the scrambling effects by the proposed scan orders (S1–S6), preliminary theoretical analysis is performed on a 4 × 4

Table 1 Encoding efficiency of the symmetrical scan order [18] and the proposed scan orders

Video	QP	Bitrate difference, %						
		Symmetrical scan order	S1	S2	S3	S4	S5	S6
Foreman	12	0.37	0.24	0.29	0.01	0.00	0.49	0.50
	18	0.45	0.34	0.45	0.04	0.03	0.70	0.77
	24	0.53	0.48	0.63	0.11	0.05	0.95	1.03
	30	0.68	0.49	0.74	0.18	0.09	0.97	1.09
	36	0.42	0.28	0.90	0.25	−0.01	0.97	1.28
Mobile	12	0.13	0.04	0.08	0.03	0.00	0.13	0.14
	18	0.21	0.09	0.16	0.06	0.03	0.22	0.26
	24	0.33	0.14	0.26	0.09	0.08	0.37	0.41
	30	0.49	0.25	0.35	0.08	0.07	0.54	0.63
	36	0.59	0.48	0.56	0.17	0.15	1.02	1.05
Coastguard	12	0.10	0.10	0.42	0.24	0.11	0.29	0.47
	18	0.51	0.18	0.61	0.33	0.19	0.44	0.70
	24	1.05	0.38	0.94	0.50	0.36	0.71	1.11
	30	1.05	0.46	1.40	0.65	0.40	0.99	1.63
	36	0.65	0.32	1.61	0.96	0.09	1.41	1.84

© The Institution of Engineering and Technology 2016

residual block. Let S1 be the example. When the conventional zigzag order is used for coefficient scanning at the decoder, the resultant coefficient matrix $\boldsymbol{Y}_1'$ can be expressed by formula (6) as follows.

$$\boldsymbol{Y}_1' = \begin{pmatrix} a_{00} & a_{10} & a_{02} & a_{03} \\ a_{11} & a_{20} & a_{12} & a_{13} \\ a_{01} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{pmatrix} \tag{6}$$

If the symmetrical scan order is used, the resultant coefficient matrix $\boldsymbol{Y}_2'$ will be the transpose of $\boldsymbol{Y}$. That is,

$$\boldsymbol{Y}_2' = \boldsymbol{Y}^T \tag{7}$$

By substituting $\boldsymbol{Y}_1'$ or $\boldsymbol{Y}_2'$ into formula (4), respectively, the decoded residual 4 × 4 block can be expressed as:

$$\begin{aligned} \boldsymbol{Z'}_1 &= \mathrm{round}(\boldsymbol{C}_i^{\mathrm{T}} \times (\boldsymbol{Y'}_1 \cdot \boldsymbol{V}) \times \boldsymbol{C}_i) \\ &= \mathrm{round}\left(\boldsymbol{C}_i^{\mathrm{T}} \times \left(\begin{pmatrix} a_{00} & a_{10} & a_{02} & a_{03} \\ a_{11} & a_{20} & a_{12} & a_{13} \\ a_{01} & a_{21} & a_{22} & a_{23} \\ a_{30} & a_{31} & a_{32} & a_{33} \end{pmatrix} \cdot \boldsymbol{V} \right) \times \boldsymbol{C}_i \right) \end{aligned} \tag{8}$$

$$\boldsymbol{Z}_2' = \mathrm{round}(\boldsymbol{C}_i^{\mathrm{T}} \times (\boldsymbol{Y}_2' \cdot \boldsymbol{V}) \times \boldsymbol{C}_i) = \boldsymbol{Z}^{\mathrm{T}} \tag{9}$$

Actually, formula (8) represents the decoded result of each 4 × 4 block when it is scanned by S1 at the encoder but decoded using the conventional zigzag scan at the decoder. Although $\boldsymbol{Y}_1' \cdot \boldsymbol{V}$ simply changes the values of matrix elements in $\boldsymbol{Y}_1'$, $\boldsymbol{C}_i^{\mathrm{T}} \times (\boldsymbol{Y}_1' \cdot \boldsymbol{V})$ changes the first two columns of the matrix values, and $\boldsymbol{C}_i^{\mathrm{T}} \times (\boldsymbol{Y}_1' \cdot \boldsymbol{V}) \times \boldsymbol{C}_i$ makes the values of all matrix elements be changed. That is, since the proposed scan order S1 only simply changes the order of 3 AC coefficients in area 1, this leads to the decoded DCT coefficients completely different. For the rest scan orders (S2–S6), the same conclusion can be made by quite similar derivation process. Formula (9) implies that a block scanned by the symmetrical scan order can be decoded as the transpose of the block, which is produced by the conventional zigzag scan order in the decoding procedure [18]. Furthermore, it can be inferred from formula (8) and (9) that the scrambling effects obtained by the proposed six scan orders have more irregular permutations when they are compared with the transpose scrambling effect by the symmetrical scan order.

Fig. 5 is a numerical example for one residual block of *Foreman* sequence after intra prediction (the first I-frame). After DCT transform and quantisation (QP = 18), the 4 × 4 residual block is scanned by the proposed set of six scan orders. The resultant coefficients after scanning are shown on the left-hand side of Fig. 5. Assuming that the security key is unknown, users can only adopt the conventional zigzag scan order for block reconstruction. The reconstruction results are shown on the right-hand side of Fig. 5. There are only more positions change of DCT coefficient by the proposed scan orders (S1–S6) than the symmetrical scan order. Meanwhile, after the conventional zigzag scanning stage, inverse quantisation and IDCT at the decoder, the reconstructed blocks are quite different from the original one. Especially, compared with the symmetrical scan order, the reconstructed blocks by the proposed scan orders are not simply the transpose of the original residual block and there have more irregular permutations among them. Thus, the proposed encryption approach has better scrambling effect than the IRZZ based encryption method [18].

