Improved Perceptual Video Encryption using Alternative Unitary Transforms

Siu-Kei Au Yeung
Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology
Hong Kong, China
Email: jeffay@ust.hk

Shuyuan Zhu
Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology
Hong Kong, China
Email: eezhshy@ust.hk

Bing Zeng
Department of Electronic and Computer Engineering
The Hong Kong University of Science and Technology
Hong Kong, China
Email: eezeng@ust.hk

***Abstract*— It has been demonstrated in our earlier work [1] that perceptual video encryption can be effectively achieved by using alternative unitary transforms. In this paper, we extent our work by studying the importance of selecting rotation angles to be used to generate the alternative transforms. We show theoretically and experimentally that by selecting a set of rotation angles properly, the resulted perceptual encryption system will be operating without affecting the coding efficiency, but with a higher encrypting capability. In additional, our earlier algorithm [1] can also be further simplified with the new set of rotation angles. Both objective and perceptive evaluation will be performed with the standard H.264 codec.**

Keywords-perceptual video encryption; alternative unitary transforms; rotation angles; H.264 codec

I. INTRODUCTION

As high-speed Internet services become more and more popular, multimedia communications, including audio, image and video data, have been widely deployed. Applications such as video-on-demand (VOD), pay-TV, real-time video broadcasting and video conferencing have closely connected people around the world. However, security issue of video transmission is becoming an important problem that needs to be handled in order to provide a safe environment to share the video data.

Based on different applications, video encryption algorithm can be divided into two main types: (1) complete encryption and (2) perceptual encryption. Application such as video conference and video telephony required fully confidential video transmission. The system should not allow any useful information to be decodable for the un-authorized person. The same requirement also occurs in military or defense related applications as well as many applications in financial business. Conventional cryptographic such as AES and DES [2] is one of the approaches to fulfill this requirement. Since AES and DES require high computation power, which is not suitable for large volume video data, researchers proposed various alternative algorithms such that computational requirement is reduced but with similar encrypting capability. An overview of those algorithms can be found in [3]. Since complete encryption is required, any proposed algorithm needs to be evaluated under many decrypting attacks such as *known-plaintext* and *chosen-plaintext* [2].

Meanwhile, we believe that a bigger portion of video transmissions is for entrainment. Applications such as video-on-demand (VoD), pay-TV or online live video broadcasting are a few typical examples. For the entrainment purpose, the encryption criteria are basically two-folded. The first one is to make sure that the cost of breaking the encryption is higher than simply buying the decoding key. We generally prefer fast and simple encryption algorithms which cannot be broken easily but may not be necessary to immune from some complicated attacking schemes. Secondly, in order to let potential customers have a quick look about the video content, it is preferable if a video preview at a significantly lower quality can be provided so as to attract customers to buy the high-quality version. This idea is referred as *secure media content preview* (SMCP) [4], and any such algorithm is called as a perceptual encryption [5]. Because of the second criterion, most of the complete encryption algorithms would definitely not be suitable for SMCP as no visual information remains after a complete encryption (some algorithms even make the video not decodable without the key).

One simple principle of perceptual encryption is to select one part of (visually) sensitive video component to do the encryption. Generally, simple and fast encryption algorithms are preferred here; whereas the visual quality would degrade to a very low level at which some of the video contents nevertheless is still visible.

Algorithms proposed so far for perceptual encryption are basically focusing on scrambling part of video signal before, during or after the encoding process [5-8] with some control parameters. Recently, we proposed to perform this encryption at a stage that had never been considered before: the transformation stage [1]. More specifically, we proposed to use alternating unitary transforms at the transformation stage based on a secret key. Further modification and analysis on this proposed algorithm is presented in [9]. The major advantage of the proposed algorithm is that scrambling process is no longer needed for both encoder and decoder, which significantly reduces the computational complexity since the transform must be done whether encryption is

CHINACOM 2010, August 25-27, Beijing, China
Copyright © 2011 ICST 973-963-9799-97-4
DOI 10.4108/chinacom.2010.58

performed or not. We have showed in those earlier works that alternative unitary transforms (that lead to an equal coding efficiency as DCT does) can be found easily by some extra rotations in the DCT's butterfly flow-graph. This encryption algorithm has been evaluated intensively with the H.264 standard system [1, 9]. One problem with this algorithm is that the sign-flip of the DC component in each block seems to overwhelm the impact of applying alternative transforms. In this paper, we first come back to have a more detailed study on the selection of rotation angles. With these newly-selected angles, we show that the sign-flip of each DC component now becomes unnecessary. At the same time, we show that the coding efficiency of using these new rotation angles remains unchanged; while a better perceptual encrypting capability in fact is yielded.