In summary, the new scan orders can achieve better scrambling effects compared with the symmetrical scan order without bigger cost of bitrate increase. In the following section, an improved perceptual video encryption approach is proposed by introducing them into video encoder. The reasons are two-folds: first, video encoder only specifies the bitstream syntax, which allow enough flexibility to replace the conventional zigzag scan with any new scan order; second, the combination of video encryption into encoder can keep both good compatibility with any video coding standards and desirable bitrate increases.

3 Proposed perceptual video encryption approach

The proposed perceptual video encryption approach is based on randomised zigzag scanning. Specifically, the six new scan orders presented in previous section are combined with the conventional zigzag scan and its symmetrical scan order to form a much bigger set of scan orders for video encoder. Then, a scan order is selected at the coefficient scanning stage for each 4 × 4 residual block after transform and quantisation, which is controlled by a random key. Thus, there are two critical steps including random key generation and alternative use of scan orders for the proposed video encryption approach.

3.1 Random key generation

Almost all video encryption approaches adopt a random key generator to obtain a sequence of pseudo-random codes. Thus, RC4 [21], which is a widely-used random key generator, is also exploited in the proposed approach. In our experiments, a key with 128-bit length and two 8-bit random pointers are used to initialise the permutation, and the key-scheduling algorithm (KSA) is used. Then, a key-stream is generated by using the pseudo-random generation algorithm. As described in [21], approximately 12 additions and 18 shifts are required by one 128-bit random generation, and two 8-bit random generations are involved in the RC4 key generator. For the KSA algorithm, it involves 1024 additions, 512 16-bit shifts, and 512 4-bit shifts. For a video frame in QCIF format, there are 11 × 9 macro-blocks. Each macro-block has 17 blocks, including one 4 × 4 DC block and 16 4 × 4 blocks. Assuming that half of these 17 blocks have non-zero residue, thus there are about $11 \times 9 \times 17 \times 0.5 \simeq 842$ blocks with non-zero residue in each frame. Because each 4 × 4 block requires 32 additions, four 1-bit shifts, and 16 $\boldsymbol{K}$-bit shifts ($\boldsymbol{K}$ depends on the QP value), the total operations can be estimated as 26,944 additions, 3368 1-bit shifts, and 13,472 $\boldsymbol{K}$-bit shifts. The computational complexity of key generation is almost negligible (~1%) when it is compared with the computational complexity of video frame decoding.

3.2 Alternating scan orders based on the random key

The proposed perceptual video encryption is based on randomly selecting one scan order from eight scan orders. The new randomised scanning (NRS) algorithm is summarised as follows:

First, one scan order is randomly chosen from the set of scan orders (with eight scan orders) by a random key, which is used for the scanning of 4 × 4 blocks. It is claimed earlier that compared with the conventional zigzag scan order and the symmetrical scan order, the proposed six new scan orders can achieve comparable or even better performances in terms of compression ratio and scrambling effect. Therefore, the proposed video encryption approach not only has much bigger encryption space, but also the scrambling effect and security are greatly enhanced.

Second, to further improve the encryption space, a sign-flip of DC component is exploited in Step 2.3, whereas the decoder can do the back-flip operation when the key is available. At the decoder side, when the pseudo-random key is reproduced and the same encryption algorithm is directly applied, the video will be decrypted correctly. This symmetrical design is suitable for practical implementation. In terms of security, Step 4 can be added to enhance the security by periodically refreshing the key (in our experiment, we update the key frame by frame). Since the key is constantly updated, the attacker is difficult to obtain enough number of ciphertexts under the same initialisation key or the same pseudo-random key.