II. OUR PROPOSED ALGORITHM FOR PERCEPTUAL ENCRYPTION

Our perceptual video encryption algorithm consists of two steps: (1) random (secret) key generation and (2) alternating transforms according to the secrete key.

A. Random key generation

Almost all video encryption algorithms rely on a random key generator to produce a sequence of pseudo-random codes. To this end, RC4 turns to be the most commonly-used random key generator [10], and it has been adopted in popular protocols such as SSL and WEP. The permutation is initialized with a variable length key which has typically 40 to 256 bits (we set it to 128 bits in our experiments) and two 8-bit internal index pointers. The key-scheduling algorithm (KSA) is first applied. The key-stream is then generated using the pseudo-random generation algorithm (PRGA). Although RC4 does not provide highly secured key generation, we aim to use simple and fast key generation algorithm for the perceptual encryption purpose. Of course, RC4 can be replaced by other more secured key generator if necessary.

To implement one 128-bit random generation in the RC4 key generator, we roughly need 8 additions and 12 shifts (including 4-bit, 16-bit, and 32-bit shifts). In addition, we also need 1,024 additions, 512 16-bit shifts, and 512 4-bit shifts for KSA (2 shifts and 1 addition are needed for each mod operation). On the other hand, during the decoding of one video frame of the CIF format, if we assume that 50% of blocks (i.e., 3168 blocks) have non-zero quantized coefficients. It is estimated that the total operations include 101,376 additions, 12,672 1-bit shifts, and 50,688 K-bit shifts (K depends on the QP value): each 4×4 block require 32 additions, 4 1-bit shifts, and 16 K-bit shifts. Clearly, the computational complexity from the key generation is negligible (~1%) when compared with that from the decoding of one video frame.

B. Encoding with alternating transforms driven by the random key

Our proposed video encryption procedure is stated as follow:

----------Encryption Algorithm----------

Step 1: Initialize the RC4 key generator by a random 128-bit key

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

- Step 2.1: Get M bits from the random generator
- Step 2.2: The first $M-1$ bits are marked as Dec; a transform is selected according to Dec
- Step 2.3: The M^{th} bit is marked as Sign; change the DC component's sign if Sign = "1"

Step 3: Perform the transformation based on Dec and Sign

Step 4: Got back to Step 1 after one frame is encrypted.

At the decoder side, the pseudo-random key will be re-produced so that the decoder can make all decisions correctly. It is determined by the user and clearly it controls the security level – more alternating transforms provides a higher security level. To further increase the security level at a given M, we can apply different transforms (alternately according to the random key) along various columns and rows within each block. For a 4x4 block, a total of 8 different transforms can be applied. In Step 2.3, we performed a sign-flip on the DC component, whereas the decoder can do the back-flip when the key is available. We will show in the next section that this sign-flip can be waived by accurately select the alternative transforms.

III. SELECTION OF NEW ROTATION ANGLE FOR ALTERNATIVE TRANSFORMS

The new perceptual encryption scheme proposed in this work tries to select some rotation angles (to be used to derive the alternative transforms) that are different from those selected earlier in [1]. In general, we need to consider several criteria for the selection:

1) Unitariness: the biggest advantage of a unitary transform is that the inverse transform matrix is simply the transpose of the forward transform matrix. Also, fast implementation can be reached for many unitary transforms.
2) Equal coding efficiency: DCT is known to be very efficient in compacting energy in the transform domain. The newly-developed unitary transforms shall offer an equal coding efficiency as much as possible.
3) Integer-based transforms: it has been employed in the H.264 video encoding standard. It would be very nice to also end up with these alternative transforms that can be implemented completely in integers.