Algorithm 1: New randomised scanning

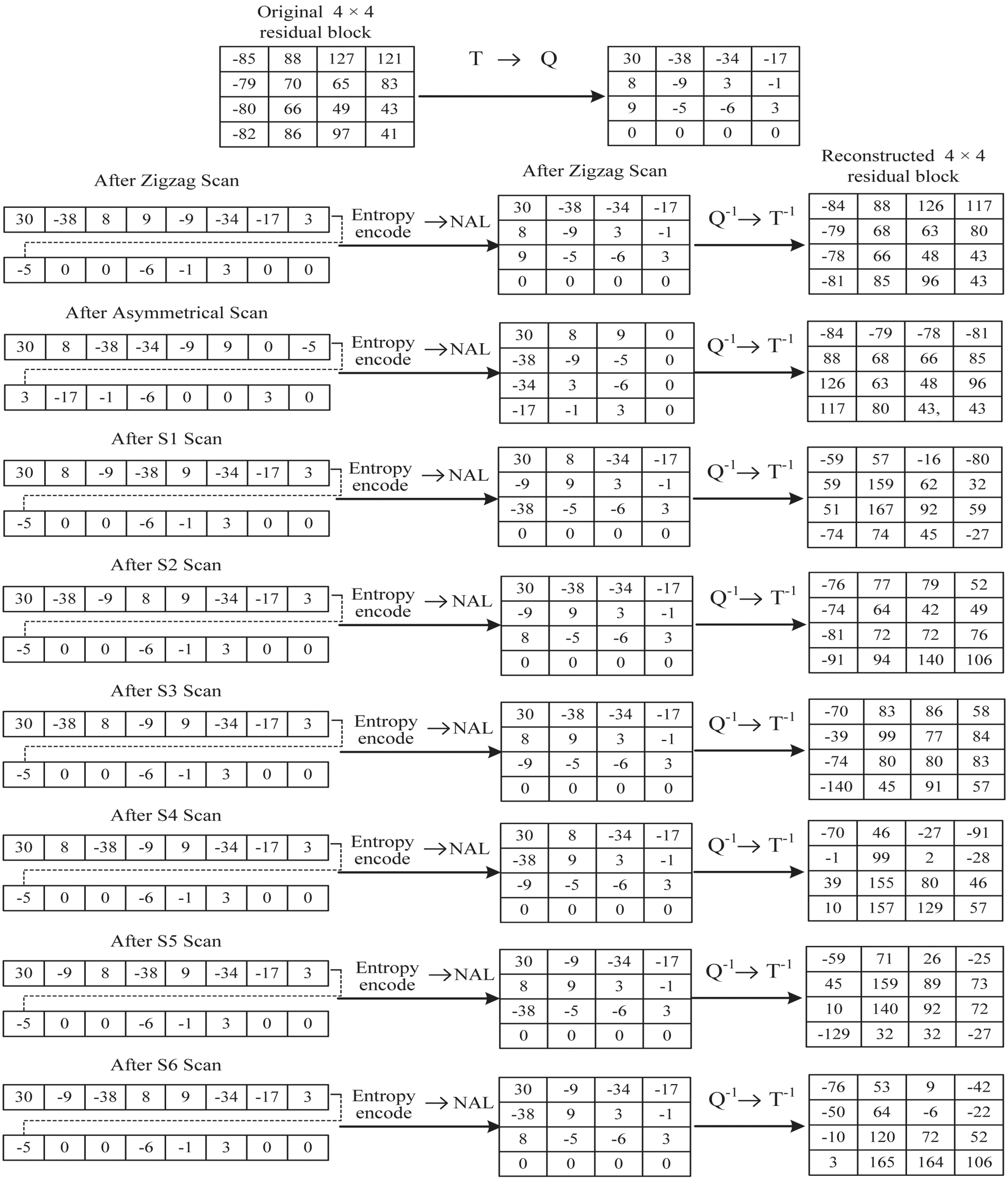


Fig. 5 *Numerical example of scan orders*

Step 1: Initialise the RC4 key generator by a random 128-bit key and produce n bits random key b_i, $i = 1, 2, \ldots, n$

Step 2: For an input residual block of size 4 × 4, do

Step 2.1: Select four bits key from random key b_i

Step 2.2: One of the eight scan orders is selected according to the first three bits

Step 2.3: The forth bit is marked as Sign; change the DC components sign if Sign = 1

Step 3: Go back to Step2 and scan next 4 × 4 DCT residual block

Step 4: Go back to Step1 after finishing one frame

4 Experimental results and analysis

To verify the performance of the proposed encryption approach, it is tested with 11 typical video sequences, which have different complexities of motion and texture and four spatial resolutions including QCIF-176 × 144 pixels, CIF-352 × 288 pixels, 4CIF-704 × 576 pixels and SD720p-1280 × 720 pixels. QCIF video sequences include *Foreman*, *Bridge* and *Salesman*; CIF video sequences include *Tempete*, *Mobile* and *Stephan*; 4CIF video sequences include *Crew*, *Ice* and *Soccer* and SD720p video sequences include *Old town* and *Tree*. The H.264/AVC reference software (JM10.2) [19] is exploited as video encoder and the proposed video encryption approach is integrated into JM10.2 with C++ programming. The hardware platform is a personal computer with an Intel Pentium G630 2.70 GHz processor and 2.0 GB memory running on Windows XP. Peak signal-to-noise ratio (PSNR) is the most popular image/video quality assessment metric, but it does not always meet people's subjective visual perception. In recent years, structural similarity index (SSIM) is widely used in image quality evaluation as well because it takes structural information into consideration. It has been reported that SSIM achieves much better results than PSNR [22, 23]. Therefore, SSIM is exploited to evaluate the visual quality of encrypted video. The smaller the SSIM value, the lower the quality of the encrypted video, and the better the scrambling effect.

© The Institution of Engineering and Technology 2016

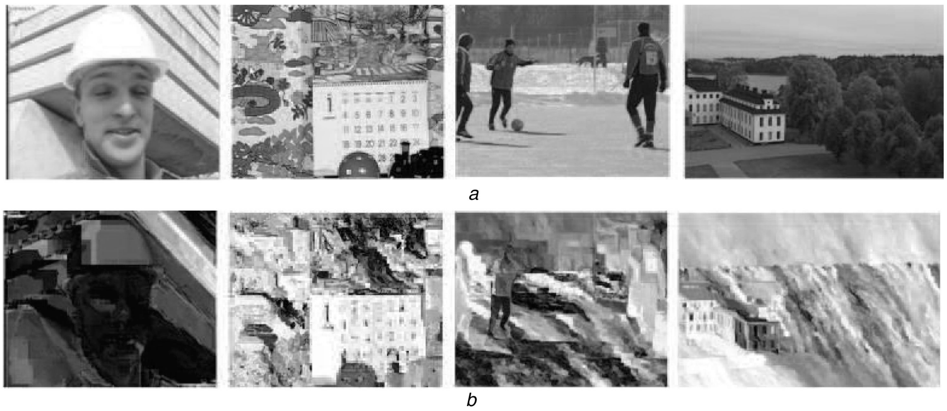


Fig. 6 *Experimental results*
a First frame of four test videos sequences
b Corresponding frame decoded without the key

4.1 Encryption performance

Since the proposed video encryption approach is integrated with the H.264/AVC video encoder, the DCT coefficients in residual blocks are scanned by randomised scan orders. The encrypted video will be decoded with a normal decoder without the encryption key, which uses the conventional zigzag scan order for coefficient scanning. Thus, the decoded video has a scrambling effect. Fig. 6 shows the experimental results, where Fig. 6*a* is the first frame of original video and Fig. 6*b* is the corresponding frame. By comparison between Figs. 6*a* and *b*, we know that their commercial values are fully destroyed because of perceptual encryption. Meanwhile, the colour distortions and scrambling effects can be easily perceived.