When DCT is used, each 1-D transform can be implemented through the flow graph as shown in Fig. 1. It is seen from Fig. 1 that the whole graph consists of three stages: (1) two butterflies that are both equivalent to the plane-based rotation of $\pi/4$; (2) two extra plane-based rotations of $\pi/4$ and $3\pi/8$, respectively; and (3) a permutation. Our proposed idea is, by keeping stages 1 and 3 unchanged, a class of new unitary transforms can be generated by using different rotation angles in Stage 2. For example, by changing both $\pi/4$ and $3\pi/8$ to

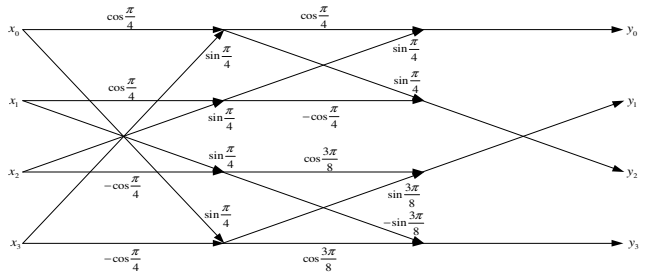


Fig. 1: The flow graph of 4-point (1-D)

$29\pi/90$ (~58°), we obtain the popular discrete sine transform (DST) of type-I, whereas the DST of type-II will be obtained by swapping $\pi/4$ with $3\pi/8$. Clearly, more alternates can be generated by selecting different rotation angles.

It is well-known that DCT outperforms other (fixed) unitary transforms in terms of the coding efficiency for most image and video signals in which the correlation among neighboring samples remains at a high level. The alternative transforms selected here are not aiming to become even better than DCT; otherwise DCT would have been replaced by one of them. In fact, we expect that these alternative transforms would be performing very similarly as DCT when used to encode the residual signal (intra-predicted I-frames and motion-compensated P-frames) in video coding

A. Theoretical analysis on coding efficiency

We follow the commonly-used circular model to assume that the correlation between two samples is calculated as

$$E\{p_{i,j}, p_{m,n}\} = \rho^{\sqrt{(m-i)^2+(n-j)^2}} \tag{1}$$

where $0 < \rho < 1$. Let $\{S_{u,v}\}$ be the set of transform coefficients and $\{c_{i,j}\}$ the $N \times N$ transform matrix ($N = 4$ in our case). Then, the variance of each transform coefficient can be calculated as:

$$\sigma_{u,v}^2 = E\{S_{u,v} \cdot S_{u,v}\} = \sum_{i,j,m,n=0}^{N-1} E\{p_{i,j} p_{m,n}\} \cdot \left(c_{u,i} c_{u,m} c_{v,j} c_{v,n}\right) \tag{2}$$

To measure the coding efficiency of a transform in the case that some quantization is carried out on transform coefficients, we can measure the energy packing efficiency (EPE) that is defined as the energy portion contained in the first M_0 transform coefficients (along the zig-zag scanning order) compared with the total energy ($M_0 = 6$ in our tests):

$$EPE = \sum_{t=0}^{5} E\{S_t^2\} / \sum_{t=0}^{15} E\{S_t^2\} \tag{3}$$

In order to evaluate various transforms with different rotation angles as described above, we set two rotation angles in Stage 2 of Fig. 1 to be $\pi/4 + \theta_1$ and $3\pi/8 + \theta_2$. We compute EPE's via adjusting both θ_1 and θ_2 from $-\pi$ to π. Figure 2 shows the results with $\rho = 0.3$, 0.5, and 0.98, respectively. It can be seen that EPE achieves its maximum in most of the case when both θ_1 and θ_2 are set to 0 - which is the original DCT case. For highly correlated signals ($\rho = 0.98$), rotation angle θ_2 seems to provide no damage on the coding efficiency; whereas the coding efficiency decreases for θ_1 in $[0, \pi/2]$ and then recovers when θ_1 falls in $[\pi/2, \pi]$. This is because that θ_1 is related to the DC component of each block – leading to a huge impact on the coding performance if it is selected badly. On the other hand, for signals with weaker correlation ($\rho = 0.5$ or 0.3), DCT still outperforms many other transforms. Furthermore, two observations can be made: First, the coding efficiency is still pretty high for θ_1 and θ_2 in $[-\pi/4, \pi/4]$. Thus, alternative transforms can be chosen from this region. However, such small angles θ_1 and θ_2 make the resulted transforms very similar to DCT as well as similar with each other – leading to a low encrypting capability (as an incorrect transform would produce a similar decoded result when the secret key is unknown). Second, it is very clear that a very similar (or an equal) coding efficiency has been yielded when θ_1 and/or θ_2 are set at π.

In particular, we believe the second observation mentioned above is very valuable. On one hand, the residual signal (obtained after the intra-prediction or motion-compensation) is indeed less correlated among its neighboring samples. To further verify it, we performed experiments on real video sequences (*Foreman*, *Mobile*, and *Stefan*, 300 frames in the CIF format). Figure 3 shows the results with QP = 26 (in H.264). It can be seen that the results from *Foreman* match closely with the theoretical results with $\rho = 0.5$; whereas the results from both *Mobile* and *Stefan* show a better match with the results with $\rho = 0.3$.

IV. EXPERIMENTAL RESULT AND SECURTIY ANALYSIS

It is always not an easy task to evaluate the effectiveness of a video encryption scheme because there are many possible attacks. In [1] and [9], we presented the security analysis for the proposed perceptual encryption system against different common attack. In this paper, we focus on the perceptual distortion performance when the encryption is decoded without the key.

In [1,9] we selected 2 alternative transforms at ($7\pi/24$, $8\pi/24$) and ($8\pi/24$, $7\pi/24$), respectively, in addition to the popular DCT and DST-2. There two new transforms are actually the equal-spaced samples between two original angles $\pi/4$ and $3\pi/8$. Results from [1] showed that the resulted encryption scheme provides a certain level of perceptual distortion when the key is missing. However, we also observed that the (random) DC sign-flip is needed in order to provide sufficient distortion (see Figs. 5a and 5b). Based on the analysis presented in the last section, we make the following selection of rotation angles: ($\theta_1 = \pi$, $\theta_2 = \pi$), ($\theta_1 = \pi$, $\theta_2 = 0$), and ($\theta_1 = 0$, $\theta_2 = \pi$). As we mentioned before, the rotation by π will not affect the EPE. In addition, integer transforms for this new set of rotation angles can be derived easily [9].

We perform the encryption based on integer transforms. In the first step of transform (vertical or horizontal), we choose one from 4 alternative transforms and apply it onto all columns or rows. In the second step of transform, we individually choose one from 4 alternative transforms for each row or column. As a result, the number of possible combinations of transforms is $4^5 = 1024$ for each 4x4 block. Totally, we have 6336 blocks

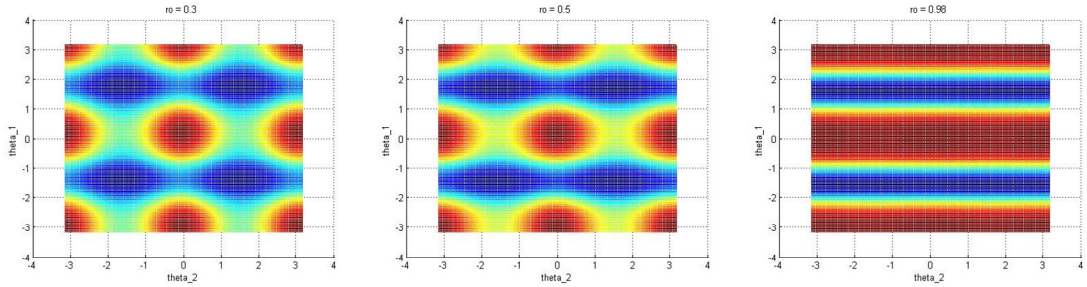


Figure 2: EPE's for different rotation angles: dark red—the highest EPE; dark blue—the lowest EPE.

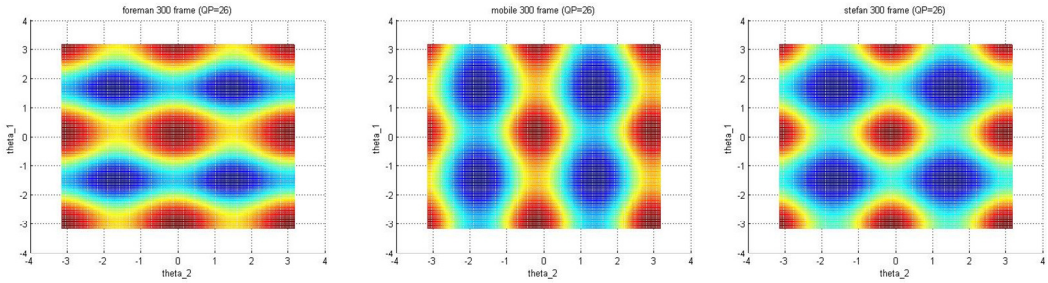


Figure 3: EPE's for different rotation angles, obtained from testing practical video sequences.

for each frame, so brute-force guessing is definitely not feasible. Notice that with this new selection of rotation angles, Step 2.3 in the aforementioned encryption algorithm in Section 2.2 is removed such that no random sign-flipping is needed and the encryption is fully controlled by the alternative transforms.