Table 2 compares the SSIM values of 30 decoded frames of test video sequences among the proposed NRS approach, the original algorithm without encryption (OAWE) and IRZZ [18]. The encrypted videos are intra-coded with QP value of 18. Table 2 lists the SSIM values of the *Y* component and YUV component. Moreover, the SSIM differences are also listed, where ‘diff 1 and 2(%)’ represent the percentage of SSIM change when the SSIM value of the NRS encrypted video are compared with the OAWE and the IRZZ, respectively. Table 3 summarises similar comparison results, where the encrypted videos are encoded with an IPPPP mode and the QP value of 18. From Tables 2 and 3, it is apparent that for intra-coded mode or IPPPP mode, a majority of SSIM values are <0.3 for the *Y* components employing the NRS but that of the IRZZ are >0.3. Meanwhile, we can observe that for intra-coded frames, the average differences of *Y* component's SSIM reduce about 70 and 39% compared with the OAWE and the IRZZ, respectively. For the frames encoded with IPPPP mode, the average differences of *Y* component's SSIM reduce about 78 and 50% compared with the OAWE and the IRZZ, respectively. In addition, it can also be observed that the SSIM values of the whole video sequence, i.e. YUV components, are very near to those of the *Y* component.

From the SSIM values reported in Tables 2 and 3, it is apparent that the visual quality of encrypted video by our proposed NRS approach is worse than that of IRZZ. Actually, this is reasonable simple because IRZZ only supports the conventional zigzag scan order and the symmetrical scan order that only transpose the residual block. For the proposed NRS approach, a set of eight scan orders is exploited for the random permutation of residue blocks, and thus better perceptual video encryption is achieved. Moreover, the proposed approach works well for video sequences with different complexities at different modes, and is independent of the spatial resolutions of video sequences as well.

Furthermore, the encrypted videos are compared when they are simultaneously encoded with different QP values. Table 4 summaries the SSIM values of decoded *Foreman* sequence after encryption, when it is intra-coded with QP values of 18, 32 and 40, respectively. For the proposed NRS approach, the average SSIM values are 0.3 and 0.26 for the *Y* and YUV components, respectively. They are just half or one third of those by IRZZ or OAWE, respectively. The results further prove that the proposed NRS approach is more effective to degrade the visual quality than IRZZ, which does not depend on the QP values.

From the above results, we can conclude that compared with the existing IRZZ approach, our proposed NRS approach has better scrambling effects for different video sequences when they are simultaneously encoded with inter/intra modes and different QP values.

4.2 Compression ratio

In this experiment, we report the bitrates of both original videos (OAWE) and encrypted videos (IRZZ and NRS). Furthermore, the bitrate changes are compared among them. In Table 5, ‘Percentage overhead (%)’ represents the percentage of bitrate change, where cases ‘1’ and ‘2'refer to the comparisons between IRZZ and OAWE, NRS and OAWE, respectively, and case ‘3’ refers to the

Table 2 SSIM values of 30 decoded I-frames employing OAWE, IRZZ [18] and NRS

Sequence	SSIM Y			SSIM difference, %		SSIM YUV			SSIM difference, %	
	OAWE	IRZZ	NRS	1	2	OAWE	IRZZ	NRS	1	2
Foreman	0.99	0.56	0.29	−71	−48	0.99	0.53	0.26	−74	−51
Bridge	0.98	0.71	0.43	−56	−39	0.97	0.70	0.40	−59	−43
Salesman	0.99	0.39	0.22	−78	−44	0.99	0.35	0.18	−82	−49
Tempete	0.99	0.34	0.18	−82	−47	0.99	0.29	0.15	−85	−48
Mobile	0.99	0.32	0.21	−79	−34	0.99	0.30	0.20	−80	−33
Stephan	0.99	0.42	0.30	−70	−29	0.99	0.40	0.28	−72	−30
Crew	0.99	0.61	0.37	−63	−39	0.99	0.56	0.31	−69	−45
Ice	0.99	0.74	0.53	−46	−28	0.99	0.70	0.48	−52	−31
Soccer	0.99	0.55	0.30	−70	−45	0.99	0.52	0.27	−73	−48
Old town	0.99	0.35	0.26	−74	−26	0.99	0.58	0.41	−59	−29
Tree	0.99	0.44	0.22	−78	−50	0.99	0.54	0.31	−69	−43

comparison between NRS and IRZZ. From Table 5, it is apparent that the bitrate of encrypted video by NRS is slightly lower than that of IRZZ. Especially, the bitrate increase of NRS is much less than IRZZ when they are compared with OAWE. As described in Section 2.2, three scan orders (S2, S5 and S6) slightly increase the bitrates of reconstructed videos. However, the rest three scan orders (S1, S3 and S4) reduce more bitrates of reconstructed videos. Therefore, if the security key follows a uniform distribution, the bitrates of reconstructed videos by using eight scan orders will be less than that of using the symmetrical scan order. In practice, since the encryption key is randomly generated, so the bitrates of encrypted video may be lower for most cases, and sometimes be a little higher.

4.3 Operating efficiency

Fig. 7 compares the processing time among NRS, IRZZ [18] and OAWE. Nine video sequences in QCIF format are encoded with intra-coded modes with QP values of 42. For video encoding and encryption, the proposed NRS approach requires 9.10 s, which is a little longer that those of OAWE (8.98 s) and IRZZ (9.04 s). For video decoding and decryption, OAWE, IRZZ and NRS take about 1.96, 1.98 and 2.02 s, respectively. Thus, the average percentage of encoding time increase by the proposed NRS approach is 0.71% at intra-coded modes compared with IRZZ. From Fig. 7, we know that NRS consumes a little more encoding and decoding times than IRZZ and OAWE. However, the time increments are within acceptable scope. Moreover, the slight increases of time consumption are straightforward to understand because extra time is needed to randomly select one from the set of eight scan orders in the coefficient scanning stage for both video encryption and decryption.

4.4 Security analysis

For perceptual video encryption approach, its aim is not to achieve a complete protection of video. Thus, the security requirement of perceptual encryption is lower than that of full video encryption. When the encryption space is sufficiently big to make the cost of attack very high, the security of perceptual encryption is acceptable.

4.4.1 Key space and encryption space: The proposed NRS approach exploits widely-used RC4 key generator, which generates

Table 3 SSIM values of decoded frames of IPPPP mode employing OAWE, IRZZ [18] and NRS

Sequence	SSIM Y			SSIM difference, %		SSIM YUV			SSIM difference, %	
	OAWE	IRZZ	NRS	1	2	OAWE	IRZZ	NRS	1	2
Foreman	0.99	0.43	0.21	−79	−51	0.99	0.32	0.13	−87	−59
Bridge	0.97	0.65	0.24	−75	−63	0.97	0.64	0.21	−78	−67
Salesman	0.99	0.36	0.16	−84	−56	0.99	0.41	0.18	−82	−56
Tempete	0.99	0.30	0.13	−87	−57	0.99	0.27	0.11	−89	−59
Mobile	0.99	0.29	0.17	−83	−41	0.99	0.28	0.17	−83	−39
Stephan	0.99	0.37	0.22	−78	−41	0.99	0.34	0.19	−81	−44
Crew	0.99	0.53	0.23	−77	−57	0.99	0.48	0.20	−80	−58
Ice	0.99	0.52	0.36	−64	−31	0.99	0.48	0.33	−67	−31
Soccer	0.99	0.48	0.24	−76	−50	0.99	0.44	0.21	−79	−52
Old town	0.99	0.38	0.21	−79	−45	0.99	0.44	0.24	−76	−45
Tree	0.99	0.40	0.19	−81	−53	0.99	0.46	0.26	−74	−43

Table 4 SSIM of decoded I-frames of *Foreman* sequence using OAWE, IRZZ [18] and NRS at different QP values

QP	SSIM Y			SSIM YUV		
	OAWE	IRZZ	NRS	OAWE	IRZZ	NRS
18	0.99	0.56	0.29	0.99	0.53	0.26
30	0.95	0.63	0.30	0.94	0.60	0.25
42	0.83	0.61	0.32	0.81	0.58	0.28
average	0.92	0.60	0.30	0.91	0.57	0.26

Table 5 Coding efficiency of IRZZ [18] and NRS

Sequences	QP	Bitrates, kbit/s			Percentage overhead, %		
		OAWE	IRZZ	NRS	1	2	3
Foreman	12	2645.22	2662.30	2653.45	0.65	0.31	−0.33
	18	1687.56	1699.26	1694.00	0.69	0.38	−0.31
	24	994.83	1002.07	998.78	0.73	0.40	−0.33
	30	570.12	575.47	573.16	0.94	0.53	−0.40
	36	320.62	322.50	322.07	0.59	0.45	−0.13
Mobile	12	4973.85	4995.71	4981.81	0.44	0.16	−0.28
	18	3716.28	3731.12	3722.94	0.40	0.18	−0.22
	24	2624.28	2636.94	2630.14	0.48	0.22	−0.26
	30	1707.69	1717.28	1712.66	0.56	0.29	−0.27
	36	966.92	974.14	971.14	0.75	0.44	−0.31
Coastguard	12	3156.14	3164.45	3168.06	0.26	0.38	0.11
	18	2140.30	2150.38	2150.18	0.47	0.46	−0.01
	24	1291.38	1306.10	1301.26	1.14	0.77	−0.37
	30	688.98	696.98	695.54	1.16	0.95	−0.21
	36	323.22	324.40	325.82	0.37	0.80	0.44

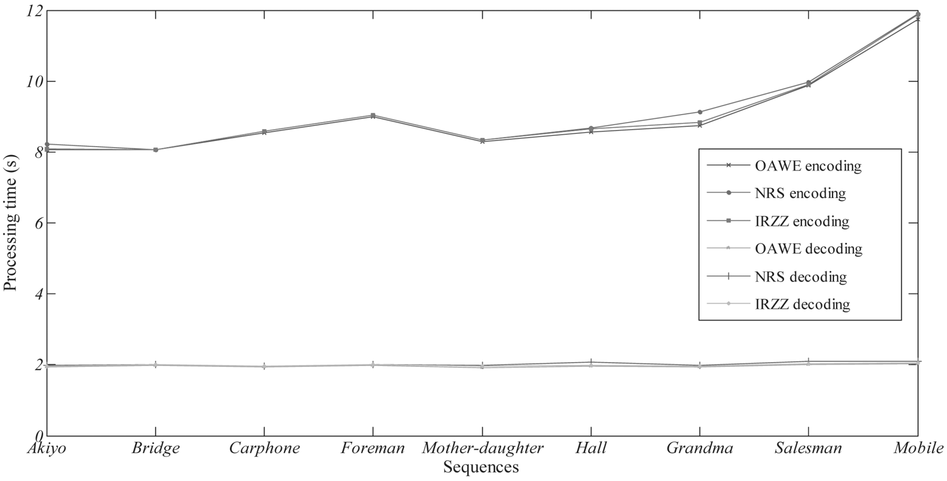


Fig. 7 *Processing time for 30 I-frames of the nine video sequences*

128-bit keys and two 8-bit pointers. Thus, the key space is 2^{128+16}, which is big enough. The 128-bit key is refreshed frame by frame, thus intruders must guess a new key for every frame. This greatly increases the complexity of attack. Therefore, the proposed perceptual encryption approach is secure in terms of key space. Moreover, since randomised scan orders and sign-flipping of DC coefficients are both adopted in the proposed approach, intruders must guess both the scan orders and the sign-flips in scanning stage for each block as well. Because there are eight scan orders and two sign-flips of DC component, the encryption space is 2^4 per block. For the existing IRZZ approach [18], there are only two scan orders for random choice, which makes its encryption space is 2^2 per block (two scan orders and two sign-flips of DC component). Meanwhile, the reconstructed block of IRZZ is either the original block or its transpose, whereas the reconstructed block of our proposed approach is more scrambled besides the original block and its transpose. Furthermore, since there are many 4 × 4 blocks within each frame, there have large enough combinational numbers to protect the blocks in each frame. That is, any incorrect guess of the scan order and sign-flipping will propagate the errors to other blocks in the same frame due to intra-mode prediction or subsequent frames due to inter-frame prediction in IPPPP coding mode.