Figure 4 present the coding performance of the proposed perceptual encryption system with and without the secret key for three video sequences. Both PSNR and *structural similarity* (SSIM) analyses are considered. We also compare our new results with the results presented in [9]. It can be seen from Fig. 4 that, if the secrete key is known, there is no coding efficiency loss for the modified system with the new set of rotation angles; whereas the coding efficiency drops very slightly when the rotation angles proposed in [9] are used. When the key is missing, the new scheme achieves a better encryption performance, which becomes more apparent when the SSIM evaluation is conducted. Notice that the curves for the case of no secret key are averaged results over 5 random trials on guessing the key. Finally, Fig. 5c shows some actual snapshots of three sequences after decoding without key. It can also be seen that the new proposed set of rotation angles provides a higher protection compared to the old set of rotation angles used Figs. 5a and 5b, especially for the *Mobile* sequence.

V. CONCLUSION

In this paper, we have presented some new results of the perceptual video encryption system that was proposed in our earlier work [1]. We keep the basic framework unchanged, i.e., it is still based on a set of unitary transforms, amongst which one is selected for each block according to a secret key. However, we have applied a new set of rotation angles to generate new unitary transforms. These new unitary transforms bring us with the following benefits: (i) the R-D performance increases slightly when the key is known; (ii) a stronger protection is achieved when the key is unknown, while still maintaining the preview purpose; and (iii) the random sign-flip on the DC component of each block becomes unnecessary. The issues that will be considered in our follow-up works include the following:

- Quality control is one important issue for perceptual video encryption, i.e., to what extent (in terms of visual quality after decoding without the key) we need to perform the encryption.
- So far, only 4x4 transforms have been considered. It would be interesting to see whether the same rotation-based scheme can be applied to the 8x8 case.

REFERENCES

[1] S. K. Au-Yeung, S. Zhu, B. Zeng, "Partial video encryption based on alternating transforms," IEEE Signal Processing Letters, pp. 893-896 ,2009

[2] D. Stinson, "Cryptography Theory and Practice," Chapman & Hall/CRC ,2006

[3] F. Liu, H. Koenig, "A survey of video encryption algorithm", Computers & Security, pp. 1-13, 2009

[4] S. Lian, "Multimedia content encryption: techniques and applications, " Boca Raton: CRC Press, 2009

[5] S. Li, G. Chen, A. Cheung, B. Bhargava and K. Lo, "On the Design of Perceptual MPEG-Video Encryption Algorithms", IEEE Trans. On Circuit and Systems For Video Technology, pp. 214-223, 2007

[6] M. Pazarci and V. Dipcin, "A MPEG-2-transparent scrambling technique," IEEE Trans. On Consumer Electronics, pp.345-355, 2002

[7] C. Wang H. Yu and M. Zheng, "A DCT-based MPEG-2 transparent scrambling algorithm," IEEE Trans. On Consumer Electronics, Vol. 49, no. 4, pp. 1208-1213, 2003

[8] F. Wang, W. Wang, J. Ma, C. Xiao, K. Wang, "Perceptual video encryption scheme for mobile application based on H.264," The Journal of China Universities of Posts and Telecommunications, pp.73-78, 2008

[9] S. K. Au-Yeung, S. Zhu, B. Zeng, "Partial video encryption based on alternating integer transforms, " accepted to IEEE ISCAS 2010 (http://ihome.ust.hk/~jeffay/detail/ChinaCom2010/ISCAS2010_01.pdf)

[10] Kaukonen, R. Thayer, "A Stream Cipher Encryption Algorithm "Arcfour", " IETF Draft, 1999