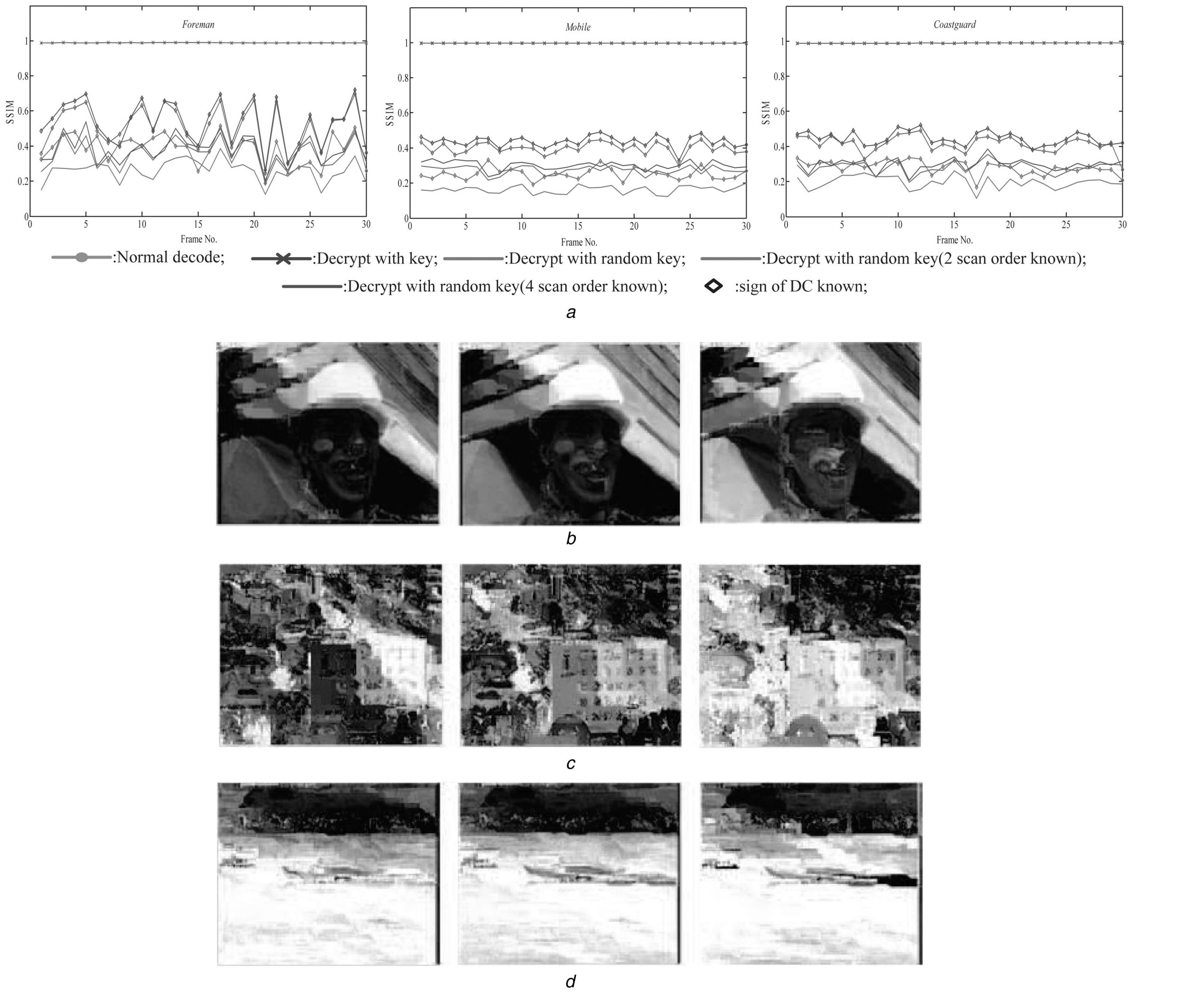
4.4.2 Known-plaintext and chosen-plaintext attacks: For the known-plaintext attack, attacks have the original video sequences and its encrypted data. For the chosen plaintext attack, attacks can obtain the corresponding encrypted video sequences if any pieces of the input video sequences are provided. For these two attacks, attacks can obtain the key generated by RC4 key generator after analysing long enough video sequences. One possible solution is to choose more secure key generator so as to provide better protection against these two attacks. However, it leads to more intensive computation to generate the key for video encryption. In this paper, the keys (128-bit keys and two 8-bit pointers) generated by RC4 are refreshed frame by frame. Thus, it invalidates the cracked keys obtained by analysing all previously-encrypted frames.

4.4.3 Ciphertext-only attacks: Ciphertext-only attack is a more realistic attack, simply because only the encrypted data is available in most cases. Thus, its security analysis is to investigate how much visual information intruders can recover under this circumstance. For ciphertext-only attacks, the most typical method is error concealment based attack, which is an effective brute force attack. However, the proposed approach has a key space of 2^{128+16}, which makes it almost impossible to be guessed. However, attackers can constantly observe the decoded frames to repeatedly guess the combination of scan orders and sign bits of DC coefficients. Let a frame in QCIF format be an example. It has 11 × 9 macro-blocks, and 80% blocks have non-zero residues. For each macro-block, it consists of 17 blocks, including one 4 × 4 block and sixteen 4 × 4 blocks. Assuming that half of the 17 blocks have non-zero residues, the number of 4 × 4 blocks with non-zero residues in one frame is $11 \times 9 \times 0.8 \times 17 \times 0.5 \simeq 673$. This implies that for the existing IRZZ approach, attackers need to guess 2^{673} times to obtain the correct scan orders for each frame at the decoder side [18]. For the proposed NRS approach, attackers need to guess $2^3 \times 2^{673} = 2^{676}$ times, which further improves the security of perceptual encryption. Apparently, it is quite difficult and time-consuming to recover each 4 × 4 block individually within a frame. Especially, since the randomised scan order technique is combined with the sign-bit flipping of DC coefficients, it will be more difficult and time-consuming. Thus, the proposed approach is robust to ciphertext-only attacks.

Fig. 8 reports the experimental results of three video sequences including *Foreman*, *Mobile*, and *Coastguard* (QCIF format, 30 frames), which are encoded with intra-coding mode. In this experiment, one scan order is selected from eight scan orders (including the conventional zigzag scan order, the symmetrical scan order and the proposed new six scan orders S1–S6) within a residual block according to the generated random key. It can be observe from Fig. 8*a* that when the security key is known there are no difference on SSIM values between the original videos and the encrypted videos. To test the security issue of the proposed NRS approach, the encrypted videos are decoded by supposing that 0, 2 or 4 scan order(s) (except the conventional zigzag scan order) is/are known in the encoder side. For those videos without knowing the scan orders, they are decoded with the conventional zigzag scan order. Experimental result shows that there is no remarkable benefit even when intruders can correctly guess four scan orders from the proposed six scan orders.

To show the visual quality of encrypted videos by the proposed NRS approach, the 15th frame of three benchmark video sequences (*Foreman*, *Mobile*, and *Coastguard*) with QCIF size for 30 I-frames at QP value 18 is presented in Figs. 8*b–d*. Subjective observations from the figures confirm that the visual quality of the decoded video frames is quite bad when the key is not completely available. Meanwhile, it is also noticed that the video quality has been improved to a certain extent but still reaches the perceptual scrambling effect even when four scan orders and the DC sign are known.

4.4.4 Trial-based attack: Note that most attacking approaches analysed above are assumed on condition that the encrypted video sequences and a part of key are known. In practice security analysis scenario, one may follow a trial-based strategy. That is to say that try a scan order in scanning stage of the H.264/AVC framework and observe/evaluate the reconstructed visual quality; try another one, and etc. Because no original video frames will be obtained to calculate the SSIM so as to select the best one among all attacks, some video subjective judge seems to be the only alternate. However, this is unnecessary that because it is very difficult and time-consuming to recover each 4 × 4 block individually at each attack. Let us assume that 2 s is needed on



a

b

c

d

Fig. 8 *Experimental results of the proposed NRS approach with different attacks*
a SSIM values for the NRS with and without the security key
b, *c* and *d* represent the perceptual scrambling effect of the NRS with random key for all scan orders and DC sign, four scan orders (out of six) known and four scan orders (out of six) and DC sign known with *Foreman*, *Mobile*, and *Coastguard*, respectively

average to subjective judge a 4×4 macroblock after each attack. According to our aforementioned discussed, NRS has an encryption space 2^4 per block. Thus, $2^4 \times 6336 \times 2$ s in total for each frame of CIF format are needed. It is worth noting that $2^4 \times 6336 \times 2 \setminus 60$ s per minute$\setminus$60 min per hour$\setminus$24 h per day is more than 2 days. However, although there are the continuous content in frames of a video and many fixed positions in the six proposed scan orders, we still believe that nobody would be willing to spend amount of time and energy trying to break that encrypted frames. As a result, we can believe that the reconstructed cost by trial-based attacking scheme can be higher than just paying for the service in entrainment video applications, especially those real-time applications such as VoD, pay-TV, and live video broadcasting.

5 Conclusions

In this paper, a set of new scan orders are designed for video encoder, which can achieve similar coding efficiency as the conventional zigzag scan order. Both theoretical analysis and experimental results show that when the proposed six scan orders are exploited in video encoders for the scanning of DCT coefficients in a 4×4 block, all the coefficients in this block will be scrambled if the encrypted video is decoded with the conventional zigzag scan order. Thus, the proposed six scan orders are combined with the conventional zigzag scan order and symmetrical scan order to form a much bigger set of scan orders, which is exploited to propose the improved perceptual video encryption approach for H.264/AVC. Specifically, one scan order is randomly selected from the set of scan orders by a security key. Moreover, the sign bits of DC coefficients are flipped to further increase the encryption space. Experimental results show that the proposed video encryption method provides better scrambling effect and higher security than existing works. Moreover, it still keeps the advantages of low computational complexity and low bitrate increase. The proposed approach shows great potentials in more popular entertainment applications such as VoD and online video streaming. In the future, we will investigate the possibility of integration new scan orders with least significant bit techniques [24, 25] for video steganography and the perceptual encryption of multi-view video coding [26].

6 Acknowledgments

The authors would like to extend sincere appreciation to the referees for their valuable comments and suggestions to improve this paper. The authors also appreciate Mrs Moses Odero John for helping us improve the English usage. This work was supported in part by the National Natural Science Foundation of China (grant nos. 61072122, 61379143, 61232016 and U1405254), the Specialized Research Fund for the Doctoral Program of Higher Education (SRFDP) under grant no. 20120161110014 and the PAPD fund.