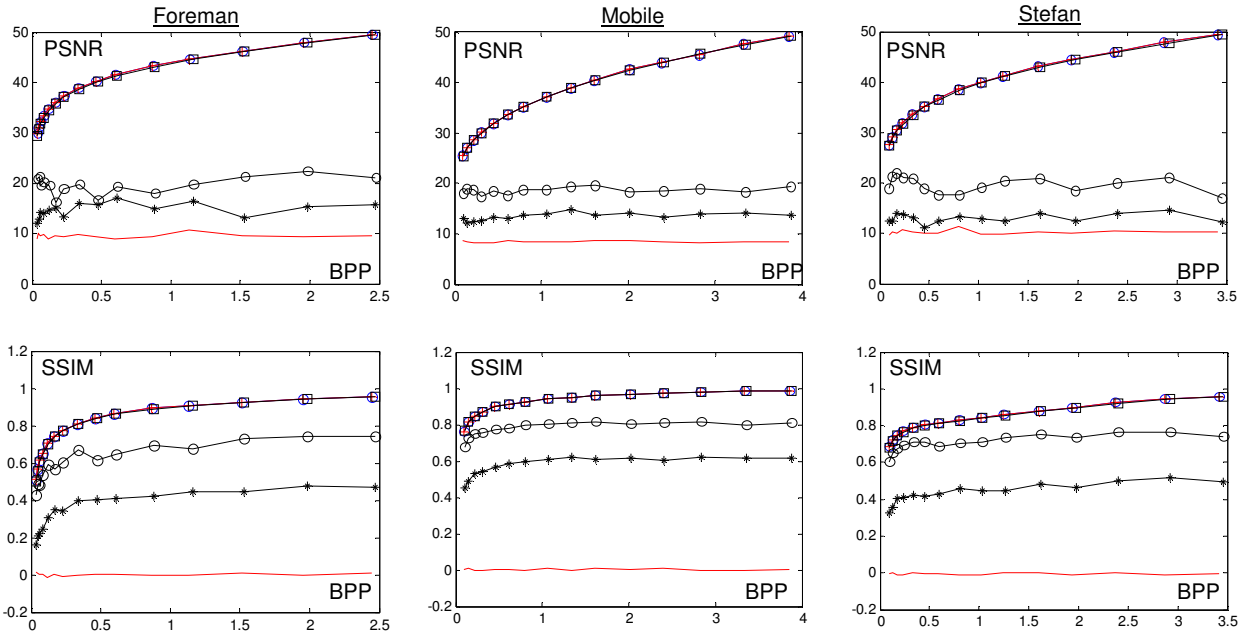
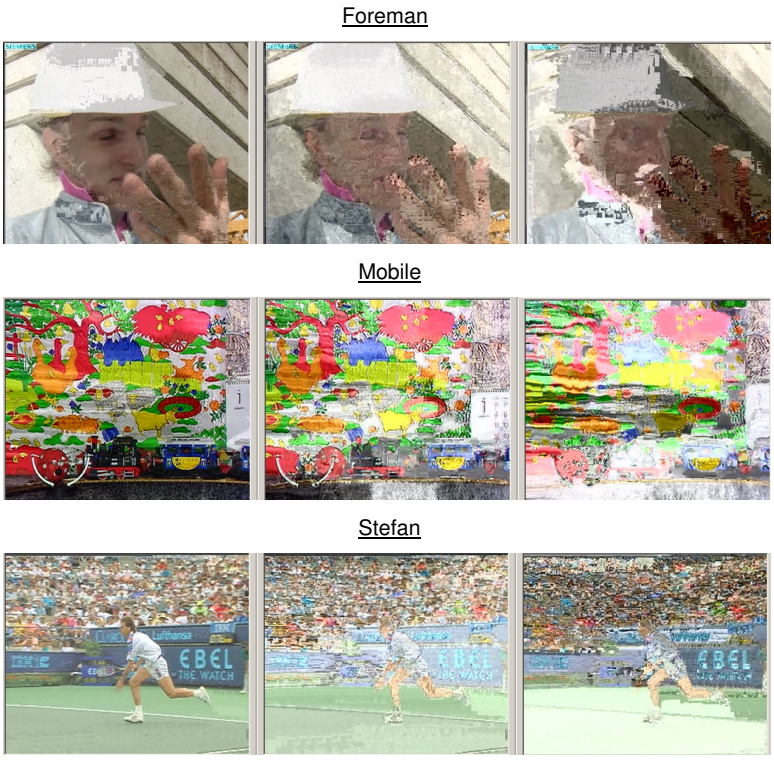


Fig. 4. R-D performances of proposed encryption system (DCT, decrypted with key (new rotation angle), decrypted with key (rotation angle proposed in [1]), decrypted without key (new rotation angle), decrypted without key (rotation angle proposed in [1]), decrypted without key (rotation angle proposed in [1] with DC sign encrypt))



(a) (b) (c)

Fig. 5. Snapshot of the proposed algorithm: (a) using rotation angle proposed in [1] -- (7π/24, 8π /24), (8π/24, 7π/24), DST-2, DCT-2 (b) using rotation angle proposed in [1] with DC sign encrypt. (c) Using rotation angle (π/4+ π, 3π/8+π), (π/4+π, 3π/8), (π/4, 3π/8+π)
(Video clip is available at http://ihome.ust.hk/~jeffay/detail/ChinaCom2010/)