7 References

[1] Liu, F., Koenig, H.: 'A survey of video encryption algorithms', *Comput. Secur.*, 2010, **29**, (1), pp. 3–15

[2] Stütz, T., Uhl, A.: 'A Survey of H.264 AVC/SVC Encryption', *IEEE Trans. Circuits Syst. Video Technol.*, 2012, **22**, (3), pp. 325–339

[3] Yeung, S.K.A., Zhu, S., Zeng, B.: 'Partial video encryption based on alternating transforms', *IEEE Signal Process. Lett.*, 2009, **16**, (10), pp. 893–896

[4] Yeung, S.K.A., Zhu, S., Zeng, B.: 'Perceptual video encryption using multiple 8 × 8 transforms in H.264 and MPEG-4'. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP), 2011, pp. 2436–2439

[5] Yeung, S.K.A., Zeng, B.: 'A new design of multiple transforms for perceptual video encryption'. IEEE Int. Conf. on Image Processing (ICIP), 2012, pp. 2637–2640

[6] Zeng, B., Yeung, S.K.A., Zhu, S., *et al.*: 'Perceptual encryption of H. 264 videos: Embedding sign-flips into the integer-based transforms', *IEEE Trans. Inf. Forensics Sec.*, 2014, **9**, (2), pp. 309–320

[7] Li, S., Chen, G., Cheung, A., *et al.*: 'On the design of perceptual MPEG-video encryption algorithms', *IEEE Trans. Circuits Syst. Video Technol.*, 2007, **17**, (2), pp. 214–223

[8] Magli, E., Grangetto, M., Olmo, G.: 'Transparent encryption techniques for H.264/AVC and H.264/SVC compressed video', *Signal Process.*, 2011, **91**, (5), pp. 1103–1114

[9] Asghar, M.N., Ghanbari, M., Fleury, M., *et al.*: 'Confidentiality of a selectively encrypted H.264 coded video bit-stream', *J. Vis. Commun. Image Represent.*, 2014, **25**, (2), pp. 487–498

[10] Shen, H., Zhuo, L., Zhao, Y.: 'An efficient motion reference structure based selective encryption algorithm for H.264 videos', *IET Inf. Sec.*, 2014, **8**, (3), pp. 199–206

[11] Joshi, J.M., Dalal, U.D.: 'Highly secure and fast video encryption using minimum overhead in H.264/AVC bitstream', *J. Test. Eval.*, 2016, **44**, (4), pp. 12–25

[12] Xu, D., Wang, R.: 'Context adaptive binary arithmetic coding-based data hiding in partially encrypted H. 264/AVC videos', *J. Electron. Imaging*, 2015, **24**, (3), pp. 033028–033028

[13] Shahid, Z., Chaumont, M., Puech, W.: 'Fast protection of H.264/AVC by selective encryption of CAVLC and CABAC for I and P frames', *IEEE Trans. Circuits Syst. Video Technol.*, 2011, **21**, (5), pp. 565–576

[14] Lui, O.Y., Wong, K.W.: 'Chaos-based selective encryption for H. 264/AVC', *J. Syst. Softw.*, 2013, **86**, (12), pp. 3183–3192

[15] Li, Z., Wang, X., Yang, W.: 'A fast selective video encryption algorithm by selecting data randomly'. Sixth Int. Conf. on Electronics and Information Engineering. Int. Society for Optics and Photonics, 2015, 97940S-97940S-6

[16] Tang, L.: 'Methods for encrypting and decrypting MPEG video data efficiently'. The fourth ACM Int. Conf. on Multimedia, 1997, pp. 219–229

[17] Qiao, L., Nahrstedt, K., Tam, M.C.: 'Is MPEG encryption by using random list instead of zigzag order secure?'. IEEE Int. Symp. on Consumer Electronics (ISCE), 1997, pp. 226–229

[18] Wang, Y., O'Neill, M., Kurugollu, F.: 'Partial encryption by randomized zigzag scanning for video encoding'. IEEE Int. Symp. on Circuits and Systems (ISCAS), 2013, pp. 229–232

[19] H.264/AVC Reference Software. Available at http://iphome.hhi.de/suehring/tml, 2014

[20] Richardson, I.E.: '*The H.264 advanced video compression standard*' (John Wiley & Sons, 2011)

[21] Thayer, R., Kaukonen, K.: 'A stream cipher encryption algorithm arcfour [J]'. *Internet Engineering Task Force*, 1999. Availabe at: http://www.mozilla.org/projects/security/pki/nss/draft-kaukonen-cipher-arcfour-03.txt

[22] Wang, Z., Bovik, A.C.: 'Mean squared error: love it or leave it? A new look at signal fidelity measures', *IEEE Signal Process. Mag.*, 2009, **26**, (1), pp. 98–117

[23] Yeung, S.K.A., Zhu, S., Zeng, B.: 'Quality assessment for a perceptual video encryption system'. IEEE Int. Conf. on Wireless Communications, Networking and Information Security(WCNIS), 2010, pp. 102–106

[24] Xia, Z., Wang, X., Sun, X., *et al.*: 'Steganalysis of least significant bit matching using multi-order differences', *Sec. Commun. Netw.*, 2014, **7**, (8), pp. 1283–1291

[25] Xia, Z., Wang, X., Sun, X., *et al.*: 'Steganalysis of LSB matching using differences between nonadjacent pixels', *Multimedia Tools Appl.*, 2014, **1**, pp. 1–16

[26] Pan, Z., Zhang, Y., Kwong, S.: 'Efficient motion and disparity estimation optimization for low complexity multiview video coding', *IEEE Trans. Broadcast.*, 2015, **61**, (2), pp. 166–176

© The Institution of Engineering and Technology 2016